How Close Are We to Understanding V1?

Bruno A. Olshausen
baolshausen@ucdavis.edu
Redwood Neuroscience Institute, Menlo Park, CA 94025, and Center for
Neuroscience, University of California at Davis, Davis, CA 95616, U.S.A.

David J. Field
dif3@cornell.edu
Department of Psychology, Cornell University, Ithaca, NY 14853, U.S.A.

A wide variety of papers have reviewed what is known about the function of primary visual cortex. In this review, rather than stating what is known, we attempt to estimate how much is still unknown about V1 function. In particular, we identify five problems with the current view of V1 that stem largely from experimental and theoretical biases, in addition to the contributions of nonlinearities in the cortex that are not well understood. Our purpose is to open the door to new theories, a number of which we describe, along with some proposals for testing them.

1 Introduction

The primary visual cortex (area V1) of mammals has been the subject of intense study for at least four decades. Hubel and Wiesel's original studies in the early 1960s created a paradigm shift by demonstrating that the responses of single neurons in the cortex could be tied to distinct image properties such as the local orientation of contrast (Hubel & Wiesel, 1959, 1968). Since that time, the study of V1 has become something of a miniature industry, to the point where the annual Society for Neuroscience meeting now routinely devotes multiple sessions entirely to V1 anatomy and physiology. Without doubt, much has been learned from these efforts. However, as we shall argue here, there remains a great deal that is still unknown about how V1 works and its role in visual system function. We believe it is quite probable that the correct theory of V1 is still far afield from the currently proposed theories.

It may seem surprising to some that we should take such a stance. V1 does, after all, have a seemingly ordered appearance: a clear topographic map and an orderly arrangement of ocular dominance and orientation columns. Many neurons are demonstrably tuned for stimulus features such as orientation, spatial frequency, color, direction of motion, and disparity. And there has even emerged a fairly well-agreed-on "standard model" for V1 in which simple cells compute a linearly weighted sum of the input over
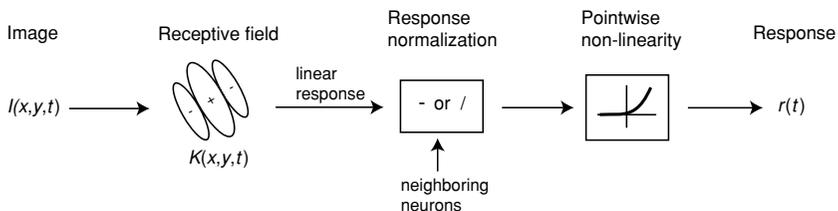
Figure 1: Standard model of V1 simple cell responses. The neuron computes a weighted sum of the image over space and time, and this result is normalized by the responses of neighboring units and passed through a pointwise nonlinearity (see e.g., Carandini, Heeger, & Movshon, 1997).

space and time (usually a Gabor-like function), which is then normalized by the responses of neighboring neurons and passed through a pointwise nonlinearity (see Figure 1). Complex cells are similarly explained in terms of a summation over the outputs of a local pool of simple cells with similar tuning properties but different positions or phases. A variety of models have been proposed for the response normalization (Heeger, 1991; Geisler & Albrecht, 1997; Schwartz & Simoncelli, 2001; Cavanaugh, Bair, & Movshon, 2002a), but the net result is often to think of V1 as a kind of "Gabor filter bank." There are numerous papers showing that this basic model fits much of the existing data well, and many scientists have come to accept this as a working model of V1 function (see, e.g., Lennie, 2003a for a discussion). Indeed, such models are widely used to predict psychophysical performance (Graham & Nachmias, 1971; Watson, Barlow, & Robson, 1983; Anderson, Burr, & Morrone, 1991), and they have been shown to provide efficient representations of natural scenes (Olshausen & Field, 1996; Bell & Sejnowski, 1997).

But behind this picture of apparent orderliness lies an abundance of unexplained phenomena, a growing list of untidy findings, and an increasingly uncomfortable feeling among many about how the experiments that have led to our current view of V1 were conducted in the first place. The main problem stems from the fact that cortical neurons are highly nonlinear—that is, they emit all-or-nothing action potentials, not analog values. They also adapt, so their response properties depend on the history of activity. Most important, cortical pyramidal cells have highly elaborate dendritic trees, and realistic biophysical models that include voltage-gated channels suggest that each thin branch could act as a nonlinear subunit, so that any one neuron could be computing many different nonlinear combinations of its inputs (Hausser & Mel, 2003; Polsky, Mel, & Schiller, 2004), in addition to being sensitive to coincidences (Softky & Koch, 1993; Azouz & Gray, 2000, 2003).

Everyone knows that neurons are nonlinear, but few have acknowledged the implications for studying cortical function. Unlike linear systems, where

there exist mathematically tractable textbook methods for system identification, nonlinear systems cannot be teased apart using some straightforward, structuralist approach. That is, there is no unique "basis set" with which one can probe the system to characterize its behavior in general.[1] Nevertheless, the structuralist approach has formed the bedrock of V1 physiology for the past four decades. Researchers have probed neurons with spots, edges, gratings, and a variety of mathematically elegant functions in the hope that the true behavior of neurons can be explained in terms of some simple function of these components. However, the evidence that this approach has been successful is lacking. We simply have no reason to believe that a population of interacting neurons can be reduced in this way.

For any complex system, it seems reasonable to begin where the system acts rationally: to study the behavior under conditions where one's models are relatively effective. But for a neural system, that leaves the question as to whether such behavior represents the relevant aspect of the neurons activity: Does this help us understand how neurons operate under natural conditions? Much of our understanding of V1 is derived from recording from one neuron at a time using simple stimuli (edges, gratings, spots). From this body of experiments has emerged the standard model that forms the basis for our conceptual understanding of V1. In recent years, a number of innovative studies have moved away from this basic approach, recording from multiple neurons with complex, ecologically relevant stimuli. Are these studies simply adding minor correction factors to our understanding, or will they require us to completely revamp the current theories? Are the current models close to accounting for the majority of responses in the majority of neurons in V1? How close are we to understanding V1?

In this review, we present our reasons for believing that we may have far to go in understanding V1. We identify five fundamental problems with the current view of V1 function that stem largely from experimental and theoretical biases, in addition to the contributions of nonlinearities in the cortex that are not well understood. Furthermore, we attempt to quantify the level of our current understanding by considering two important factors: an estimate of the fraction of V1 neuron types that are typically characterized in experimental studies and the fraction of variance explained in the responses of these neurons under natural viewing conditions. Together, these two factors lead us to conclude that at present, we can rightfully claim to understand only 10% to 20% of how V1 actually operates under normal conditions.

Our aim in pointing these things out is not simply to tear down the current framework. We ourselves have attempted to account for some aspects of the

---

[1] The Volterra series expansion is often touted as a general approach for characterizing nonlinear systems, but it has been of little practical value in analyzing neural systems because it requires estimating many higher-order moments. In addition, it is an overly general "black box" approach that does not easily allow one to incorporate prior knowledge about the types of nonlinearities known to exist in the nervous system.

standard model in terms of efficient coding principles (sparse coding), so obviously we believe that we have made a good start. Rather, our goal is to show how much room there is for new theories and where the weaknesses in the current theories might lie. In the second half of the review, we describe a few of our favorite alternatives to the standard theories. A central conclusion that emerges from this exercise is that we need to begin seriously studying how V1 behaves with natural scenes, using multiunit recording techniques, in addition to explicitly describing any potential biases in the gathering of data. We believe this approach can help to reveal not just how much we know about neural coding in the visual pathway but also how much we do not know.

2 Five Problems with the Current View

2.1 Biased Sampling of Neurons. The vast majority of our knowledge about V1 function has been obtained from single unit recordings in which a single microelectrode is brought into close proximity with a neuron in cortex. Ideally, when doing this, one would like to obtain an unbiased sample from any given layer of cortex. But some biases are difficult to avoid. For instance, neurons with large cell bodies will give rise to extracellular action potentials that have larger amplitudes and propagate over larger distances than neurons with small cell bodies. Without careful spike sorting, the smaller extracellular action potentials may easily become lost in the background when in the vicinity of neurons with large extracellular action potentials. This creates a bias in sampling that is not easy to dismiss.

Even when a neuron has been successfully isolated, detailed investigation of the neuron may be bypassed if it does not respond "rationally" to standard test stimuli or fit the stereotype of what the investigator believes the neuron should do. This is especially true for higher visual areas such as V4, but it is also true for V1. Such neurons are commonly regarded as "visually unresponsive." It is difficult to know how frequently such neurons are encountered because often they simply go unreported, or else it is simply stated that only visually responsive units were used for analysis.

While it is admittedly difficult to characterize the information processing capabilities of a neuron that seems unresponsive, it is still important to know in what way these neurons are unresponsive. What are the statistics of activity? Do they tend to appear bursty or tonic? Do they tend to be encountered in particular layers of cortex? And most important, are they merely unresponsive to bars and gratings, or are they also equally uninterpretable in their responses to a wider variety of stimuli, such as natural images? A seasoned experimentalist who has recorded from hundreds of neurons would probably have some sense of these things. But for the many readers not directly involved in collecting the data, there is no way of knowing these unreported aspects of V1 physiology. It is possible that someone may eventually come up with a theory that could account for some of these unresponsive neurons, but this cannot happen if no one knows they are there.
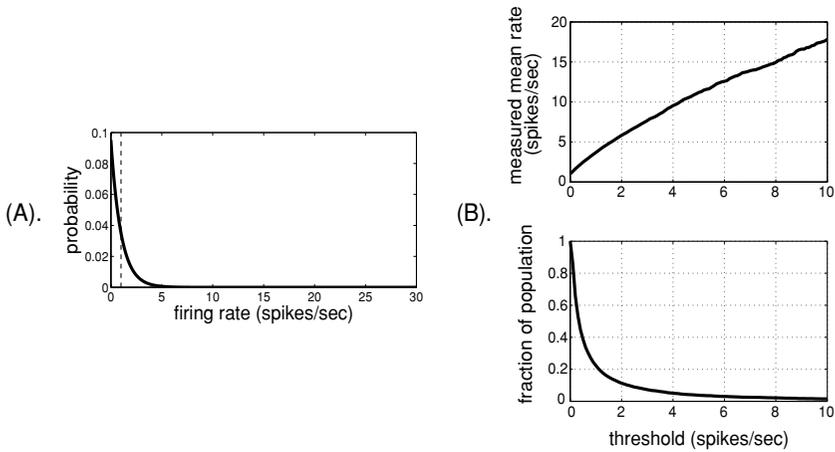
Figure 2: (A) Exponential firing rate distribution with a mean of 1 Hz (dashed line denotes mean). (B) Resulting overall mean rate of the measured population (top) and fraction of the population captured (bottom) as a result of recording from neurons only above a given mean firing rate (threshold). A log-normal distribution of mean firing rates was assumed, with rate $r = 10^u$, $u \sim N(-0.55, 0.5)$.

A related bias that arises in sampling neurons is that the process of hunting for neurons with a single microelectrode will typically steer one toward neurons with higher firing rates. One line of evidence suggesting that this is a significant bias comes from work estimating mean firing rates in the cortex based on energy consumption. Attwell and Laughlin (2001) and Lennie (2003b) calculate that the average activity must be relatively low—less than 1 Hz in primate cortex. However, in the single-unit litera-ture, one finds many studies in which even the spontaneous or background rates are well above 1 Hz. This suggests that the more active neurons are substantially overrepresented (Lennie, 2003b). What makes matters worse is that if we assume V1 neurons exhibit a roughly exponential firing rate distribution, as has been demonstrated for natural scenes and other stimuli (Baddeley et al., 1997), then a mean firing rate of 1 Hz would yield the distribution shown in Figure 2A.[2] With such a distribution, only a small fraction of neurons would exhibit the sorts of firing rates normally associated with a robust response. For example, the total prob-ability for firing rates of even 5 Hz and above is 0.007, meaning that

---

[2] Our own analysis of the firing rate distribution, as measured from the PSTH in re-sponse to repeated presentation of natural movies (measured from V1 of anaesthetized cats—J. Baker, S. C. Yen, C. M. Gray, personal communication to the authors, 2004), sug-gests that the distribution is actually power-law, which would mean that it is even more heavily skewed toward zero.

one would have to wait 1 to 2 minutes on average in order to observe a 1-second interval containing five or more spikes. It seems possible that such neurons could either be missed altogether or else purposely bypassed because they do not yield enough spikes for data analysis. For example, the overall mean firing rate of V1 neurons in the Baddeley et al. study was 4.0 Hz (SD 3.6 Hz), suggesting that these neurons constitute a subpopulation that were perhaps easier to find but not necessarily representative of the population as a whole. Interestingly, the authors point out that even this rate is considered low (which they attribute to anaesthesia), as previous studies (Legendy & Salcman, 1985) report the mean firing rate to be 8.9 Hz (SD 7.0 Hz).

Given the variety of neurons in V1, it seems reasonable to presume there exists a heterogeneous population of neurons with different mean firing rates. If we assume some distribution over these rates, then it is possible to obtain an estimate of the fraction of the population characterized given a particular criterion response. And from that, we can calculate what the observed mean rate would be for that fraction. The result of such an analysis, assuming a log-normal distribution of mean rates with an overall mean of 1 Hz, is shown in Figure 2B. As one can see, an overall mean of 4 Hz implies that the selection criterion was somewhere between 1 and 2 Hz, which would capture less than 20% of the population.

Neurophysiological studies of the hippocampus provide an interesting lesson about the sorts of biases introduced by low firing rates. Prior to the use of chronic implants, in which the activity of neurons could be monitored for extended periods while a rat explored its environment, the granule cells of the dentate gyrus were thought to be mostly high-rate "theta" cells (e.g., Rose, Diamond, & Lynch, 1983). But it eventually became clear that the majority are actually very low-rate cells (Jung & McNaughton, 1993) and that for technical reasons only high-rate interneurons were being detected in the earlier studies (W. E. Skaggs, personal communication to the authors, January 2004). In fact, Thompson and Best (1989) found that nearly two-thirds of all hippocampal neurons that showed activity under anaesthesia became silent in the awake, behaving rat. This overall pattern appears to be upheld in macaque hippocampus, where the use of chronic implants now routinely yields neurons with overall firing rates below 0.1 Hz (Barnes et al., 2003), which differs by nearly two orders of magnitude from the "low baseline rates" of 8.1 Hz reported by Wirth et al. (2003) using acutely implanted electrodes.

The dramatic turn of events afforded by the application of chronic implants combined with natural stimuli and behavior in the hippocampus can only make one wonder what mysteries could be unraveled when similar techniques are applied to visual cortex. What is the natural state of activity during free viewing of natural scenes, where the animal is actively exploring its environment? What are the actual average firing rates and other statistics of activity among layer 2/3 pyramidal cells? What are the huge numbers

of granule cells in macaque layer 4, which outnumber the geniculate fiber inputs by 30 to 1, doing? Do they provide a sparser code than their geniculate counterparts? And what about the distribution of actual receptive field sizes? Current estimates show that most parafoveal neurons in V1 have receptive field sizes on the order of 0.1 degree. But based on retinal anatomy and psychophysical performance, one would expect to find a substantial number of neurons with receptive fields an order of magnitude smaller, ca. 0.01 degree (Olshausen & Anderson, 1995). Such receptive field sizes are extremely rare, if not nonexistent, in the existing data on macaque V1 neurons collected using acute recording techniques (De Valois, Albrecht, & Thorell, 1982; Parker & Hawken, 1988).

Overall, then, one can identify at least three different biases in the sampling of neurons:

1. Preference for neurons with large cell bodies and large extracellular action potentials

2. Preference for "visually responsive" neurons

3. Preference for neurons with high firing rates

So where does this leave us? Let us be conservative. If we assume that 5% to 10% of neurons are missed because they have weak extracellular action potentials, another 5% to 10% are discarded because they are not visually unresponsive, and 50% to 60% are missed because of low firing rates (assuming a conservative threshold of 0.5 Hz in Figure 2), then even allowing for some overlap among these populations would yield the generous estimate that 40% of the population has actually been characterized (although we would not be surprised if that number is as low as 20%).

2.2 Biased Stimuli. Much of our current knowledge of V1 neural response properties is derived from experiments using reduced stimuli. Often these stimuli are ideal for characterizing linear systems—spots, white noise, or sine wave gratings—or else they are designed around preexisting notions of how neurons should respond. The hope is that the insights gained from studying neurons using these reduced stimuli will generalize to more complex situations, such as natural scenes. But of course there is no guarantee that this is the case. And given the nonlinearities inherent in neural responses, we have every reason to be skeptical.

Sine wave gratings are ubiquitous tools in visual system neurophysiology and psychophysics. In fact, the demand for using these stimuli is so high that some companies produce lab equipment with specialized routines designed for this purpose (e.g., Cambridge Research Systems). But sine waves are special only because they are eigenfunctions of linear, time- or space-invariant systems. For nonlinear systems, they bear no particular meaning and occupy no special status. In the auditory domain, sine waves could be justified from the standpoint that many natural sounds are produced by

oscillating membranes. However, in the visual world, there are few things that naturally oscillate either spatially or temporally. The Fourier basis set is just one of many possible basis sets, and if the system is nonlinear, no one basis set will necessarily provide a proper account of the system.

Bars of light, Gabor functions, Walsh patterns, or any other basis set will suffer from similar problems requiring assumptions of the types of nonlinearities that are present. The Gabor function has been argued to provide a good model of cortical receptive fields (Field & Tolhurst, 1986; Jones & Palmer, 1987). However, the methods used to measure the receptive field in the first place generally search for the best-fitting linear model. They are not tests of how well the receptive field model actually describes the response of the neuron. Not until recent work by Gallant and colleagues (David, Vinje, & Gallant, 2004) have these models been tested in ecological conditions. And as we discuss below, the results demonstrate that these models often fail to adequately capture the actual behavior of neurons.

The use of white noise and m-sequences can provide some advantage over the traditional linear systems approach, as they can provide a wider range of stimuli than a simple basis set and are thus capable of mapping out the nonlinearity of a system if the nonlinearities take on particular forms (e.g., Nykamp & Ringach, 2002). In addition, by analyzing the eigenvectors of the spike-triggered covariance matrix, one can recover fairly complex nonlinear models, such as the hypothetical subunits composing a complex cell, or suppressive dimensions in the stimulus space (Touryan, Lau, & Dan, 2002; Rust, Schwartz, Movshon, & Simoncelli, 2004).

However, there is only one way to map a nonlinear system with complete confidence: present the neuron with all possible stimuli. The scope of this task is truly breathtaking. Even an 8×8 pixel patch with 6 bits of gray level requires searching $2^{384} > 10^{100}$ possible combinations (a google of combinations). If we allow for temporal sensitivity and include a sequence of 10 such patches, we are exceeding 10^{1000}. With the estimated number of particles in the universe estimated to be in the range of 10^{80}, it should be clear that this is far beyond what any experimental method could explore. In theory, a nonlinear neuron could behave quite rationally for all but a handful of these stimuli, so unless this handful has been measured, there is no way to be certain the neuron has been adequately characterized. The use of independent white noise can theoretically present a neuron with all possible stimuli. However, 10 hours of recording from a single neuron with a patch like that above at 30 frames per second will present just 10^6 out of the 10^{1000} possible stimuli. Using such a tiny fraction of the possible stimuli allows mapping of the nonlinearities only if the nonlinearities are quite smooth.

The deeper question is whether one can predict the responses of neurons from some combinatorial rule of the responses derived from a reduced set of stimuli. The response of the system to any reduced set of stimuli cannot be guaranteed to provide the information needed to predict the response to an arbitrary combination of those stimuli. Of course, we will never know

this until it is tested, and that is precisely the problem: the central assumption of the elementwise, reductionist approach has yet to be thoroughly tested.

We believe that the solution to these problems is to turn to natural scenes. Our intuitions for how to reduce stimuli should be guided by the sorts of structure that occur in natural scenes, not arbitrary (or even elegant) mathematical functions or stimuli that are conceptually simple or happen to be easy to generate on a monitor. Since it is impossible to map out the response to all possible stimuli, some assumptions about the nature of the nonlinearity and the stimulus space must be made. The assumption we believe is appropriate is that the nonlinearities relevant to visual processing are most likely to be revealed when the system is presented with ecologically relevant stimuli.

Traditionally, experimentalists have been reluctant to use natural scenes as stimuli because they seem highly variable and uncontrolled. But in recent years there has been significant progress in modeling the structure of natural images (Simoncelli & Olshausen, 2001), and already a number of studies have used some of the basic properties of natural scenes ($1/f^2$ power spectrum, contrast distributions, texture statistics, etc.) to develop parametric descriptions of natural images that can be used to generate experimental stimuli (e.g., Knill, Field, & Kersten, 1990; Heeger & Bergen, 1995). In addition, there have been some recent attempts to map out the nonlinearities in response to natural images (Sharpee, Rust, & Bialek, 2004). And the development of several adaptive stimulus techniques looks to be a promising avenue for determining the relevant stimulus for sensory neurons (Foldiak, Xiao, Keysers, Edwards, & Perrett, 2004; Edin, Machens, Schutze, & Herz, 2004; O'Connor, Petkov, & Sutter, 2004).

