2009

Perceptual Learning, Long-Range Horizontal Connections And Top-Down Influences In Primary Visual Cortex

Valentin Piëch

PERCEPTUAL LEARNING, LONG-RANGE HORIZONTAL CONNECTIONS AND TOP-DOWN INFLUENCES IN PRIMARY VISUAL CORTEX

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Valentin Piëch

June 2009

PERCEPTUAL LEARNING, LONG-RANGE HORIZONTAL CONNECTIONS AND TOP-DOWN INFLUENCES IN PRIMARY VISUAL CORTEX

Valentin Piëch, Ph.D.

The Rockefeller University 2009

The earliest cortical stage of visual processing, the primary visual cortex, has long been seen as a static preprocessor that finds local edges and their orientation like a linear filter bank, and passes this information on to downstream visual areas. This view has been challenged in recent years since the discovery of contextual influences, that is, interactions between the responses of neurons that encode for non-overlapping adjacent areas of visual space, and their anatomical substrate, long-range horizontal connections. These contextual interactions have been shown in awake behaving primates to be modulated depending on the task the animals are performing. A first set of electrophysiological experiments has shown with the help of information theory that when an animal performed one of two tasks on the same visual display, the contextual modulations of the task-

relevant parts of the visual display contained more information about the stimulus position than when the same elements were task-irrelevant. A second set of experiments on contour integration was analyzed with ROC analysis to show that an ideal observer could predict the presence of an embedded contour from the spike count of a single neuron on a single trial as well as the animal's behavioral performance. A final set of experiments showed that prior to learning the same contour integration task, the responses did not contain any information about the stimulus position, that the information in the response increased in parallel with the animals performance during learning, and that the enhanced response after learning disappeared during anesthesia, but is only weakened when performing an irrelevant task in a different part of visual space. Last, a neural network is presented that allows gating of long-range horizontal connections by top-down feedback. The stability and the dynamic behavior of the network have been established with phase-plane analysis. Large-scale simulations have been performed to confirm the stability and show the enhanced contour integration of realistic stimuli as a function of feedback gain. This model has fit quantitatively the electrophysiological experiments of contour integration.

Für meine Mutter und für Max

in Liebe und mit grösster Dankbarkeit

ACKNOWLEDGMENTS

First and foremost I would like to thank my advisors Charles Gilbert and George Reeke for their excellent mentorship. Thank you for always taking interest, and having an open door at any time. I was very lucky to have guidance and feedback at every level, from how to ask and answer scientific questions to how to communicate results more clearly in a talk or paper to how to improve the efficiency of my simulation code, yet was allowed and encouraged to explore new ideas even if they failed. I could not have asked for a better scientific education.

From the Gilbert laboratory, I would first like to thank my collaborator Wu Li who is now at Beijing Normal University. Without his extraordinary experimental skills and dedication, and his willingness to share his understanding and insights in many discussions, and being able to witness the experiments in person to understand the technical limitations, this thesis would not have been possible in this form – thank you so much, Wu! I am also very grateful to Doruk Golcu and Justin McManus, Doruk particularly for his collaboration on an fMRI project, and his willingness to be my first

and very extensively used test subject providing feedback on how to improve the experiment, and to Justin for many scientific discussions over the years and for classifying neurons as excitatory and inhibitory based on the extracellular action potential wave form.

From the Reeke laboratory, I would like to thank Allan Coop and Cameron Wellock for many scientific discussions, Allan particularly on Information Theory, and Cameron for his excellent insights into computer science and the use of Mathematica.

From the Citigroup Biomedical Imaging Center at Weill Cornell Medical College, I would like to particularly thank Henning U. Voss for his help in setting up, troubleshooting and improving the fMRI acquisition sequences to allow me to develop the use of fieldmaps for multi-channel receiver coils, and for becoming a collaborator based on this work together with Karsten Tabelow and Joerg Polzehl from the Weierstrass Institute for Applied Analysis and Stochastics. I would also like to thank Muc Du and Josefina Borja for their assistance with the fMRI experiments.

Importantly, I would like to thank my Faculty Advisory Committee Leslie Vosshall and Marcelo Magnasco for their advice, their guidance, and for

their encouragement even when experiments did not work as planned. I am also very grateful to my external thesis advisor Kenneth Miller for his interest and constructive criticism to improve my thesis and manuscript.

I would also like to thank my friends back in Switzerland and here in the United States, including those from our graduate program. I am particularly grateful to Dirk Hockemeyer and Chad Euler for many discussions on science and social distractions from it, but even more importantly, their friendship.

My special thanks go to my girlfriend Jia Li for her support, encouragement and love, particularly during the crucial final months of the thesis.

Last, I am forever grateful to my family for their unwavering support and love during all these years. Both of my brothers crossed the oceans to surprise me on the day of my thesis defense! And without my mother and Max, I would have never been able to achieve the dream to get a glimpse at the limit of scientific knowledge and at how the brain composes the world inside of our head – thank you with all my heart!

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

1.1 PERCEPTUAL LEARNING AS A MODEL FOR NONDECLARATIVE LEARNING IN CORTEX

We often associate learning and memory with being able consciously to recall information that we have learned at some earlier time (declarative learning). Yet many forms of learning do not require conscious awareness (nondeclarative learning).

Amongst them is one's ability to improve at discriminating attributes of simple stimuli by repeating a task, such as judging the orientation of a line. This form of learning is called perceptual learning (Gibson, 1963), and involves changes at the earliest stage of cortical processing, the primary sensory cortex of the corresponding sensory modality (for review, see Gilbert et al., 2001)

While the primary sensory cortices are the most extensively studied and well characterized cortical areas, the observed changes are likely to involve mechanisms that are more general to learning in cortex, and may help us to understand the neural basis of wide variety of learning phenomena.

1.2 PERCEPTUAL LEARNING PARADIGMS FOR EARLY VISUAL SYSTEM

Perceptual learning has been shown for a diversity of visual discrimination tasks (for review, see Gilbert et al., 2001). One line of experiments uses short oriented bars as stimuli. These elicit strong responses in neurons sensitive to the corresponding orientation and retinal location in primary visual cortex (Hubel and Wiesel, 1959). They include vernier discrimination (judging the direction of a small offset of two nearly collinear bars), bisection discrimination (determining the direction of offset of the middle of three parallel bars), and orientation discrimination. Also, oriented bars can be arranged to explore the Gestalt rules of smooth contour continuation (Wertheimer, 1923): if a number of short bar segments is arranged

according to a set of geometric rules to form a smooth, continuous contour, it visually "pops out" amongst a set of distracting bars at random orientations (Li and Gilbert, 2002). These tasks have the additional advantage that they are simple enough to be performed by awake behaving monkeys, and thus allow complementing the psychophysical data from human observers with physiological recordings from single (and multiple) units and optical imaging.

1.3 PROPERTIES OF PERCEPTUAL LEARNING IN THE VISUAL SYSTEM

Studies of perceptual learning can be distinguished by the time scale over which learning occurs. A first group of studies is concerned with the fast learning component, measured between blocks within a training session of a few hundred to a few thousand trials. While the threshold improvements are statistically significant, they are in the lower double digit percentage range, and quite noisy. Typically, these experiments are performed with human subjects and stimuli presented to the fovea. A second group measures

perceptual learning over several days to a few weeks, which reduces the threshold several fold (Crist et al., 1997; Crist et al., 2001). These experiments are performed in the visual periphery, as the threshold to foveal perceptual learning stops decreasing after a few hundred trials (Bennett and Westheimer, 1991). This work is concerned with the latter paradigm, as it is much more robust, and allows one to acquire physiological data from awake behaving monkeys (Crist et al., 2001).

Perceptual learning is specific for the location in visual space: Human observers trained in a bisection discrimination task over several weeks in a particular location at 5 degrees eccentricity showed a higher threshold when the stimulus was moved as little as 1 degree (Crist et al., 1997). Yet it is not completely local either, as the performance was still better then prior to learning when the stimulus was relocated as far as 8 degrees (Crist et al., 1997). Learning is specific to the orientation of the stimulus: Performing the same bisection discrimination when the display is rotated by 90 degrees shows much less improvement compared to that in the trained orientation (Crist et al., 1997). It is specific to the task: isection training does not

improve the threshold for an orientation discrimination in the same location (Crist et al., 1997). And finally, it is specific to the context: Doing a vernier discrimination at the same location and orientation where the bisection task was trained does not enhance vernier performance (Crist et al., 1997), even though both tasks require judging offsets of parallel lines.

One surprising result is that even vernier discrimination is not improved by prior orientation discrimination training at the same location (Crist et al., 1997), as most neural network models of vernier discrimination proposed that units sensitive to different orientations covering the two bars are the basis for the discrimination (Herzog and Fahle, 1998; Poggio et al., 1992).

Perceptual learning seems to be under top-down control, as attention is necessary for memory consolidation (Ahissar and Hochstein, 1993; Braun, 1998; Crist et al., 1997; Fahle and Morgan, 1996; Ito et al., 1998; Schoups et al., 2001; Shiu and Pashler, 1992; Treisman et al., 1992; Watanabe et al., 2001; but see Watanabe et al., 2002). Even more, attention seems to directly influence the way information is integrated: While surrounding a central

Gabor patch with aligned flankers on each side lowers the contrast detection threshold, they have no such facilitating effect if the subjects are forced to attend to other patches along the orthogonal axis in order to solve a vernier discrimination at the same time (Freeman et al., 2001).

Perceptual learning is long-lasting: the improved performance after training does not deteriorate for at least several months (Crist et al., 1997). Also, there is no indication that perceptual learning of one task interferes with performance on a different one (Crist et al., 1997).

Slow, peripheral perceptual learning occurs even if the subjects do not receive any feedback on their performance (Crist et al., 1997). While most studies of the fast, foveal learning component are in agreement with this result (Fahle and Edelman, 1993; McKee and Westheimer, 1978), one detailed study of vernier discrimination indicated that learning is at least severely impaired, if not abolished, if all learning trials are just below the threshold level, instead of being interspersed with trials having larger offsets (Herzog and Fahle, 1997). Thus, it cannot be ruled out that learning relies on

internal feedback on the easy trials where the subjects can be certain to have guessed correctly, and that the resulting learning generalizes to smaller offsets.

The same study (Herzog and Fahle, 1997) showed that uncorrelated feedback prevents learning, but partial feedback (an incorrect answer was reported with 50% probability) and block feedback (reporting the number of correct trials after each block of 80 trials) is almost as effective as feedback for every trial.

1.4 PROPERTIES OF V1 SUPERFICIAL LAYER CELLS

The high spatial resolution and the specificity for location suggest that perceptual learning at least involves neurons early in the visual pathway where the receptive fields are smallest, and that the arrangement is retinotopic. The specificity for orientation seems to preclude neurons in the retina and the lateral geniculate nucleus as substrates, as their receptive fields are not sensitive to orientation.

Neurons of the next processing stage, the primary visual cortex, are highly selective for the orientations of contours presented in their receptive fields (Hubel and Wiesel, 1959). They are divided into two classes, simple cells, which have a distinct on- and off-subregion in their receptive field, and complex cells, whose receptive fields do not allow such a distinction (Hubel and Wiesel, 1959, 1968). While there has been a longstanding controversy whether the tuning properties of simple and complex cells are the result of a hierarchical feedforward connectivity (LGN to simple cells, and simple cells to complex cells, as originally proposed by Hubel and Wiesel, (1962), parallel input from LGN to simple and to complex cells, or recurrent connections within the cortex (for review, see Martinez and Alonso, 2003), recent evidence seems to support the original feedforward model (Martinez and Alonso, 2001).

The cells of V1 layers 2+3 (referred to as superficial layers) have complex receptive fields (Hubel and Wiesel, 1962) and are of particular interest for modeling perceptual learning. First, the pyramidal cells of the superficial

layers are the principal output neurons to higher cortical levels, so that the information from the retina depicting the stimulus configuration must pass through them on the way to higher level cortical areas. Second, while these cells can only be driven by a stimulus inside of their receptive field, their firing rate can be modulated by placing oriented bars in the vicinity of the receptive field. When two parallel bars are placed along the axis of the bar in the receptive field, the effect is often excitatory, particularly at low contrasts, but inhibitory if the added bars are placed to the sides of the receptive field (Kapadia et al., 2000). The excitatory influence of colinearity decreases when the distance between the bars increases, the bars are displaced from colinearity, or the relative orientation is changed (Kapadia et al., 1995). Inserting an orthogonal line between the two colinear ones abolishes the facilitation, both for complex cell responses, and for a psychophysical contrast detection task (Kapadia et al., 1995). This seems to be in agreement with the Gestalt rules of smooth contour continuation (Wertheimer, 1923), and to reflect the statistics of natural images (Sigman et al., 2001). The likely substrate of these interactions is the long-range horizontal connections that link orientation columns of similar orientation, which can span 6 to 7

mm from end to end (Gilbert and Wiesel, 1979, 1989; Rockland and Lund, 1982; Stettler et al., 2002).

1.5 EFFECTS OF PERCEPTUAL LEARNING ON V1 SUPERFICIAL LAYER NEURONS

In the somatosensory and the auditory systems, perceptual learning leads to cortical recruitment, a remapping of the cortical surface such that a much larger area is responsive to the trained stimulus (Recanzone et al., 1992a; Recanzone et al., 1993). In the visual system, no such remapping has been observed (Crist et al., 2001; Schoups et al., 2001). The receptive field size and position are not changed by training of a bisection task (Crist et al., 2001). The sharpness of tuning was is increased by training of an orientation discrimination task (Schoups et al., 2001), but not by practicing bisection orientation (Crist et al., 2001).

Bisection training did have a profound influence on the contextual influences. When a monkey performed a fixation task, placing a parallel bar

to the side of one in the classical receptive field led to a slight reduction of the firing rate (Crist et al., 2001), as expected from previous experiments (Kapadia et al., 2000). When the same animal was performing a trained bisection task in the vicinity of the bar in the classical receptive field, the influence changed to excitatory (Crist et al., 2001). Even more, changing the distance between the two bars modulated the response significantly more than when performing a fixation task, providing evidence that the contextual interactions change in a way consistent with the task, and this change is only present when the task is being performed (Crist et al., 2001)

Compared to higher-level cortical areas, attention has little influence on the responses of V1 superficial layer neurons when only the classical receptive field is stimulated (Motter, 1993), and the influence can be either facilitatory or inhibitory (Haenny and Schiller, 1988; Motter, 1993), for review, see (Posner and Gilbert, 1999). Placing additional stimuli in the surround potentiates this effect (Ito and Gilbert, 1999; Ito et al., 1998; Motter, 1993). In a brightness discrimination task, distributed attention led to a much stronger facilitation by an aligned bar than did focal attention before

practice, but the difference diminished after training. This suggests that the benefit of focal attention is reduced with training (Ito et al., 1998). Thus, attention can lead to an increase in contextual influences per se without training, and lead after training to contextual influences that are particularly useful to solving the task at hand.

Our first study (Li et al., 2004) demonstrated the usefulness of contextual influences directly by having an animal perform one of two different tasks, and showed that the neuronal responses contain more information about the task that the animal is performing.

During the experiment, electrophysiological recordings from the superficial layers of primary visual cortex were made while awake behaving primates performed either a vernier or a bisection task on the same visual display consisting of five parallel bars arranged like a cross while recordings were made from cells responsive to the central bar. The contextual influences, that is the modulation of the firing rate as a function of the position of the surrounding bars, was clearly different in the two tasks. By calculating the

mutual information between the stimuli that were presented to the animal and the spike count distribution that each stimulus evoked, we were able to show that the flanking bars relevant to the vernier task conveyed more information during the vernier task than during the bisection task, and vice versa.

In the second set of experiments (Li et al., 2006, 2008), awake behaving primates had to find an embedded contour consisting of several collinear line segments in a patch of randomly oriented lines whose perceptual saliency could be parametrically varied. We were able to use ideal observer analysis to show that the mean spike count in a single trial from a neuron with its receptive field in the middle of the contour was sufficient to predict the animal's behavioral performance quantitatively over a wide range of perceptual saliencies and performance levels. This effect depends on the animal doing the contour detection task: The facilitation by the presence of the contour is only half as strong if the animal is doing an attentionally demanding bisection task in the opposite hemisphere. Last, if the same patterns are presented to the animal before training, the contour does not

modulate the response even if the animal is attending to the location of the central bar of the contour (Li et al., 2008). Only after the animal has been trained to detect the contour does the previously observed facilitation occur. During training, the facilitation increases in parallel with the improvement in behavioral performance, though the behavioral performance is higher than predicted by the ideal observer analysis of the single spike count until the animal's performance saturates. This indicates that only in the fully trained state does a single neuron encode the presence of the contour on average as reliably as the animal's performance. And finally, the facilitation disappears in the anesthetized animal even after training. Thus, perceptual learning is not only specific to the task at hand, but even the basic property of contour integration – which was previously assumed to be hard-wired – heavily depends on top-down influences.

1.6 NEURAL MODELS OF HORIZONTAL CONNECTIONS, PERCEPTUAL LEARNING, AND FEEDBACK

None of the published neuronal network models can gate horizontal connections for perceptual learning in a feedback-dependent manner. This section will give a quick overview of models incorporating elements that are related to the subject.

A first category of models uses horizontal connections to explain how surround facilitation at low contrasts and surround suppression at high contrast can emerge from a network with fixed connections, and how the orientation of the surround influences the results (Dragoi and Sur, 2000; Series et al., 2001; Somers et al., 1998; Stemmler et al., 1995). From a computational point of view, they are all based on integrate and fire neurons, with local excitatory and inhibitory connections, and long-range excitatory connections between columns with similar orientation preference.

A related group is concerned with explaining the same experimental data, but using cell models that allow an analytical treatment of the mathematical properties (Bressloff and Cowan, 2002; Stetter et al., 2000).

A different class of models tries to use long-range horizontal connections to model one specific phenomenon observed during perceptual learning experiments involving oriented stimuli, such as non-linearities in flanker-facilitation with an increasing number of flankers (Adini et al., 1997), modifying the context allowing perceptual learning (Adini et al., 2002), different speed of apparent motion as a function of orientation (Series et al., 2002), optimal solutions for bisection discrimination in the presence of eye movements (Zhaoping et al., 2002), fast perceptual learning of a vernier task in the hyperacuity range with radial basis functions (Poggio et al., 1992), and a network with self-supervision to explain learning without feedback (Herzog and Fahle, 1998). Mathematically, they all have in common that they are modeled in one dimension only.

The next category can be described as “optimization” models, which are based on information theoretical arguments that the nervous system should try to remove redundancies in the encoding of sensory stimuli. According to this view, perceptual learning adjusts the encoding of sensory stimuli to the statistics of natural images, but not by putting emphasis on often encountered examples, but by removing responses to them. Thus, they perform predictive encoding, and only signal where statistical expectation and visual inputs do not match (Rao and Ballard, 1999; Schwartz and Simoncelli, 1999; for review, seeSimoncelli and Olshausen, 2001).

The final category uses long-range horizontal connections to perform perceptual grouping (Grossberg and Mingolla, 1985b; Li, 1998b). They use bow-tie shaped kernels to allow the enhancement of responses for elements forming smooth contours according to the Gestalt rules (Wertheimer, 1923),. The model by Grossberg and Mingolla (1985b) proposes this linking stage to be part of V2, as cells responsive to illusory contours have been found in area 18, but not in area 17 of cat visual cortex (von der Heydt et al., 1984). The model of Li (1998b) includes linking as part of V1, and has the

additional advantage that it does not only perform contour linking, but also segmentation and pop-out based on inhomogeneities in orientation space (Li, 1999b). Later versions of Grossberg's model (for example, see Raizada and Grossberg, 2003) also include long-range horizontal connections in V1.

The major difficulty with all models of long-range horizontal connections is that the maximum strength of long-range horizontal connections is quite limited because the "tension" between excitation and inhibition to uniform input will lead to visual hallucinations due to the spatial anisotropy of the long-range horizontal connections (Ermentrout and Cowan, 1979; Li, 2001) and the recurrency of the network dictated by the anatomical constraints. Additionally, it is very difficult to ensure that the activity of the network nodes is bound for all geometric configurations, yet still leads to the desired pop-out of more than 2-fold for embedded contours (Li et al., 2006). One solutions to this problem has been to choose network parameters that lead to very strong oscillations so that the mean activity still oscillates around an unstable equilibrium (Li, 1998b). Unfortunately, this can lead to strong all-

or-none oscillations that need to be bound by a saturating response function, and also requires a strong non-linear normalizing current (Li, 1998b).

A second limitation is that previous models cannot selectively gate the long-range horizontal connections: The feedback always affects the responses to stimuli only presented to the classical receptive field as well.

Our final modeling study (Piech et al., 2009) overcomes these problems by using a local circuit that consists of a coupled excitatory and a divisive-inhibitory network node. We analyze the local circuit with phase-plane analysis and demonstrate that this local circuit will always have bound responses because the excitatory nullcline has a horizontal asymptote. We then extend the model to include long-range horizontal connections that are in agreement with the statistics of natural images (Sigman et al., 2001), the contextual influences measured electrophysiologically in V1 and psychophysically (Kapadia et al., 2000), and the anatomical constraints quantified by the labeling of long-range horizontal connection axons and

feedback projections by recombinant adenovirus expressing GFP (Stettler et al., 2002).

We then propose a mechanism by which the gain to either both the bottom-up and the long-range horizontal connections, or the long-range horizontal connections alone, can be controlled, and thus how the long-range horizontal connections can be effectively gated. This feedback gain controls the strength of the local coupling of the excitatory neurons with themselves, or, in other words, the local self-excitation.

Next, we show the effectiveness of the circuit with large-scale simulations on a naturalistic stimulus for different levels of feedback gain.

We then quantitatively match the time course of eletrophysiological responses of V1 superficial layer neurons to pop-out experiments using textured backgrounds consisting of oriented bars (Li et al., 2006; Li et al., 2000). And last, we demonstrate the robustness of the network to the presence of noise.

2 PERCEPTUAL LEARNING AND TOP-DOWN INFLUENCES IN PRIMARY VISUAL CORTEX

Wu Li, Valentin Piëch & Charles D Gilbert

Nature Neuroscience 7, 651 - 657 (2004)

2.1 SUMMARY

Neuronal responses at early stages in visual cortical processing, including those in primary visual cortex (V1), are subject to the influences of visual context, experience and attention. Here we show that for monkeys trained in a shape discrimination task, V1 neurons took on novel functional properties related to the attributes of the trained shapes. Furthermore, these properties depended on the perceptual task being performed; neurons responded very differently to an identical visual stimulus under different visual discrimination tasks. These top-down influences were seen from the very beginning and throughout the entire time course of the neural responses. Information theoretic analysis showed that neurons carried more information about a stimulus attribute when the animals were performing a task related

to that attribute. Our findings suggest that the output from V1 reflects both sensory and behavioral context.

2.2 INTRODUCTION

Perceptual learning, a form of implicit memory, is an improvement in sensory discrimination after an extended period of practice. In the visual system, training can improve the discriminability of spatial resolution (Crist et al., 1997; McKee and Westheimer, 1978; Poggio et al., 1992; Saarinen and Levi, 1995), orientation (Shiu and Pashler, 1992; Vogels and Orban, 1985), direction of motion (Ball and Sekuler, 1982), the waveform of sinusoidal stimuli (Fiorentini and Berardi, 1980), texture (Karni and Sagi, 1991) and depth (Ramachandran and Braddick, 1973), (Fendick and Westheimer, 1983). An increasing body of evidence implicates plasticity of early sensory cortex in perceptual learning. The response properties of neurons, the structure of receptive fields (RFs) and the topographical map of the cortex can be altered by sensory experience (for review, see refs. (Gilbert, 1998; Weinberger and Bakin, 1998)). In primary somatosensory

and auditory cortex, perceptual learning may lead to 'cortical recruitment,' a remapping of the cortical representation of the sensory surface that results in larger representation of the trained area (Recanzone et al., 1992b), (Recanzone et al., 1993). In primary visual cortex (V1), on the other hand, after extensive training in a visual discrimination task, there is no significant change in the cortical area representing the trained part of visual space, nor in certain simple RF attributes (Crist et al., 2001; Ghose et al., 2002). Training does have profound effects in V1 (Schoups et al., 2001), however, notably on the contextual influences mediating the selectivity of neurons to complex stimuli (Crist et al., 2001).

Response properties of V1 neurons are dynamic. They can be modified by stimulus context (Blakemore and Tobin, 1972; DeAngelis et al., 1994; Gilbert and Wiesel, 1990; Kapadia et al., 1995; Knierim and van Essen, 1992; Lamme, 1995; Li and Li, 1994; Li et al., 2000; Maffei and Fiorentini, 1976; Nelson and Frost, 1978; Rossi et al., 1996), visual experience (Crist et al., 2001; Schoups et al., 2001), and attentional state (Crist et al., 2001; Haenny and Schiller, 1988; Ito and Gilbert, 1999; Motter, 1993; Roelfsema

et al., 1998). The dynamic properties and plasticity of V1 neurons suggest a direct involvement of V1 in analyzing more complex features of the visual environment and in mediating higher-order cognitive influences. Our previous work has shown an interaction in V1 among spatial attention, visual context and perceptual learning (Crist et al., 2001). In the current study, to isolate the contribution of perceptual task, we trained the monkeys to do two different discrimination tasks with the same visual stimulus at the same visual field location. We observed striking task-related effects on response properties of V1 neurons.

2.3 RESULTS

2.3.1 TASK MODULATION OF CONTEXTUAL INFLUENCES

Our experimental design allowed a dual dissociation between two perceptual tasks and two RF properties (Figure 1a). The visual stimuli consisted of five line segments, a central line fixed in the RF center and paralleled side-by-side and end-to-end by four additional flanking lines surrounding the RF. Two monkeys were trained to perform two visual discrimination tasks based

on different attributes of the five-line stimuli. In the bisection task, the task-relevant components were the three side-by-side lines. In different trials, the two side flanks were positioned randomly at one of five possible positions. In one position, the three lines were equidistant, and in the other four positions, either flank was placed closer to the central line with varying separation. The task was to determine to which of the two flanks the central line was closer. In the vernier task, the task-relevant components in the five-line stimuli were the three lines placed end-to-end. In different trials, both end flanks were positioned randomly at one of five positions, one collinear with the central line and two positions on either side of the central line with varying displacement. The task was to determine to which side of the flanks the central line was displaced. The animal was cued to the task by color difference, with task-relevant components shown in green and task-irrelevant components shown in white (insets in Figure 1b,c). During V1 recording sessions, the contrasts of the three task-relevant green lines and the two task-irrelevant white lines were adjusted so that a single green line and a single white line presented in the RF alone elicited the same mean responses. The animal indicated its choice in the task by making a saccade to

either of two targets displayed after the five-line stimulus was extinguished. From the psychometric curve (Figure 1b,c), we were able to determine that the animal was performing the cued task and was not influenced by the positions of the task-irrelevant lines.

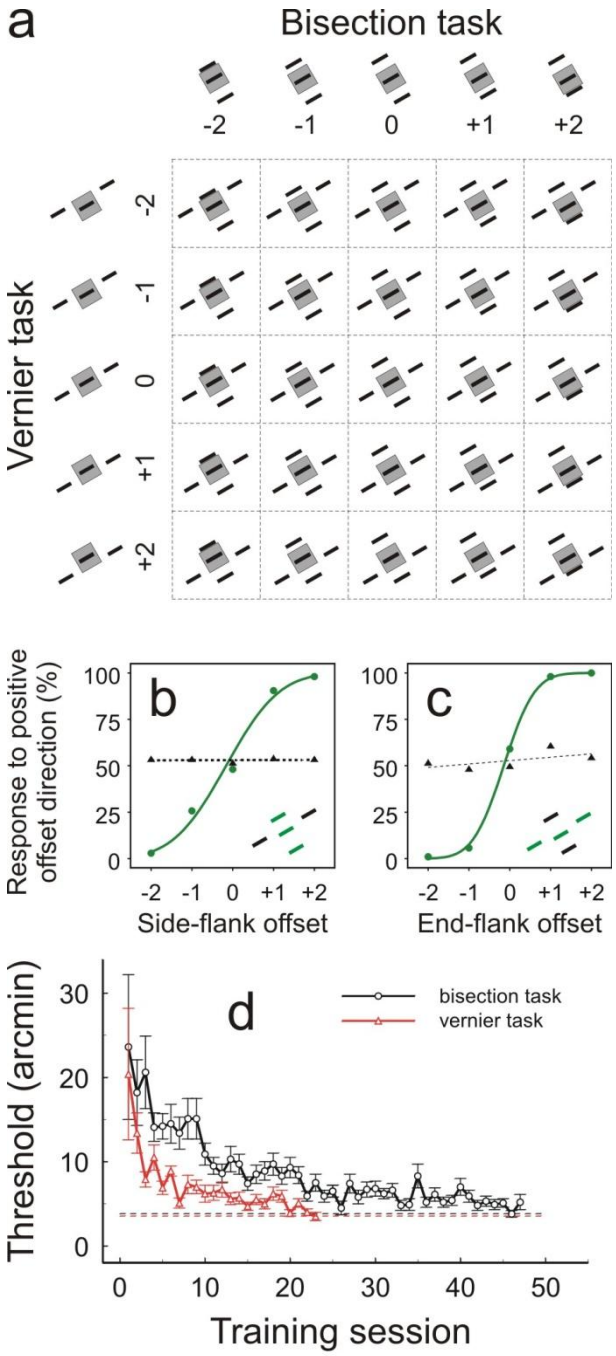

Figure 1. Stimulus and behavior paradigms. (a) The 5 × 5 stimulus array upon which two different discrimination tasks were performed. (b,c) Typical psychometric function (green curve) measured for the bisection (b) and vernier (c) tasks, using the five-line stimuli. The black dashed curve was constructed by

assuming that the animal responded to the offsets of the task-irrelevant flanks rather than the task-relevant flanks. Note that the monkey's behavioral responses were strongly associated with the cued task, but were not affected by the irrelevant stimulus flanks. (d) Learning curves of a monkey in a simplified bisection and vernier task, where either three side-by-side or three end-to-end lines were displayed horizontally at 1.4° eccentricity. Each data point was based on 300 responses. The bottom horizontal lines indicate thresholds measured 3 months after training.

The monkeys were initially trained with simplified displays of three side-by-side lines or three end-to-end lines. After their thresholds for bisection and vernier discrimination reached asymptote (Figure 1d), the two sets of three-line stimuli were combined to form the five-line array, and the monkey was trained to perform the cued tasks with the presence of both task-relevant and task-irrelevant lines. Further training generalized the task to different orientations. Repeating bisection and vernier tasks at the same spatial location and on the same stimulus patterns allowed a dual dissociation of contextual tuning (response modulation by side flanks versus end flanks) and behavioral task (bisection versus vernier discrimination). This allowed us to explore the effect of the discrimination task independent of stimulus context and of attention to spatial location.

We measured neuronal responses as a function of side-flank position (referred to as 'side-flank offset tuning') or end-flank position ('end-flank offset tuning'). When generating a side-flank offset tuning function, at each position of the side flanks, we pooled and averaged the five conditions with different positions of the end flanks (a column in the 5 times 5 stimulus array, Figure 1a). Similarly, when generating an end-flank offset tuning function, we pooled the five conditions sharing the same end-flank position but with different side-flank positions (a row in the 5 times 5 stimulus array, Figure 1a). To determine the modulation of neuronal responses by the offset of stimulus flanks, the central position of the tuning curve was used as a reference, where the central line of the five-line stimulus array was equidistant from the two side flanks in the side-flank offset tuning or was collinear with the two end flanks in the end-flank offset tuning. Any significant deviation of the neuronal response in other offset conditions relative to the zero offset condition provided a measure of the modulation induced by varying the offset of stimulus flanks. The side- or end-flank offset tuning function was compared under both bisection and vernier task

conditions where the examined flanks were task-relevant in one condition but not the other.

When the animal performed the bisection task where the side flanks were task-relevant, some cells showed substantial positive modulation by the offset of the side flanks (Figure 2a). In contrast, when the animal performed the vernier task where the side flanks were task-irrelevant, there was weak negative modulation for the same cells. This task-related effect can be clearly seen in the V-shaped differential tuning curve (Figure 2e), which provides a measure of the magnitude and shape of the difference between the offset tuning curves for the same contextual influence under the two behavioral conditions.

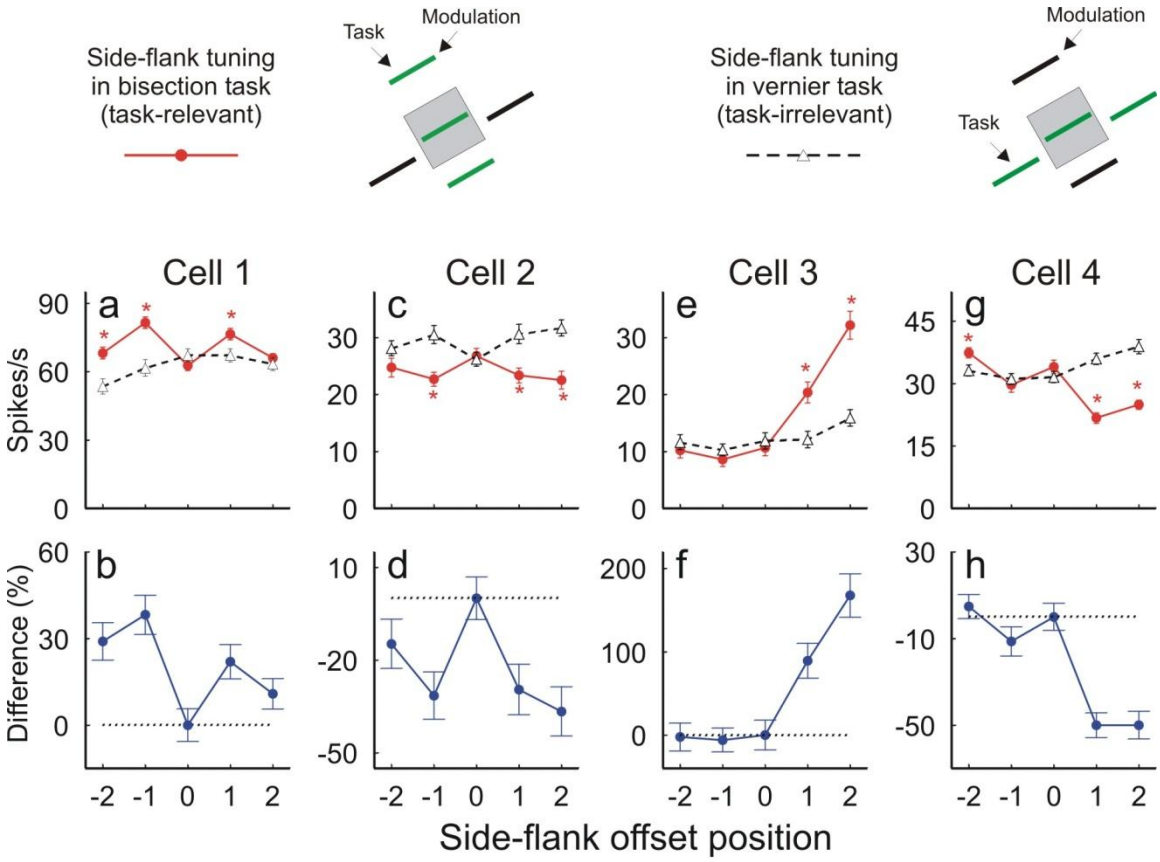

Figure 2. Task modulation of side-flank offset tuning.

The top panel shows the stimulus paradigms used in task-relevant and task-irrelevant conditions.

(a–d) Side-flank offset tuning curves of four examples of cells when the side flanks were task-relevant (red curves) and task-irrelevant (black dashed curves). Asterisks indicate that the difference in mean responses between the two tasks was significant (t-test, $P < 0.05$).

(e–h) Differential side-flank offset tuning curves of each cell. The 'differential tuning' function was constructed by subtracting the normalized curve (with response at 0 offset set to 100%) of the task-irrelevant condition from that of the task-relevant condition.

Error bars represent ± s.e.m.

The task-related effect varied from cell to cell. Whereas the neuron in Figure 2a showed positive modulation by lateral offsets of the side flanks in the task-relevant condition, others showed negative modulation (Figure 2b), and still others showed asymmetric influences, with positive Figure 2c) or negative (Figure 2d) modulation when the offset of side flanks was to one side or the other in the task-relevant condition. When the side flanks were task-irrelevant, the side-flank offset tuning curves were generally flatter, indicating a weaker modulation, than those in task-relevant conditions (Figure 2a,c,d). Even when the modulation amplitude looked similar under both task conditions, the tuning curve frequently changed shape (Figure 2b).

Similar task dependency was observed in the end-flank offset tuning function (Figure 3). For the neurons shown, when the animal performed the vernier task where the end flanks were task-relevant, the shape of the end-flank offset tuning curve differed markedly from the curve obtained when the animal performed the bisection task, where the end flanks were task-irrelevant. The differential end-flank offset tuning curve could be V-shaped (Figure 3e), caret-shaped (Figure 3f) or asymmetric (Figure 3g,h).

Furthermore, when the end flanks were task-irrelevant, the end-flank offset tuning curves were generally flatter than those in task-relevant conditions (Figure 3a,c,d).

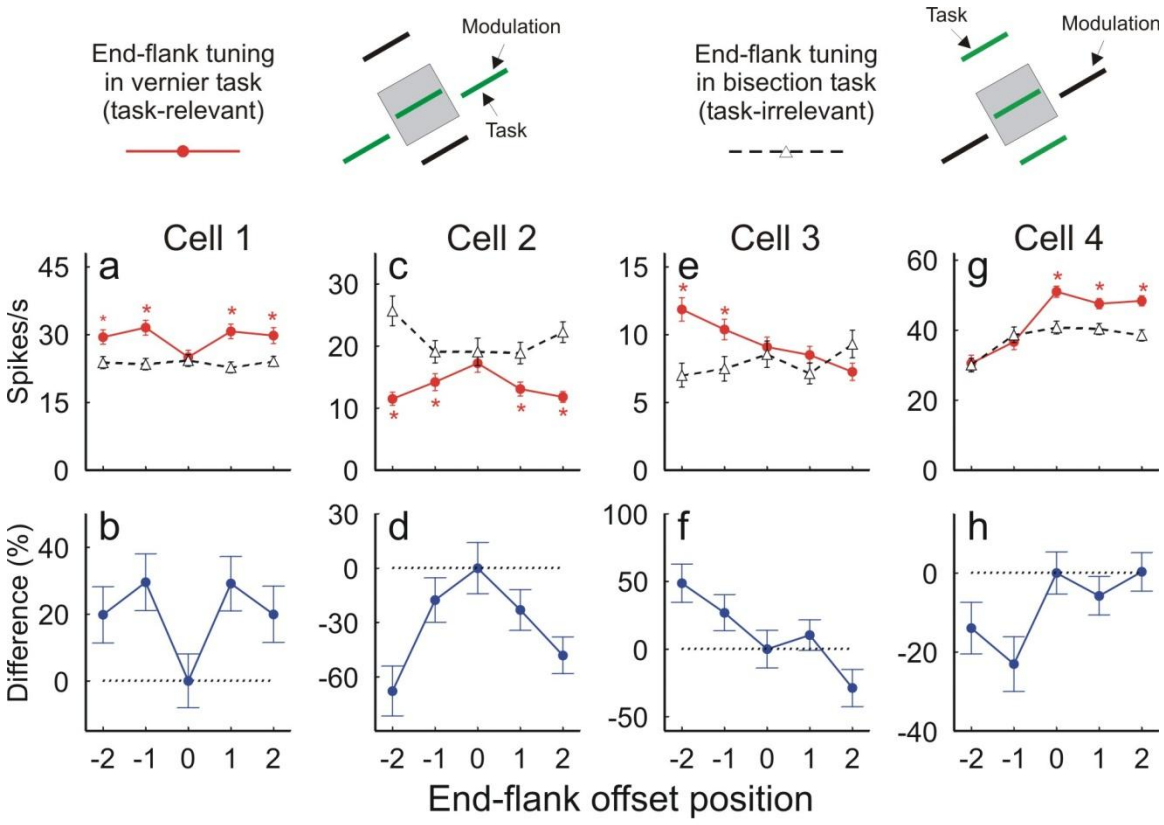

Figure 3. Task modulation of end-flank offset tuning.

The top panel shows the stimulus paradigms used in task-relevant and task-irrelevant conditions.

(a–d) End-flank offset tuning curves of four examples of cells when the end flanks were task-relevant (red curves) and task-irrelevant (black dashed curves). Asterisks indicate that the difference in mean responses was significant between the two tasks (t-test, $P < 0.05$).

(e–h) Differential end-flank offset tuning curves of each cell (for details, see Fig. 2 legend).

Error bars represent ± s.e.m.

No obvious relationship was observed for individual neurons in the task-related effects between the side- and end-flank offset tuning functions. The shape of the offset tuning curves and the differential tuning curves or the strength of the task-related effect could be different for the same cell in the side- and end-flank offset tuning functions.

To quantify the tendency for offset-tuning to be flatter in the task-irrelevant condition (Figure 2 and Figure 3), we used three measures addressing different aspects of the task-related effects: the modulation index, the asymmetry index and the average 'mutual information' between the number of spikes elicited in a trial and the task stimulus presented (see Methods).

The modulation index provides a measure of the magnitude of change of neuronal responses induced by the lateral displacement of the stimulus flanks16. The mean modulation index across the population of cells was higher in the task-relevant than task-irrelevant condition (Figure 4a). The two clouds in Figure 4a are the modulation indices generated by 1,000 Monte Carlo simulations (see Methods) for side- and end-flank modulation,

respectively. Since the mean modulation index was above the diagonal and clearly separated from the simulated clouds of data points, the same stimulus context did produce significantly higher modulation of neuronal responses when they were task-relevant than when they were irrelevant.

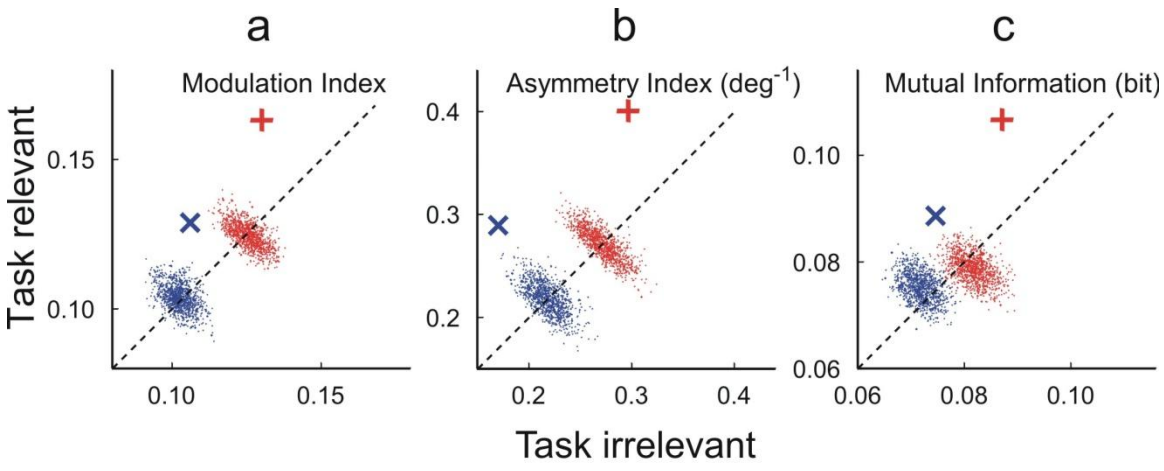

Figure 4. Population analysis of the task-related effect (n = 51).

Comparisons between task-relevant and task-irrelevant conditions of (a) mean modulation index, (b) mean asymmetry index and (c) mean mutual information.

Red +: side-flank offset tuning. Blue ×: end-flank offset tuning. The red and blue clouds of dots are Monte Carlo simulations of data for side- and end-flank modulation, respectively.

A third measure of the difference in tuning under the task-relevant and task-irrelevant conditions as well as the utility of the changes in performing the task was that of mutual information. Information theory (Shannon, 1948)

provides a rigorous measure of the changes in offset tuning functions between task-relevant and irrelevant conditions, and also tells us to what extent an ideal observer could categorize the stimulus from knowing the spike count of one cell during one trial. The corresponding measure is the average mutual information between the response of a cell in a trial and the task stimulus eliciting it.

For both side- and end-flank modulation, the mutual information was significantly higher in the task-relevant than task-irrelevant condition (Figure 4c). This analysis indicates that the offset-tuning function in the task-relevant condition was not only significantly different from the task-irrelevant condition, but also could have been useful to the animal in performing the discrimination task.

One should be aware that, when examining the side-flank (or end-flank) modulation, stimulus conditions sharing the same side-flank (or end-flank) position but different end-flank (or side-flank) positions were pooled (Figure 1a). This would have the effect of decreasing the mutual information due to

an increase of the variability of neuronal responses induced by position shifts of the flanks not relevant to the analysis, and thus the values shown in Figure 4c are only conservative estimates.

2.3.2 EFFECTS OF SPATIAL ATTENTION

In the experiments described so far, the animal did the discrimination tasks by attending to the five-line stimulus at the RF location (Figure 5a, 1 and 3). To explore the influence of the locus of spatial attention on contextual tuning, we introduced two 'attending-away' conditions (Figure 5a, 2 and 4). In the attending-away conditions, the animal did the bisection or vernier task in the hemifield contralateral to the RF, while the five-line stimulus array was still presented at the RF location as a test stimulus. Because of the number of stimulus conditions in this experiment (4 behavioral conditions times 25 stimulus patterns), we were able to complete this set on a small proportion of recorded single units.

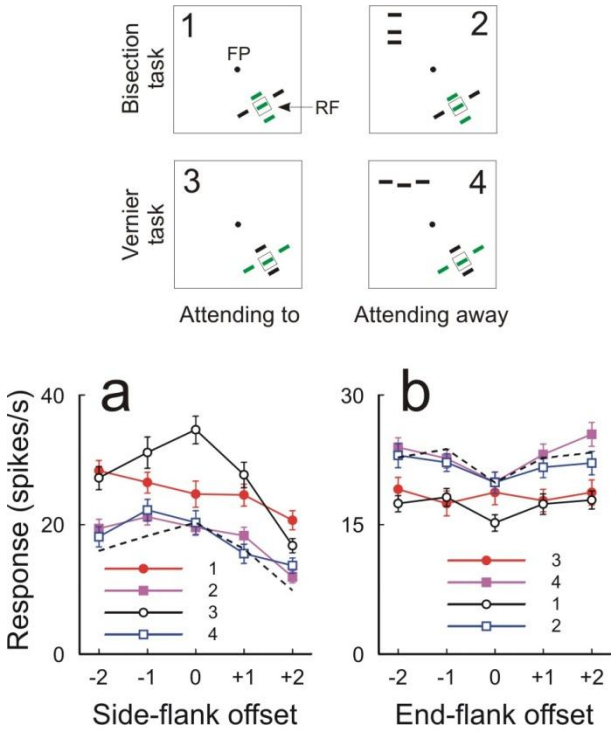

Figure 5. Effects of spatial attention.

(a) The cartoons illustrate the stimulus paradigms: 1, bisection task at the RF location; 2, bisection task in hemifield opposite to the RF of the recorded neuron; 3, vernier task at the RF location; 4, vernier task in the hemifield opposite to the RF.

(b) Side-flank offset tuning curves from a cell under the four test conditions. Each curve is numbered by the test condition shown in a. The dashed black curve represents curve 3 scaled to the level of curves 2 and 4. (c) End-flank offset tuning curves from another cell. The dashed black curve represents curve 1 scaled to the level of curves 2 and 4.

Error bars represent ± s.e.m.

The side-flank offset tuning functions under the four behavioral conditions are shown for a V1 neuron in Figure 5b. In the attending-to conditions,

consistent with Figure 2, the task-relevant tuning curve was very different from the task-irrelevant one. By contrast, the two curves in the attending-away conditions were very similar to each other, and both were substantially below the two curves made from the attending-to conditions. When the monkey did the vernier task at the RF location (where the side flanks were task-irrelevant), although the side-flank offset tuning curve was substantially above the two curves from the attending-away conditions, their shapes were similar to each other when scaled to the same level. These curves were very different, on the other hand, from the trials when the animal performed the bisection task at the RF location (where the side flanks were task-relevant). For this neuron, therefore, attending to the RF location by itself generally increased neuronal responses relative to the attending-away condition. At the higher response levels seen under the attending-to condition, performing the bisection task led to a further differential modulation of the side-flank offset tuning function relative to when the animal performed the vernier task.

Attending to the RF location did not always facilitate responses, as seen earlier (Haenny and Schiller, 1988; Motter, 1993). For some cells, we

observed a general suppression in the attending-to condition (Figure 5b), where the two end-flank offset tuning curves in the attending-away conditions were elevated substantially above the two curves in the attending-to conditions. The two curves in the attending-away conditions and the curve in the attending-to but task-irrelevant condition had similar shapes, as seen by rescaling, but all differed from the curve in the task-relevant condition.

The color change used to cue the task in the five-line stimulus did not interfere with the observation of a task-related effect (Figure 5) because the tuning curves in the two attending-away conditions were very close to each other despite the color difference in the five-line stimuli. Although the color change by itself in the absence of attention was insufficient to cause the differential effect that we observed, one might argue that an interaction between the color of the stimulus and attending to color could produce such a difference. To address this issue, we conducted a control experiment in which the monkey performed a simple three-line bisection task at the RF location. Here the two end flanks in the five-line stimuli were removed and only the three side-by-side lines were shown (Figure 6b insets). While the

color of the central line was kept green, the two side flanks were either green or white, as in the other experiments with five-line stimuli. For two representative cells (Figure 6a) the shapes of the offset tuning curves were almost identical regardless of the color difference of the stimulus flanks. The averaged mutual information across the population (Figure 6b) lay on the diagonal and within the cloud of Monte Carlo simulations, indicating no significant difference in mutual information between the two test conditions, where the task was the same three-line bisection discrimination, but the colors of the stimulus flanks were different. Here, because no pooling over end-flank positions was required, there was lower variability of neuronal responses at each side-flank position and thus higher mutual information than that seen in the experiments involving a five-line stimulus configuration.

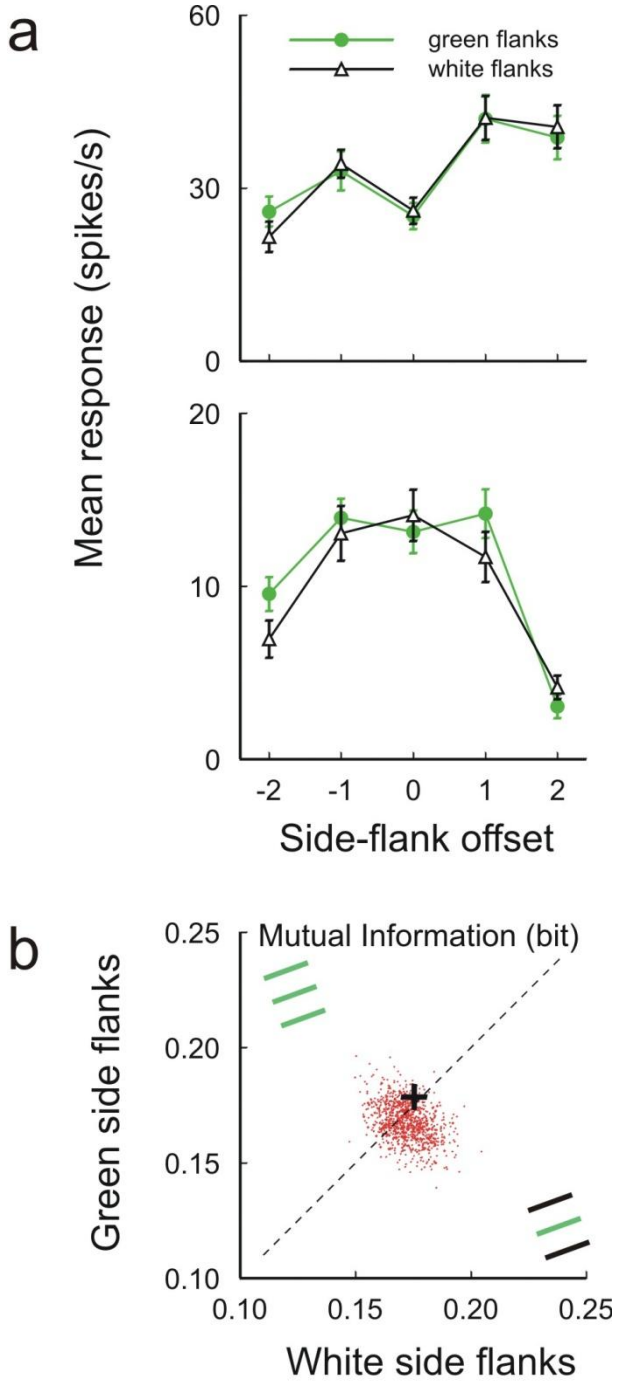

Figure 6. The color change had no significant effect on flank offset tuning function.

(a) Two examples of cells recorded while the animal did the bisection task at the RF location with a central green and two flanking white lines (black curve) and with three green lines (green curve).
(b) Comparison of the mean mutual information across the population of cells (n = 24) between two test conditions (as illustrated in insets). +: the mean value; red dots: Monte Carlo simulations.

2.3.3 TIMING OF THE TASK-DEPENDENT EFFECTS

One indication of the locus of interaction between the top-down influence of behavioral context and the lateral influences of sensory context can be obtained from the timing of the top-down influence relative to stimulus onset. To address whether the differential effect of task operated differently over the time course of neuronal responses, we compared the post-stimulus time histograms (PSTHs) of neuronal responses between the conditions when the stimulus flanks were task-relevant and task-irrelevant. The difference started at the beginning and lasted through the full time course of the neuronal responses (Figure 7).

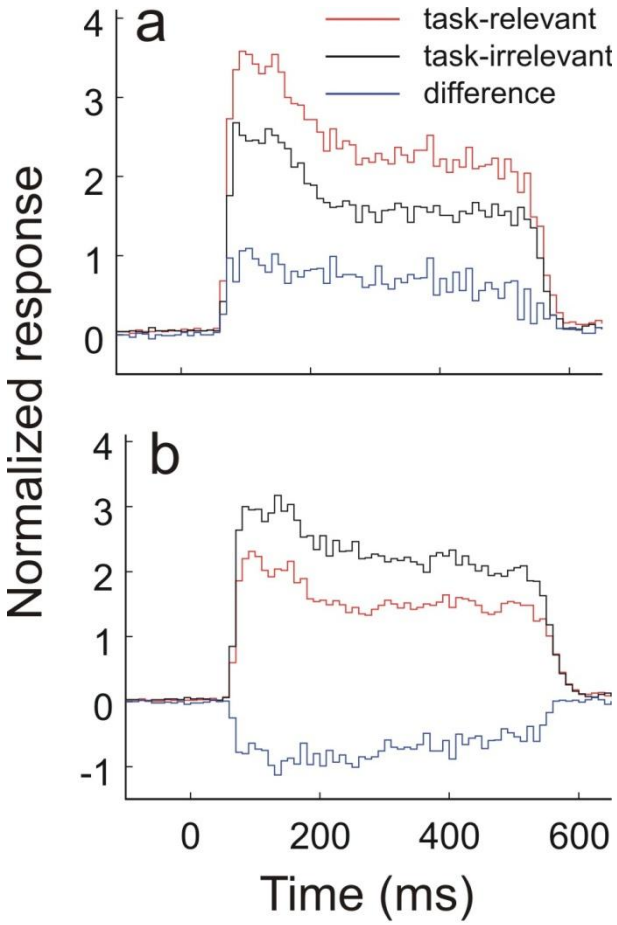

Figure 7. Population analysis of the timing of task-related effect.

Normalized and averaged PSTHs showing the time course of positive (a) and negative (b) differential effects. Time 0 represents stimulus onset. Time bin is 10 ms. Data from side- and end-flank offset tuning were pooled in this analysis. The original PSTHs corresponding to each data point on the offset tuning curve were normalized so that the area under the PSTH at zero offset was set to 100%. For the population analysis, we selected for each cell the pairs of normalized PSTHs in task-relevant and task-irrelevant conditions that gave significantly different mean responses (t-test, $P < 0.01$). At a given offset position of stimulus flanks the differential effects could be positive or negative (see Figs. 2 and 3 for examples), so the selected pairs of PSTHs across the population were separately pooled and averaged for the positive (a) and negative (b) cases, respectively. For

both positive and negative differential effects, the difference started at the beginning and lasted through the full time course of the neuronal responses.

Normalized and averaged PSTHs showing the time course of positive (a) and negative (b) differential effects. Time 0 represents stimulus onset. Time bin is 10 ms. Data from side- and end-flank offset tuning were pooled in this analysis. The original PSTHs corresponding to each data point on the offset tuning curve were normalized so that the area under the PSTH at zero offset was set to 100%. For the population analysis, we selected for each cell the pairs of normalized PSTHs in task-relevant and task-irrelevant conditions that gave significantly different mean responses (t-test, $P < 0.01$). At a given offset position of stimulus flanks the differential effects could be positive or negative (see Figure 2 and Figure 3for examples), so the selected pairs of PSTHs across the population were separately pooled and averaged for the positive (a) and negative (b) cases, respectively. For both positive and negative differential effects, the difference started at the beginning and lasted through the full time course of the neuronal responses.

2.4 DISCUSSION

We have previously shown that in monkey area V1, performing a bisection task near the RF in the trained visual field changes contextual influences with respect to a simple fixation task (Crist et al., 2001). In the present study, centering the task stimulus directly on the RF enabled us to disambiguate the influence of spatial attention per se (attending to or away from the RF location) from the influence of the perceptual task. As with our earlier work on top-down effects (Crist et al., 2001; Ito and Gilbert, 1999), the task-related effect operated most powerfully on contextual influences. In the attending-away conditions, however, the task-related effect was absent (Figure 5). These results suggest that the modulation of response properties of V1 neurons by perceptual task is neither global, because it was not seen when the animal performed the task in the opposite hemifield several degrees away from the RF, nor completely local, because it is seen when the task is done near the RF (Crist et al., 2001).

The diversity of profiles of differential tuning curves is reminiscent of the diversity of disparity-tuned cells in visual cortex. Disparity-tuned neurons in the striate and prestriate cortex have disparity tuning curves whose shapes

provide the information necessary for making fine-tuned depth judgments, including tuned-excitatory (caret-shaped), tuned-inhibitory (V-shaped) and near and far (asymmetric) (Poggio, 1995). The differential effects of task on offset tuning properties in V1 had profiles that recall those seen for disparity tuning. We propose that these neurons may collectively encode not only the absolute spatial displacement information but also the sign (direction) of the offset relative to the RF, which is a requirement for performing the discrimination tasks. This is reinforced by the increase in the asymmetry index for the task-relevant conditions (Figure 4b).

The finding that the task-relevant context produced tuning curves with greater modulation than the task-irrelevant context provides a possible neural basis for the discrimination involved in the trained tasks. In one sense, this observation is consistent with the 'competition model' of selective attention (for review, see refs. (Desimone and Duncan, 1995; Kastner and Ungerleider, 2000; Treue, 2001)). However, the task-related modulation did not simply change the gain of the cell, but instead changed the shape of the contextual tuning curves. This differs from the multiplicative effects of

space- or feature-based attention (McAdams and Maunsell, 1999; Treue and Martinez Trujillo, 1999). One might still consider this to be an example of object-oriented attention (Roelfsema et al., 1998), where the 'object' is either the three side-by-side lines or the three end-to-end lines. Whether or not one wants to include these task-dependent effects under the rubric of attention, the observation that the tuning properties of neurons were so markedly changed suggests that these neurons were altering their function to fit the requirements of the immediate perceptual task, and in essence changing their 'line-label' according to top-down signals. The broader implication is that, within certain limits, V1 acts as an adaptive processing unit, executing different algorithms according to the behavioral context. This idea is supported by the observation that neurons carried more information about a stimulus attribute when the animal performed a task related to that attribute (Figure 4c). A further implication is that the improvement in perceptual discriminability with training (Figure 1d) might be also associated with an increase in the mutual information between V1 neuronal responses and the task stimuli.

That the differential effects of task began from the outset of the neural responses suggests that the interaction between context and task was unlikely to occur in higher areas and subsequently to be fed back to V1, but rather occurred within V1. In our experimental paradigm, the top-down influence was likely to begin before stimulus presentation because a block design with some leading cuing trials was used. The top-down signal enables V1 neurons to multiplex visual signals for different perceptual tasks. Between the alternative possibilities of multiplexing (all cells participate in all perceptual tasks) and compartmentalization (cells become specialized in one task or another), our evidence suggests the former, with the same cells capable of coding information about vernier or bisection depending on the discrimination task. In fact, among the neurons responding to the orientation of the lines in the stimulus array, roughly half showed significant changes in their tuning according to the current task. The small increase in mutual information in the task-relevant condition relative to the irrelevant condition (Figure 4c) implies that a large number of cells may be required to obtain enough information to reliably solve the task. A rough estimation of the number of cells needed by an ideal observer on average to distinguish

among the five different offset positions in stimuli is made as follows. The five offset positions are equivalent to $\log_2 5 = 2.32$ bit in the binary system in which mutual information is usually quantified. Let us assume that the spike counts of different cells are statistically independent, that is, the mutual information contributed by each cell is linearly additive, and that the difference between task-relevant and irrelevant conditions is used in the bisection or vernier discrimination task. For an average increase of 0.02 bit in mutual information in side-flank modulation and 0.014 bit in end-flank modulation (Figure 4c), a total of 116 neurons for the bisection task and 166 for the vernier task would acquire the necessary 2.32 bit for reliable performance in the task for an ideal observer. This estimate is an upper bound for several reasons. As pointed out in (Figure 4c, our estimation of the mutual information is conservative as a result of pooling conditions. Moreover, even in the task-irrelevant condition, the mutual information in neuronal responses was not zero.

Our findings provide further evidence for the involvement of early visual cortical areas in representing learned information. V1 might be involved in

this learning because of the high-resolution requirements of the task (a few arc minutes of visual angle). V1 has been shown to become involved in learned search tasks, whereby the learned shapes have to be identified rapidly and in parallel within numerous distractors, requiring representation in retinotopically mapped areas (Sigman and Gilbert, 2000) (also Sigman, M. and C.D.G., unpublished data). Here, we found that the RF properties related to a task were present only when the animal was executing the learned task. This would indicate that the information related to a given shape would be represented at the level of subsets of inputs to a cell, and that multiple learned shapes can be represented by the same cells, greatly expanding the information carrying capacity of V1.

2.5 METHODS

2.5.1 STIMULUS AND BEHAVIOR PROTOCOLS.

Visual stimuli were generated by a VSG2/5 stimulus generator (Cambridge Research Systems) and consisted of oriented anti-aliased 0.5 times 0.08° light bars. The background was gray of 6.5 cd/m2. Bar contrast ranged

between 15% and 60%. In the five-line stimulus array (Figure 1a), the central line was fixed in the RF center and the flanks were placed near the edge of the RF. In different trials, the stimulus flanks were laterally displaced in varying steps with step size of 0.1° or 0.15°. All lines were oriented at the preferred orientation of the recorded neuron.

In stimulus presentation, all stimuli in the same task were randomized within a block of trials, and different tasks were arranged in different blocks and were preceded by a couple of cuing trials with only task-relevant lines shown. Each stimulus was repeated 8−20 times in different trials. At trial outset, a 0.08° fixation point (FP) was displayed in the center of the screen. Eye positions were sampled at 30 Hz by an infrared tracking system (Matsuda, K.et al., Soc. Neurosci. Abstr. 26, 744.2, 2000). Within 600 ms after FP presentation, the animal was required to fixate within an invisible circular window of 0.4−0.5° in radius around the FP. After the animal maintained fixation for 191 ms (20 CRT frames at 105 Hz), the stimulus was presented for 476 ms. Another 191 ms later, the FP was extinguished and two 0.15° saccade targets were presented for 600 ms symmetrically around

the screen center on an axis orthogonal to the orientation of stimulus lines. The animal indicated its choice by a saccade toward either target.

2.5.2 ANIMAL PREPARATION AND ELECTROPHYSIOLOGICAL RECORDINGS.

Two adult monkeys (Macaca mulatta) were trained to perform the discrimination tasks (Figure 1, see ref. (Crist et al., 2001) for details). Single units were recorded by a spike sorting and acquisition system (Plexon Inc.) with glass-coated platinum-iridium microelectrodes (FHC, impedance 1−3 MOmega at 1 KHz). All recordings were made from the opercular surface of the striate cortex with eccentricities between 1.8 and 3.5° and RF sizes between 0.5 and 1.5°, and were restricted to the superficial layers. Recording tracks were skipped where cells were weakly orientation-selective or strongly color-selective. All procedures were conducted in compliance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and under approval of the Institutional Animal Care and Use Committee at Rockefeller University.

2.5.3 DATA ANALYSIS.

The thresholds for bisection and vernier discrimination were determined by the method of probits (Finney, 1952), with thresholds defined as the half-distance between the 75% and 25% points on the psychometric curve (Figure 1b,c).

The responses of individual cells were analyzed by comparing tuning curves, while the behavior of the population was characterized by comparing the averages of three measures for the task-relevant and task-irrelevant condition: the modulation index, the asymmetry index and the mutual information between neuronal responses and the stimuli. The modulation index is the difference between the highest and lowest point of the tuning curve divided by their sum. The asymmetry index is the absolute value of the average slope on the tuning curve divided by the mean firing rate. The mutual information (Shannon, 1948), I(R;S), was calculated with the probabilities of presenting a stimulus $p(s_j)$, the probabilities of observing a particular number of spikes in the response $p(r_i)$ and the conditional

probabilities of observing a particular number of spikes in the response for a given stimulus $p(r_i|s_j)$:

To calculate the probabilities $p(r_i)$ and $p(r_i|s_j)$, the spike count was binned at the standard deviation of the response to all stimulus conditions for a given task and cell, rounded to the next integer. While changing the bin size from 30 to 300% of the standard deviation also changed the mutual information, it had little effect on the difference between task-relevant and task-irrelevant conditions because both had very similar limited sampling biases and regularization losses from binning (Panzeri and Treves, 1996).

After the neuronal response latency (50−70 ms after stimulus onset), a 400 ms window was used for the spike count. The population analysis used 51 of the 66 recorded cells with at least 8 trials per stimulus pattern and thus 40 trials when pooled over the positions irrelevant to the performed analysis.

The Monte Carlo simulation represents the null hypothesis that the difference in the population measures between two experimental conditions

(e.g., the mean modulation index between task-relevant and task-irrelevant conditions, (Figure 4a) resulted from random fluctuation of neuronal responses rather than the difference in experimental conditions. In the simulation, for each stimulus pattern and for each cell, individual trials were randomly assigned to either of the two compared conditions, resulting in a binomial distribution of trials from each condition.

2.6 ACKNOWLEDGMENTS

This work was supported by US National Institutes of Health grant EY07968. We are grateful to K. Matsuda for generously sharing the eye tracking software and G. N. Reeke for valuable comments on the data analysis with information theory. We also thank J. Jones, K. Hazleton and N. Lingenhol for technical assistance.

3 CONTOUR SALIENCY IN PRIMARY VISUAL CORTEX

Wu Li, Valentin Piëch and Charles D. Gilbert

Neuron 50, 951-962 (2006)

3.1 SUMMARY

Contour integration is an important intermediate stage of object recognition, in which line segments belonging to an object boundary are perceptually linked and segmented from complex backgrounds. Contextual influences observed in primary visual cortex (V1) suggest the involvement of V1 in contour integration. Here, we provide direct evidence that, in monkeys performing a contour detection task, there was a close correlation between the responses of V1 neurons and the perceptual saliency of contours. Receiver operating characteristic analysis showed that single neuronal responses encode the presence or absence of a contour as reliably as the animal's behavioral responses. We also show that the same visual contours

elicited significantly weaker neuronal responses when they were not detected in the detection task, or when they were unattended. Our results demonstrate that contextual interactions in V1 play a pivotal role in contour integration and saliency.

3.2 INTRODUCTION

To recognize an object in visual scenes, contour elements that belong together must be correctly linked and segregated from other contours, a process known as contour integration. Contour integration follows the Gestalt law of "good continuation" (Wertheimer, 1923), by which discrete contour elements positioned and oriented along a smooth path are readily grouped together (such as the straight contour shown in Figure 8C), forming a continuous visual contour that is globally salient and "pops out" from its background (Field et al., 1993). However, depending on the spatial configuration of contour and background elements, the perceptual saliency of contours can be different (compare the three contours shown in Figure 8; see also (Field et al., 1993; Li and Gilbert, 2002).

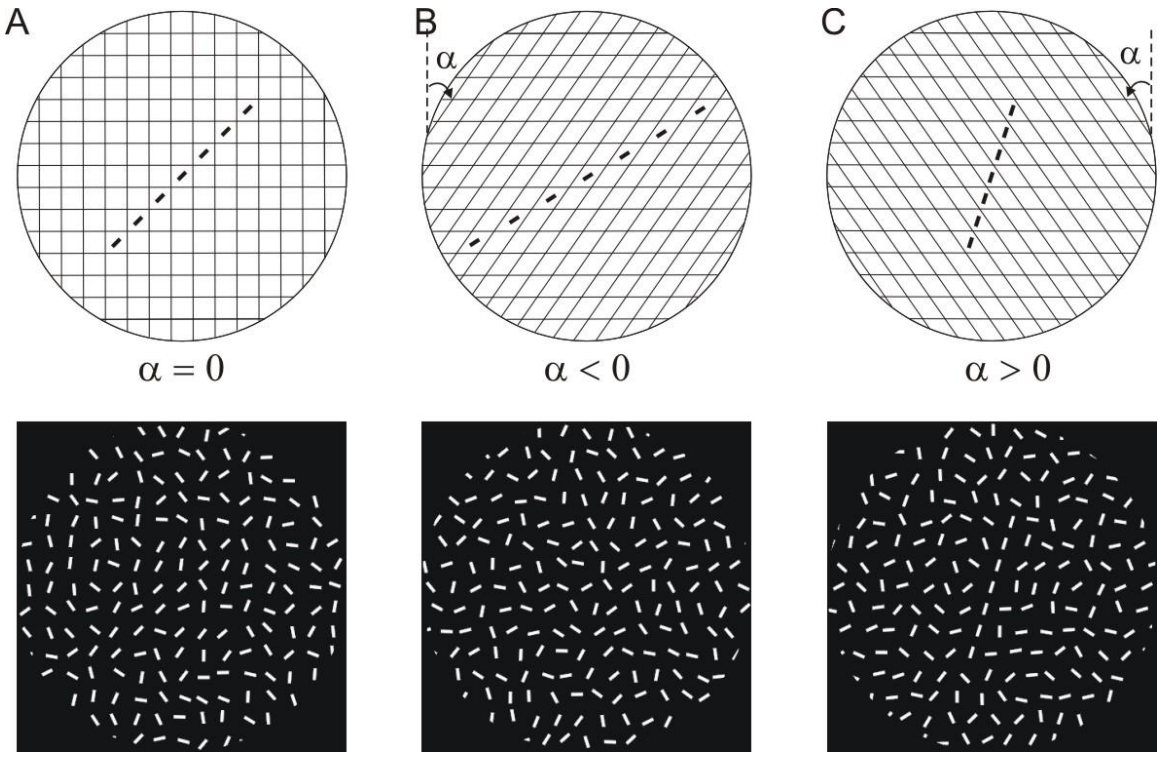

Figure 8. Design of Visual Contours.

(A) A circular area of 5.2° in diameter was divided into 0.4° squares (top panel). Each square contained a randomly oriented line segment of 0.2° by 0.05° (bottom panel), whose position was slightly jittered within the small compartment. The orientation and position jitter of each line was re-randomized in each trial. By aligning some adjacent line segments along a diagonal of the grids in a collinear, evenly spaced arrangement, a straight contour was generated in the center of stimulus pattern.

(B and C) To vary the spacing between collinear elements comprising the contour while keeping the density of line segments in the full stimulus pattern unchanged, a skew angle "α" was introduced to the grids. A clockwise skew ([B], $\alpha < 0$) increased the spacing between contour elements, while a counterclockwise skew ([C], $\alpha > 0$) decreased the spacing. Contour saliency was altered by changing either the number of collinear lines forming the contour or the spacing between them. The "average spacing" was defined as the width or height of an individual square or diamond compartment, which was fixed at 0.4°. The "(relative) collinear spacing," which was defined as the center-to-center distance between two adjacent contour elements, was calculated relative to the average spacing. Therefore, a relative collinear spacing of 1.0 represents an absolute spacing of 0.4°.

Evidence from physiological (Bauer and Heinze, 2002; Kapadia et al., 1995; Kourtzi et al., 2005; Kourtzi et al., 2003; Polat et al., 1998), psychophysical (Adini et al., 1997; Dresp, 1993; Field et al., 1993; Kapadia et al., 1995; Li and Gilbert, 2002; Polat and Sagi, 1994a), and anatomical (Gilbert and Wiesel, 1979, 1983, 1989; Rockland et al., 1982; Schmidt et al., 1997; Stettler et al., 2002) studies, as well as computational approaches (Ernst et al., 2004; Li, 1998b; Ullman, 1992; VanRullen et al., 2001; Yen and Finkel, 1998), suggests that contour integration, a form of intermediate-level vision, can be mediated by primary visual cortex (V1), the first stage in visual cortical processing. Support for the involvement of V1 in contour integration comes from parallel psychophysical measurements in human subjects (Kapadia et al., 1995; Li and Gilbert, 2002; Polat and Sagi, 1994b) and quantification of neuronal response properties (Kapadia et al., 1995) and circuits (Stettler et al., 2002) within V1. These experiments show that, though contour integration can extend across large areas of visual space, this global integration capability is based on a cascade of interactions of more limited spatial extent, and that the visuotopic extent of these interactions is comparable to that of the intrinsic horizontal interactions in V1.

Although the substrate for contour integration has been suggested to lie in V1 on the basis of the earlier studies, no direct correlation has yet been established between V1 responses and perceptual saliency of contours. In the current study, by taking advantage of the graded nature of contour saliency (Li and Gilbert, 2002), we examined in behaving monkeys the relationship between the responses of V1 neurons and the performance of the animals on detection of contours of various saliencies.

3.3 RESULTS

Two monkeys (MA and MB) were trained in a contour detection task, in which straight contours of various saliencies were made of a series of collinear line segments embedded in the center of an array of randomly oriented and positioned line segments (Figure 8).

3.3.1 CONTOUR SALIENCY AND NEURONAL RESPONSES IN V1

Within a trial in the contour detection task (Figure 9A), a contour pattern was presented randomly either at the receptive field (RF) location of the recorded cell or in the opposite hemifield. Simultaneously, a similar noise pattern without any embedded contour was displayed in the hemifield opposite to the contour pattern. After stimulus exposure, the animal reported which of the two patterns contained the embedded contour by making a saccadic eye movement to either of two dot targets displayed at the end of the trial. Responses of single orientation-selective neurons in V1 were recorded together with the animal's behavioral responses. The orientation of the contour in the contour pattern was adjusted to the preferred orientation of the recorded cell and so was the central line segment in the noise pattern.

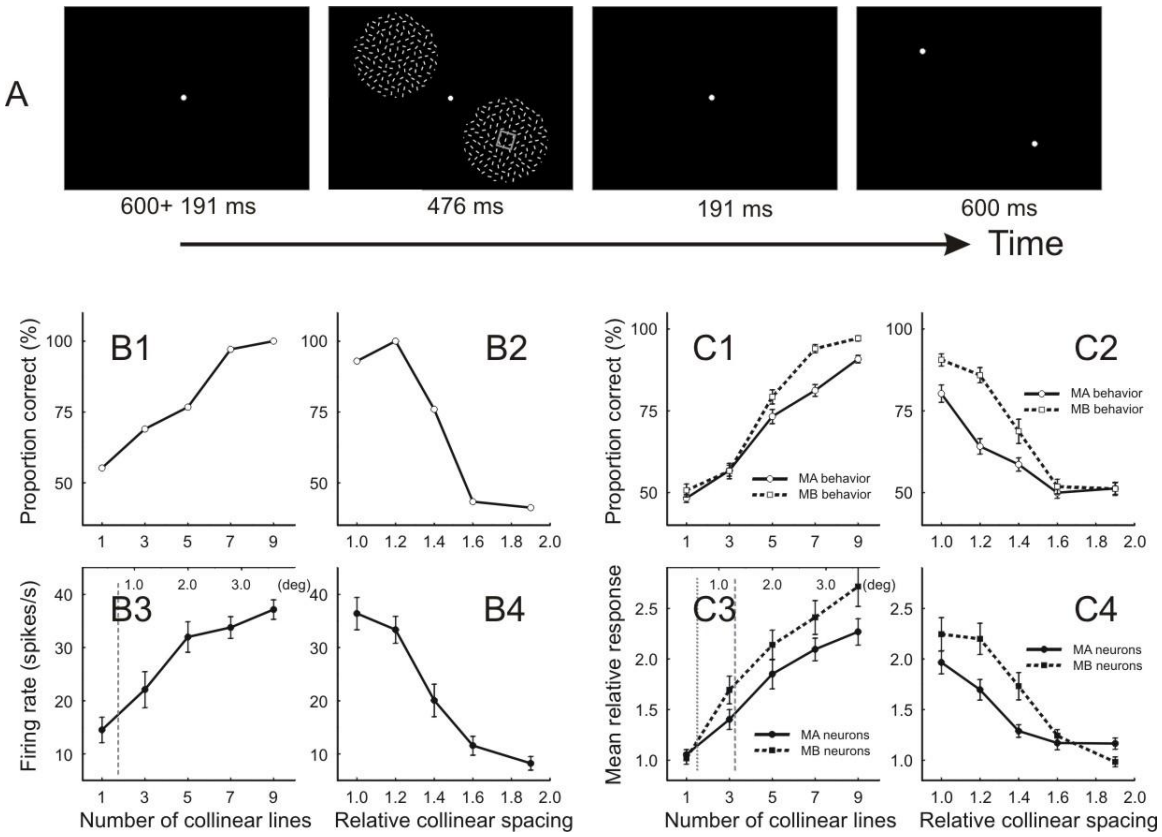

Figure 9. Behavioral and Neuronal Responses in the Contour Detection Task.

(A) Stimulus and behavior paradigm. A trial consisted of four consecutive intervals: fixation, stimulus exposure (the small gray square denotes the RF), continuing fixation, and choice (see Experimental Procedures for details).

(B1-B4) Data from subject MB in a typical recording session. (B1) The animal's performance on contour detection as a function of the number of collinear lines. When the number was 1, the two stimulus patterns were identical noise patterns, and the "contour pattern" was randomly assigned to either of them. The relative collinear spacing (for definition see Figure 1 legend) was fixed at 1.0 in this experiment. (B2) The animal's performance as a function of the relative collinear spacing. The contours consisted of seven collinear lines in this experiment. (B3 and B4) The mean responses of a V1 cell to the contour patterns recorded in the session shown in (B1) and (B2), respectively. This V1 cell had a RF of 0.55° by 0.70° in width and length, and orientation tuning width of 47° (see Experimental Procedures for definitions). The upper abscissa in (B3) indicates the

corresponding contour length in degrees of visual angle, and the RF length of this cell is indicated by the dashed vertical line.

(C1-C4) Population analysis. (C1) Proportion correct in contour detection, averaged across all recording sessions for MA (n = 30) and MB (n = 24) respectively, as a function of the number of collinear lines. (C2) Proportion correct for MA (n = 25) and MB (n = 20) as a function of the relative collinear spacing. (C3) The mean relative responses of all the recorded cells (n = 30 for MA; n = 24 for MB) as a function of the number of collinear lines. The "relative response" is defined as the ratio between the mean neuronal responses to the contour patterns and to the noise patterns. The mean and maximum RF lengths of the recorded cells are indicated by the dotted and dashed vertical lines. (C4) The mean relative neuronal responses (n = 25 for MA; n = 20 for MB) as a function of the relative collinear spacing. Error bars in all figures represent ±SEM.

Data collected from a typical recording session in MB are shown in Figure 9B1–Figure 9B4. We examined the animal's behavioral response as a function of two independent variables in two sets of experiments. In one set of experiments, contour saliency was increased by increasing the number of collinear lines (Figure 9B1). In the other set of experiments, contour saliency was decreased by increasing the spacing between collinear lines (Figure 9B2; see Figure 8 for a demonstration of the saliency change). The change of contour saliency was reflected in the change of the animal's performance on contour detection, which was around 50% (the chance level) for the least salient contours and was close to 100% for the most salient contours. These

observations are consistent with our previous study on human subjects (Li and Gilbert, 2002). The animal's behavioral performance, calculated as "proportion correct," was our psychophysical measure of perceptual saliency of contours in the current study. To explore the relationship between performance on contour detection and V1 responses, we recorded from neurons in the superficial layers of V1 during the same trials when the animal was performing the discrimination task (Figure 9B3 and Figure 9B4). Changes in the perceptual saliency of contours were closely correlated with the responses of the recorded cells. When contours were rendered more salient by increasing the number of collinear lines, the neuronal responses were increased monotonically (Figure 9B3). Conversely, when contour saliency was reduced by increasing the spacing between collinear lines, the neuronal responses were correspondingly reduced (Figure 9B4).

In both MA and MB, the effects of parametrically changing the contour settings on the animal's behavioral responses and on the neuronal responses were qualitatively similar in all recording sessions and in all recorded cells (Figure 9C1–Figure 9C4). Thus, the behavioral data from each experiment

were averaged across all recording sessions for the same animal (Figure 9C1 and Figure 9C2). To determine the mean response of all neurons, we calculated, for individual cells, the ratio between the mean responses to the contour patterns and to the noise patterns. This ratio is referred to as the "relative (neuronal) response" throughout the text. A relative response greater than 1.0 indicates facilitation of neuronal responses induced by the contours. The relative responses were averaged across all recorded cells in the same animal for the same test conditions (Figure 9C3 and Figure 9C4). Very similar to the single example shown in Figure 9B1–Figure 9B4, for both animals systematic changes in contour parameters resulted in parallel changes in the animal's behavioral performance and the neuronal responses (compare Figure 9C1 with Figure 9C3 and Figure 9C2 with Figure 9C4). In addition, for the most salient contours (detection performance close to 100%) neuronal responses were facilitated on average by more than a factor of two relative to the responses to the noise patterns. Note that the same visual contours seemed to be more salient on average to subject MB, whose overall performance was better than MA's (Figure 9C1 and Figure 9C2). Correspondingly, the facilitation of neuronal responses by the same visual

contours was stronger on average in MB than in MA (Figure 9C3 and Figure 9C4).

Note that the facilitatory effect induced by the embedded contours extended well beyond the classically defined RF. The RF lengths of all recorded cells were in the range of 0.2° and 1.3° with a mean of 0.6°, while nine collinear lines, the largest number of collinear lines tested in our experiments, extended 3.4° in visual space. Even at this contour length the collinear interactions had not reached a plateau, suggesting that the facilitatory interaction could extend over larger distances (Figure 9B3 and Figure 9C3; the vertical lines indicate the sizes of classical RF). While for cells with bigger RFs integration of contour elements could start within the classical RF, our data indicate that the boundary of the classical RF was not a property that was relevant to contour integration, since there was no correlation between the RF size and the facilitation strength (Figure S3 in the Supplemental Data available with this article online). Moreover, for all recorded cells, without exception, neuronal responses to very salient contours were enhanced relative to the noise pattern. This suggests that

contour integration is a generic feature of orientation-selective neurons in superficial layers of V1.

The data shown in Figure 9 imply a close correlation between neuronal responses and contour saliency. Note that saliency is a perceptual phenomenon that is not solely determined by stimulus configurations. Even for the same contour patterns the perceptual saliency can be different for different subjects (Figure 9C1 and Figure 9C2; see also (Li and Gilbert, 2002). Contour saliency can also be profoundly affected by practicing the detection task (Kovacs et al., 1999; Li and Gilbert, 2002) and can vary across experimental sessions depending on other factors, such as the observer's cognitive state. On the other hand, contours that are embedded in the same noise context but that have different configurations can be equally salient in perception. For example, a contour made up of five collinear lines can be as salient as seven collinear lines spaced further apart. Considering the combination of properties that contribute to contour saliency, a more appropriate analysis of the correlation between neuronal responses and contour saliency is to directly compare the neuronal responses with the

behavioral performance (Figure 10). To this end, we performed correlation analysis on the raw data obtained from individual recording sessions, where saliency was adjusted by changing the number of contour elements (Figure 10A1) or by changing the spacing between them (Figure 10B1). Each data point in the figure corresponds to a given contour configuration in a given recording session in which a pair of mean neuronal and behavioral responses was collected. Due to the response variability across and within sessions, the data points are scattered. Nevertheless, a strong and highly significant correlation is evident between neuronal responses and the animal's performance, as indicated by the large correlation coefficients (the r values in the figure). Note that, in Figure 10A1 and Figure 10B1, the relative neuronal response was plotted as a function of the animal's performance on contour detection, which is a psychophysical measure of contour saliency. In this case, the results obtained from changing the number of collinear lines (Figure 10A1) or changing the collinear spacing (Figure 10B1) could be related to a common measure, allowing us to pool the data from the two experiments (Figure 10C1). Moreover, the difference in the mean relative neuronal responses between MA and MB, which is substantial in Figure 9C3

and Figure 9C4, is negligible when the neuronal response is measured as a function of contour saliency rather than as a function of the stimulus parameters themselves (see the two nearly superimposed clouds of data points and regression lines in Figure 10A1, Figure 10B1, and Figure 10C1). All these data demonstrated a close correlation between perceptual saliency of contours and neuronal responses in V1. To further demonstrate this close relationship, for the same data shown in Figure 10A1, Figure 10B1, and Figure 10C1, respectively, we grouped the data points by dividing the animal's performance into five bins (Figure 10A2, Figure 10B2, and Figure 10C2). The data points falling within the same performance level were pooled and averaged to generate a single data point representing the averaged neuronal response and averaged performance. The binned and averaged data showed closer correlation than the raw data.

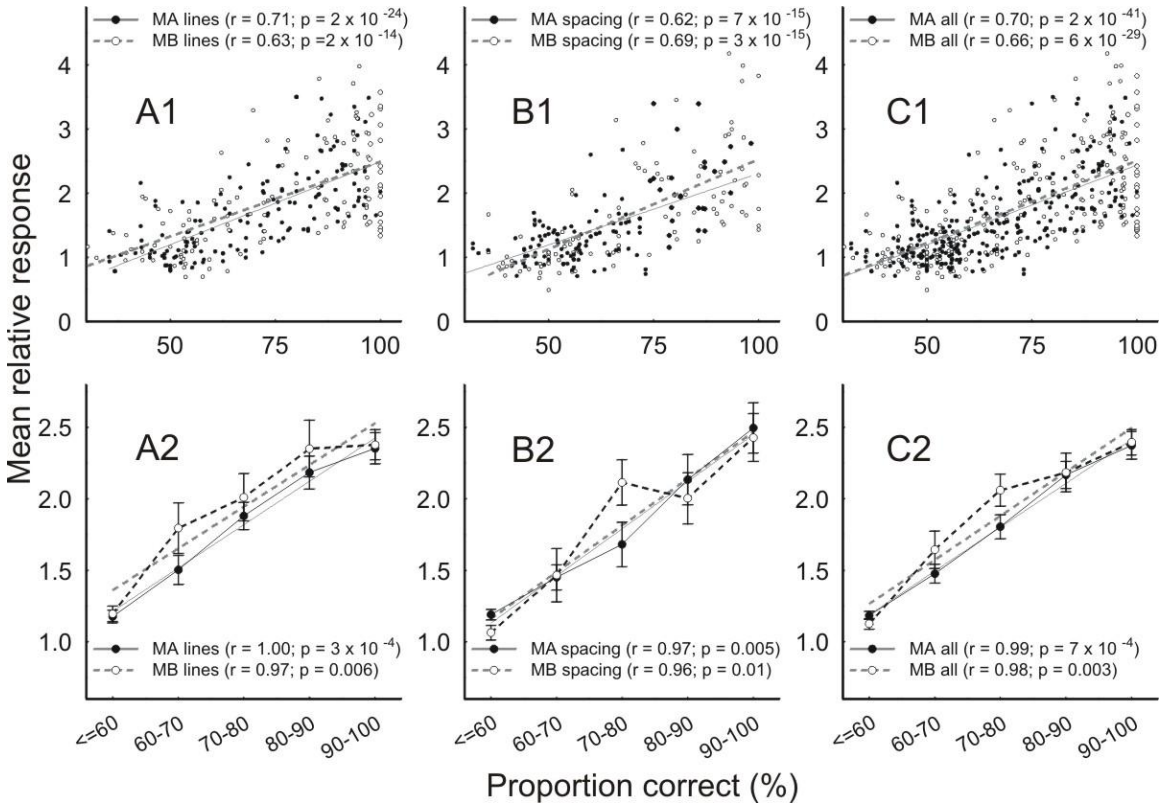

Figure 10. Correlation between Neuronal Responses and Contour Saliency

(A1, B1, and C1) The mean relative neuronal response obtained on a session by session basis for each animal is plotted against the animal's behavioral performance on contour detection, which is a psychophysical measure of contour saliency. Data are shown for the experiment of changing the number of collinear lines (A1), and for the experiment of changing the spacing between the contour elements (B1), and for both experiments pooled together (C1). The straight lines are from linear regression. The r value indicates the correlation coefficient, and the p value gives the probability under the null hypothesis that the observed correlation was due to random variation of responses.

(A2, B2, and C2) The same data shown in (A1), (B1), and (C1) were binned and averaged according to the specified levels of performance. Error bars represent ±SEM.

Psychophysical evidence has shown that contour integration involves complex interactions between contextual stimuli (Field et al., 1993; Li and Gilbert, 2002). In this regard, contour integration is a stimulus-driven, bottom-up process. To further explore the process of contour integration in V1 in the absence of top-down control, we compared neuronal responses to the same contour patterns for the same cells when the animal performed a task unrelated to contour detection: MA did a three-line bisection discrimination task presented in the visual field opposite to the RF (Figure 11A1; see Experimental Procedures), and MB did a simple fixation task (Figure 11A2). In these conditions neuronal responses were mainly driven by the stimuli, and any potential top-down components were largely removed. We plotted the mean relative population response to the contour patterns for correct trials in the contour detection task as a function of contour saliency and compared these responses to trials when the animals performed a task unrelated to contour detection (Figure 11B). In this figure contour saliency was determined as the animal's performance in the detection task and was binned as indicated in the figure. The same saliency

values were assigned to the corresponding contour patterns used in the unattended condition in order to compare neuronal responses between contour-attended and contour-unattended conditions. Data from both animals in both experiments were pooled, since the neuronal responses were very similar in the detection task when binned according to contour saliency (refer to Figure 10).

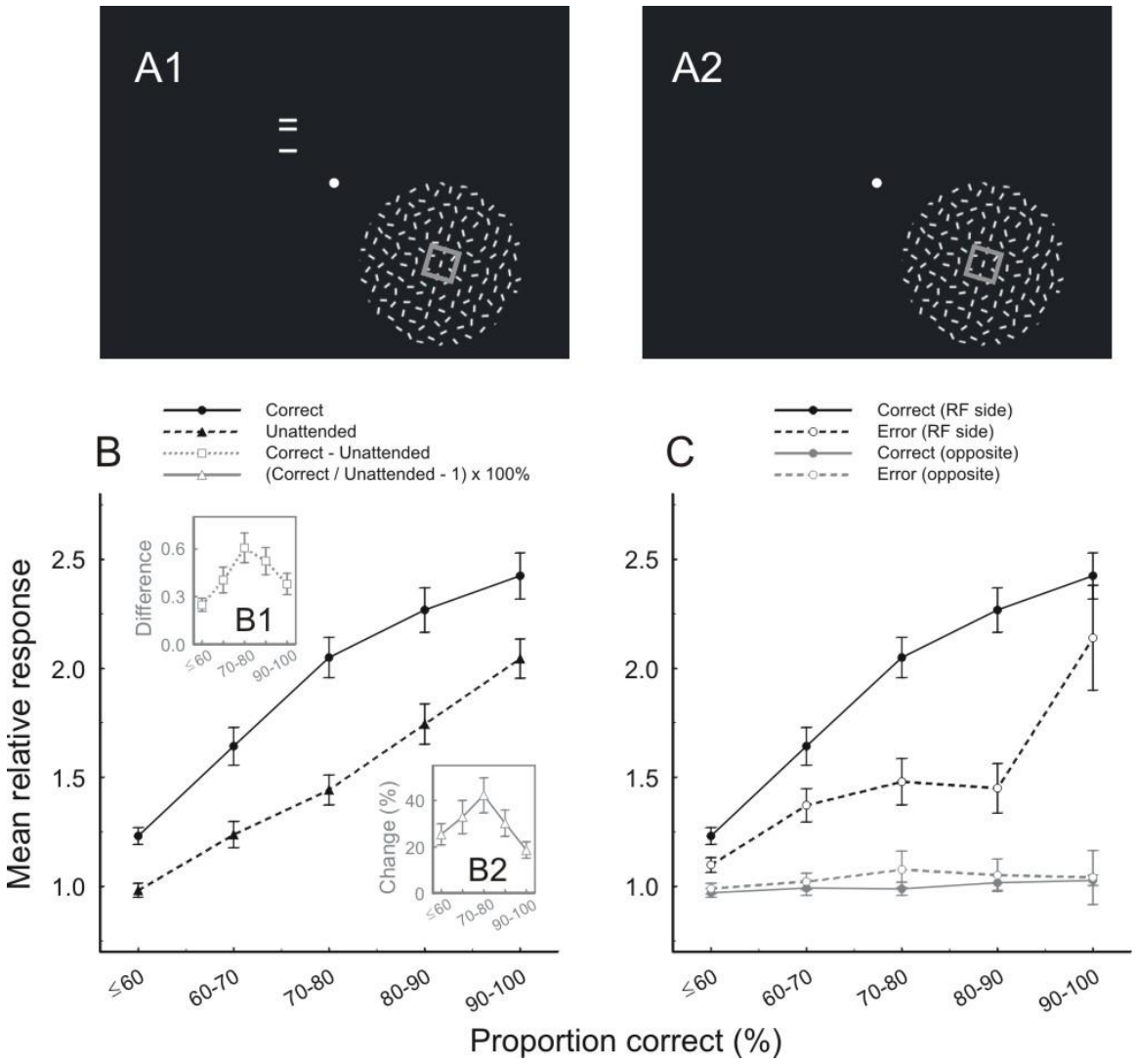

Figure 11. Contour Integration in the Absence of Top-Down Influences, and in Correct versus Error Trials

(A1 and A2) Stimulus paradigm. During V1 recordings, MA did not do the contour detection task at the RF location (denoted by the small gray square) but instead performed a three-line bisection discrimination task in the opposite hemifield (A1). MB performed a fixation task (A2).

(B) The mean relative responses of the neuronal population, for correct trials in the detection task and for the trials in the unattended condition, are plotted against contour saliency, which was measured as the animal's performance in the detection task but was also assigned to the same stimulus patterns used in the unattended condition. Data from both animals in both experiments were pooled and bootstrapped to calculate the mean

responses and SEM. The differences between the two curves were highly significant (correct versus unattended; $p < 10^{-5}$ for all data points, determined by bootstrapping). Inset (B1) shows the difference in mean relative responses between correct trials and unattended trials. The middle data point (corresponding to 70%-80% correct) is significantly different from the leftmost data point (≤60% correct; $p < 10^{-4}$ determined by bootstrapping) and the rightmost data point (90%-100% correct; $p < 0.02$). Inset (B2) shows the percentage change of neuronal responses in correct trials relative to the unattended trials. The middle data point is significantly different from the leftmost data point ($p < 0.02$) and the rightmost data point ($p < 10^{-3}$). (C) Comparisons of the mean relative responses of the neuronal population between correct and error trials for the condition when the contour patterns were displayed on the RF side and the noise patterns on the opposite hemifield (dark curves), and for the condition when the positions of the contour and noise patterns were switched with the noise patterns on the RF (gray curves). Error bars in all figures represent ±SEM obtained by bootstrapping.

As shown in Figure 11B, neuronal responses to the same contour patterns in the unattended trials were significantly lower than those in correct trials in contour detection task. Moreover, the difference was neither additive (inset B1) nor multiplicative (inset B2). At an intermediate level of contour saliency (corresponding to 70%–80% correct) the absolute (inset B1) or relative (inset B2) difference between attended and unattended conditions increased by about a factor of two relative to the lowest (≤60% correct) or highest (90%–100% correct) level of saliency. These differences were statistically significant as determined by bootstrapping (for details see Figure 11 legend and Experimental Procedures). Although attention greatly boosted

neuronal responses, there was still a clear correlation between responses and saliency in the unattended condition, indicating an important role of stimulus-driven, bottom-up processes in contour integration.

In some trials of the contour detection task, the animals failed to detect the embedded contours. Comparing correct and error trials when the contour patterns were displayed on the RF side showed a significant difference in firing rates (Figure 11C; correct [RF side] versus error [RF side]). In contrast, if the noise was presented at the RF location and the contour pattern was on the opposite side, no significant difference was observed between correct and error trials (Figure 11C, correct [opposite] versus error [opposite]). Thus, withdrawal of attention has a much greater effect on the responses to embedded contours than on the responses induced by the noise pattern.

3.3.2 ANALYSIS OF RELATION BETWEEN BEHAVIORAL AND NEURONAL RESPONSES

The relation between V1 responses and contour saliency allowed us to perform additional analysis to determine the degree to which V1 responses were predictive of contour saliency. Receiver operating characteristic (ROC) analysis (Parker and Newsome, 1998; Tolhurst et al., 1983) was employed to calculate the probability that an ideal observer can correctly discriminate the contour pattern from the noise pattern using only a single neuronal spike count. By calculating this probability as a function of the stimulus parameters, (the number of collinear lines or the spacing between them), we obtained a neurometric curve, which indicates how well the contour and the noise pattern could be differentiated at different contour settings simply based on a single spike count. Comparing this neurometric curve with a psychometric curve (the animal's performance on contour detection as a function of the stimulus parameters) gives an indication of how useful a single neuronal response could be for making a behavioral decision.

The neurometric curves (averaged across all cells) and the psychometric curves (averaged over the corresponding recording sessions) in the contour detection task were largely in agreement (Figure 12). Therefore, given the optimal strategy (which is implied by "ideal observer"), a single spike count would be generally predictive of the animal's performance in the contour detection task. Detailed examination of Figure 12 reveals two noticeable discrepancies between neurometric and psychometric curves. One discrepancy is that the actual behavioral performance in the most salient condition is somewhat better than predicted from single neuronal responses using ROC analysis (for example, the rightmost data points in Figure 12A1 and Figure 12B1). A situation similar to this has been explained by more cells sampling the visual space outside the RF of the recorded single unit(Parker and Newsome, 1998). Thus, this is not unexpected in view of the fact that there are many additional neurons with RFs lying along the contour that can contribute to the perceptual salience of the contour. The other discrepancy is that in Figure 12B1 but not in Figure 12A1 the neurometric curve is significantly elevated above the psychometric curve at the data point corresponding to three collinear lines. This may be explained by a bias by

one of the subjects (MB) toward selecting the RF side when presented with low-saliency contours. Note that the neurometric curves from the contour detection task are elevated above those from contour-unattended conditions and are closer to the psychometric curves. This indicates that neuronal responses in the contour detection task were better predictive of the animal's behavioral performance than those in unattended conditions.

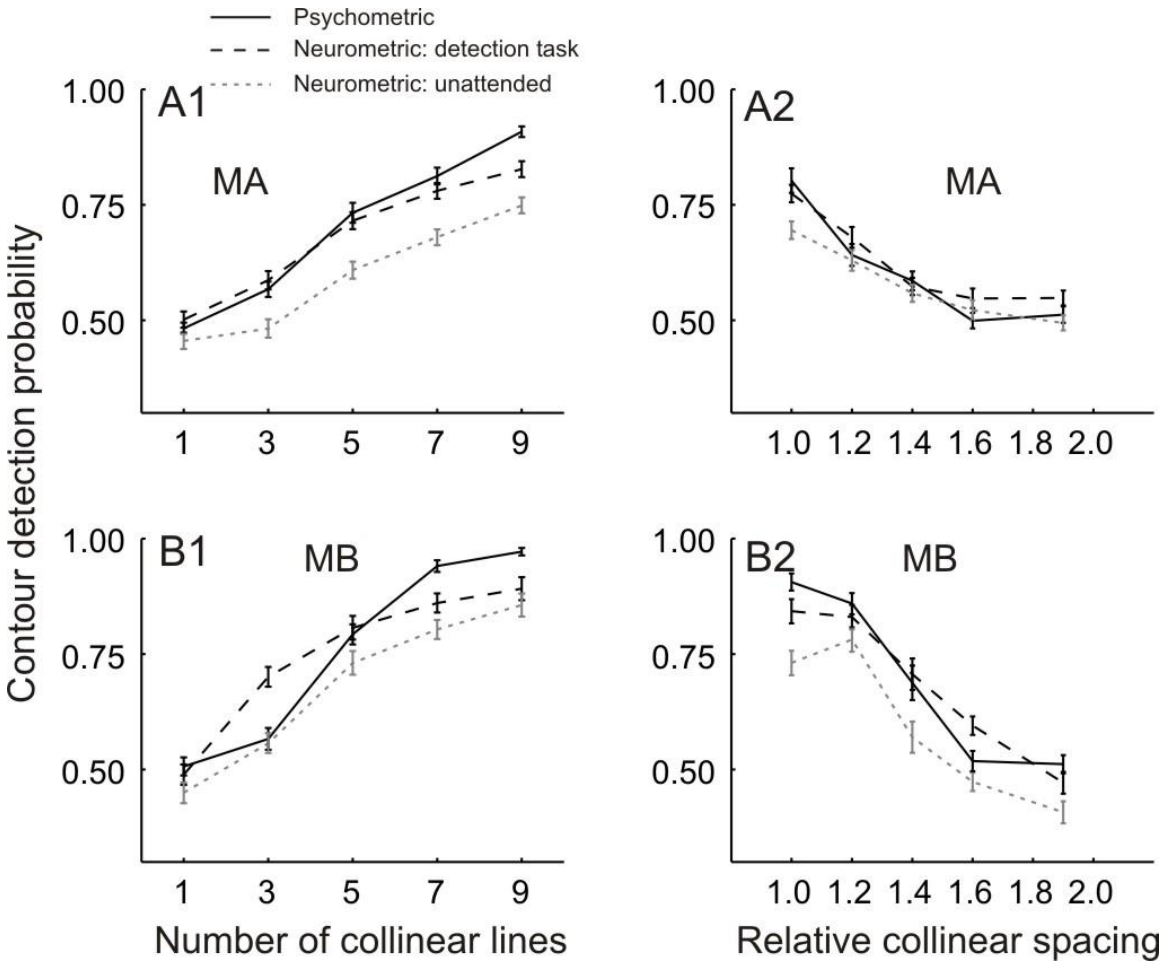

Figure 12. ROC Analysis

The animal's performance on contour detection (psychometric curve) and the neurometric curve (see Experimental Procedures) are compared. Data from MA and MB are shown in the top (A1 and A2) and bottom (B1 and B2) panels, respectively. The left and right panels show the results from two different experiments, changing the number of collinear lines (A1 and B1) and changing the collinear spacing (A2 and B2). Neurometric curves are shown for both the contour detection task condition and the contour-unattended condition. The neurometric curve was averaged across all cells, and the psychometric curve was averaged over the corresponding recording sessions. Error bars represent ±SEM.

Information theory was also used to calculate the mutual information between the stimuli and neuronal responses, which tells us how much the uncertainty whether the contour or the noise pattern was displayed over the RF location was reduced by knowing the spike count of one cell during one trial. Results from mutual information analysis, either directly based on the spike count of neurons (Figure S4), or as a control based on the probabilities from ROC analysis (Figure S5), were in close agreement with the ROC analysis.

3.3.3 FIGURE-GROUND EFFECTS

An isolated visual contour without any context is always conspicuous, but within a complex environment perceptual saliency of the contour depends on the global context within which the contour appears (Li and Gilbert, 2002). For the figure-ground segregation involved in the presence of a complex background, one might expect that even without top-down control the background itself influences the characteristics of neuronal responses and is critical for the aforementioned facilitatory effects in V1 to take place. To

measure center-surround interactions in the absence of top-down influences, particularly as they pertain to the influence of a complex background on contour integration, we recorded neuronal responses when the animals performed a three-line bisection task in the opposite hemifield (MA; Figure 13A1) or a simple fixation task (MB; Figure 13A2). The collinear lines were presented at the RF location, either with the surrounding background of randomly positioned and oriented lines (Figure 11A1 and Figure 11A2), or in isolation with the background removed (Figure 13A1 and Figure 13A2). By comparing the contour-unattended conditions with and without the complex context, we observed that the background itself produced very strong inhibition of neuronal responses, up to a 60% reduction in firing rates when the number of collinear lines was small (Figure 13B) or when the spacing between collinear lines was large (Figure 13C). Moreover, the effect of increasing the number of collinear lines forming the contour was strikingly different with and without the presence of the complex environment (Figure 13B). Without the background, at the contrast used (Michelson contrast 50%), a small proportion of cells showed facilitation by three collinear lines compared with a single line in the RF. Adding more

collinear lines tended to inhibit neuronal responses as the contour was increased in length. In contrast, in the presence of the complex background, neuronal responses increased monotonically with increasing contour length for nearly all recorded cells. This finding highlights the complexity of contextual effects in V1: the contextual interaction between collinear lines could be either inhibitory or facilitatory, depending on the greater stimulus context.

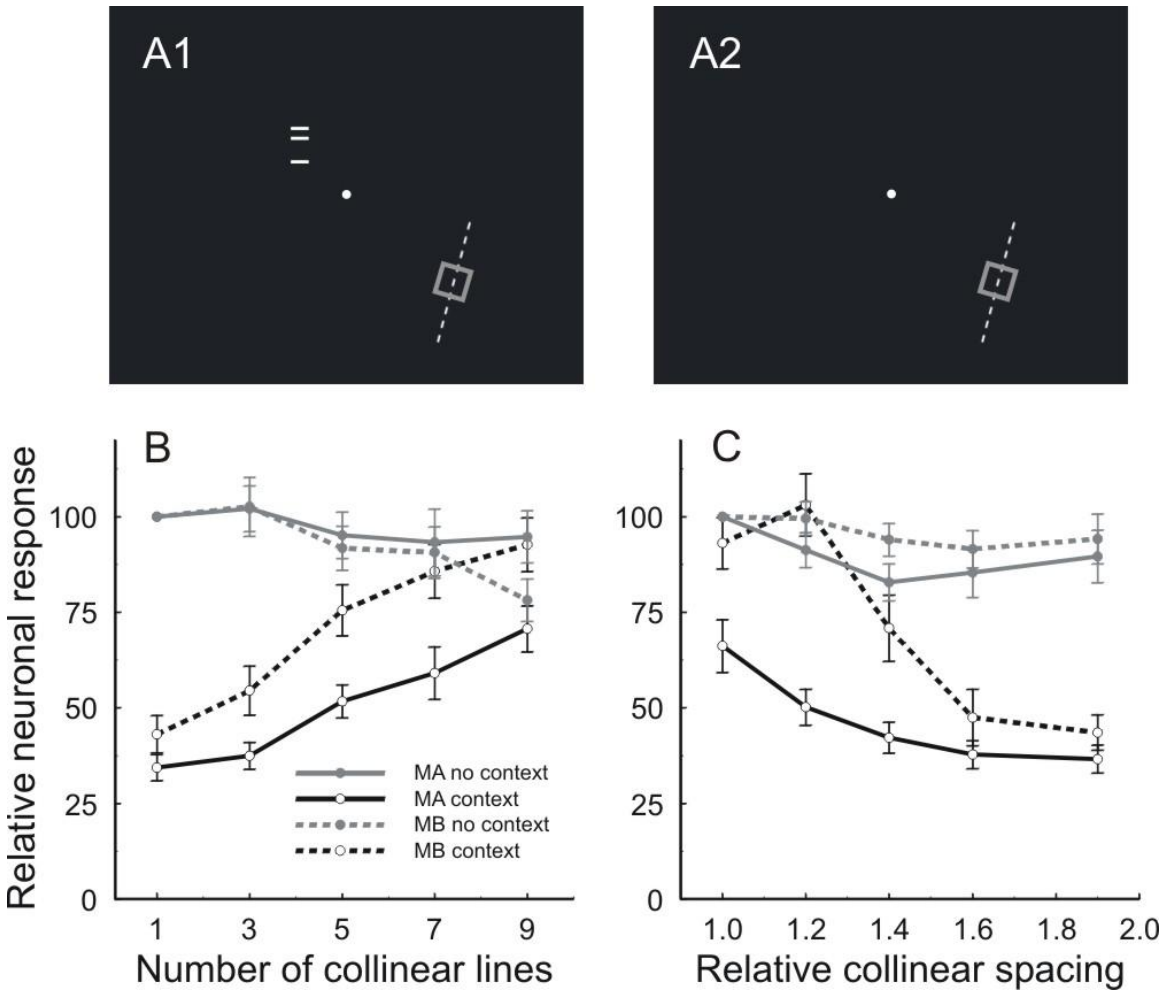

Figure 13. Figure-Ground Interactions

(A1 and A2) Stimulus paradigm. MA performed a three-line bisection discrimination task (A1) and MB a fixation task (A2), while the collinear lines alone were presented at the RF location (denoted by the small gray square). (B) Mean responses of neurons (n = 30 for MA; n = 24 for MB) as a function of the number of collinear lines. In this particular figure, a relative response of 100 represents the mean response of a neuron to a single line element in the RF. Data from the conditions with (dark curves) and without (gray curves) the complex background are compared for each subject. Error bars represent ±SEM. (C) Mean responses of neurons (n = 25 for MA; n = 20 for MB) as a function of the relative collinear spacing. In this figure, a relative response of 100 represents the mean response of a neuron to the contour alone with seven collinear lines and a relative collinear spacing of 1.0.

3.3.4 TIME COURSE OF SALIENCY EFFECTS

By examining the neuronal responses over time after stimulus onset (peristimulus time histogram [PSTH]), we observed that the neural signal for contour saliency was delayed relative to the outset of neuronal responses. The amplitude of the initial peak of the PSTHs recorded during the detection task was independent of the number of collinear lines within the contours (Figure 14A1 and Figure 14B1). The significant facilitation with increasing number of collinear lines was seen after the initial response peak and was maintained for the remaining time course of the response. The onset of facilitation was somewhat different for the two animals, beginning 150–160 ms after stimulus onset for MA and 90–100 ms for MB (see Experimental Procedures for statistical significance analyses). The response latencies (response outset after stimulus display) of neurons in MA and MB were 60 ms and 50 ms, respectively. Thus, V1 neurons in MB not only had a 10 ms shorter response latency and stronger collinear facilitation effect, but also a 60 ms smaller latency for the onset of facilitation. Analysis of PSTHs under different experimental conditions also showed that the late response components associated with contour saliency in the detection task were

stronger than those seen in the unattended condition, while the initial burst peak of neuronal responses was little affected by attentional state (no statistically significant difference for the first 150 ms of neuronal responses for MA and the first 100 ms for MB). On the other hand, the inhibitory influence of the complex background, as compared with the no-background condition, started in the initial phase of neuronal responses and rapidly reached a maximum within the period of the initial burst (Figure 14A2 and Figure 14B2; statistically significant inhibition could arise as early as the very beginning of the neuronal responses).

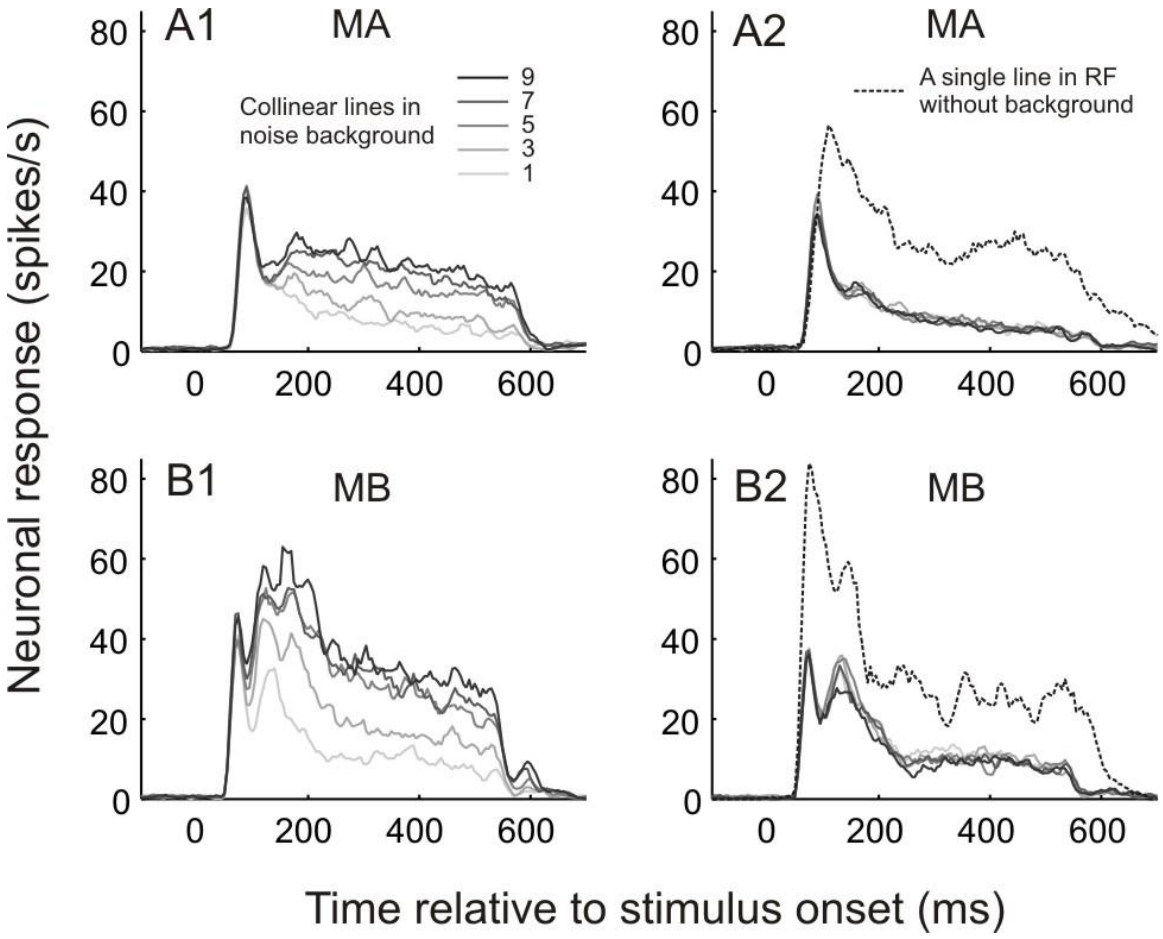

Figure 14. Time Course of Neuronal Responses

Peristimulus time histograms (PSTHs) were constructed by binning neuronal spikes at 5 ms resolution over time and averaged for all trials in all recorded cells in each animal. The PSTHs were smoothed using a rectangular window of 20 ms (boxcar filter). Time 0 indicates stimulus onset. (A1 and B1) PSTHs, for MA and MB, respectively, from neuronal responses in the detection task to contours made up of 1, 3, 5, 7, and 9 collinear lines embedded in the complex background, showing the delayed response components associated with contour saliency. (A2 and B2) The five continuous PSTH curves at the bottom, which are nearly superimposed, were constructed from neuronal responses to the noise patterns when the contour patterns corresponding to the five conditions described in (A1) and (B1) were displayed in the opposite hemifield. Thus, the stimuli in these instances were composed of a single optimally oriented line in the RF surrounded by the complex background of random lines. The single dotted curve shows the PSTH from the condition of a single line in the RF alone

without any context when MA did the bisection task in the opposite hemifield and MB did the fixation task. By comparing the solid curves with the dotted one we see that the inhibition induced by the complex background arises very early in the neurons' responses.

3.4 DISCUSSION

3.4.1 NEURAL BASIS OF CONTOUR INTEGRATION

The Gestalt law of "good continuation" that governs contour integration may have its neural substrate in V1. Earlier supporting evidence is that pyramidal cells in V1 with nonoverlapping RFs but similar orientation selectivity are wired together via intrinsic horizontal connections (Bosking et al., 1997; Gilbert and Wiesel, 1979, 1983, 1989; Rockland et al., 1982; Schmidt et al., 1997; Stettler et al., 2002) and that responses of orientation-selective neurons in V1 to an oriented visual stimulus in the RF can be facilitated by collinear context placed outside the RF (Ito and Gilbert, 1999; Kapadia et al., 1995; Polat et al., 1998). Previous studies from our laboratory (Ito and Gilbert, 1999; Kapadia et al., 1995) showed lateral interactions between collinear line segments with respect to brightness induction rather than to contour saliency per se. Other studies have investigated the neural correlates

of perceptual pop-out in V1 based on orientation contrast (Knierim and van Essen, 1992; Nothdurft et al., 1999), texture difference (Lamme, 1995), or more complex attributes(Lee et al., 2002), which base saliency on discontinuities rather than continuity.

Somewhat more direct evidence that V1 cells can mediate contour integration comes from studies using elongated contours displayed in complex environments, where neuronal responses were enhanced by visual contours (Bauer and Heinze, 2002; Kapadia et al., 1995; Roelfsema et al., 1998, 2004). These studies did not allow examination of the relationship between perceptual saliency and neuronal responses. Despite the implications from previous studies on the role of V1 in contour integration, the strongest evidence is to establish a direct correlation between neuronal responses and the perception of the animal in which and at the time the recordings are made, as done in the current study.

The nature of the neural code for contour saliency has been a matter of some debate. As suggested by some studies, contour saliency could be derived

from an increase of neuronal responses through facilitatory horizontal interactions (Bauer and Heinze, 2002; Kapadia et al., 1995; Kapadia et al., 1999; Li, 1998b; Polat and Bonneh, 2000; Roelfsema et al., 1998, 2004). Other studies suggest that temporal encoding could also be involved in representation of contour saliency (Gray et al., 1989; Li, 1998b; Singer, 1999; Yen and Finkel, 1998). However, recordings in monkeys trained in a contour-tracking task indicate that response synchrony is not related to contour perception (Roelfsema et al., 2004). Our current study clearly showed that the mean firing rate is at least one form of the neural code for contour saliency.

The contrast detection threshold for an oriented target is decreased by the presence of collinear flankers (Dresp, 1993; Kapadia et al., 1995; Polat and Sagi, 1993), and this enhancement increases with the number of collinear elements (Adini et al., 1997; Bonneh and Sagi, 1998). This perceptual phenomenon has its neural correlate in V1. The interaction between collinear line segments in V1 tends to be facilitatory at low contrast but inhibitory at high contrast (Kapadia et al., 1999, 2000; Polat et al., 1998). Our results are

in accord with and extend these earlier findings. At the contrast level used in the current study (50%), and in the absence of the complex background, a small proportion of neurons showed facilitation by a pair of flanking collinear lines, and adding additional collinear lines invariably led to a decline in response. However, by presenting a complex background around the contour, nearly all cells' responses increased monotonically with an increasing number of collinear lines. Our results support the notion that embedding a target in a complex environment leads to a shift toward a dominance of facilitation in collinear interactions, equivalent to the effect of a reduction in contrast (Kapadia et al., 1999; Polat and Bonneh, 2000). The complex interactions among the contour elements and the global contextual elements shown in the current study are also consistent with our previous psychophysical observation that contour saliency depends on not only the geometry of contour elements but also on the surrounding context (Li and Gilbert, 2002). In integrating the contour elements and generating the saliency map, there is a cascade of nonlinear interactions among stimulus elements across visual field areas much larger than the RF in V1. Therefore the ultimate output from a V1 neuron with its RF lying on the contour path

conveys information not only about the contour segment within its RF but also the global structure of visual stimuli extending over much larger areas of visual space.

3.4.2 DELAYED SALIENCY EFFECTS

Our data showed that neuronal responses to a line were greatly suppressed when the line was placed in a complex environment and even tended to be suppressed when the line was flanked by a stack of collinear lines. These observations are consistent with the evidence for inhibitory modulation from outside the classical RF (Bair et al., 2003; Bishop et al., 1973; Hubel and Wiesel, 1965; Knierim and van Essen, 1992; Li et al., 2000; Nothdurft et al., 1999; Walker et al., 2000). Moreover, this contextual inhibitory effect started in the initial phase of neuronal responses and reached a maximum very quickly, as reported previously (Bair et al., 2003; Knierim and van Essen, 1992; Li et al., 2000; Nothdurft et al., 1999). Though much of the earlier work emphasizes inhibitory contextual influences, the current study reinforces that the sign of these influences depends strongly on the position

of the stimuli (e.g., collinear) and on the greater context (contours embedded in a complex background). Collinear inhibition can be reversed to a powerful facilitation in the presence of the fast and strong inhibition induced by the complex background, in contrast to that seen without the background. Under these circumstances, the collinear facilitation associated with contour saliency developed much later than the background induced inhibition. This is reminiscent of earlier findings of delayed components in V1 responses related to higher-order, contextually dependent properties (Bauer and Heinze, 2002; Kinoshita and Komatsu, 2001; Lamme, 1995; Lee et al., 2002; Li et al., 2000, 2001; Roelfsema et al., 1998; Rossi et al., 2001; Super et al., 2001; Zipser et al., 1996), where complex center-surround interactions or top-down influences are involved. The source for these delayed components is a matter of debate. Some have speculated that delayed facilitatory response components are due to feedback (Lamme, 1995; Lee et al., 2002; Zipser et al., 1996), and others have argued that feedback is associated with early, inhibitory influences on neuronal responses (Bair et al., 2003). Alternatively, it is equally possible that, in the presence of a complex stimulus context, considerable delays could arise from an iteration

of cascading processes within V1 (Bauer and Heinze, 2002). In this view, recurrent interactions between V1 neurons via horizontal connections could require a period of time to reach a stable state, with horizontal propagation of activity in V1 being rather slow (Bringuier et al., 1999). Further support of figure-ground interactions arising within V1 comes from their persistence after inactivation of V2 (Hupe et al., 2001a)and the relative rapidity of feedback effects from higher cortical areas to V1 (Hupe et al., 2001b).

In our study, the delayed responses occurred both when the animal was performing the contour detection task, as well as when the animal did an unrelated task. While the delay in a component of neuronal responses could be explained in terms of the time required for directing attention to the stimulus, as is seen in extrastriate cortical areas (for review see (Treue, 2001), the fact that one still sees the delayed facilitation in the absence of attention suggests the involvement of the complex lateral interactions described above. But the conclusion that we can draw from the current study is that information about contour saliency is represented in V1. The connectivity underlying this interaction remains to be determined, but it is

tempting to speculate that it involves an interaction between feedback connections from higher-order visual areas to V1 and long-range horizontal connections intrinsic to V1. One may ask which components of the computation are contributed by V1. In this regard, it is worth considering the properties of contour integration in terms of its spatial extent and orientation dependence, which are consistent with the circuits and functional architecture of V1 (Gilbert and Wiesel, 1989; Li and Gilbert, 2002; Stettler et al., 2002). It has been suggested, however, that contour integration operates at multiple spatial scales, each of which engages different cortical areas, including V1 (Kourtzi et al., 2003).

3.4.3 BOTTOM-UP, TOP-DOWN, AND PERCEPTUAL LEARNING

In the current study we showed that information about contour saliency was encoded by V1 neurons even if the stimulus was not the focus of attention. This observation is in agreement with other studies, in which collinear facilitation (Kapadia et al., 1995) and contour-associated enhancement (Bauer and Heinze, 2002) of neuronal responses in V1 is observed in a

simple fixation task. Collinear facilitation is seen even in the primary visual cortex of anesthetized animals (Polat et al., 1998), although the degree of facilitation and the proportion of neurons showing facilitation are reduced in the anesthetized state. These results suggest that contour integration contains a bottom-up process driven by the configuration of visual stimuli.

In addition to this bottom-up contribution to contour integration, our data showed profound differences in neuronal responses when animals performed the contour detection task as compared with the contour-unattended condition. With our current experimental design we were unable to identify what forms of top-down influences were responsible for these differences. The range of possibilities includes spatial attention, object-based attention, the perceptual task of contour detection per se, or a combination of them. A clear answer to this question requires further investigation. But in any event, our data revealed striking top-down modulation in V1 with amplitude comparable to that reported in higher visual areas (see review (Maunsell and Cook, 2002)). While earlier studies have suggested that V1 only shows minimal attentional effects, with a hierarchy of attentional modulation along

the visual pathway, the effects presented here emphasize the importance of stimulus geometry and behavioral context in eliciting the strongest top-down effects. To see the attentional effects that are appropriate to any given cortical area, it is of key importance to explore them in the context of the specific integrative functions of that area. To wit, based on our finding that V1 is intimately involved in linking contour elements within a complex background for the purpose of contour integration and saliency, the strongest top-down influences in V1 are seen with a stimulus context that adheres to the rules of contour saliency (Field et al., 1993; Sigman et al., 2001; Wertheimer, 1923) and a behavioral context involving either attending to or detecting contours.

In the current study, the most significant top-down modulation was seen for contours of intermediate saliency rather than high saliency. This may be due to the difficulty in drawing attention away from a highly salient stimulus even in the unattended condition. As a result, the difference in neuronal responses is smaller between attended and unattended conditions for the most salient contours. On the other hand, one may need an effective stimulus

to deploy spatial attention, so that when no detectable contour is embedded in the complex background, the top-down modulatory effect is lessened due to a decrease in attentional resources dedicated to the target. The strongest top-down influences are consequently observed for contours of intermediate saliency.

Given the findings that performance on contour detection can be changed by experience (Kovacs et al., 1999; Li and Gilbert, 2002), it is likely that different observers may have different contour integration capability. In the current study subject MB's performance in contour detection was better than MA's for the same contour patterns. Interestingly, V1 neurons in MB also exhibited, on average, stronger collinear facilitation by the same visual contours as well as a shorter latency for the effect to develop. This suggests an experience-dependent difference in V1 responses between the two subjects, as supported by the findings that past experience with some pop-out targets enhances perceptual detection (Lee et al., 2002; Sigman et al., 2001; Wang et al., 1994) as well as neuronal responses (Lee et al., 2002) and fMRI signals (Kourtzi et al., 2005; Sigman et al., 2005) in V1. Moreover,

our earlier studies have also shown that perceptual learning can result in profound long-term changes in response properties of V1 cells (Crist et al., 2001; Li et al., 2004). The difference in facilitation and contour detection between the two experimental animals likely reflects different stages of perceptual learning in the contour detection task, but interestingly, despite their difference in performance, the two subjects were very similar in the relationship between behavioral performance and neuronal responses (Figure 10). Thus, the learning is unlikely to involve a change in the significance attributed to a given level of firing in V1 by a higher stage, but rather an alteration in the level of facilitation generated by a given stimulus. This could be achieved by a change in the interaction between top-down influences and lateral interactions within V1, as well as by a change in the lateral interactions themselves. The representation of learned information in lower visual cortical areas in other perceptual learning paradigms has also been shown to involve top-down influences, leading to global alterations in the representation of learned shapes across the visual pathway (Sigman et al., 2005).

In summary, our current study highlights the role of V1 in contour integration and the neural code for contour saliency. Taken together with our earlier studies (Crist et al., 2001; Li et al., 2004), our findings suggest that V1 acts as an adaptive processor whose processing is profoundly influenced by context, top-down control and training. Our findings have a broader implication for the general mechanisms underlying early cortical processing of visual information in a natural environment, where visual stimuli are rarely isolated but rather embedded in a rich context, and viewing is usually driven by behavioral goals and shaped by past experiences.

3.5 EXPERIMENTAL PROCEDURES

3.5.1 STIMULUS AND BEHAVIOR PROTOCOLS

Stimuli were generated by a visual stimulus generator (VSG2/5, Cambridge Research Systems) on a monitor (NANAO FlexScan F2-21) at a resolution of 1024 by 769 pixels and a refresh rate of 105 Hz. The viewing distance was 145 cm. The stimuli consisted of an array of randomly oriented white (19.5 cd/m2) line segments displayed on a gray (6.5 cd/m2) background.

The positions of line segments were defined by geometric rules that allowed precise control of stimulus parameters (Figure 1; for more details see (Li and Gilbert, 2002)).

Two adult male monkeys (Macaca mulatta, referred to as MA and MB) were trained in a contour detection task. A two-alternative forced choice (2AFC) task was used to measure the animal's ability to detect contours of various saliencies (Figure 9A). A stimulus pattern containing an embedded contour (contour pattern) and a similar pattern without any embedded contour (noise pattern) were displayed simultaneously at two locations symmetrical around the fixation point. During V1 recordings the RF center of the recorded cell was one of these two locations. The probability that a given pattern was display at a given location was 50%. The task was to report which of these two patterns contained a contour. Note that the noise pattern was derived from the contour pattern simply by randomizing the orientations and position jitters of the collinear lines forming the contour, leaving the central line segment in the pattern and all background noise elements unchanged. Therefore, the very central line segment and all the noisy contextual lines in

the contour pattern were identical to those in the noise pattern. The central line segment in both contour and noise pattern was set at the optimal orientation of the recorded cell and was centered in the RF for the stimulus pattern presented at the RF location. Different stimulus conditions in an experiment were randomized, and different experiments were arranged in different blocks. Each stimulus condition was repeated five to ten times in a block of trials, and typically two to four blocks of trials were repeated for the same experiment. At trial outset a 0.08° fixation point (FP) was displayed in the screen center. Eye positions were sampled at 30 Hz by an infrared tracking system (K. Matsuda et al., 2000, Soc. Neurosci., abstract). Within 600 ms after FP presentation the animal was required to fixate within an invisible circular window of 0.5° in radius around the FP. After the animal maintained fixation for 191 ms, the contour pattern and noise pattern were presented for 476 ms. Another 191 ms later, the FP was extinguished and two 0.15° saccade targets were presented for 600 ms at the locations where the two stimulus patterns were centered (Figure 9A). The animal indicated the hemifield in which the contour pattern was presented by executing a saccadic eye movement to either target.

One animal (MA) was used in our previous study and had been trained to perform a three-line bisection discrimination task (Li et al., 2004). In the experiments examining neuronal responses in the absence of potential top-down influences, MA performed the bisection task while exposed to the contour patterns at the RF location (Figure 11A1 and Figure 13A1). The task stimulus was three parallel lines displayed in the hemifield opposite to the RF. In different trials the positions of two flanking lines were fixed while the middle line was displaced up or down to varying extents from the center. The monkey reported to which of the two flankers the middle line was closer by making a saccade to either of two targets (for more details see (Li et al., 2004)). Since training the bisection task is time consuming, in the other animal (MB), a fixation task was performed in the unattended condition (Figures Figure 11A2 and Figure 13A2), which had the desired effect of directing the animal's attention away from the contour display. The animals were cued to the to-be-performed task by a few leading trials in which only the task-relevant stimuli were presented and all irrelevant stimuli were omitted. Only if the animal made a few correct responses to the task did the

actually experimental trials start. Moreover, the first few experimental trials were used as practice trials, and only if the animal made correct choices in the practice trials did data collection begin.

3.5.2 ANIMAL PREPARATION AND ELECTROPHYSIOLOGICAL RECORDINGS

Details were described elsewhere (Li et al., 2004). Orientation-selective cells in V1 superficial layers were recorded by a spike sorting and acquisition system (Plexon Inc.) with platinum-iridium or tungsten microelectrodes (impedance 0.5–2 MΩ at 1 KHz). RF eccentricities ranged between 2.5° and 4.5°, RF lengths between 0.2° and 1.3° (0.64° ± 0.28°; mean ± SD), RF width between 0.3° and 1.2° (0.66° ± 0.21°). The RF length or width was determined using an optimally oriented bar, 0.25° by 0.1° in size, presented 0.25° apart along or orthogonal to the orientation axis across the RF. The mean neuronal responses corresponding to different bar positions were fitted with a Gaussian function. The RF center was defined as the position corresponding to Gaussian peak, and the RF length or width was defined by

the interval containing the central 95% of the area under the Gaussian (2 × 1.96 SD) with the length or width, respectively, of the mapping bar subtracted. The orientation tuning curve was quantitatively determined by using a bar, 0.5°–0.8° long and 0.1° wide, placed in the RF center at varying orientations, and the optimal orientation and tuning width were measured in a similar way by Gaussian fitting except that orientation tuning width was defined as the full width at half height (2 × 1.17 SD). The orientation tuning widths measured in this way for recorded neurons were between 33° and 94° (54° ± 16°, mean ± SD). Neurons that were not responsive to single bars or did not show a clear orientation tuning preference were skipped. All procedures were conducted in compliance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and under approval of the Institutional Animal Care and Use Committee at Rockefeller University.

3.5.3 DATA ANALYSIS

Two measures were taken to ensure that our results were not contaminated by eye movements. First, we discarded all those trials in which the animal's eye position changed more than 0.3° in either the vertical or horizontal direction during the trial. All data analyses were based on the remaining trials (about 70% left), though the results were similar to those from the full set of trials. Second, we provide Supplemental Data to demonstrate that our eye tracking system has sufficient resolution to track biased movements smaller than 0.05° (3 arcmin) and that no such movements occurred in any condition during the recordings (see Figures S1 and S2).

Following the neuronal response latencies (about 60 ms for MA and 50 ms for MB after stimulus onset), a 400 ms window was used for calculating the mean firing rates of a cell. Mean responses averaged across all recorded cells were used to examine the change of neuronal responses associated with different stimuli or behaviors. To calculate the mean response across all recorded cells, the mean neuronal responses from individual neurons were normalized to the responses to the noise patterns. The spike counts within

the same 400 ms window were used to calculate the neurometric curve with ROC analysis (Parker and Newsome, 1998; Tolhurst et al., 1983). In brief, the ROC curve for a given stimulus condition was calculated by plotting the hit rate (the percentage of spike counts larger than a given decision threshold when the stimulus was presented to the RF) against the false alarm rate (the percentage of spikes larger than the same threshold when the noise pattern was shown) while varying the decision threshold from the smallest to the largest spike count. The area under the ROC curve corresponds to the probability that an ideal observer can correctly identify whether the contour or noise pattern was presented for the given stimulus condition and thus gives a data point on the neurometric curve.

To determine the time point at which the mean firing rate started to deviate between contour and noise pattern, and the onset time of contextual suppression, the mean response of each cell was determined in successive windows of 10 ms width. These values were compared between the corresponding conditions with a paired nonparametric statistical test (Wilcoxon signed rank test), since they were significantly different from a

Gaussian distribution (assessed with Lilliefors test). The onset time was defined as the first of three successive intervals significant at the 5% level.

To calculate meaningful error bars and statistical significance for the non-Gaussian distributed data pooled across animals and conditions (Figure 11), the cell population was resampled by bootstrapping method (Effron and Tibshirani, 1993); this technique allows significance tests without assuming distribution shapes or equal variance of the data sets being compared), and then the spike counts for each stimulus from a single unit were resampled as well (10^5 resampling iterations were used in both cases). This resampling method accounts for both the variability due to the limited number of trials per stimulus display and that due to the difference between individual cells, and thus allows rigorous statistical inference about the population of cells from which the recorded cells were selected (i.e., orientation-selective cells in the superficial layers of V1).

3.6 ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant EY07968. We are grateful to K. Matsuda for sharing the eye tracking software and G.N. Reeke for valuable comments on the data analysis. We also thank J. McManus for helpful discussion, and J. Jones, C. Marney, T. Hartman, and P. Gogia for technical assistance.

4 LEARNING TO LINK VISUAL CONTOURS

Wu Li, Valentin Piëch and Charles D. Gilbert

Neuron 57, 442-451 (2008)

4.1 SUMMARY

In complex visual scenes, linking related contour elements is important for object recognition. This process, thought to be stimulus driven and hard wired, has substrates in primary visual cortex (V1). Here, however, we find contour integration in V1 to depend strongly on perceptual learning and top-down influences that are specific to contour detection. In naive monkeys, the information about contours embedded in complex backgrounds is absent in V1 neuronal responses and is independent of the locus of spatial attention. Training animals to find embedded contours induces strong contour-related responses specific to the trained retinotopic region. These responses are most robust when animals perform the contour detection task but disappear under anesthesia. Our findings suggest that top-down influences dynamically adapt neural circuits according to specific perceptual tasks. This may serve as a

general neuronal mechanism of perceptual learning and reflect top-down mediated changes in cortical states.

4.2 INTRODUCTION

Parsing visual scenes into different objects involves grouping processes, including contour integration, whereby contour elements belonging to the same object are perceptually linked and segregated from other scene components. Contour integration is generally characterized as a bottom-up process conforming to the rules of natural scene geometries. From the point of view of Gestalt psychology, the visual system has built-in functionality to connect line elements that form continuous and smooth contours (Field et al., 1993; Li and Gilbert, 2002; Wertheimer, 1923). This rule of continuity in perceptual organization is ecologically correlated with two statistical regularities commonly seen in natural scenes—collinearity and cocircularity (Geisler et al., 2001; Sigman et al., 2001). With respect to the underlying cortical circuitry, the long-range horizontal connections formed by the axons of pyramidal cells in the primary visual cortex (area V1) tend to link cells

with nonoverlapping receptive fields (RFs) but with similar orientation preference(Gilbert and Wiesel, 1979, 1983, 1989; Rockland et al., 1982; Schmidt et al., 1997; Stettler et al., 2002). This hard-wired intracortical connectivity is ideally suited for mediating contour integration, both in terms of its orientation specificity and its spatial extent (Li and Gilbert, 2002; Stettler et al., 2002). The involvement of V1 in contour integration is supported by physiological evidence that collinearly arranged line segments can facilitate V1 neuronal responses (Bauer and Heinze, 2002; Kapadia et al., 1995; Li et al., 2006; Polat et al., 1998). Computational models that simulate interconnected V1 neurons also demonstrate the capability of extracting global visual contours out of complex backgrounds without the intervention of top-down influences (Ernst et al., 2004; Li, 1998b; Ullman, 1992; VanRullen et al., 2001; Yen and Finkel, 1998) . As hardware encoding of contour integration is given particular emphasis in the literature, the roles of top-down influences and perceptual learning are largely overlooked.

Within a background of randomly oriented lines (Figure 15A), those discrete line segments following the Gestalt law of "good continuation" are easily

grouped together, forming a global visual contour. The perceptual saliency of contours in a complex environment depends on the spatial arrangement of contour and background elements (Field et al., 1993; Kovacs et al., 1999; Li and Gilbert, 2002). A contour made up of more collinear lines is easier to detect, or more salient, than a shorter contour within the same background (compare Figure 15A with Figure 15B), and the same array of collinear lines forming the contour appears less salient when they are spaced further apart (compare Figure 15B with Figure 15C). On the other hand, contour detectability depends not only on the geometry of visual stimuli but also on perceptual learning. In particular, the contours that are originally hidden or less salient Figure 15C) become increasingly easier to detect with training due to an enhancement in interactions between contour elements (Kovacs et al., 1999; Li and Gilbert, 2002). It has also been shown that perceptual learning leads to increased strength and effective distance of facilitation between collinearly arranged targets (Polat and Sagi, 1994b).

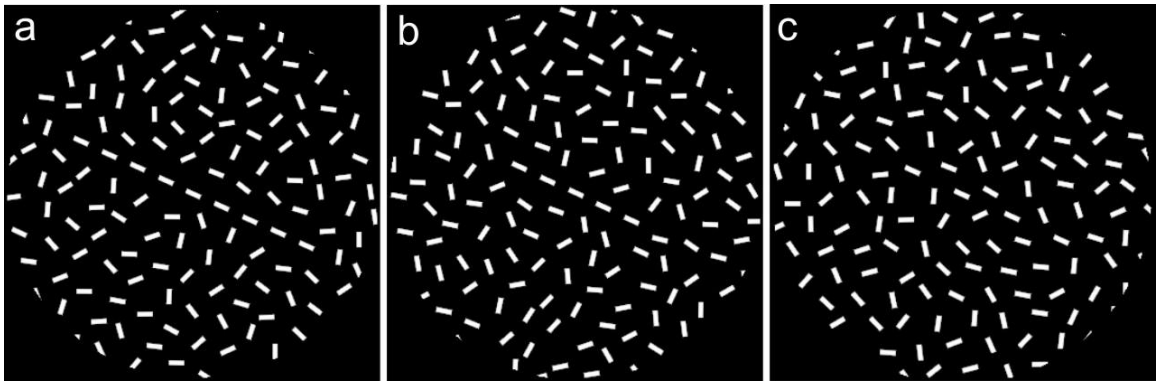

Figure 15. Visual Contours Formed by Collinear Line Segments Embedded in a Background of Randomly Oriented Lines

Within the same background, perceptual saliency of contours varies with the number of collinear lines (A) and (B) and with the spacing between them (B) and (C).

In a recent study, we have shown that in monkeys performing a contour detection task, responses of V1 neurons are predictive of the animals' behavioral performance on contour detection and therefore are closely correlated with perceptual saliency of contours (Li et al., 2006). The correlation is seen in the facilitation of neuronal responses by collinear line segments lying outside the classical RF within a background of randomly oriented lines. This facilitation is stronger with more salient contours and is present in virtually all orientation-selective neurons in superficial layers of V1, indicating that V1 is intimately involved in linking visual contours and in mediating contour saliency. However, the facilitation of neuronal

responses by the same contours is significantly weakened if the animals perform a different task irrelevant to contour detection. This implies a degree of top-down influence on V1 responses, but the nature of this influence is not obvious. The top-down effects could be due to differences in either the spatial locus of attention or the task of contour detection itself.

In the current study, in order to examine whether the neural substrate of the improvement in contour detection with practice is found in area V1 and also to explore the nature of top-down influences on contour integration, we trained monkeys on a succession of tasks, beginning with fixation, then attending to the location of a contour, and then detecting the contour embedded in a complex background. Neuronal responses were recorded from the same V1 regions during different experimental stages. This design enabled us to follow the emergence of contour-related responses in V1 as the animals learned to detect contours. These experiments revealed striking experience- and task-dependent changes in V1 responses associated with training on contour detection.

4.3 RESULTS

Two naive adult monkeys (Macaca mulatta, male, 6–7 kg) were used in all experiments. As the results from the two animals were qualitatively similar (see Figure S1 available online), data from both subjects were pooled in most analyses to guarantee fair sample size in statistical tests. A total of 275 cells were recorded from V1 in the four hemispheres of the two animals.
Perceptual Learning in Contour Detection

For simple discrimination and detection tasks, repeated execution of the same task usually leads to a remarkable improvement in a subject's performance on the task. This perceptual phenomenon, known as perceptual learning, has been reported in contour detection tasks in human subjects (Kovacs et al., 1999; Li and Gilbert, 2002). The two animals used in the current study showed similar learning effects.

The visual stimuli used in the contour detection task consisted of a series of collinear line segments embedded in the center of an array of randomly oriented and positioned lines (Figure 15A and Figure 15B). In addition to

this contour pattern, a second pattern without any embedded contour (referred to as the noise pattern) was simultaneously presented as a distracter in the visual-field quadrant opposite to the contour pattern (Figure 16A). The perceptual task was to indicate which of the two stimulus patterns contained a straight contour. From trial to trial the contour pattern and the noise pattern were randomly interchanged so that the overall detection ratio simply by guessing was 50%. Within a block of trials, five stimulus conditions with visual contours made up of one, three, five, seven, or nine collinear lines were interleaved, with line spacing kept constant. To eliminate any possible cues, other than the embedded contour, for the two stimulus patterns shown in each trial, all the noncontour elements and the very central line segment in the contour pattern were made identical to those in the noise pattern (compare the corresponding line segments between the two stimulus patterns shown in Figure 16A). With this design, when the embedded contour consisted of only a single line segment, the contour pattern and the noise pattern were identical. This particular condition was included as a baseline control. After 476 ms exposure of the stimuli, the contour pattern and the noise pattern were replaced with two dots. The animals indicated the

location of the pattern containing the embedded contour by making a saccadic eye movement toward either of the two dots.

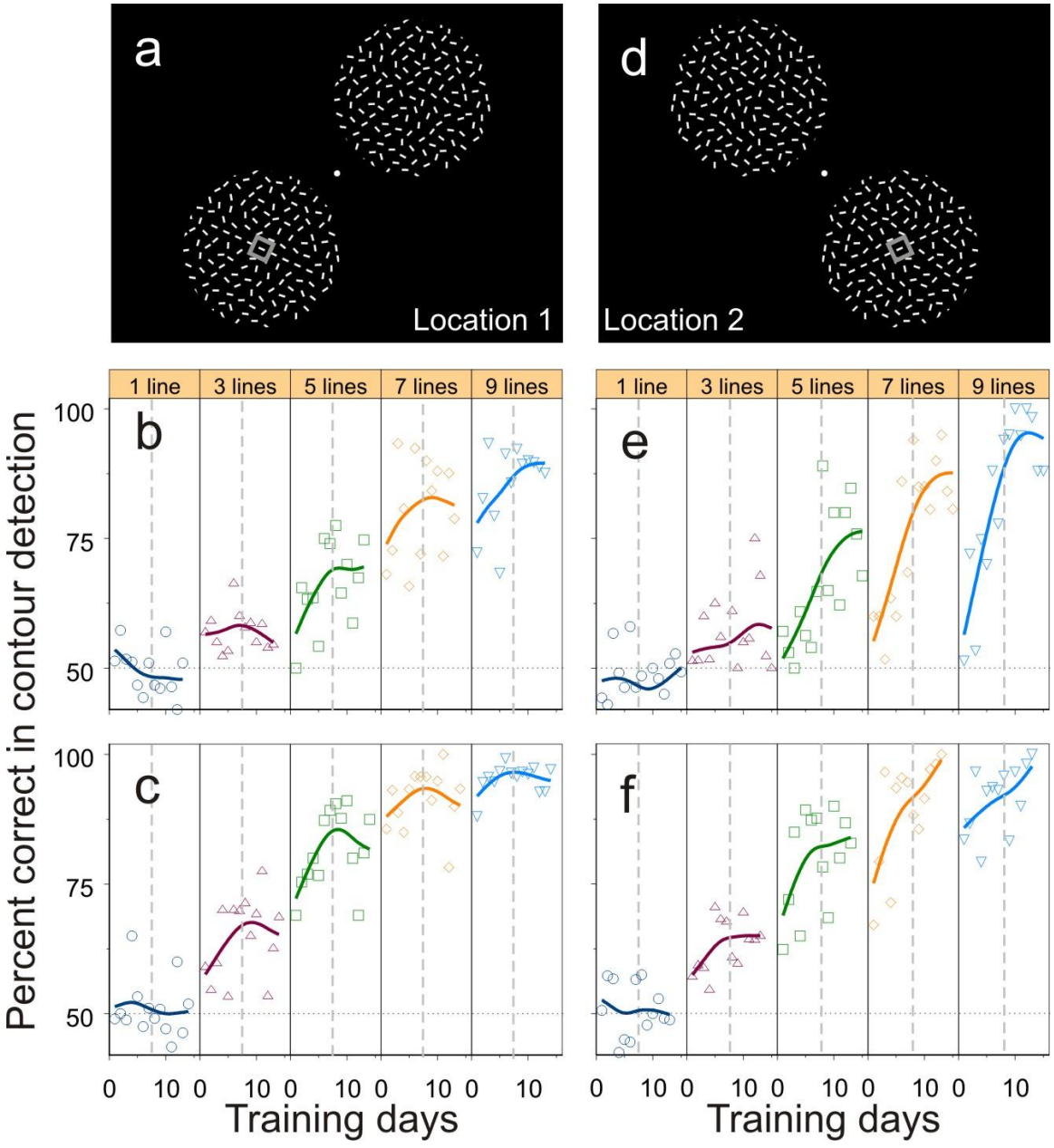

Figure 16. Learning Curves in the Contour Detection Task

(A) The visual field locations trained first (named location 1). The animal maintained its fixation at the central spot throughout the trial. The positions of the contour and noise patterns were randomly interchanged from trial to trial. The animal indicated the location of the contour pattern by making a saccadic eye movement to either of two target dots displayed at the end of the trial. The small gray square denotes the receptive field during V1 recordings.

(B and C) The learning curves of the two animals, respectively, at visual-field location 1. The five panels correspond to the five stimulus conditions in which the number of collinear lines forming the embedded contour was one, three, five, seven, or nine, respectively. Each dot represents the mean performance within a day based on 20-160 trials (with a median of 90 trials), and each curve was generated by a cubic spline with 3 degrees of freedom. The horizontal dotted line denotes the chance level (50%); the vertical dashed lines indicate the arbitrary division of the whole training period into the early phase and the late phase.

(D) The new visual field location (named location 2) tested after all the experiments at location 1 were completed.

(E and F) The learning curves of the two animals, respectively, at location 2.

In order to let the animals understand and execute this two-alternative-forced-choice (2AFC) task, a preliminary procedural or operational training was conducted before data collection. The main training procedure is as follows. At the training outset, the complex background surrounding the contour elements, as well as the whole noise pattern, was hidden by setting the luminance of line segments to that of the display background, leaving the collinear lines shown in isolation at either of the two locations. It took a couple of days, about 1500 trials a day, for the animals to learn, by trial and error for a reward, to associate the location of the stand-alone contours with

the direction of saccades. Subsequently, the hidden line elements were rendered visible by gradually increasing their luminance above the display background, but these noncontour elements were kept noticeably dimmer than the contour. It took another several days for the animals to reliably choose, between the two stimulus patterns, the one with a highlighted contour of different lengths at any given orientation. Then the luminance difference between contour and background lines was removed and data collection was started. For each session of the contour detection task, the orientation of the embedded contour was determined by the preferred orientation of the recorded V1 cell; therefore, the contour orientation was the same within a session but could be different across sessions.

Daily training and recordings began with the lower-left and upper-right quadrants in the visual field (Figure 16A; defined as location 1 throughout the text). Both animals' performance on contour detection was improved with practice and reached different plateaus for different contour lengths, a typical effect of perception learning (Figure 16B and Figure 16C).

A prominent characteristic of perceptual learning is its specificity to visual-field location. The learning-induced improvement is usually restricted to the area where the task stimuli are displayed (for example see (Crist et al., 1997; Sigman and Gilbert, 2000)). The improved contour detection ability in these two animals was also specific to the trained visual-field location. After the animals' performance had reached plateaus at location 1, we moved the stimulus patterns to untrained areas, the lower-right and upper-left quadrants of the visual field (Figure 16D; defined as location 2 throughout the text). Before data collection, we had ensured that the animals still understood the detection task at the new stimulus location. This procedure was done with highlighted contours as in the initial operational training described previously, but it was much quicker and took only a couple of hundred trials.

Compared with the final performance level at the trained location 1, the animals' performance on contour detection dropped substantially in the first couple of days at the new location for the same contour length (compare Figure 16E with Figure 16B, and Figure 16F with Figure 16C, for the two animals, respectively). Subsequent training greatly enhanced the animals'

contour detection ability (Figure 16E and Figure 16F), indicating a repetition of the perceptual learning process at the new visual field location. Note that the initial performance at location 2 could be even worse than the initial performance at location 1, suggesting that a certain degree of perceptual learning had already taken place during the preliminary procedural training at the first location (location 1). But in any event, the changes in the animals' behavioral responses along with the location specificity of these changes clearly demonstrated perceptual learning in contour detection. As the animals learned to detect the embedded contours, recordings from area V1 showed parallel changes in neuronal responses that were related to contour integration.

4.3.1 ABSENCE OF CONTOUR INFORMATION IN UNTRAINED V1

As a first stage of the experimental series and before training the monkeys on the contour detection task, we recorded responses of V1 neurons to the embedded contours when the naive animals were only trained on a simple central fixation task. In this task, the animals were required to detect the

dimming of a small fixation target. They indicated the dimming by releasing a lever that they held from the outset of the trial. The visual stimuli were identical to those described previously in the contour detection task, except that the contour patterns were always displayed over the RF location.

While the monkeys performed this central-dimming task and were passively exposed to the stimuli in the periphery, responses of single orientation-selective neurons to the contour patterns were recorded from superficial layers of V1. The orientation of the contour was adjusted to the preferred orientation of the recorded cell, and the central line element of the contour was centered in the RF. Although the visual contours consisting of different numbers of collinear lines were displayed at the RF location, the mean neuronal responses to these stimuli were indistinguishable and were identical to the responses to the noise pattern (Figure 17A; Kruskal-Wallis test, $p = 0.987$). This experiment showed that the information about the embedded contours was absent in the firing rates of V1 neurons in monkeys naive to the contour detection task.

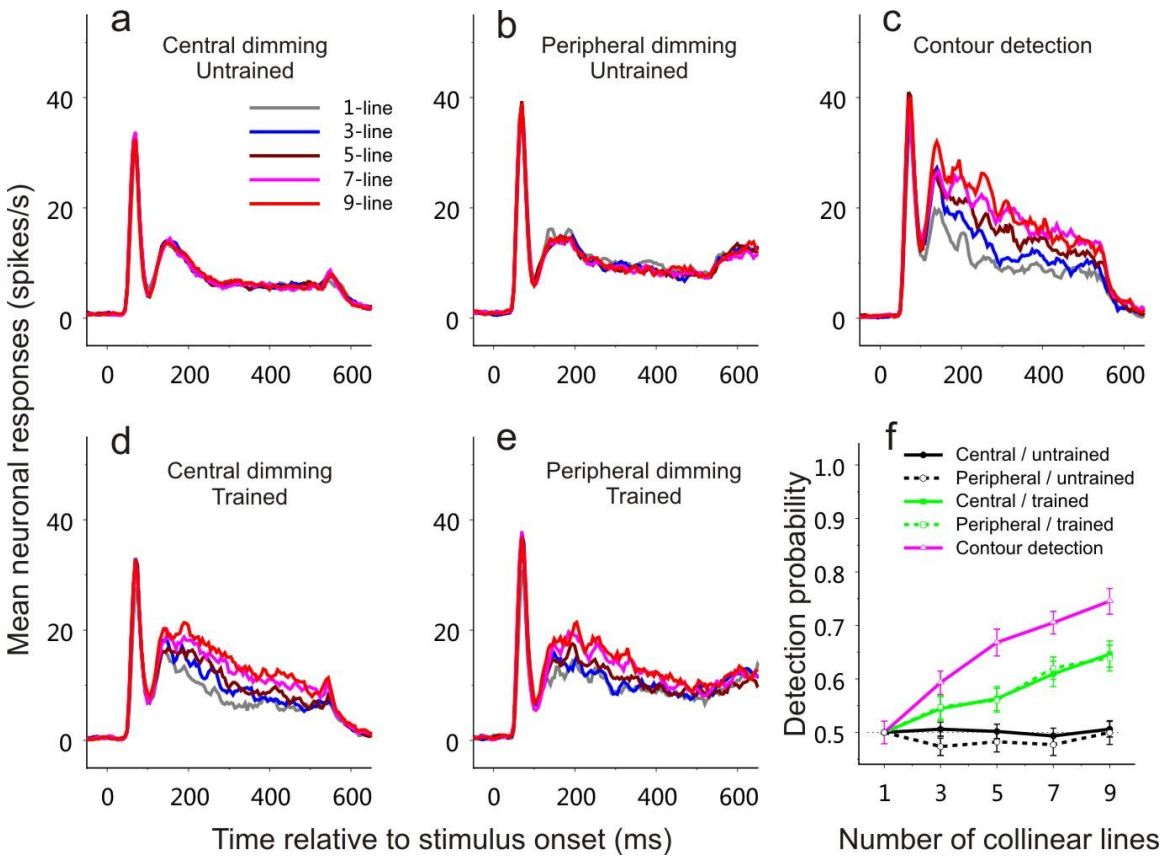

Figure 17. Emergence of Contour-Related Responses in V1 with Training on Contour Detection

(A-E) PSTHs were constructed by binning neuronal spikes at 5 ms resolution over time and by averaging across all recorded cells for each experimental condition. The PSTHs were smoothed using a rectangular window of 20 ms (boxcar filter). Time 0 indicates stimulus onset. Five curves in each panel represent neuronal responses to contours made up of one, three, five, seven, or nine collinear lines, respectively. Different panels show different experimental stages. All data were collected from a 5 × 5 mm region in the right hemisphere, and the visual stimuli were presented at visual field location 1 (Figure 2A).

(A) Central-dimming task before training on contour detection. The task was to respond to a slight dimming of the fixation point. n = 73 (52 and 21 cells, respectively, from the two animals).

(B) Peripheral-dimming task before training on contour detection. The task was to maintain fixation at the fixation point and respond to a dimming of the line segment in the RF, which was also the central element of the collinear contour. n = 45 (25 and 20, respectively, from the two animals).

(C) Contour detection task. The animal had to indicate the location of the embedded contour in a two-alternative-forced-choice task. Neuronal responses recorded in both correct and error trials during the late training phase are included (refer to Figure 2 for the division of early and late phases). n = 32 (13 and 19, respectively, from the two animals).

(D) Central-dimming task after training on contour detection. n = 47 (24 and 23, respectively, from the two animals).

(E) Peripheral-dimming task after training on contour detection. n = 48 (26 and 22, respectively, from the two animals).

(F) Neurometric curves based on ROC analysis of the data shown in (A)-(E), respectively. Each curve was averaged across all the recorded cells for each condition. Error bars represent ± SEM calculated by the method of bootstrapping (see Experimental Procedures).

In the central-dimming task, the animals were not required to pay attention to the spatial location where the contours were embedded. To examine whether spatial attention by itself would differentiate neuronal responses to contours of different lengths, in the second stage of experiment the monkeys were trained to maintain their fixation at the fixation point while attending to the middle of the contour pattern, where the central line element of the contours was slightly dimmed at times randomly distributed between 476 and 1476 ms after stimulus onset. The animals had to respond to the dimming by releasing the lever within 600 ms. We refer to this task as the peripheral-dimming task, as distinguished from the central-dimming task in the previous stage. Even though the animals attended to the center of contour

patterns and to the RF location, no significant difference was observed in the mean V1 responses to contours of different lengths (Figure 17B; Kruskal-Wallis test, $p = 0.997$). Therefore, for monkeys that had never been trained to detect the embedded visual contours, V1 responses were independent of the presence and length of contours, regardless of whether or not the animals' attention was directed to the target location.

4.3.2 LEARNING-INDUCED V1 CHANGES RELATED TO CONTOUR INTEGRATION

In an earlier study, we observed a close correlation between V1 responses and contour lengths in monkeys well trained on contour detection (Li et al., 2006). The absence of contour-related responses here in animals naive to the contour detection task suggests that either the process of training on contour detection, or the specific action of conducting the contour detection task, results in neuronal responses to the visual contours. To differentiate between these possibilities, in the next stage of the experiments the animals were trained to perform the contour detection task (Figure 16A–Figure 16C). After the animals' performance had reached plateaus and when the animals

were actively detecting the embedded contours, a late response component associated with contour length emerged—the longer the contours, the stronger the neuronal responses (Figure 17C). These contour-related responses started about 100–120 ms after stimulus outset. Statistical tests showed that the contour length became a significant factor influencing neuronal responses (Kruskal-Wallis test, $p = 0.000019$) and that the responses in the nine-line condition were significantly stronger than the one-line condition (Wilcoxon rank-sum test, $p = 0.000018$). Note that the initial response peak in the peristimulus-time histogram (PSTH) did not systematically change with the contour length.

With the monkeys already trained on the contour detection task, we recorded again neuronal responses to the contour patterns while the animals performed the central (Figure 17D) and peripheral (Figure 17E) dimming tasks in which the visual contours were task-irrelevant. Longer contours still evoked stronger neuronal responses: contour length remained a significant influencing factor (Kruskal-Wallis test, $p = 0.0038$ in Figure 17D; $p = 0.045$ in Figure 17E); responses in the nine-line condition were significantly

stronger than in the one-line condition (Wilcoxon rank-sum test, $p = 0.00054$ in Figure 17D; $p = 0.012$ in Figure 17E). This is in sharp contrast with the experimental stages in which the animals did the same dimming tasks but had never been trained to perform the contour detection task (compare Figure 17D with Figure 17A, and Figure 17E with Figure 17B). On the other hand, the amount of facilitation with contour length was significantly weaker in both dimming tasks than the contour detection task (compare Figures Figure 17D and Figure 17E with Figure 17C; Wilcoxon rank-sum test, $p < 0.05$ for all five-, seven-, and nine-line conditions).

To examine the changes in neuronal responses in relation to contour integration quantitatively, receiver-operating-characteristic (ROC) analysis (Parker and Newsome, 1998; Tolhurst et al., 1983) was used to calculate the probability that an ideal observer can distinguish the contour pattern from the noise pattern simply based on a single neuronal spike count. By calculating this probability as a function of the number of collinear lines forming the contour, we obtained a neurometric curve, which indicates how well the contour and noise patterns could be differentiated at different

contour lengths based on neuronal responses. In the central- and peripheral-dimming tasks before training on contour detection, the neurometric curves averaged across all recorded cells were at the chance (50%) level, independent of the contour length (Figure 17F); therefore, the embedded contours could not be detected simply based on V1 firing rates. After training, however, the discrimination ability of V1 neurons increased with contour length, and this emergent response property was most robust when animals did the contour detection task.

4.3.3 SPECIFICITY OF LEARNING-INDUCED CHANGES TO VISUAL FIELD LOCATION

The specificity of perceptual learning to visual-field location suggests learning-induced changes in early visual cortical areas where the visual field is represented at the finest resolution by a topographic map. To examine whether the emergence of V1 responses to the camouflaged contours was specific to the trained retinotopic region, in the next experimental stage, we flipped the contour and noise patterns around the vertical meridian, changing their display locations to the untrained lower-right and upper-left quadrants

(location 2; Figure 16D). We first repeated the central- and peripheral-dimming experiments by recording from the V1 region representing the untrained lower-right visual field. The differences in the mean neuronal responses to contours of different lengths were significantly smaller than those at the trained location 1 for the same central- or peripheral-dimming tasks (compare Figure 18A with Figure 17D, and Figure 18B with Figure 17E). Statistical tests showed that, at the new visual-field location, contour length was no longer a significant factor affecting neuronal responses in both the central-dimming (Kruskal-Wallis test, $p = 0.42$) and peripheral-dimming (Kruskal-Wallis test, $p = 0.74$) tasks. Subsequent training on contour detection in this untrained area induced the profound changes in neuronal responses that were comparable to those observed in the previously trained area (Figure 18C–Figure 18F). In detail, the contour length became again a significant modulatory factor in the contour detection (Kruskal-Wallis test, $p = 0.0000017$), central-dimming ($p = 0.00018$), and peripheral-dimming ($p = 0.0019$) experiments. Moreover, the contour-related responses in the nine-line condition were significantly stronger than in the one-line condition in the contour detection (Wilcoxon rank-sum test, $p = 0.000012$), central-

dimming ($p = 0.000089$), and peripheral-dimming ($p = 0.00028$) tasks. However, the amount of facilitation of neuronal responses with contour length was significantly weaker in both central- and peripheral-dimming tasks than the contour detection task (Wilcoxon rank-sum test, $p < 0.05$ for all three-, five-, seven-, and nine-line conditions).

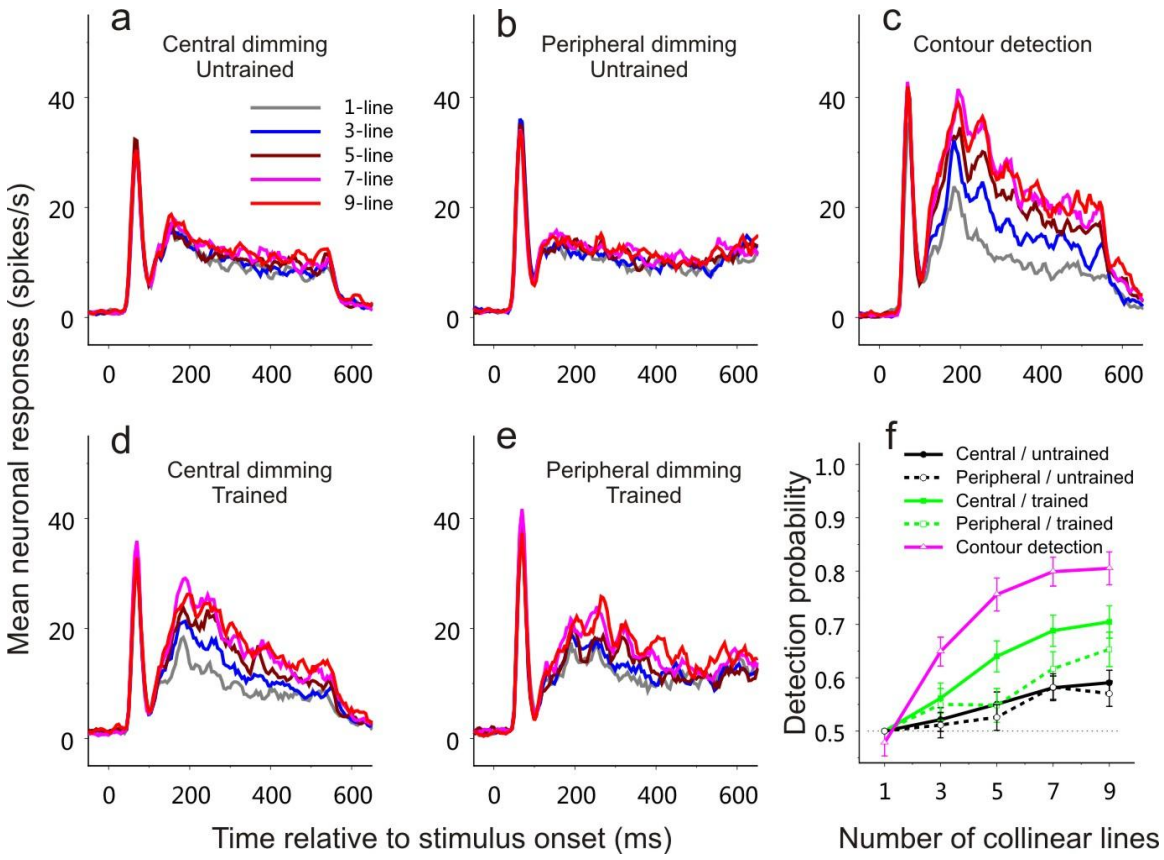

Figure 18. Results from a New Visual Field Location

After all the experiments shown in Figure 3 had been completed at visual-field location 1 (Figure 2A), the same series of experiments were repeated at a new stimulus location (location 2; Figure 2D). All data were collected from a 5 × 5 mm V1 region in the left hemisphere.

(A-E) PSTHs based on averaged neuronal responses to contours consisting of one, three, five, seven, or nine collinear lines in different experimental stages.

(A) Central-dimming task before training on contour detection. n = 50 (26 and 24 cells, respectively, from the two animals).

(B) Peripheral-dimming task before training on contour detection. n = 47 (24 and 23, respectively, from the two animals).

(C) Contour detection task during the late training phase. n = 28 (16 and 12, respectively, from the two animals).

(D) Central-dimming task after training on contour detection. n = 37 (22 and 15, respectively, from the two animals).

(E) Peripheral-dimming task after training on contour detection. n = 36 (22 and 14, respectively, from the two animals).

(F) Averaged neurometric curves based on ROC analysis of the data shown in (A)-(E), respectively.

Refer to Figure 3 for more details about each panel.

4.3.4 PARALLEL EFFECTS OF TRAINING ON PERCEPTION AND NEURONAL RESPONSES

In perceptual learning, the improvement in perceptual skills evolves with practice and time (Figure 16; see also (Crist et al., 1997; Crist et al., 2001; Li et al., 2004). To examine the parallel effects of training on the animal's ability in contour detection and on V1 responses, we arbitrarily divided the data collected from the contour detection task into two chronological groups: those collected before the animals' performance reached plateaus (referred to as the early phase) versus those collected afterwards (referred to as the late phase; the dividing time point is indicated by the vertical dashed lines in Figure 16). A quantitative comparison of the results from the late and early

training phases showed that the animals' performance on contour detection was significantly improved in the late phase, resulting in a steepening of the psychometric curve that indicates the animals' detection performance as a function of the number of collinear lines (Figure 19A; unpaired Student's t test, $p = 0.43$, 0.12, 0.0037, 0.0042, and 0.00040, respectively, for the one-, three-, five-, seven-, and nine-line conditions). The neurometric curves averaged across the cells recorded within the same periods of time showed parallel changes, increasing steepness with training.

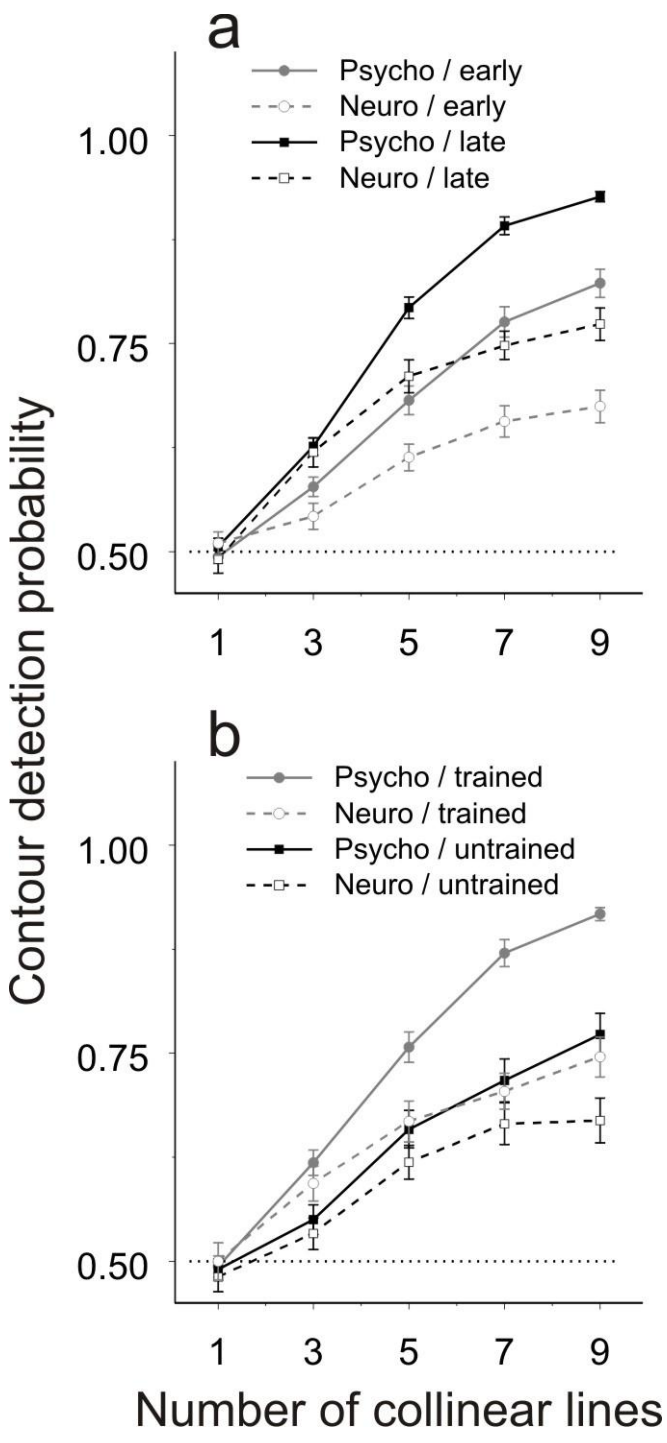

Figure 19. Parallel Effects of Learning on Behavioral and Neuronal Responses

(A) The neurometric curves (averaged across all cells) and the psychometric curves (averaged over the corresponding recording sessions) are shown for two different training phases: the early phase before the animal's performance reached plateaus (n = 66; 31 and 35 cells from the two animals, respectively) and the following late phase (n = 60; 29 and 31 from the two animals, respectively). The dividing time point between the early and late phases is indicated by the vertical dashed lines in Figure 2. Data were pooled from the four hemispheres of the two animals within each training phase. A significant steepening in both psychometric and neurometric curves was observed with training. Error bars represent ± SEM calculated by the method of bootstrapping (see Experimental Procedures).

(B) Data collected in the early training phase at visual-field location 2 (untrained; n = 32; 13 and 19 cells from the two animals, respectively) are compared with those in the late training phase at the previously trained location 1 (trained; n = 34; 15 and 19 from the two animals, respectively). A significant drop in steepness of psychometric and neurometric curves was seen in the untrained area relative to the trained.

To examine whether the parallel improvement in both psychometric and neurometric functions was restricted to the trained area, we compared the averaged psychometric and neurometric functions at location 1 in the late training phase with those at location 2 in the early phase of training (Figure 19B). When the stimuli were moved from the trained (location 1) to untrained (location 2) visual field location, the animal's performance on contour detection, the psychometric curve, was significantly lower (unpaired Student's t test, p = 0.99, 0.11, 0.027, 0.0087, and 0.0018, respectively, for the one-, three-, five-, seven-, and nine-line conditions). The contour detectability based on ROC analysis of V1 responses, the neurometric

function, was also lowered. These results indicate a lack of generalization of the learning effects to the new visual-field location and to the corresponding retinotopic region in V1.

4.3.5 DISAPPEARANCE OF LEARNING-INDUCED CHANGES IN V1 UNDER ANESTHESIA

The experiments described previously indicate that contour-related responses in V1 are driven by a number of factors, including visual stimuli, training experience, and top-down control. To obtain a measure of responses that are wholly independent of behaviorally mediated top-down influences, in the last experimental stage we recorded neuronal responses to the embedded contours in the trained V1 regions when the same animals anesthetized. The late response components associated with the embedded contours completely disappeared (Figure 20; Kruskal-Wallis test, $p = 0.997$), suggesting that even in the trained V1 area the stimulus-driven process by itself does not suffice to extract the embedded visual contours. In addition to the absence of contour-related responses, the overall neuronal responses to

the complex stimuli were much weaker in anesthetized state than in awake state, as has been reported earlier (Lamme et al., 1998).

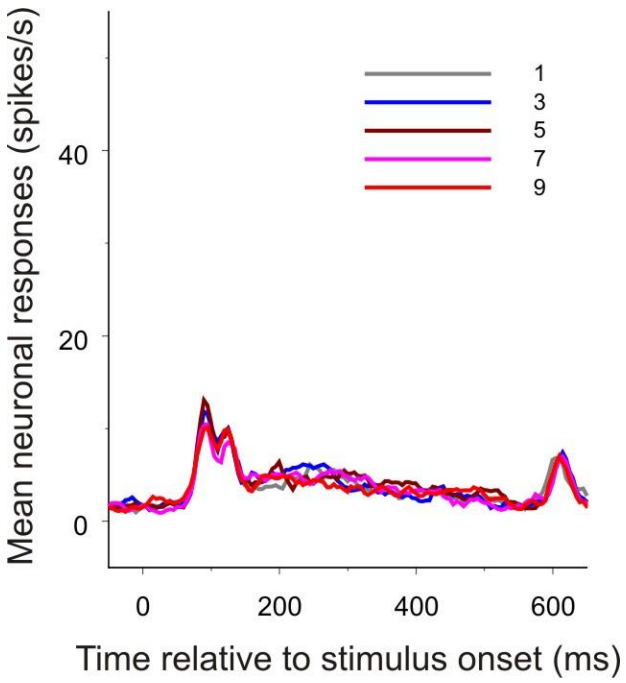

Figure 20. The Effect of Anesthesia

When all behaviorally related top-down influences were eliminated by anesthesia, contour-related responses disappeared in the trained V1 regions (n = 22). The PSTHs were constructed as described in Figure 3.

4.4 DISCUSSION

It has been shown that training monkeys on some visual discrimination tasks can increase the selectivity of V1 neurons for the trained visual stimuli (Crist et al., 2001; Li et al., 2004; Schoups et al., 2001). Moreover, training can enhance V1 responses to familiar targets within distractors (Kourtzi et al., 2005; Lee et al., 2002). Our current study is consistent with these observations, supporting the notion of adaptive cortical processing of visual information even at the earliest stage. The most important finding of the current study is that even the process of contour integration in V1, which is generally believed to be a hard-wired process, can strongly depend on perceptual learning and top-down influences.

4.4.1 TASK-SPECIFIC TOP-DOWN INFLUENCES AND EXPERIENCE-DEPENDENT CHANGES

Psychophysical studies have shown that for very similar (Fahle, 1997; Fahle and Morgan, 1996; Sigman and Gilbert, 2000) or even completely identical (Ahissar and Hochstein, 1993; Saffell and Matthews, 2003; Shiu and

Pashler, 1992) visual stimuli, a subject's discrimination ability for a given stimulus attribute can be improved only if the attribute is attended and related to the trained task. Repeated passive exposure to the stimuli rarely induces any learning effect. Our current study demonstrated that top-down influences specific to the task of contour detection were responsible for the emergence of contour-related responses in V1, as passive exposure to the embedded contours did not alter V1 responses, nor did attending to the contour location.

As different forms of top-down influences could be mingled in any perceptual task, it is arguable that the late response components that we observed in V1 of the trained animals could be attributed to attentional load or task difficulty rather than the contour integration task itself. One would then, however, expect a negative correlation between V1 responses and the contour length, because longer or more salient contours are much easier to detect, and therefore less attentionally demanding, than short ones. To the contrary, we show that longer contours are associated with better detection performance and stronger neuronal responses. Moreover, in the periphery

dimming task, which is more difficult than simple fixation (the central dimming task), neuronal responses were actually somewhat weaker (compare Figure 17D with Figure 17E, and Figure 18D with Figure 18E). Our findings therefore suggest that top-down influences are not limited to spatial attention but can convey much more information, including information about specific perceptual tasks. By using perceptual learning to engage different components of top-down control, first spatial attention and then task, we were able to separately see the effects of these components. Consistent with our previous studies showing task-specific modifications of V1 response properties (Crist et al., 2001; Li et al., 2004), our findings have important implications for the neuronal mechanisms underlying the task specificity of perceptual learning, in that learning may involve the way top-down influences engage certain cortical areas. A recent fMRI study has shown that task-specific top-down influences can trigger a large-scale reorganization of cortical activity that can account for the improvement with training in detecting a particular shape within similar distracters (Sigman et al., 2005). That study is consistent with the current study in that both the

overall behavioral performance and V1 activity increase significantly with perceptual learning.

The parallel changes of psychometric and neurometric functions indicate that the improvement in the animal's performance on contour detection is associated with the enhancement of V1 signals for the embedded contours. On the other hand, the neurometric curve was significantly below the psychometric curve over the same training period (Figure 19), indicating that the prediction of contour detection performance based on single spike count is generally worse than the animals' actual performance. A possible explanation is that several V1 neurons with RFs lying along the contours contribute to detection of an extended contour. Interestingly, our earlier study has shown that, for monkeys extensively trained on contour detection, the difference between psychometric and neurometric functions can be much smaller (Li et al., 2006). Taken together, these observations suggest that, although both the overall behavioral and neuronal responses are gradually enhanced in parallel with training, the behavioral performance reaches a plateau earlier than the average changes in V1. It may be that different

subsets of neurons exhibit different learning speeds, some more rapid, and thus more contemporaneous with the behavioral change, than the others. Alternatively, perceptual learning may first involve engagement of a cortical area by feedback to that area, and then a consolidation and increased efficiency in the representation of the trained stimulus.

Once the animals had been trained, a measure of contour-related responses was observed even in the central- and peripheral-dimming tasks in which the contours were task irrelevant. This indicates that the emergence of contour-related responses in V1 is associated with the animals' particular experience with the contour detection task. It is likely that contour integration is turned into a rather automatic process due to repetitive practice of the detection task. This might be interpreted in one of two ways. Either a component of the contour-related responses becomes mediated entirely by local circuits without top-down control or the contours become more salient, attracting attention to involuntarily engage the animal in the contour detection task. