

What is the other 85% of V1 doing?

Bruno A. Olshausen
Redwood Neuroscience Institute
and Center for Neuroscience, UC Davis

David J. Field
Department of Psychology
Cornell University

July 22, 2004

Abstract

This article will pose the following challenge: that despite four decades of research characterizing the response properties of V1 neurons, we still do not have a decent picture of how V1 really operates—i.e., how a population of its neurons represents natural scenes under realistic viewing conditions. We identify five problems with the current view that stem largely from biases in the design and execution of experiments, in addition to the contributions of non-linearities in the cortex that are not well understood. Our purpose is to open the window 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 1960's 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. 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. Given the magnitude of these efforts, one might reasonably expect that we would by now have a fairly concrete grasp of how V1 works and its role in visual system function. However, as we shall argue here, there still remains so much unknown that, for all practical purposes, we stand today in nearly the same state of ignorance as did Hubel and Wiesel 40 years ago.¹

¹We do not mean to imply here that nothing has been learned, but rather that what has been learned is but a small fraction of what lies ahead and still needs to be understood.

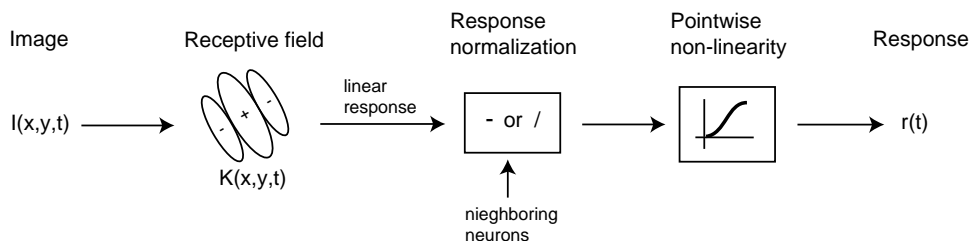


Figure 1: Standard model of V1 simple cell responses. The neuron computes a weighted sum of the image over space and time, and this result is normalized by the responses of neighboring units, and passed through a pointwise non-linearity (see e.g., Carandini et al., 1997)

It may seem surprising to some that we should take such a stance. V1 does afterall 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 upon “standard model” for V1 in which simple-cells compute a linearly weighted sum of the input over space and time (usually a Gabor-like function) and the output is passed through a pointwise nonlinearity, in addition to being subject to contrast gain control to avoid response saturation (Figure 1). Complex cells are similarly explained in terms of summing the outputs of a local pool of simple-cells with similar tuning properties but different positions or phases. The net result is to think of V1 roughly as a “Gabor filter bank.” There are now many 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)

But behind this picture of apparent orderliness, there 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—i.e., they emit all-or-nothing action potentials, not analog values. They also adapt, so their response properties depend upon the history of activity. Cortical pyramidal cells have highly elaborate dendritic trees, and realistic biophysical models suggest that each thin branch could act as a non-linear subunit, so that any one neuron could be computing many different non-linear combinations of its inputs (Hausser & Mel, 2003), in addition to being sensitive to coincidences (Softky & Koch, 1993; Azouz & Gray, 2000, 2003). Everyone knows that neurons are non-linear, but few have acknowledged the implications for studying cortical function. Unlike linear systems, where there exist mathematically tractable, textbook methods for system identification, non-linear systems can not be teased apart using some straightforward, reductionist approach. In other words, *there is no general method for characterizing non-linear systems.*²

²The Volterra series expansion is often touted as a general approach for characterizing non-linear

The reductionist approach has formed the bedrock of V1 physiology for the past four decades. Indeed, it would seem necessary given the stunning complexity of the brain to tease apart one chunk at a time. But whether or not a reductionist approach tells you anything useful depends entirely on how you reduce. Some modes of interaction may be crucial to the operation of the system, and so cutting them out—either in theories or experiments—may give a misleading picture of how the system actually works. Obviously, if one knew in advance what the important modes of interaction were then one could choose to reduce appropriately. But when it comes to the brain we really haven't a clue. V1 physiologists have for the most part chosen one particular way to reduce complexity by using highly simplified stimuli and recording from only one neuron at a time, and from this body of experiments has emerged the standard model which forms the basis for our conceptual understanding of V1. But whether or not the physiologists chose correctly is anyone's guess. The best-case scenario is that they did, and that the standard model is more or less correct. The worst-case scenario, which we lean towards, is that they chose inappropriately in many cases, that the standard model is but one small part of the full story, and that we still have much to learn about V1.

In this chapter we lay out the reasons for our skepticism by identifying five fundamental problems with the reductionist approach that have led us to the current view of V1 as a Gabor filter bank. In addition, 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 have actually been characterized and the fraction of variance explained in the responses of these neurons under natural viewing conditions. Together, these lead us to conclude that at present we can rightfully claim to understand only 10-20% of how V1 actually operates under normal conditions.

Our aim in pointing out these things is not to simply tear down the current framework. Indeed, we ourselves have attempted to account for some aspects of the standard model in terms of efficient coding principles (sparse coding), so obviously we buy into at least part of the story. We are also not claiming that reductionism in general is flawed. Rather, our goal is to make room for new theories that we believe are essential for understanding V1 and its relation to perception, and we shall present a few candidates in the second part of this chapter. A central conclusion that emerges from this exercise is that we need to begin seriously studying how V1 behaves using natural scenes. Based on these observations, we will then be in a more informed position when it comes to making choices about how to reduce complexity to tease apart the fundamental components of the system.

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 non-linearities known to exist in the nervous system.

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 micro-electrode 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 unavoidable. 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 the neuron does not respond “rationally” to the investigators stimuli or fit the stereotype of what the experimenter believes the neuron should do. This is especially true for higher visual areas like area V4, but it is also true for V1. Such neurons are commonly regarded as “visually unresponsive.” It is difficult to know how often such neurons are encountered because oftentimes they simply go unreported, or else it is simply stated that only visually responsive units were used for analysis.

While it is certainly 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 importantly, 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 feel for 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 can't happen if no one knows they are there.

Another bias that arises in sampling neurons, perhaps unintentionally, is that the process of hunting for neurons with a single micro-electrode will almost invariably steer one towards neurons with higher firing rates. This is especially disturbing in light of recent analyses showing that, based on estimates of energy consumption, the average firing rates of neurons in cortex must be rather low—i.e., less than 1 Hz (Attwell & Laughlin 2001; Lennie 2003b). One finds many neurons in the literature for which even the spontaneous or background rates are well above 1 Hz, suggesting then that they are likely to be substantially overrepresented (Lennie 2003b). What makes matters worse is that if we assume that most neurons exhibit an exponential distribution of firing rates, as has been demonstrated for natural scenes (Baddeley et al., 1997), then a mean firing rate of 1 Hz would yield the distribution shown in

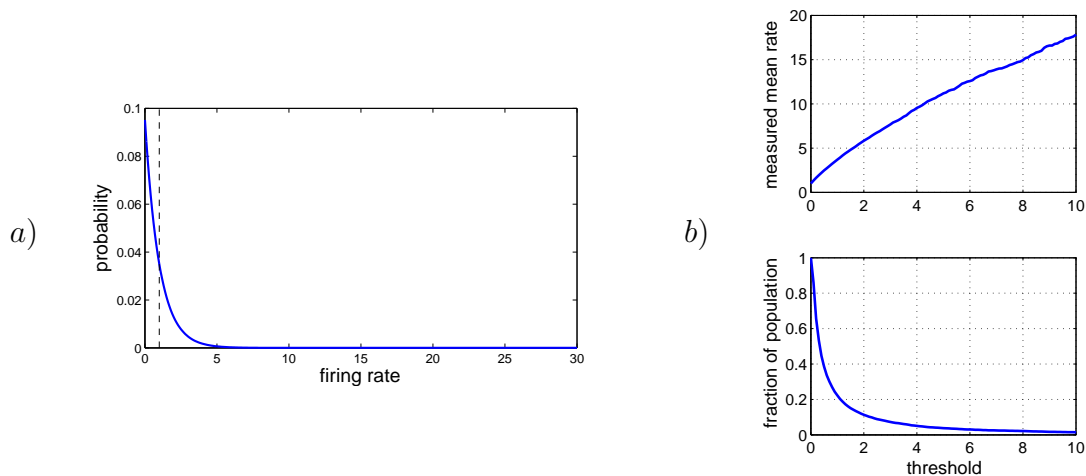


Figure 2: a) Exponential firing rate distribution with a mean of 1 Hz (dashed line denotes mean). b) resulting overall mean-rate of the population (top), and fraction of the population captured (bottom), as a result of recording from neurons only above a given mean firing-rate (threshold).

Figure 2a. Under natural conditions, then, only a small fraction of neurons would exhibit the sorts of firing rates normally associated with a robust response. For example, the total probability for firing rates of even 5 Hz and above is 0.007, meaning that one would have to wait 1-2 minutes on average in order to observe a one-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 (std. 3.6 Hz), suggesting that these neurons constitute a sub-population 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 & Salzman, 1985) report the mean firing rate to be 8.9 Hz (s.d. 7.0 Hz).

Presumably V1 contains 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, and the subsequent observed mean rate, if one were to selectively record from neurons only above a certain mean rate. The result of such an analysis, assuming a log-normal distribution of mean-rates so as to yield 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 1-2 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 et al., 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 (Skaggs, personal communication). In fact, Thompson & Best (1989) found that nearly two-thirds of all hippocampal neurons which 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 is in stark contrast to 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 are the patterns of activity that occur during free-viewing of natural scenes? What are the actual average firing rates and other statistics of activity in layers 2/3? What are the huge numbers of granule cells in macaque layer 4, which outnumber the geniculate fiber inputs by 30:1, 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 deg. 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 deg. (Olshausen & Anderson, 1995). Such receptive field sizes are extremely rare, if not non-existent, in the existing data on macaque V1 neurons collected using acute recording techniques (De Valois et al., 1982; Parker & Hawken, 1988).

Overall, then, one can identify at least three different biases in the sampling of neurons: 1) neurons with large cell bodies, 2) “visually responsive” neurons, and 3) neurons with high firing rates. So where does this leave us? If we assume that 5-10% of the population is missed because of weak extracellular action potentials, another 5-10% is discarded for being visually unresponsive, and 50-60% is missed because of low firing rates (as demonstrated in Figure 2), then even allowing for some overlap among these populations would yield the generous estimate that 40% of the population has been adequately sampled.

2.2 Biased stimuli

Much of our current knowledge of V1 neural response properties is derived from experiments using reduced stimuli. Oftentimes these stimuli are ideal for characterizing linear systems—i.e., spots, white noise, or sinewave gratings—or else they are designed around pre-existing 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 situation—e.g., natural scenes. But of course there is no guarantee that this is the case. And given the non-linearities inherent in neural responses, we have every reason to be skeptical.

Sinewave gratings are ubiquitous tools in visual system neurophysiology and psy-

chophysics. 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 sinewaves are special only because they are eigenfunctions of linear, time- or space-invariant systems. For non-linear systems, they bear no particular meaning, nor do they occupy any special status. In the auditory domain, sinewaves 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. Given these observations, one is led to the unavoidable conclusion that *there is no principled reason for using sinewaves to study vision*.

White noise, m-sequences, and spots suffer from similar problems. They are informative stimuli only to the extent that the system is linear, or contains rather simplistic nonlinearities (such as squaring, or response normalization) that can be characterized via 2nd-order methods (Touryan et al. 2002; Rust et al. 2004). Otherwise, they are no more valid than any other stimulus. Although it is true that an orthonormal basis (which could comprise any of these stimuli) can fully describe any pattern, characterizing the responses to each basis function in isolation is of limited value when the system is strongly non-linear.

What about bars of light or Gabor patches? The use of these stimuli also makes some assumptions about linearity. However, this approach primarily assumes that neurons are selective or tuned to localized, oriented, bandpass structure, and that the appropriate parameters for characterizing them are properties such as position, length/width, orientation, spatial-frequency, etc. This may seem reasonable given the fact that images also contain localized, oriented structure (i.e., edges), but how do we really know this is the right choice?

The use of reduced stimuli is sometimes justified by the fact that one would actually like to know how a neuron responds to a single point of light in an image, or to a specific spatial-frequency. In this case, one does need to construct a controlled stimulus. But we would argue that such questions are misplaced to begin with. Given the nonlinearities inherent in real neurons, there is every reason to believe that neurons are selective to certain *combinations* of pixel values or spatial-frequencies, and so probing the system's response to one element at a time will not necessarily tell you how it responds to particular combinations. Of course, we will never know this until it is tested, and that's precisely the problem—the *central assumption of the elementwise, reductionist approach has yet to be thoroughly tested*.

The brute force solution would be to exhaustively search the stimulus space. However, even an 8×8 patch with 6 bits of grey level requires searching $2^{384} > 10^{100}$ possible combinations. Needless to say, this is far beyond what any experimental method could explore. Therefore, the hope is that the non-linearities are smooth enough to allow predictions from a smaller set of stimuli.

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 mathematical functions or stimuli that are simple to think about or happen to be easy to generate on a monitor. If neurons are selective to specific combinations of stimuli, then we will need to explore their

responses to the sorts of combinations that occur in natural scenes. And at the same time, we will need to put more effort into mathematically characterizing the structure of natural scenes so as to better understand the forms of structure contained in them. No matter what stimuli one uses to characterize the response of a neuron, the true test that the characterization is correct is to demonstrate that one can predict the neuron’s behavior in ecological conditions.

2.3 Biased theories

Currently in neuroscience there is an emphasis on “telling a story.” This often encourages investigators to demonstrate when a theory explains data, not when a theory provides a poor model. We therefore have theories of grating responses and line weighting functions which may predict grating responses and line responses with some success, but which appear to provide poor predictions for natural scene data (as noted below). In addition, editorial pressures often 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—i.e., how does this help the brain to solve difficult problems in vision?

Hubel and Wiesel introduced the classification of neurons into categories of *simple*, *complex*, and *hypercomplex* or *end-stopped* based on their investigations using stimuli composed largely of bars of light. 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 less sensitive to position and the sign of contrast (black-white edge vs. 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 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 but having different positions or phases—i.e., the so-called “energy model” (Adelson & Bergen, 1985). This theory is consistent with measurements of response modulation in response to drifting sinewave 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 shown that this particular nonlinear measure tends to exaggerate or even introduce bimodality rather than revealing an actual, intrinsic property of the data (Mechler & Ringach, 2002). When receptive fields are characterized instead by the degree of overlap between zones activated by increments or decrements in contrast, one still obtains a bimodal distribution, in addition to a clear separation according to layer, but the F1/F0 ratio does a poor job at predicting which cells are complex according to the overlap criterion (Kagan et al., 2002). In addition, the energy model

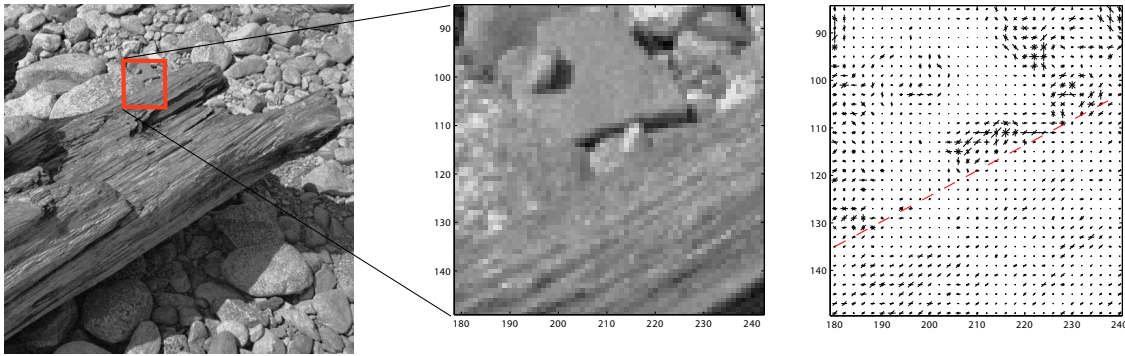


Figure 3: A natural scene (left), and an expanded section of it (middle). Far right shows the information conveyed by an array of complex cells at four different orientations. The length of each line indicates the strength of response of a model complex cell at that location and orientation. The red dashed line shows the location of the boundary of the log in the original image.

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 “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 for example Figure 3. Indeed, at the scale of a V1 receptive field, the structures in this image are quite complex and they 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 2D shape features in images. But years of research in computer vision have demonstrated that it is impossible to compute the presence even of simple edges in an image in a purely bottom-up fashion (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 confusing array of activity from which it would be quite difficult to make sense of what is going on. Our perception of crisp contours, corners, and junctions in images is largely a posthoc phenomenon that is the result of massive *inferential* computations performed by the cortex. In this sense then, our initial introspections about scene structure may actually be a poor guide as to the actual problems faced by the cortex.

In order to properly understand V1 function, we will need to go beyond bottom-

up filtering models and think about the ‘priors’ used by V1, or fed back from higher areas, for interpreting images (Olshausen, 2003; Lee & Mumford, 2003). Our theories need to be guided by functional considerations and an appreciation for the ambiguities contained in natural images, rather than appealing to simplistic notions of feature detection that are suggested by a select population of recorded neurons using reduced stimuli.

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 LGN, with the majority resulting from intracortical inputs (Peters & Payne 1993; Peters et al. 1994). Thalamocortical synapses have been found to be stronger, making them more likely to be effective physiologically (Ahmed et al., 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 neurons response. This leaves 65% of the response determined by factors outside of the direct feedforward input. Using optical imaging methods, Arieli et al. (1996) show that the population ongoing activity can account for 80% of an individual V1 neuron’s response variance. Thus, we are left with the real possibility that somewhere between 60-80% 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 multi-modal input. Sadato et al (1996) and Amedi et al (2003) demonstrated that both tactile Braille reading and verbal material can activate visual cortex in those that 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—i.e., the spatio-temporal structure outside of the classical receptive field (CRF)—is by now well known and has been the subject of many investigations over the past decade. Knierim & 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 et al. (1995) have shown that one can introduce quite dramatic changes in orientation tuning based on the orientation of gratings outside the CRF (see Figure 4). But these investigations have likely tapped only a portion of the interdependencies and contextual effects that actually exist.

The problem in teasing apart contextual effects in a reductionist fashion is that one faces a combinatorial explosion in the possible spatial/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 & Van Essen (1993), or Sillito et al. (1995), 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

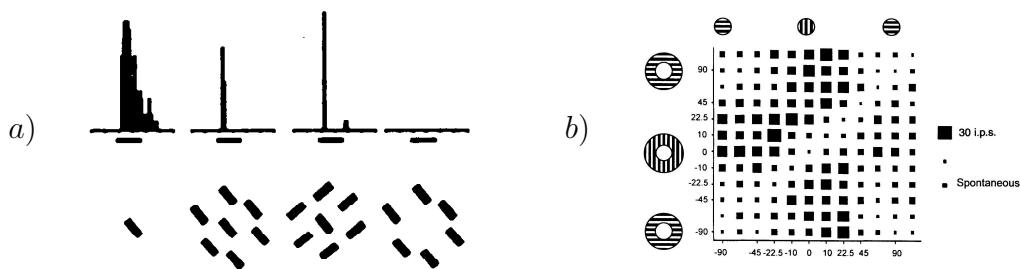


Figure 4: a) Knierim & Van Essen (1992) showed that many neurons in V1 of the alert macaque monkey are inhibited by stimuli outside their classical receptive field, even though these stimuli by themselves elicit little or no response from the neuron. Shown for each configuration is the peristimulus time histogram (PSTH) of the neuron’s response (bar denotes stimulus duration, which was 500 ms.). b) Sillito et al. (1995) showed that the presence of a grating stimulus outside the classical receptive field (CRF) could have a profound influence on the orientation tuning to a grating inside the CRF. Shown are the responses—denoted by the size of the square—to different combinations of grating orientation in the center and surround.

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 importantly, we need to understand how context is doing this.

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 EEG implies that large numbers of neurons must be acting together. Gray et al. (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 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 et al. (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

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.

The Gallant lab at UC Berkeley has for the past several years attempted to see 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 two reasons. One is that responses to repeated trials vary, and so if one wishes to attribute the variability to noise, then one must measure the inter-trial variability and discount this from the unexplained variance. The other problem is that whatever model is chosen, one is always subject to the criticism that it is not sufficiently elaborate, and thus any inability to predict the neuron’s response is simply due to some missing element in the model.

For example, David, Vinje & Gallant (1999) have explored two different types of models—a linear model in which the neuron’s response is a weighted sum of the image pixels, and a “Fourier power” model in which the neuron’s response is a weighted sum of the local power spectrum (which is capable of capturing the position or phase invariance non-linearity of a complex cell). After correcting for inter-trial variability, which is only approximate due to the limited number of trials, these models can explain approximately 20-30% of the response variance. It is possible that with more trials, and with the addition of non-linearities such as contrast normalization, adaptation, and response saturation, the fraction of variance explained would rise considerably. However, our own view is that these are well-established models that have been given a fair run for their money, and the addition of simple response non-linearities such as these is unlikely to improve matters much. Moreover, we contend that one can not easily dismiss inter-trial variability as “noise”—e.g., it could well be due to internally generated activity that plays an important role in information processing that we simply do not as yet understand (Arieli et al., 1996; see also Section 3.1). Given these results with both linear and Fourier power models, we conjecture the best-case scenario is that the percentage of variance explained is likely to reach an asymptote of 30-40% with the standard model.

One of the reasons for our pessimism is due to the way in which these models fail. For example, Figure 5 shows data collected from the Gray lab at Montana State University, Bozeman, in which the responses of V1 neurons to repeated presentations of a natural movie are recorded using tetrodes. Shown (Fig. 5a) is the peristimulus time histogram (PSTH) of one neuron, in addition to the predicted response generated by convolving the neuron’s space-time receptive field (obtained from an M-sequence) with the movie. The receptive field model does a poor job predicting the neuron’s actual response, especially in capturing the brief, punctate episodes of activity. Moreover, the responses of nearby neurons (recorded off the same tetrode) having similar receptive fields are extremely heterogeneous, much more so than expected from a linear receptive field model (Fig. 5b). It quickly becomes evident from looking at these and many other such neurons that adding further pointwise non-linearities or simple network nonlinearities such as contrast normalization is going to be of limited use. (Recently, Machens et al. (2004) came to a similar conclusion for neurons in auditory cortex.) What seems to be suggested by this behavior is a fairly complex network non-linearity involving interactions among neurons that are not yet fully understood.

An important lesson of these findings is that simply mapping out receptive fields

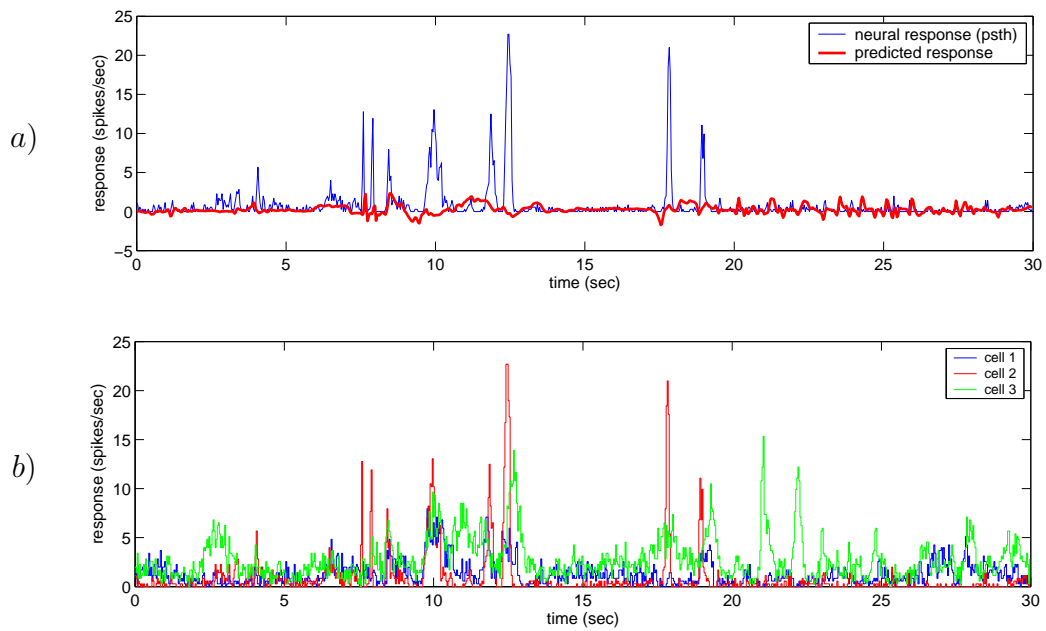


Figure 5: Activity of V1 neurons in anaesthetized cat in response to a natural movie. a) The PSTH of one neuron’s response (blue), together with the predicted response (red) generated from convolving the space-time receptive field with the movie. b) Simultaneous responses of three different neurons, with similar receptive fields, to the same movie. Spike rates were obtained by averaging across trials in 30 ms bins.

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 non-linear effects such as cross-orientation inhibition. 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 what these results really mean.

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 which works well. However, in this case the models being tested are well established in the literature, and up to now they have been largely *assumed* to work. Thus, these sorts of data are crucial to presenting a complete picture of V1 function, even without a competing model in hand. 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 multi-unit data (stimuli and neural responses) that would allow modelers to test their best theory under ecological conditions.

Finally, it should be noted that much 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). This would seem to suggest that the difficulty in predicting responses in cortex has to do with the effects of the massive, recurrent intra-cortical circuitry that is engaged during the natural dynamics of vision.

2.6 Summary

Table 1 presents a summary of the five problems we have identified with the current, established view of V1 as a “Gabor filter bank.” Given these factors, is it possible to quantify how well we currently understand V1 function? We attempt to estimate this as follows:

$$[\text{fraction understood}] = \left[\begin{array}{l} \text{fraction of variance ex-} \\ \text{plained from neurons} \\ \text{recorded} \end{array} \right] \times \left[\begin{array}{l} \text{fraction of population} \\ \text{recorded} \end{array} \right]$$

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

³We have primarily drawn upon 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.

Biased sampling	Biased stimuli	Biased theories	Interdependence & context	Ecological deviance
large neurons; visually responsive neurons; neurons with high firing-rates	use of reduced stimuli such as bars, gratings & spots	simple/complex cells; data-driven theories vs. functional theories	influence of intra-cortical input, context in natural scenes. synchrony	responses to natural scenes deviate from predictions of standard models

Table 1: Five problems with the current view of V1.

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—e.g., fan-in/fan-out, the spatial extent of axonal and dendritic arbors, etc.). In the sections below, we discuss some of the theories that are plausible given our current data. However, 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 than those theories presented here.

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 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 non-linear 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 multi-electrode

recording to attempt to account for the remaining variance. For example, Arieli et al. (1996) have shown that much of the inter-trial variability may be explained in terms of large-scale fluctuations in ongoing activity of the surrounding population of neurons measured using optical recording. 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 non-linearities found in V1 is to argue that neurons are attempting to achieve some degree of gain control (e.g., Heeger, 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 non-linearities of V1 neurons (further details are provided by Olshausen & Field, 1997, and Field & Wu, 2004). We argue that the spatial non-linearities primarily serve 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 takes, such data never requires more than D vectors to represent it. A system where data with dimensionality D is 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 ICA also attempt to find a critically sampled basis which minimizes the dependencies among the vectors, and the result is a wavelet-like code with tuning much like the neurons in V1 (Bell and 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 50:1. 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 non-linear transform. One of the non-linear 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 non-linearity. For example, consider the responses of a unit to two stimuli: the first stimulus aligned with the neuron’s weight vector, and a second stimulus separated by 90 degrees. The second stimulus

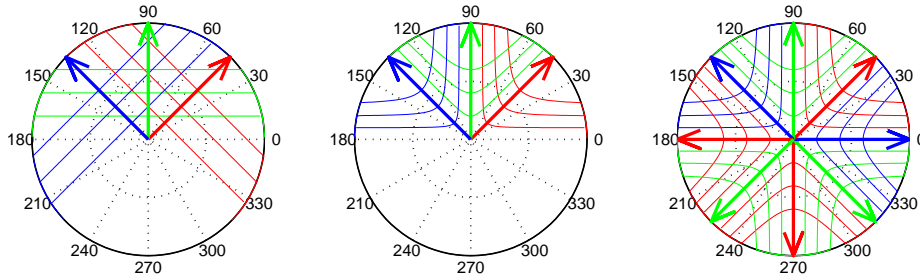


Figure 6: Overcomplete representation. a) Shown are the iso-response contours of three linear neurons (with half-wave rectification) having linearly dependent weight vectors. A stimulus falling anywhere along a given contour will result in the same response from the neuron. A stimulus falling in the 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 non-linear response.

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 non-linearity with the characteristic “non-classical,” suppressive behavior: Stimuli which on their own have no affect on the neuron (stimuli orthogonal to the principal direction of the neuron), can modulate the behavior of an active neuron. This general non-linearity comes in several forms and includes end-stopping and cross orientation inhibition, and is what is typically meant by the term “non-classical surround.” Indeed, as Zetsche et al (1999) note, this curvature is simply a geometric interpretation of such behaviors. With the addition of a compressive non-linearity this curvature results in the behavior described as “contrast normalization.”

In contrast to the gain control or divisive normalization theory, we argue that the non-linearities 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 non-orthogonal code.

Importantly, this model predicts that the non-linearities are a function of the angle between the neuron’s weight vector and those surrounding it. Future multi-electrode 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 non-linearities much like those proposed. It seems possible, then, that the

family of non-linearities found in V1 can eventually be explained within one general framework of efficient coding.

3.3 Contour integration

If the contrast normalization model were a complete account of V1 neurons, then we might expect the surround suppression to be relatively unspecific. However, the physiological and anatomical evidence implies that V1 neurons have a rather selective connection pattern both within layers 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 et al., 1993; Yoshioka et al., 1996). The short range projections, by contrast, do not show such specificity. 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 and Wiesel, 1979), and studies on the tree shrew (Rockland and Lund, 1982; Bosking et al., 1997), primate (e.g., Malach et al., 1993; Sincich & Blasdel, 2001), ferret (Ruthazer & Stryker, 1996), and cat (e.g., Gilbert & Wiesel, 1989) have all demonstrated significant specificity in the projection of these lateral connections.

A number of neurophysiological studies also show that co-linearly oriented stimuli presented outside of the classical receptive field have a facilitatory effect (Kapadia et al., 1995; Kapadia et al., 2000; Polat et al. 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 et al 1993). This follows from work showing that the amplification of locally co-aligned, oriented elements provides an effective means of identifying contours in natural scenes (Parent & Zucker, 1989; Sha’ashua & Ullman, 1990). This type of mechanism could work in concert with the sparsification non-linearities mentioned above, since the facilitatory interactions would primarily occur among elements that are non-overlapping—i.e., receptive fields whose weight vectors are orthogonal.

An alternative theoretical perspective is that the effect of the 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 re-

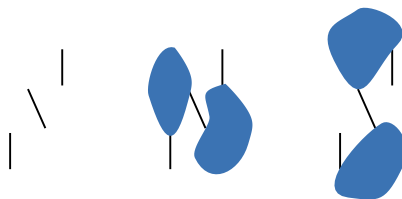


Figure 7: The three line-strokes at left are interpreted as different objects depending on the arrangement of occluders. Thus, pattern completion depends on resolving figure-ground relationships. At what level of processing is this form of completion taking place? Since it would seem to demand access to high-resolution detail in the image, it can not simply be relegated to high-level areas.

duction 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 connectons 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—i.e., surface structure, foreground/background relationships, etc. 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. Recently, a few studies along these lines have begun to yield interesting findings in extra-striate cortex (Nguyenkim & DeAngelis, 2003; Zhou, 2000; Bakin et al., 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, nor how figure-ground relationships of the sort depicted in Figure 7 are resolved.

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 do

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 vs. background in a scene of texture elements (Lamme 1995; Zipser et al., 1996), and a substantial fraction of neurons in V1 are selective to border ownership (Zhou, 2000). In addition, Lee et al. (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 such effects are just the tip of the iceberg, and there could be even more effects lurking.

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 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 now shown that human observers are quite poor judges of what constitutes a boundary or junction when these features are shown in isolation (Elder et al., 1999; McDermott, 2003). 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-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, 1997; Rao & Ballard, 1999; Lee & Mumford, 2003).

An example of disambiguation at work in the visual cortex can be seen in the resolution of the aperture problem in computing direction of motion. Because receptive fields limit the field of a 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 & 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 anaesthesia. A similar delayed response effect has been demonstrated in end-stopped V1 neurons as well (Pack et al., 2003).

Recent evidence from fMRI points to a disambiguation process occurring in V1 during shape perception (Murray et al., 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 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, 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 upon activity in lower areas seem to suggest that these connections play a role in enhancing the salience of stimuli (Hupe et al., 1998). 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. (See also Young (2000) for similar arguments to those presented here.)

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—i.e., 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 leaves much to be desired—namely, there is no provision for how phase, or information about relative spatial relationships, is preserved. Clearly, though, we have conscious access to this information. In addition, resolving figure/ground relationships and occlusion demands that higher levels of analysis have access to this information 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 et al. 1993). Rather than passively pooling, information is dynamically linked from one stage to the next by a set of control neurons that progressively remaps information into an object-centered reference frame. Thus, it is proposed that there are two distinct classes of neurons—those conveying image/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 mathematically it is equivalent to a bilinear model (e.g., Tenenbaum & Freeman, 2000; Grimes & Rao, 2003).

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* which 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. Whereas if information is routed then 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 (Zetsche & 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 et al., 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 & 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 Conclusions

Our goal in this article has been to point out that there are still substantial gaps in our knowledge of V1 function, and more importantly, that there is more room for new theories to be considered than the current conventional wisdom allows. We have identified five specific problems with the current view of V1, emphasizing the need for using natural scenes in experiments, in addition to multi-unit recording methods, in order to obtain a more representative picture of V1 function. While the single-unit, reductionist 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 nearly been 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 reductionist 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.

5 Acknowledgements

We thank Bill Skaggs for discussions on hippocampal physiology, Charlie Gray and Jonathan Baker for sharing preliminary data, Jack Gallant for clarifying the issues involved in predicting neural responses, and Jeff Johnson and Issac Trotts for comments on the manuscript.

6 References

- Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America, A*, 2, 284-299.
- Ahmed B, Anderson JC, Douglas RJ, Martin KA, Nelson JC (1994). Polynuclear 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, 758-66.
- Arieli A, Sterkin A, Grinvald A, Aertsen A (1996). Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science*, 273, 1868-71.
- Attwell D, Laughlin SB.(2001) An energy budget for signaling in the grey matter of the brain. *J Cereb Blood Flow Metab.* 21, 1133-45.

- Azouz R, Gray CM (2000) Dynamic spike threshold reveals a mechanism for synaptic coincidence detection in cortical neurons in vivo. *Proc Natl Acad Sci U S A*, 97, 8110-5.
- Azouz R, Gray CM (2003) Adaptive coincidence detection and dynamic gain control in visual cortical neurons in vivo. *Neuron* 37, 513-23.
- Baddeley R, Abbott LF, Booth MCA, Sengpiel F, Freeman T, Wakeman EA, Rolls ET (1997) Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc R. Soc. Lond. B*, 264, 1775-1783.
- Bakin JS, Nakayama K, Gilbert CD (2000) Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. *J Neurosci*, 20, 8188-98.
- Barlow HB (2001) Redundancy reduction revisited. *Network: Computation in Neural Systems*, 12, 241-253.
- Barnes CA, Skaggs WE, McNaughton BL, Haworth ML, Permenter M, Archibeque M, Erickson CA (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 AJ, Sejnowski TJ (1997) The independent components of natural images are edge filters, *Vision Research*, 37, 3327-3338.
- Bosking WH, Zhang Y, Schofield B, and Fitzpatrick D (1997) Orientation selectivity and the arrangement of horizontal connections in Tree Shrew striate cortex. *The Journal of Neuroscience* 17, 2112-2127.
- Carandini M, Heeger DJ, Movshon JA (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. *The Journal of Neuroscience*, 17, 8621-8644.
- Chung S, Ferster D (1998) Strength and orientation tuning of the thalamic input to simple cells revealed by electrically evoked cortical suppression. *Neuron* 20, 1177-89.
- Connor CE, Preddie DG, Gallant JL, Van Essen DC (1997). Spatial attention effects in macaque area V4. *J. Neurosci.* 17, 3201-3214.
- David SV, Vinje WE, Gallant JL (1999). Natural image reverse correlation in awake behaving primates. *Society for Neuroscience Abstracts*, 25, 1935.
- De Valois RL, Albrecht DG, Thorell LG (1982) Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res*, 22: 545-559.
- Elder JH, Benjaminov D, Pinteric G (1999) Edge classification in natural images. *Investigative Ophthalmology and Visual Science*, 40, S357.
- Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am, A*, 4: 2379-2394.
- Field DJ, Hayes A, and Hess RF (1993) Contour integration by the human visual system: evidence for a local 'association field'. *Vision Research* 33, 173-193.

- Field DJ, Wu M (2004). submitted.
- Gilbert CD, Wiesel TN (1979). Morphology and intracortical projections of functionally characterised neurones in the cat visual cortex. *Nature* 280, 120-125.
- Gilbert CD, Wiesel TN (1989). Columnar specificity of intrinsic horizontal and cortico-cortical connections in cat visual cortex. *J.Neurosci.*, 9, 2432-2442.
- Gray CM, Konig P, Engel AK, Singer W (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* 338, 334-7.
- Grimes DB, Rao RPN (2003) A Bilinear Model for Sparse Coding. In: *Advances in Neural Information Processing*, 15.
- Gur M, Snodderly DM (1997) Visual receptive fields of neurons in primary visual cortex (V1) move in space with the eye movements of fixation. *Vision Res.*, 37, 257-65.
- Hausser M, Mel B (2003). Dendrites: bug or feature? *Current Opinion in Neurobiology*, 13, 372-83.
- Heeger DJ (1992) Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9, 181-198
- Hupe, JM, James, AC, Payne, BR, Lomber, SG, Girard, P, and Bullier, J. (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* 394, 784-787.
- Jung MW, McNaughton BL (1993) Spatial selectivity of unit activity in the hippocampal granular layer. *Hippocampus*, 3, 165-182.
- Kagan I, Gur M, Snodderly DM (2002) Spatial organization of receptive fields of V1 neurons of alert monkeys: Comparison with responses to gratings. *J Neurophysiol*, 88, 2557-2574.
- Kapadia MK, Ito M, Gilbert CD, 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 MK, Westheimer G, Gilbert CD (2000) Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Journal of Neurophysiology* 84, 2048-2062.
- Knierim JJ, Van Essen DC (1992) Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J Neurophys.*, 67, 961-980.
- Lamme VA (1995) The neurophysiology of figure-ground segregation in primary visual cortex. *J Neurosci.*,15, 1605-15.
- Lee TS, Mumford D, Romero R, Lamme VA (1998) The role of the primary visual cortex in higher level vision. *Vision Res.*, 38, 2429-54.
- Lee TS, Mumford D (2003) Hierarchical Bayesian inference in the visual cortex. *J Opt Soc Am A*, 20, 1434-48.

- Legendy CR, Salzman M (1985) Bursts and recurrences of bursts in spike trains of spontaneously active striate cortex neurons. *J. Neurophysiol.* 53, 926-39.
- Lennie P (2003a) Receptive fields. *Curr Biol.* 13, R216-9.
- Lennie P (2003b) The cost of cortical computation. *Curr Biol.* 13, 493-7.
- Lewicki MS, Sejnowski TJ (1996) Bayesian unsupervised learning of higher order structure, In: *Advances in Neural Information Processing Systems*, 9, MIT Press.
- Machens CK, Wehr MS, Zador AM (2004) Linearity of cortical receptive fields measured with natural sounds. *J Neurosci.*, 24, 1089-100.
- 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, Grun S (2004). Synchrony and oscillations in primary visual cortex of monkeys viewing natural images. Submitted.
- McDermott J (2004) Psychophysics with junctions in real images. *Perception*. Submitted.
- Mechler F, Ringach DL (2002) On the classification of simple and complex cells. *Vision Res.*, 42, 1017-1033.
- Motter BC, Poggio GF (1990) Dynamic stabilization of receptive fields of cortical neurons (VI) during fixation of gaze in the macaque. *Exp Brain Res.*, 83, 37-43.
- Motter BC (1995) Receptive field border stabilization during visual fixation. *Investigative Ophthalmology and Visual Science*, 36, S691.
- Mumford D (1994) Neuronal architectures for pattern-theoretic problems. In: *Large Scale Neuronal Theories of the Brain*, Koch C, Davis, JL, eds., MIT Press, pp. 125-152.
- Murray SO, Kersten D, Olshausen BA, Schrater P, Woods DL (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 ZJ, Shimojo S (1995) Visual surface representation: a critical link between lower-level and higher level vision. In: *An Invitation to Cognitive Science*. S.M. Kosslyn, D.N. Osherson, Eds., M.I.T. Press, p. 1-70.
- Nguyenkim JD, DeAngelis GC. (2003) Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. *J Neurosci.* 23(18), 7117-28.
- Olshausen BA, Anderson CH, Van Essen DC (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *The Journal of Neuroscience*, 13, 4700-4719.
- Olshausen BA, Field DJ (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381, 607-609.
- Olshausen BA, Field DJ (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37, 3311-3325.

- Olshausen BA, Anderson CH (1995) A model of the spatial-frequency organization in primate striate cortex. *The Neurobiology of Computation: Proceedings of the Third Annual Computation and Neural Systems Conference*. J.M. Bower, Ed., Boston: Kluwer Academic Publishers, pp. 275-280.
- Olshausen BA (2003) Principles of image representation in visual cortex. In: *The Visual Neurosciences*, L.M. Chalupa, J.S. Werner, Eds. MIT Press. pp. 1603-15.
- Pack CC, Berezovskii VK, Born RT (2001) Dynamic properties of neurons in cortical area MT in alert and anaesthetized macaque monkeys. *Nature*, *414*, 905-908.
- Pack CC, Livingstone MS, Duffy KR, Born RT (2003) End-stopping and the aperture problem: Two-dimensional motion signals in macaque V1. *Neuron*, *39*, 671-680.
- Parent P, Zucker S (1989) Trace inference, curvature consistency and curve detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* *11*, 823-839.
- Parker AJ, Hawken MJ (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 BR (1993) Numerical relationships between geniculocortical afferents and pyramidal cell modules in cat primary visual cortex. *Cereb Cortex*. *3*, 69-78.
- Peters A, Payne BR, Budd J (1994) A numerical analysis of the geniculocortical input to striate cortex in the monkey. *Cereb Cortex*. *4*, 215-29.
- Polat U, Mizobe K, Pettet MW, Kasamatsu T, and Norcia AM (1998) Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature* *391*, 580-584.
- 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 KS, Lund JS (1983) Intrinsic laminar lattice connections in primate visual cortex. *Journal of Comparative Neurology* *216*, 303-18.
- Rose G, Diamond D, Lynch GS (1983) Dentate granule cells in the rat hippocampal formation have the behavioral characteristics of theta neurons. *Brain Res.* *266*, 29-37.
- Rust NC, Schwartz O, Movshon JA, Simoncelli EP (2004) Spike-triggered Characterization of Excitatory and Suppressive Stimulus Dimensions in Monkey V1. *Neurocomputing*, *58-60C*, 793-799.
- Ruthazer ES, Stryker MP (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 MP, Dold G, Hallett M (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature*, *380*, 526-8.
- Schwartz O, Simoncelli EP (2001) Natural signal statistics and sensory gain control. *Nat Neurosci*, *4*, 819-25.

- Sha'ashua A, Ullman S (1988) Structural saliency. *Proceedings of the International Conference on Computer Vision, Tampa, Florida*, 482-488.
- Sillito AM, Grieve KL, Jones HE, Cudeiro J, Davis J (1995) Visual cortical mechanisms detecting focal orientation discontinuities. *Nature*, *378*, 492-496.
- Sincich LC, Blasdel GG (2001) Oriented axon projections in primary visual cortex of the monkey. *Journal of Neuroscience* *21*, 4416-4426.
- Skottun BC, De Valois RL, Grosof DH, Movshon JA, Albrecht DG, Bonds AB (1991) Classifying simple and complex cells on the basis of response modulation. *Vision Res.*, *31*, 1079-86.
- Smyth D, Willmore B, Baker GE, Thompson ID, Tolhurst DJ (2003) The receptive-field organization of simple cells in primary visual cortex of ferrets under natural scene stimulation. *J Neurosci.* *23*, 4746-59.
- Softky WR, Koch C (1993) The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J Neurosci.* *13(1)*:334-50.
- Tarr MJ (1999) News on views: pandemonium revisited. *Nat Neurosci.*, *2*, 932-5.
- Tenenbaum JB, Freeman WT (2000) Separating style and content with bilinear models. *Neural Computation* *12*, 1247-1283.
- Thompson LT, Best PJ (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*. Harvard University Press.
- 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 JH, 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 WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, *287*, 1273-76.
- Wirth S, Yanike M, Frank LM, Smith AC, Brown EN, Suzuki WA (2003) Single neurons in the monkey hippocampus and learning of new associations. *Science*, *300*, 1578-81.
- Worgotter F, Suder K, Zhao Y, Kerscher N, Eysel UT, Funke K (1998) State-dependent receptive-field restructuring in the visual cortex. *Nature*, *396*, 165-168.
- Yoshioka T, Blasdel GG, Levitt JB, and Lund JS (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 MP (2000). The architecture of visual cortex and inferential processes in vision. *Spatial Vision*, 13, 137-46.
- Zetsche C, Krieger G, Wegmann B (1999) The Atoms of Vision: Cartesian or Polar? *J. Opt. Soc. Am. A*.
- Zetsche C, Rohrbein F (2001) Nonlinear and extra-classical receptive field properties and the statistics of natural scenes. *Network*, 12, 331-50.
- Zhou H, Friedman HS, and von der Heydt R (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience*, 20, 6594-6611.
- Zipser K, Lamme VA, Schiller PH (1996) Contextual modulation in primary visual cortex. *J Neurosci.*, 16, 7376-89.