In summary, then, there are two main reasons for using natural scenes as stimuli: (1) by devoting resources to relevant ecological stimuli, the experimentalist has a greater chance of finding and mapping the nonlinearities relevant to the function of neurons, and (2) the responses to natural scenes provide an ecologically meaningful test of any neural model. Even if nonecological stimuli are used to map a neuron's behavior, the true test that the characterization is correct is to demonstrate that one can predict the neurons behavior in ecological conditions.

2.3 Biased Theories. Currently in neuroscience, there is an emphasis on telling a simple story. This often encourages investigators to demonstrate when a theory explains data, not when a theory provides a poor model. In addition, editorial pressures can encourage one to make a tidy picture out of data that may actually be quite messy. This, of course, runs the risk of forcing a picture that does not actually exist. Theories then emerge that are centered around explaining a particular subset of published data, or which can be conveniently proven rather than being motivated by functional considerations: How does this help the brain to solve the real problems of vision?

For instance, early work demonstrating the spatial frequency selectivity of neurons (e.g., Blakemore & Campbell, 1969) led a number of investigators

toward a Fourier view of the cortex. Such work led to thousands of studies devoted to questions regarding frequency tuning and the relevance of this tuning to the human detection and discrimination of sinusoidal gratings. This left us with complex theories for how we detect gratings, but with little understanding of how such a system would function in the natural world.

Another example is the classification of V1 neurons into the categories of simple, complex, and hypercomplex or end-stopped. Simple cells are noted for having oriented receptive fields organized into explicit excitatory and inhibitory subfields, whereas complex cells are tuned for orientation but are relatively insensitive to position and the sign of contrast (black-white edge versus white-black edge). Hypercomplex cells display more complex shape selectivity, and some appear most responsive to short bars or the terminations of bars of light (so-called end-stopping). Are these categories real, or a result of the particular way neurons were stimulated and the data analyzed?

A widely accepted theory that accounts for the distinction between simple and complex cells is that simple cells compute a (mostly linear) weighted sum of image pixels, whereas complex cells compute a sum of the squared and half-rectified outputs of simple cells of the same orientation—the so-called energy model (Adelson & Bergen, 1985). This theory is consistent with measurements of response modulation in response to drifting sine wave gratings, otherwise known as the F1/F0 ratio (Skottun et al., 1991). From this measure, one finds clear evidence for a bimodal distribution of neurons, with simple cells having ratios greater than one and complex cells having ratios less than one. Recently, however, it has been argued that this particular nonlinear measure tends to exaggerate or even introduce bimodality rather than reflecting an actual intrinsic property of the data (Mechler & Ringach, 2002). When receptive fields are instead characterized by the degree of overlap between zones activated by increments or decrements in contrast, one obtains a continuous, unimodal distribution when the overlap is expressed as the normalized distance between the zones, but a bimodal distribution when expressed as an overlap index (sum of widths minus the separation divided by sum of widths plus the separation) (Mata & Ringach, 2005; Kagan, Gur, & Snodderly, 2002). In addition, the energy model of complex cells does a poor job accounting for complex cells with a partial overlap of activating zones. Thus, the way in which response properties are characterized can have a profound effect on the resulting theoretical framework that is adopted to explain the results. The notion of two classes of neurons, simple and complex, has been firmly planted in the minds of modelers and experimentalists alike, but a closer examination of the data reveals that this classification scheme may actually be an artifact of the lens through which we view the data.

The notion of end-stopped neurons introduces even more questions when one considers the structure of natural images. Most natural scenes are not littered with line terminations or short bars (see Figure 3, middle). Indeed, at the scale of a V1 receptive field, the structures in this image are
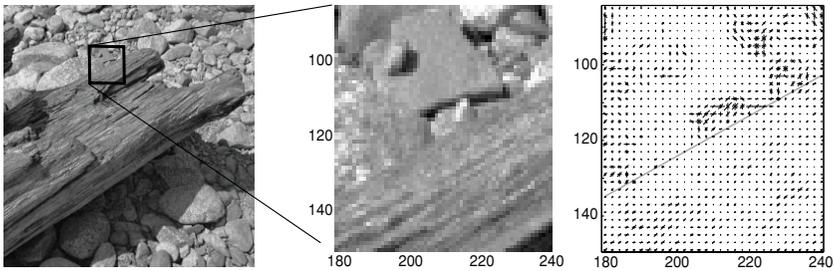
Figure 3: A natural scene (left) and an expanded section of it (middle). Far right shows the information conveyed by an array of complex cells at four different orientations. The length of each line indicates the strength of response of a complex cell at that location and orientation. The solid black line shows the location of the boundary of the log in the original image. Note that very few complex cells of the appropriate orientation are responding along this contour.

quite complex and defy the simple line-drawing-like characterization of a "blocks world." Where in such an image would one expect an end-stopped neuron to fire? By asking this question, one could possibly be led to a more ecologically relevant theory of these neurons than suggested by simple laboratory stimuli.

Another theory bias often embedded in investigations of V1 function is the notion that simple cells, complex cells, and hypercomplex cells are actually coding for the presence of edges, corners, or other two-dimensional (2D) shape features in images. However, much of this thinking is derived from a rather cartoon view of images. Computer vision studies provide clear evidence of the fallacy of the purely bottom-up approach. One cannot compute the presence even of simple edges of an object purely from the luminance discontinuities (i.e., using a filter such as a simple or complex cell model). As an example, Figure 3 demonstrates the result of processing a natural scene with the standard energy model of a complex cell. Far from making contours explicit, this representation creates a cluttered array of orientation signals that make it difficult to discern what is actually going on in the scene. Our perception of crisp contours, corners, and junctions in images is largely a post hoc phenomenon that is the result of massive inferential computations performed by the cortex, which are heavily informed by context and high-level knowledge. It could well be that our initial introspections about scene structure are a poor guide as to the actual problems faced by the cortex.

In order to properly understand V1 function, our theories will need to be guided by functional considerations and an appreciation for the ambiguities contained in natural images rather than being biased by simplistic notions of feature detection that are suggested by the responses of a select population

of neurons recorded using simplified stimuli. One of the most challenging problems facing the cortex is that of inferring a representation of 3D surfaces from the 2D image (Nakayama, He, & Shimojo, 1995; see also section 3.4). This is not an easy problem to solve and still lies beyond the abilities of modern computer vision. It seems quite likely that V1 plays a role in solving this problem, but understanding how it does so will require going beyond bottom-up filtering models to consider how top-down information is used in the interpretation of images (Olshausen, 2003; Lee & Mumford, 2003; see also section 3.5 below).

2.4 Interdependence and Contextual Effects. It has been estimated that roughly 5% of the excitatory input in layer 4 of V1 arises from the lateral geniculate nucleus (LGN), with the majority resulting from intracortical inputs (Peters & Payne, 1993; Peters, Payne, & Budd, 1994). Thalamocortical synapses have been found to be stronger, making them more likely to be effective physiologically (Ahmed, Anderson, Douglas, Martin, & Nelson, 1994). Nevertheless, based on visually evoked membrane potentials, Chung and Ferster (1998) have argued that the geniculate input is responsible for just 35% of a layer 4 neuron's response. This leaves 65% of the response determined by factors outside the direct feedforward input. Using optical imaging methods, Arieli, Sterkin, Grinvald, and Aertsen (1996) showed that the ongoing population activity can account for 80% of an individual V1 neuron's response variance, and recent work using multielectrode arrays has shown that the ongoing activity V1 neurons is only slightly modified by visual input (Fiser, Chiu, & Weliky, 2004). Thus, we are left with the real possibility that somewhere between 60% and 80% of the response of a V1 neuron is a function of other V1 neurons, or inputs other than those arising from LGN.

It should also be noted that recent evidence from the early blind has demonstrated that primary visual cortex has the potential for a wide range of multimodal input. Sadato et al. (1996) and Amedi, Raz, Pianka, Malach, and Zohary (2003) demonstrated that both tactile braille reading and verbal material can activate visual cortex in those who have been blind from an early age, even though no such activation occurs in those with normal sight. This implies that in the normal visual system, primary visual cortex has the potential for interactions with quite high-level sources of information.

That V1 neurons are influenced by context—the spatiotemporal structure outside the classical receptive field (CRF)—is by now well known and has been the subject of many investigations over the past decade. Knierim and Van Essen (1992) showed that many V1 neurons are suppressed by a field of oriented bars outside the classical receptive field of the same orientation, and Sillito, Grieve, Jones, Cudeiro, and Davis (1995) have shown that one can introduce quite dramatic changes in orientation tuning based on the orientation of gratings outside the CRF. Other investigators have probed the spatial specificity of the surround using grating patches and demonstrated fairly specific zones of suppression (Walker, Ohzawa, & Freeman, 1999; Cavanagh,

Bair, & Movshan, 2002b). And these studies, in addition to others (see Series, Lorenceau, & Frégnac, 2003, for a review), have likely tapped only a portion of the interdependencies and contextual effects that actually exist.

The problem in teasing apart contextual effects in such a piecemeal fashion is that one faces a combinatorial explosion in the number of possible spatial and featural configurations of surrounding stimuli such as bars or gratings. What we really want to know is how neurons respond within the sorts of context encountered in natural scenes. For example, given the results of Knierim and Van Essen (1992) using bar stimuli, or Sillito et al. (1995) using gratings, what should we reasonably expect to result from the sorts of context seen in the natural scene of Figure 3? Indeed, it is not even clear whether one can answer the question since the contextual structure here is so much richer and more diverse than that which has been explored experimentally. Some of the initial studies exploring the role of context in natural scenes have demonstrated pronounced nonlinear effects that tend to sparsify activity in a way that would have been hard to predict from the existing reductionist studies (Vinje & Gallant, 2000). More studies along these lines are needed, and most important, we need to understand how and why the context in natural scenes produces such effects.

Another striking form of interdependence exhibited by V1 neurons is in the synchrony of activity. Indeed, the fact that one can even measure large-scale signals such as the local field potential or electroencephalogram (EEG) implies that large numbers of neurons must be acting together. Gray, König, Engel, and Singer (1989) demonstrated gamma band synchronization between neurons in cat V1 when bars moved through their receptive fields in similar directions, suggesting that synchrony is connected to a binding or segmentation process. More recently, Wörgötter et al. (1998) have shown that receptive field sizes change significantly with the degree of synchrony exhibited in the EEG, and Maldonado, Babul, Singer, Rodriguez, and Grün (2004) have shown that periods of synchronization preferentially occur during periods of fixation as opposed to during saccades or drifts. However, what role synchrony plays in the normal operation of V1 neurons is entirely unclear, and it is fair to say that this aspect of response variance remains a mystery.

2.5 Ecological Deviance. We have argued above for experiments that measure the responses of neurons in ecological conditions even when no model is capable of predicting the results—or, we should say, especially if no model can predict the results. Publishing findings only in conditions when a particular model works would be poor science. It is important to know not only where the current models can successfully predict neural behavior, but also under what conditions they break down and why. And as we have emphasized above, it is most important to know how they fare under ecological conditions. If the current models fail to predict neural responses under such conditions, then the literature should reflect this.

In the past few years, a number of labs have begun using natural scenes as stimuli when recording from neurons in the visual pathway (Dan, Atick, & Reid, 1996; Baddeley et al., 1996; Keysers, Xiao, Foldiak, & Perrett, 2001; Vinje & Gallant, 2002; Ringach, Hawken, & Shapley, 2002; Smyth, Willmore, Baker, Thompson, & Tolhurst, 2003; David et al., 2004). In particular, the Gallant lab at UC Berkeley has taken the approach of attempting to determine how well one can predict the responses of V1 neurons to natural stimuli using a variety of different models. However, assessing how well these models fare, and what it implies about our current understanding of V1, is difficult for at least three reasons.

First, one must make several assumptions (either implicitly or explicitly) regarding what aspects of the response are relevant to the model. Spike counts will show significant variability over repeated trials (Tolhurst, Movshon, & Dean, 1983). One can take the average over a number of presentations, but this implicitly assumes that the variability can be attributed to noise. This can be questioned, especially considering that in many cases, individual spikes have been shown to have relatively high reliability (Rieke, Warland, de Ruyter van Stevenick, & Bialek, 1997). The trial-to-trial variability could well be due to internally generated dynamics that plays an important role in information processing that we simply do not as yet understand (Arieli et al., 1996; Fiser et al., 2004; see also section 3.1). Furthermore, to take averages, one must make assumptions regarding the temporal window over which the average is computed.

Second, these studies are best performed with an awake, behaving animal. In such conditions, there are limitations to the spatial and temporal accuracy with which the gaze can be measured. When averaging across presentations of stimuli, one must ascertain whether or not the same stimulus was actually presented. Again, one must make an assumption as to what spatiotemporal window to use.

The third problem is that whatever model is chosen, one is always subject to the criticism that the model is not sufficiently elaborate. Thus, any inability to predict the neuron's response might be argued to be simply due to some missing element in the model.

For example, David et al. (2004) have explored two different types of models: a linearized spatiotemporal receptive field model, in which the neuron's response is essentially a weighted sum of the image pixels over space and time, and a phase-separated Fourier model, which allows one to capture the phase invariance nonlinearity of a complex cell. These models can typically explain between 20% and 40% of the response variance. Correcting for intertrial variability improves matters somewhat and it is possible that with more trials and the addition of other nonlinearities such as contrast normalization, adaptation, and response saturation, the fraction of variance explained could rise even more above these levels (and this is a current direction of these studies).

(A).

(B).
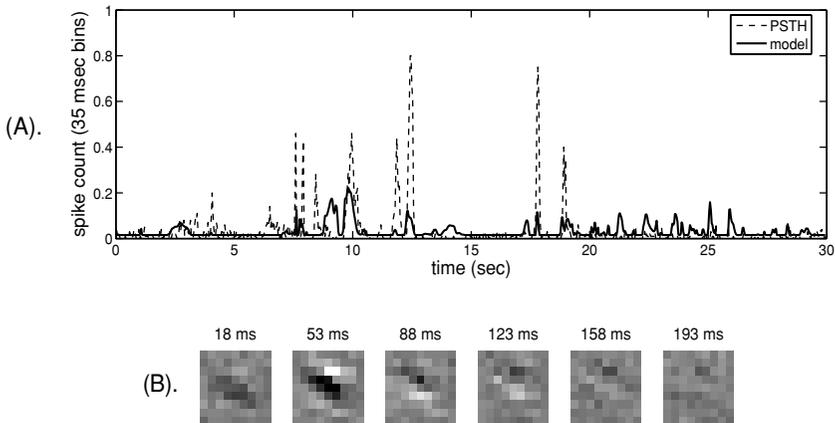
Figure 4: Activity of a V1 neuron in anesthetized cat in response to a natural movie. (A) The PSTH of the neuron's response (dashed line), together with the predicted response (solid line) generated from the model: $r(t) = \alpha\, h(\sum_{\mathbf{x}} k(\mathbf{x}, t) * I(\mathbf{x}, t) + \theta)^p + r_0$. The function $h(\,)$ is a half-wave rectifying function, and the parameters α, p, θ, and r_0 are fit to minimize the squared error with the data. The resulting correlation coefficient in this case is 0.36. Average spike counts were obtained by averaging across 100 trials in 35 ms bins (corresponding to the frame rate). (B) The kernel $k(\mathbf{x}, t)$ was measured via reverse correlation with an m-sequence and is shown here as a series of frames in 35 ms intervals, with the center time of the interval displayed above each frame.

We believe such reports are critically important for several reasons. First, such results create a benchmark for showing how well the standard or basic models actually predict ecologically relevant data. Second, these are well-established models that have been given a fair run for their money. One could imagine any number of improvements to these models, and it will be interesting to see if they fare better, but in the meantime, these results provide a useful baseline for comparison. Furthermore, these are the data that represent the ultimate goal of any computational model, and so they are crucial to presenting a complete picture of V1 function. Given the nature of the errors, we do not believe that the addition of simple response nonlinearities such as contrast normalization is likely to improve matters much. Given these results with both linear and Fourier power models, our conjecture is that the best-case scenario is that the percentage of variance explained is likely to asymptote at 30% to 40% with the standard model.

One of the reasons for our pessimism is due to the way in which these models fail. For example, Figure 4 shows data collected from the laboratory of Charles Gray at Montana State University, Bozeman, in which the activities of V1 neurons in anesthetized cat are recorded in response to repeated presentations of a natural movie (C. M. Gray, J. Baker, & S. C. Yen personal

communication to the authors, 2004). Figure 4A shows the peristimulus time histogram (PSTH) of a typical V1 simple cell, whose receptive field as measured from an M-sequence kernel is similar to those found in the literature—a Gabor-like function that translates over time (i.e., space-time inseparable). Superimposed on this is the predicted response generated by convolving the neuron's space-time receptive field (see Figure 4B) with the movie, and putting the result through a point-wise nonlinearity (including a gain factor and offset term). The neuron tends to exhibit sparse, punctate responses, some of which are predicted by the receptive field model and others not. In most cases, the model response undershoots the PSTH, and this cannot simply be addressed by increasing the gain or narrowing the response of the model, because there are many other episodes where the model predicts responses of equal magnitude in which there is little or no response from the neuron. One could possibly obtain a better fit to the data by including additional terms modeling suppression (Rust et al., 2004) and temporal adaptation (Lesica, Boloori, & Stanley, 2003), or even a spiking mechanism (Paninski, Pillow, & Simoncelli, 2004), but we believe it is useful to see how much the linear, driving term of the model alone fares under these circumstances. Moreover, these additions are essentially single-neuron mechanisms. What seems to be suggested by our initial informal observations of multiple simultaneously recorded units is that a more complex network nonlinearity is at work here, and that describing any one neuron's behavior will require including the influence of other simultaneously recorded neurons.

An important lesson of these findings is that simply mapping out receptive fields does not provide a complete understanding of V1 response properties. For example, Ringach et al. (2002) have shown that it is possible to map out receptive fields using natural scenes, and they show that it is even possible to recover some nonlinear effects such as cross-orientation inhibition with this technique. However, the resulting receptive field models were not tested by comparing their predictions to the actual activity of neurons in response to natural movies. Without doing so, it is difficult to assess how well such models capture the function of the neuron.

Unfortunately, journals are often unprepared to publish results when a study demonstrates the failure of a model, unless the study also presents a competing model that works well. Part of this may seem understandable since a model might fail for a variety of reasons. However, until a benchmark is placed in the literature, it is impossible to determine how good a model actually is. And given the magnitude of the task before us, it could take years before a good model emerges. In the meantime, what would be most helpful is to accumulate a database of single-unit or multiunit data (stimuli and neural responses) that would allow modelers to test their best theory under ecological conditions.

Finally, it should be noted that better success has been obtained in using receptive field models to predict the responses of neurons to natural scenes

in the LGN (Dan et al., 1996), or the response of cortical neurons to purely static images (Smyth et al., 2003), although they are still far from making perfect predictions. This would seem to suggest that much of the difficulty in predicting responses in cortex has to do with the effects of the massive, recurrent intracortical circuitry that is engaged during natural vision.

2.6 Summary. Table 1 presents a summary of the five problems we have identified with the current view of V1 that has emerged from the data collected to date, along with some of the solutions that we have suggested could possibly help in obtaining a more complete picture of V1 function.

Given the limitations described above, is it possible to quantify how well we currently understand V1 function? We attempt to estimate this as follows:

$$[\text{Fraction understood}] = \begin{bmatrix} \text{Fraction of variance explained} \\ \text{from neurons recorded} \end{bmatrix}$$
$$\times\ [\text{Fraction of population recorded}].$$

If we consider that roughly 40% of the population of neurons in V1 has actually been recorded from and characterized, together with our conjecture that 30% to 40% of the response variance of these neurons can be explained under natural conditions using the currently established models, then we are left to conclude that we can currently account for 12% to 16% of V1 function. Thus, approximately 85% of V1 function has yet to be explained (see Figure 5).[3]

3 New theories

Given the above observations, it becomes clear that there is so much unexplored territory that it is very difficult to rule out theories at this point (although there are some obvious bounds dictated by neural architecture, such as fan-in/fan-out and the spatial extent of axonal and dendritic arbors). In the sections below, we discuss some of the theories that are plausible given our current data. The goal here is not to provide a detailed review of the theories currently in the literature. Rather, it is to provide a few examples of the range of theories that are consistent with the experimental data. It must be emphasized that considering that there may exist a large family of neurons with unknown properties and given the low level of prediction for the neurons studied, there is still considerable room for theories dramatically different from those theories presented here.

---

[3] We have primarily drawn on the Gallant lab's data for obtaining the percentage of variance explained, and so we are assuming that their methods for isolating neurons are subject to the same biases in sampling discussed earlier.

Table 1: Five Problems with the Current view of V1 and Some Possible Solutions for Obtaining a More Complete Picture.

	Biased Sampling	Biased Stimuli	Biased Theories	Interdependence and Context	Ecological Deviance
Problem	Large neurons; visually responsive neurons; neurons with high firing rates	Use of reduced stimuli such as bars, spots, and gratings	Simple/complex cells; data-driven theories	Influence of intracortical input; effect of context; synchrony	Responses to natural scenes deviate from predictions of standard models
Solution	Use chronically implanted electrodes; parallel recording arrays	Use natural scenes, ecologically relevant stimuli	Consider more functional/computational theories that solve problems of vision	Examine how context affects responses in natural scenes	Develop models that can account for responses to natural images
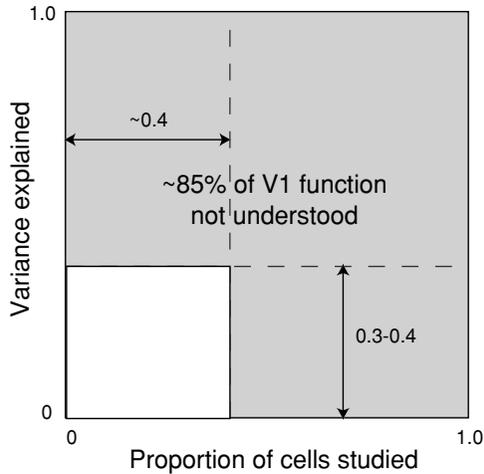
Figure 5: 85% of V1 function remains to be understood.

3.1 Dynamical Systems and the Limits of Prediction. Imagine tracking a single molecule within a hot gas as it interacts with the surrounding molecules. The particular trajectory of one molecule will be erratic and fundamentally unpredictable without knowledge of all other molecules with potential influence. Even if we presumed that the trajectory of the particular molecule was completely deterministic and following simple laws, in a gas with large numbers of interacting molecules, one could never provide a prediction of the path of a single molecule except over very short distances.

In theory, the behavior of single neurons may have similar limitations. To make predictions of what a single neuron will do in the presence of a natural scene may be fundamentally impossible without knowledge of the surrounding neurons. The nonlinear dynamics of interacting neurons may put bounds on how accurately the behavior of any neuron can be predicted. And at this time, we cannot say where that limit may be.

What is fascinating in many ways, then, is that neurons are as predictable as they are. For example, work from the Gallant lab has shown that under conditions where a particular natural scene sequence is repeated to a fixating macaque monkey, a neuron's response from trial to trial is fairly reliable (e.g., Vinje & Gallant, 2000). This clearly suggests that the response is dependent in large part on the stimulus, certainly more than a molecule in the gas model. So how do we treat the variability that is not explained by the stimulus? We may find that the reliability of a local group of neurons is more predictable than a single neuron, which would then require multielectrode recording to attempt to account for the remaining variance. For example, Arieli et al. (1996) have shown that much of the intertrial variability may be explained in terms of large-scale fluctuations in ongoing activity

of the surrounding population of neurons measured using optical recording, and Fiser et al. (2004) have similarly shown that ongoing population activity as measured with multielectrode arrays is only loosely modulated by visual input. However, what role these large-scale fluctuations play in the normal processing of natural scenes has yet to be investigated.

3.2 Sparse, Overcomplete Representations. One effort to explain many of the nonlinearities found in V1 is based on the idea that neurons are attempting to achieve some degree of gain control (Geisler & Albrecht 1992). Because any single neuron lacks the dynamic range to handle the range of contrasts in natural scenes, it is argued, the contrast response must be normalized. Here we provide a different line of reasoning to explain the observed response nonlinearities of V1 neurons (further details are provided by Olshausen & Field, 1997, and Field & Wu, 2004). We argue that the spatial nonlinearities serve primarily to reduce the linear dependencies that exist in an overcomplete code, and as we shall see, this leads to a fundamentally different set of predictions about the population activity.

Consider the number of vectors needed to represent a particular set of data with dimensionality D (e.g., an 8×8 pixel image patch would have $D = 64$). No matter what form the data take, such data never require more than D linearly independent vectors to represent it. A system where data with dimensionality D are spanned by D vectors is described as critically sampled. Such critically sampled systems (e.g., orthonormal bases) are popular in the image coding community as they allow any input pattern to be represented uniquely, and the transform and its inverse are easily computed. The wavelet code, for example, has seen widespread use, and wavelet-like codes similar to that of the visual system have been shown to provide very high efficiency in terms of sparsity when coding natural scenes (e.g., Field, 1987). Some basic versions of independent component analysis (ICA) also attempt to find a critically sampled basis that minimizes the dependencies among the vectors, and the result is a wavelet-like code with tuning much like the neurons in V1 (Bell & Sejnowski, 1997; van Hateren & van der Schaaf, 1998).

However, the visual system is not using a critically sampled code. In cat V1, for example, there are 25 times as many output fibers as there are input fibers from the LGN, and in macaque V1, the ratio is on the order of 50 to 1. Such overcomplete codes have one potential problem: the vectors are not linearly independent. Thus, if neurons were to compute their output simply from the inner product between their weight vector and the input, their responses will be correlated.

Figure 6A shows an example of a two-dimensional data space represented by three neurons with linearly dependent weight vectors. Even assuming the outputs of these units are half-rectified so they produce only positive values, the data are redundantly represented by such a code. The only way to remove this linear dependence is through a nonlinear transform.
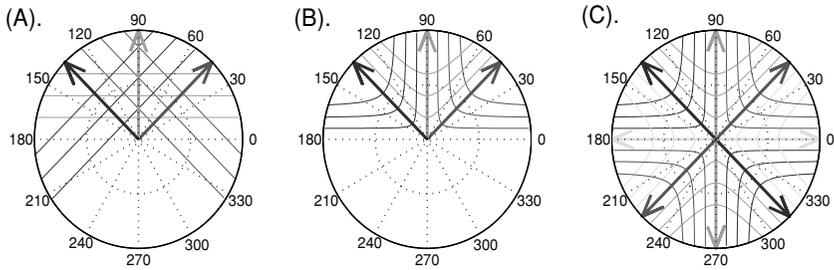
Figure 6: Overcomplete representation. (A) The iso-response contours of three linear neurons (with half-wave rectification) having linearly dependent weight vectors. A stimulus falling anywhere along a given contour will result in the same response from the neuron. A stimulus falling in the upper half-plane will result in responses on all three neurons, even though only two would be required to uniquely determine its position in the space. (B) Curving the response contours removes redundancy among these neurons. Now only two neurons will code for a stimulus anywhere in this space. (C) A full tiling of the 2D stimulus space now requires eight neurons, which would be overcomplete as a linear code, but critically sampled given this form of nonlinear response.

One of the nonlinear transforms that will serve this goal is shown in Figure 6B. Here, we show the iso-response curves for the same three neurons. This curvature represents an unusual nonlinearity. For example, consider the responses of a unit to two different stimuli: the first stimulus aligned with the neuron's weight vector and a second stimulus separated by 90 degrees. The second stimulus will have no effect on the neuron on its own since its vector is orthogonal to that of the neuron. However, when added to the first vector, the combined stimulus will be on a lower iso-response curve (i.e., the neuron will have reduced its activity). In other words, the response curvature of the neuron results in a nonlinearity with the characteristic non-classical, suppressive behavior: stimuli that on their own have no effect on the neuron (stimuli orthogonal to the principal direction of the neuron) can modulate the behavior of an active neuron. This general nonlinearity comes in several forms and includes end-stopping and cross-orientation inhibition, and is what is typically meant by the term *nonclassical surround*. Indeed, as Zetzsche, Krieger, and Wegmann (1999) note, this curvature is simply a geometrical interpretation of such behaviors. With the addition of a compressive nonlinearity, this curvature results in the behavior described as contrast normalization.

In contrast to the gain control or divisive normalization theory, we argue that the nonlinearities observed in V1 neurons are present primarily to allow a large (overcomplete) population of neurons to represent data using a small number of active units, a process we refer to as sparsification. The goal is not to develop complete independence, as the activity of any neuron

partially predicts the lack of activity in neighboring neurons. However, the code allows for expanding the dimensionality of the representation without incurring the linear dependencies that would be present in a nonorthogonal code.

Importantly, this model predicts that the nonlinearities are a function of the angle between the neuron's weight vector and those surrounding it. Future multielectrode recordings may provide the possibility to test this theory. From the computational end, we have found that our sparse coding network (Olshausen & Field, 1996, 1997) produces nonlinearities much like those proposed. Our hope, then, is that many of the nonlinearities that have been observed in V1 can eventually be explained within one general framework of efficient coding.

3.3 Contour Integration. There is now considerable physiological and anatomical evidence showing that V1 neurons have a rather selective connection pattern both within and between layers. For example, research investigating the lateral projections of pyramidal neurons in V1 has shown that the long-range lateral connections project primarily to regions of the cortex with similar orientation columns, as well as to similar ocular dominance columns and cytochrome oxidase blobs (Malach, Amir, Harel, & Grinvald, 1993; Yoshioka, Blasdel, Levitt, & Lund, 1996). Early studies exploring the horizontal connections in V1 discovered that selective long-range connections extend laterally for 2 to 5 mm parallel to the surface (Gilbert & Wiesel, 1979), and studies on the tree shrew (Rockland & Lund, 1983; Bosking, Zhang, Schofield, & Fitzpatrick, 1997), primate (e.g., Malach et al., 1993; Sincich & Blasdel, 2001), ferret (Ruthazer & Stryker, 1996), and cat (e.g., Gilbert & Weisel, 1989) have all demonstrated significant specificity in the projection of these lateral connections. A number of neurophysiological studies also show that colinearly oriented stimuli presented outside the classical receptive field have a facilitatory effect (Kapadia, Ito, Gilbert, & Westheimer, 1995; Kapadia, Westheimer, & Gilbert, 2000; Polat, Mizobe, Pettet, Kasamatsu, & Norcia, 1998). The results demonstrate that when a neuron is presented with an oriented stimulus within its receptive field, a second collinear stimulus will sometimes increase the response rate of the neuron while the same oriented stimulus presented orthogonal to the main axis of orientation (displaced laterally) will produce inhibition, or at least less facilitation.

These results suggest that V1 neurons have an orientation- and position-specific connectivity structure beyond what is usually included in the standard model. One line of research suggests that this connectivity helps resolve the ambiguity of contours in scenes and is involved in the process of contour integration (e.g., Field, Hayes, & Hess, 1993). This follows from work showing that the amplification of locally coaligned, oriented elements provides an effective means of identifying contours in natural scenes (Parent & Zucker, 1989; Sha'ashua & Ullman, 1988; Ben-Shahar & Zucker, 2004). This

type of mechanism could work in concert with the sparsification nonlinearities mentioned above, since the facilitatory interactions would primarily occur among elements that are nonoverlapping—that is, receptive fields whose weight vectors are orthogonal.

An alternative theoretical perspective is that the effect of these orientation- and position-specific connections should be mainly suppressive, with the goal of removing dependencies among neurons that arise due to the structure in natural images (Schwartz & Simoncelli, 2001). In contrast to the contour integration hypothesis, which proposes that the role of horizontal connections is to amplify the structure of contours, this model would attempt to attenuate the presence of such structure in the V1 representation. Although this may be a desirable outcome in terms of redundancy reduction, we would argue that the cortex has objectives other than redundancy reduction per se (Barlow, 2001). Chief among these is to provide a meaningful representation of image structure that can be easily read out and interpreted by higher-level areas.

Finally, it is important to note, with respect to the discussion in the previous section, that the type of redundancy we are talking about here is due to long-range structure in images beyond the size of a receptive field, not that which is simply due to the overlap among receptive fields. Thus, we propose that the latter should be removed via sparsification, while the former should be amplified by the long-range horizontal connections in V1.

3.4 Surface Representation. We live in a three-dimensional world, and the fundamental causes of images that are of behavioral relevance are surfaces, not two-dimensional features such as spots, bars, edges, or gratings. Moreover, we rarely see the surface of an object in its entirety. Occlusion is the rule, not the exception, in natural scenes. It thus seems quite reasonable to think that the visual cortex has evolved effective means to parse images in terms of the three-dimensional structure of the environment: surface structure, foreground-background relationships, and so forth. Indeed, there is now a strong body of psychophysical evidence showing that 3D surfaces and figure-ground relationships constitute a fundamental aspect of intermediate-level representation in the visual system (Nakayama et al., 1995; see also Figure 7).

Nevertheless, it is surprising how little V1 physiology has actually been devoted to the subject of three-dimensional surface representation. Some recent studies in extrastriate cortex have begun to yield interesting findings (Nguyenkim & DeAngelis, 2003; Zhou, Friedman, & von der Heydt, 2000; Bakin, Nakayama, & Gilbert, 2000), but V1's involvement in surface representation remains a mystery. Although many V1 neurons are disparity selective, this by itself does not tell us how surface structure is represented or how figure-ground relationships of the sort depicted in Figure 7 are resolved.
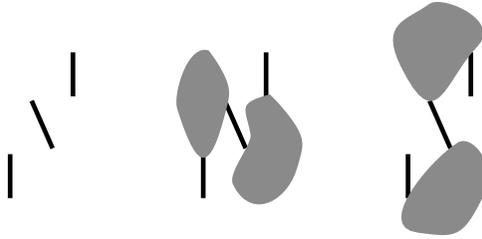
Figure 7: The three line strokes at left are interpreted as different objects depending on the arrangement of occluders. Thus, pattern completion depends on resolving figure-ground relationships. At what level of processing is this form of completion taking place? Since it would seem to demand access to high-resolution detail in the image, it cannot simply be relegated to high-level areas.

At first sight, it may seem preposterous to suppose that V1 is involved in computing three-dimensional surface representations. But again, given how little we actually know about V1, combined with the importance of 3D surface representations for guiding behavior, it is a plausible hypothesis to consider. In addition, problems such as occlusion demand resolving figure-ground relationships in a relatively high-level representation where topography is preserved (Lee & Mumford, 2003). There is now beginning to emerge physiological evidence supporting this idea. Neurons in V1 have been shown to produce a differential response to the figure versus background in a scene of texture elements (Lamme, 1995; Zipser, Lamme, & Schiller, 1996), and a substantial fraction of neurons in V1 are selective to border ownership (Zhou et al., 2000). In addition, Lee, Mumford, Romero, and Lamme (1998) have demonstrated evidence for a medial axis representation of surfaces in which V1 neurons become most active along the skeletal axis of an object. It seems quite possible that such findings are just the tip of the iceberg.

3.5 Top-Down Feedback and Disambiguation. Although our perception of the visual world is usually quite clear and unambiguous, the raw image data that we start out with is not. Looking back at Figure 3, one can see that even the presence of a simple contour can be ambiguous in a natural scene. The problem is that information at the local level is insufficient to determine whether a change in luminance is due to an object boundary, simply part of a texture, or a change in reflectance. Although boundary junctions are also quite crucial to the interpretation of a scene, a number of studies have shown that human observers are poor judges of what constitutes a boundary or junction when these features are shown in isolation (Elder, Beniaminov, & Pintilie, 1999; McDermott, 2004). Thus, the calculation of what forms a boundary is dependent on the context, which provides

information about the assignment of figure and ground, surface layout, and so forth.

Arriving at the correct interpretation of an image, then, constitutes something of a chicken- and-egg problem between lower and higher levels of image analysis. The low-level shape features that are useful for identifying an object—edges, contours, surface curvature, and the like—are typically ambiguous in natural scenes, so they cannot be computed directly based on a local analysis of the image. Rather, they must be inferred based on global context and higher-level knowledge. However, the global context itself will not be clear until there is some degree of certainty about the presence of low-level shape features. A number of theorists have thus argued that recognition depends on information circulating through cortico-cortical feedback loops in order to disambiguate representations at both lower and higher levels in parallel (Mumford, 1994; Ullman, 1995; Lewicki & Sejnowski, 1996; Rao & Ballard, 1999; Young, 2000; Lee & Mumford, 2003; Hawkins & Blakeslee, 2004).

An example of disambiguation at work in the visual cortex can be seen in the resolution of the aperture problem in computing the direction of motion. Because receptive fields limit the field of view of a neuron to just a portion of an object, it is not possible for any one neuron to signal with certainty the true direction of the object in a purely bottom-up fashion. Pack, Berezovskii, and Born (2001) have shown that the initial phase of response of neurons in MT signals the direction of motion directly orthogonal to a contour and that the latter phase of the response reflects the actual direction of the object that the contour is part of, presumably from the interaction with other neurons viewing other parts of the object. Interestingly, this effect does not occur under anesthesia. A similar delayed-response effect has been demonstrated in end-stopped V1 neurons as well (Pack, Livingstone, Duffy, & Born, 2003).

Recent evidence from fMRI points to a disambiguation process occurring in V1 during shape perception (Murray, Kersten, Olshausen, Schrater, & Woods, 2002). Subjects viewed a translating diamond that was partially occluded so that the vertices are invisible, resulting in a bistable percept in which the line segments forming the diamond are seen moving independently in one case, and coherently in the direction of the object motion in the other case. When subjects experience the coherent motion and shape percept, activity in lateral occipital cortex (LOC) increases while activity in V1 decreases. This is consistent with the idea that when neurons in LOC are representing the diamond, they feed back this information to V1 so as to refine the otherwise ambiguous representations of contour motion. If the refinement of activity attenuates the many incorrect responses while amplifying the few that are consistent with the global percept, the net effect could be a reduction as seen in the BOLD signal measured by fMRI. An alternative interpretation for the reduction in V1 is based on the idea of predictive coding (Rao & Ballard, 1999), in which higher areas actually subtract their predictions from lower areas.

There exists a rich set of feedback connections from higher levels into V1, but little is known about the computational role of these connections. Recent experiments in which higher areas are cooled to look at the effect on activity in lower areas seem to suggest that these connections play a role in enhancing the salience of stimuli (Hupe et al., 1998), and Shapley (2004) has concluded that top-down feedback is necessary to account for the spatial extent of surround inhibition. But we would argue that feedback has a far more important role to play in disambiguation, and as far as we know, no one has yet investigated the effect of feedback using such cooling techniques under normal conditions that would require disambiguation (e.g., natural scenes).

3.6 Dynamic Routing. A challenging problem faced by any visual system is that of forming object representations that are invariant to position, scale, rotation, and other common deformations of the image data. The currently accepted, traditional view is that complex cells constitute the first stage of invariant representation by summing over the outputs of simple cells whose outputs are half-rectified and squared—the classical "energy model" (Adelson & Bergen 1985). In this way, the neuron's response changes only gradually as an edge is passed over its receptive field. This idea forms the basis of so-called Pandemonium models, in which a similar feature extraction and pooling process is essentially repeated at each stage of visual cortex (see Tarr, 1999, for a review).

However, the Pandemonium model cannot provide a complete account of perception because it does not preserve information about relative phase or the spatial relationships among features. Clearly, though, we have conscious access to this information. The ability to navigate, grasp, and interact with foreign objects implies that we have the ability to perceive spatial relationships among features without ever doing "object recognition." In addition, resolving figure-ground relationships and occlusion demands that higher levels of analysis have access to information about spatial relationships as well.

One of us has proposed a model for forming invariant representations that preserves relative spatial relationships by explicitly routing information at each stage of processing (Olshausen, Anderson, & Van Essen, 1993). Rather than passively pooling, information is dynamically linked from one stage to the next by a set of control neurons that progressively remap information into an object-centered reference frame. It is thus proposed that there are two distinct classes of neurons: those conveying image and feature information and those controlling the flow of information. The former corresponds to the invariant part, the latter to the variant part. The two are combined multiplicatively, so that mathematically it is equivalent to a bilinear model (e.g., Tenenbaum & Freeman, 2000; Grimes & Rao, 2005).

Is it possible that dynamic routing occurs in V1 and underlies the observed shift-invariant properties of complex cells? If so, there are at least

two things we would expect to see: (1) that at any given moment, a complex cell is effectively connected to only one or a small fraction of simple cells to which it is physically connected, and (2) that there are control neurons that dynamically gate these connections. Interestingly, the observed invariance properties of complex cells are just as consistent with the idea of routing as they are with pooling. What could possibly distinguish between these models is to look at the population activity: if the complex cell outputs are the result of passive pooling, then one would expect a dense, distributed representation of contours among the population of complex cells. If information is dynamically routed, though, the representation at the complex cell level would remain sparse. The control neurons, on the other hand, would look something like contrast normalized simple cells, which represent phase independent of magnitude (Zetzsche & Rohrbein, 2001).

One of the main predictions of the dynamic routing model is that the receptive fields of the invariant neurons would be expected to shift depending on the state of the control neurons. Such effects have been seen in V4, where some neurons shift their receptive fields depending on where the animal is directing its attention (Moran & Desimone, 1985; Connor, Preddie, Gallant, & Van Essen, 1997. And in V1, Brad Motter has shown that neurons appear to shift their receptive fields in order to compensate for the small eye movements that occur during fixation (Motter & Poggio, 1990; Motter, 1995), although Gur and Snodderly (1997) provide evidence to the contrary. Thus, there exists some evidence for dynamic routing in visual cortex, but further experiments are needed in order to characterize how and to what extent this occurs in V1 under normal viewing conditions.

4 Conclusion

Our goal in this review has been to point out that there are still substantial gaps in our knowledge of V1 function and, more important, that there is more room for new theories to be considered than the current conventional wisdom might allow. We have identified five specific problems with the current view of V1, emphasizing the need for using natural scenes in experiments, in addition to multiunit recording methods, in order to obtain a more representative picture of V1 function. While the single-unit, structuralist approach has been a useful enterprise for getting a handle on basic response properties, we feel that its usefulness as a tool for investigating V1 function has been nearly exhausted. It is now time to dig deeper, using richer, ecologically relevant experimental paradigms, and developing theories that can help to elucidate how the cortex performs the computationally challenging problems of vision.

As we explore the response properties of V1 neurons using natural scenes, we are likely to uncover some interesting new phenomena that defy explanation with current models. It is at this point that we should be prepared to revisit the structuralist approach in order to tease apart what is going on.

Reductionism does have its place, but it needs to be motivated by functionally and ecologically relevant questions, similar to the European tradition in ethology (Tinbergen, 1972).

At what point will we actually understand V1? This is obviously a difficult question to answer, but we believe at least three ingredients are required: (1) an unbiased sample of neurons of all types, firing rates, and layers of V1; (2) the ability to observe simultaneously the activities of hundreds of neurons in a local population; and (3) the ability to predict, or at least qualitatively model, the responses of the population under natural viewing conditions. Given the extensive feedback connections into V1, in addition to the projections from pulvinar and other sources, it seems unlikely that we will ever understand V1 in isolation. Thus, our investigations must also be guided by how V1 fits into the bigger picture of thalamo-cortical function.

Acknowledgments

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America, A, 2*, 284–299.

Ahmed, B., Anderson, J. C., Douglas, R. J., Martin, K. A., & Nelson, J. C. (1994). Polyneuronal innervation of spiny stellate neurons in cat visual cortex. *J. Comp. Neurol., 341*, 39–49.

Amedi, A., Raz, N., Pianka, P., Malach, R., & Zohary, E. (2003). Early 'visual' cortex activation correlates with superior verbal memory performance in the blind. *Nat. Neurosci., 6*, 758–766.

Anderson, S. J., Burr, D. C., & Morrone, M. C. (1991). Two-dimensional spatial and spatial-frequency selectivity of motion-sensitive mechanisms in human vision. *Journal of the Optical Society of America A, 8*, 1340–1351.

Arieli, A., Sterkin, A., Grinvald, A., & Aertsen, A. (1996). Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science, 273*, 1868–1871.

Attwell, D., & Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab., 21*, 1133–1145.

Azouz, R., & Gray, C. M. (2000). Dynamic spike threshold reveals a mechanism for synaptic coincidence detection in cortical neurons in vivo. *Proc. Natl. Acad. Sci. U.S.A.*, *97*, 8110–8115.

Azouz, R., & Gray, C. M. (2003). Adaptive coincidence detection and dynamic gain control in visual cortical neurons in vivo. *Neuron*, *37*, 513–523.

Baddeley, R., Abbott, L. F., Booth, M. C. A., Sengpiel, F., Freeman, T., Wakeman, E. A., & Rolls, E. T. (1997). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. R. Soc. Lond. B*, *264*, 1775–1783.

Bakin, J. S., Nakayama, K., & Gilbert, C. D. (2000). Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. *J. Neurosci.*, *20*, 8188–8198.

Barlow, H. B. (2001). Redundancy reduction revisited. *Network: Computation in Neural Systems*, *12*, 241–253.

Barnes, C. A., Skaggs, W. E., McNaughton, B. L., Haworth, M. L., Permenter, M., Archibeque, M., & Erickson, C. A. (2003). Chronic recording of neuronal populations in the temporal lobe of awake young adult and geriatric primates. *Program No. 518.8. Abstract Viewer/Itinerary Planner*. Washington, DC: Society for Neuroscience.

Bell, A. J., & Sejnowski, T. J. (1997). The independent components of natural images are edge filters. *Vision Research*, *37*, 3327–3338.

Ben-Shahar, O., & Zucker, S. W. (2004). Geometrical computations explain projection patterns of long-range horizontal connections in visual cortex. *Neural Computation*, *16*, 445–476.

Blakemore, C., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *J. Physiol.*, *203*, 237–260.

Bosking, W. H., Zhang, Y., Schofield, B., & Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *Journal of Neuroscience*, *17*, 2112–2127.

Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience*, *17*, 8621–8644.

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002a). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *J. Neurophys.*, *88*, 2530–2546.

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002b). Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. *J. Neurophys.*, *88*, 2547–2556.

Chung, S., & Ferster, D. (1998). Strength and orientation tuning of the thalamic input to simple cells revealed by electrically evoked cortical suppression. *Neuron*, *20*, 1177–1189.

Connor, C. E., Preddie, D. G., Gallant, J. L., & Van Essen, D. C. (1997). Spatial attention effects in macaque area V4. *J. Neurosci.*, *17*, 3201–3214.

Den, Y., Atick, J. J., & Reid, R. C. (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computational theory. *Journal of Neuroscience*, *16*, 3351–3362.

David, S. V., Vinje, W. E., & Gallant, J. L. (2004). Natural stimulus statistics alter the receptive field structure of V1 neurons. *J. Neurosci.*, *24*, 6991–7006.

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

Edin, F., Machens, C. K., Schutze, H., & Herz, A. V. (2004). Searching for optimal sensory signals: Iterative stimulus reconstruction in closed-loop experiments. *J. Comput. Neurosci.*, *17*, 47–56.

Elder, J. H., Beniaminov, D., & Pintilie, G. (1999). Edge classification in natural images. *Investigative Ophthalmology and Visual Science*, *40*, S357.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, *4*, 2379–2394.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field." *Vision Research*, *33*, 173–193.

Field, D. J., & Tolhurst, D. (1986). The structure and symmetry of simple-cell receptive-field profiles in the cat's visual cortex. *Proc. R. Soc. Lond. B. Biol. Sci.*, *228*, 379–400.

Field, D. J., & Wu, M. (2004). An attempt towards a unified account of non-linearities in visual neurons. *Journal of Vision*, *4*, 283a.

Fiser, J., Chiu, C., & Weliky, M. (2004). Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature*, *431*, 573–578.

Foldiak, P., Xiao, D., Keysers, C., Edwards, R., & Perrett, D. I. (2004). Rapid serial visual presentation for the determination of neural selectivity in area STSa. *Prog. Brain Res.*, *144*, 107–116.

Geisler, W. S., & Albrecht, D. G. (1992). Cortical neurons: Isolation of contrast gain control. *Vision Research*, *32*, 1409–1410.

Geisler, W. S., & Albrecht, D. G. (1997). Visual cortex neurons in monkeys and cats: Detection, discrimination and identification. *Visual Neuroscience*, *14*, 897–919.

Gilbert, C. D., & Wiesel, T. N. (1979). Morphology and intracortical projections of functionally characterised neurones in the cat visual cortex. *Nature*, *280*, 120–125.

Gilbert, C. D., & Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *J. Neurosci.*, *9*, 2432–2442.

Graham, N., & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: A test of single-channel and multiple-channels models. *Vision Research*, *11*, 251–259.

Gray, C. M., König, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, *338*, 334–337.

Grimes, D. B., & Rao, R. P. (2005). Bilinear sparse coding for invariant vision. *Neural Comput.*, *17*, 47–73.

Gur, M., & Snodderly, D. M. (1997). Visual receptive fields of neurons in primary visual cortex (V1) move in space with the eye movements of fixation. *Vision Res.*, *37*, 257–265.

Hausser, M., & Mel, B. (2003). Dendrites: Bug or feature? *Current Opinion in Neurobiology*, *13*, 372–383.

Hawkins, J., & Blakeslee, S. (2004). *On Intelligence.* New York: Holt.

Heeger, D. J. (1991). Computational model of cat striate physiology. In M. S. Landy & A. Movshan (Eds.), *Computational models of visual perception* (pp. 119–133). Cambridge, MA: MIT Press.

Heeger, D. J., & Bergen, J. R. (1995, August). Pyramid based texture analysis/synthesis. *Computer Graphics Proceedings*, 229–238.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.*, *148*, 574–591.

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

Hupe, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, *394*, 784–787.

Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.*, *58*, 1233–1258.

Jung, M. W., & McNaughton, B. L. (1993). Spatial selectivity of unit activity in the hippocampal granular layer. *Hippocampus*, *3*, 165–182.

Kagan, I., Gur, M., & Snodderly, D. M. (2002). Spatial organization of receptive fields of V1 neurons of alert monkeys: Comparison with responses to gratings. *J. Neurophysiol.*, *88*, 2557–2574.

Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. *Neuron*, *15*, 843–856.

Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (2000). Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Journal of Neurophysiology*, *84*, 2048–2062.

Keysers, C., Xiao, D. K., Foldiak, P., & Perrett, D. I. (2001). The speed of sight. *J. Cogn. Neurosci.*, *13*, 90–101.

Knierim, J. J., & Van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophys.*, *67*, 961–980.

Knill, D. C., Field, D., & Kersten, D. (1990). Human discrimination of fractal images. *J. Opt. Soc. Am. A*, *7*, 1113–1123.

Lamme, V. A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci.*, *15*, 1605–1615.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A*, *20*, 1434–1448.

Lee, T. S., Mumford, D., Romero, R., & Lamme, V. A. (1998). The role of the primary visual cortex in higher level vision. *Vision Res.*, *38*, 2429–2454.

Legendy, C. R., & Salcman, M. (1985). Bursts and recurrences of bursts in spike trains of spontaneously active striate cortex neurons. *J. Neurophysiol.*, *53*, 926–939.

Lennie, P. (2003a). Receptive fields. *Curr. Biol.*, *13*, R216–219.

Lennie, P. (2003b). The cost of cortical computation. *Curr. Biol.*, *13*, 493–497.

Lesica, N. A., Boloori, A. S., & Stanley, G. B. (2003). Adaptive encoding in the visual pathway. *Network*, *14*, 119–135.

Lewicki, M. S., & Sejnowski, T. J. (1996). Bayesian unsupervised learning of higher order structure. In M. Mozer, M. Jordan, & T. Petsche (Eds.), *Advances in neural information processing systems*, *9*. Cambridge, MA: MIT Press.

Malach, R., Amir, Y., Harel, M., & Grinvald, A. (1993). Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex. *Proc. Natl. Acad. Sci. USA*, *90*, 10469–10473.

Maldonado, P., Babul, C., Singer, W., Rodriguez, E., & Grün, S. (2004). *Synchrony and oscillations in primary visual cortex of monkeys viewing natural images*. Manuscript submitted for publication.

Mata, M. L., & Ringach, D. L. (2005). Spatial overlap of "on" and "off" subregions and its relation to response modulation ratio in macaque primary visual cortex. *J. Neurophys.*, *93*, 919–928.

McDermott, J. (2004). Psychophysics with junctions in real images. *Perception*, *33*, 1101–1127.

Mechler, F., & Ringach, D. L. (2002). On the classification of simple and complex cells. *Vision Res.*, *42*, 1017–1033.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extra striate cortex. *Science*, *229*, 782–784.

Motter, B. C. (1995). Receptive field border stabilization during visual fixation. *Investigative Ophthalmology and Visual Science*, *36*, S691.

Motter, B. C., & Poggio, G. F. (1990). Dynamic stabilization of receptive fields of cortical neurons (VI) during fixation of gaze in the macaque. *Exp. Brain Res.*, *83*, 37–43.

Mumford, D. (1994). Neuronal architectures for pattern-theoretic problems. In C. Koch, & J. L. Davis (Eds.), *Large scale neuronal theories of the brain* (pp. 125–152). Cambridge, MA: MIT Press.

Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences, USA*, *99*(23), 15164–15169.

Nakayama, K., He, Z. J., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher level vision. In S. M. Kosslyn & D. N. Osherson (Eds.), *In an invitation to cognitive science* (pp. 1–70). Cambridge, MA: MIT Press.

Nguyenkim, J. D., & DeAngelis, G. C. (2003). Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. *J Neurosci.*, *23*(18), 7117–7128.

Nykamp, D. Q., & Ringach, D. L. (2002). Full identification of a linear-nonlinear system via cross-correlation analysis. *Journal of Vision*, *2*, 1–11.

O'Connor, K. N., Petkov, C. I., & Sutter, M. L. (2004). Stimulus optimization for auditory cortical neurons. *Society for Neuroscience Abstracts*, *529.14*.

Olshausen, B. A. (2003). Principles of image representation in visual cortex. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 1603–1615). Cambridge, MA: MIT Press.

Olshausen, B. A., & Anderson, C. H. (1995). A model of the spatial-frequency organization in primate striate cortex. In J. M. Bower (Ed.), *The neurobiology of computation: Proceedings of the Third Annual Computation and Neural Systems Conference* (pp. 275–280). Norwell, MA: Kluwer.

Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, *13*, 4700–4719.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, *37*, 3311–3325.

Pack, C. C., Berezovskii, V. K., & Born, R. T. (2001). Dynamic properties of neurons in cortical area MT in alert and anaesthetized macaque monkeys. *Nature*, *414*, 905–908.

Pack, C. C., Livingstone, M. S., Duffy, K. R., & Born, R. T. (2003). End-stopping and the aperture problem: Two-dimensional motion signals in macaque V1. *Neuron*, *39*, 671–680.

Paninski, L., Pillow, J. W., & Simoncelli, E. P. (2004). Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. *Neural Comput.*, *16*, 2533–2561.

Parent, P., & Zucker, S. (1989). Trace inference, curvature consistency and curve detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*, 823–839.

Parker, A. J., & Hawken, M. J. (1988). Two-dimensional spatial structure of receptive fields in monkey striate cortex. *Journal of the Optical Society of America A*, *5*, 598–605.

Peters, A., & Payne, B. R. (1993). Numerical relationships between geniculocortical afferents and pyramidal cell modules in cat primary visual cortex. *Cereb. Cortex*, *3*, 69–78.

Peters, A., Payne, B. R., & Budd, J. (1994). A numerical analysis of the geniculocortical input to striate cortex in the monkey. *Cereb. Cortex.*, *4*, 215–229.

Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T., & Norcia, A. M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature*, *391*, 580–584.

Polsky, A., Mel, B. W., & Schiller, J. (2004). Computational subunits in thin dendrites of pyramidal cells. *Nat. Neurosci.*, *7*, 621–627.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.*, *2*, 79–87.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: MIT Press.

Ringach, D., Hawken, M., & Shapley, R. (2002). Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. *Journal of Vision*, *2*, 12–24.

Rockland, K. S., & Lund, J. S. (1983). Intrinsic laminar lattice connections in primate visual cortex. *Journal of Comparative Neurology*, *216*, 303–318.

Rose, G., Diamond, D., & Lynch, G. S. (1983). Dentate granule cells in the rat hippocampal formation have the behavioral characteristics of theta neurons. *Brain Res.*, *266*, 29–37.

Rust, N. C., Schwartz, O., Movshon, J. A., & Simoncelli, E. P. (2004). Spike-triggered characterization of excitatory and suppressive stimulus dimensions in monkey V1. *Neurocomputing*, *58-60C*, 793–799.

Ruthazer, E. S., & Stryker, M. P. (1996). The role of activity in the development of long-range horizontal connections in area 17 of the ferret. *Journal of Neuroscience*, *16*, 7253–7269.

Sadato, N., Pascual-Leone, A., Grafman, J., Ibanez, V., Deiber, M. P., Dold, G., & Hallett, M. (1996). Activation of the primary visual cortex by braille reading in blind subjects. *Nature*, *380*, 526–528.

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nat. Neurosci.*, *4*, 819–825.

Series, P., Lorenceau, J., & Frégnac, Y. (2003). The "silent" surround of V1 receptive fields: Theory and experiments. *J. Physiol. Paris.*, *97*, 453–474.

Sha'ashua, A., & Ullman, S. (1988). Structural Saliency: The detection of globally salient structures using a locally connected network. In *Proceedings of the International Conference on Computer Vision, Tampa, Florida* (pp. 321–327). Washington, DC: IEEE Computer Society Press.

Shapley, R. (2004). A new view of the primary visual cortex. *Neural Networks*, *17*, 615–623.

Sharpee, T., Rust, N. C., & Bialek, W. (2004). Analyzing neural responses to natural signals: Maximally informative dimensions. *Neural Computation*, *16*, 223–250.

Sillito, A. M., Grieve, K. L., Jones, H. E., Cudeiro, J., & Davis, J. (1995). Visual cortical mechanisms detecting focal orientation discontinuities. *Nature*, *378*, 492–496.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annu. Rev. Neurosci.*, *24*, 1193–1216.

Sincich, L. C., & Blasdel, G. G. (2001). Oriented axon projections in primary visual cortex of the monkey. *Journal of Neuroscience*, *21*, 4416–4426.

Skottun, B. C., De Valois, R. L., Grosof, D. H., Movshon, J. A., Albrecht, D. G., & Bonds, A. B. (1991). Classifying simple and complex cells on the basis of response modulation. *Vision Res.*, *31*, 1079–1086.

Smyth, D., Willmore, B., Baker, G. E., Thompson, I. D., & Tolhurst, D. J. (2003). The receptive-field organization of simple cells in primary visual cortex of ferrets under natural scene stimulation. *J. Neurosci.*, *23*, 4746–4759.

Softky, W. R., & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci.*, *13*(1), 334–350.

Tarr, M. J. (1999). News on views: Pandemonium revisited. *Nat. Neurosci.*, *2*, 932–935.

Tenenbaum, J. B., & Freeman, W. T. (2000). Separating style and content with bilinear models. *Neural Computation*, *12*, 1247–1283.

Thompson, L. T., & Best, P. J. (1989). Place cells and silent cells in the hippocampus of freely-behaving rats. *J. Neurosci.*, *9*, 2382–2390.

Tinbergen, N. (1972). *The animal in its world: Explorations of an ethologist.* Cambridge, MA: Harvard University Press.

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.*, *23*, 775–785.

Touryan, J., Lau, B., & Dan, Y. (2002). Isolation of relevant visual features from random stimuli for cortical complex cells. *J. Neurosci.*, *22*, 10811–10818.

Ullman, S. (1995). Sequence seeking and counter streams: A computational model for bidirectional information flow in the visual cortex. *Cereb. Cortex.*, *5*, 1–11.

van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. Lond. B*, *265*, 359–366.

Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, *287*, 1273–1276.

Vinje, W. E., & Gallant, J. L. (2002). Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *J. Neurosci.*, *22*, 2904–2915.

Walker, G. A., Ohzawa, I., & Freeman, R. D. (1999). Asymmetric suppression outside the classical receptive field of the visual cortex. *Journal of Neuroscience*, *19*, 10536–10553.

Watson, A. B., Barlow, H. B., & Robson, J. G. (1983). What does the eye see best? *Nature*, *302*, 419–422.

Wirth, S., Yanike, M., Frank, L. M., Smith, A. C., Brown, E. N., & Suzuki, W. A. (2003). Single neurons in the monkey hippocampus and learning of new associations. *Science*, *300*, 1578–1581.

Wörgötter, F., Suder, K., Zhao, Y., Kerscher, N., Eysel, U. T., & Funke, K. (1998). State-dependent receptive-field restructuring in the visual cortex. *Nature*, *396*, 165–168.

Yoshioka, T., Blasdel, G. G., Levitt, J. B., & Lund, J. S. (1996). Relation between patterns of intrinsic lateral connectivity, ocular dominance, and cytochrome oxidase-reactive regions in macaque monkey striate cortex. *Cerebral Cortex*, *6*, 297–310.

Young, M. P. (2000). The architecture of visual cortex and inferential processes in vision. *Spatial Vision*, *13*, 137–146.

Zetzsche, C., Krieger, G., & Wegmann, B. (1999). The atoms of vision: Cartesian or polar? *J. Opt. Soc. Am. A*, *16*, 1554–1565.

Zetzsche, C., & Rohrbein, F. (2001). Nonlinear and extra-classical receptive field properties and the statistics of natural scenes. *Network*, *12*, 331–350.

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

Zipser, K., Lamme, V. A., & Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *J. Neurosci.*, *16*, 7376–7389.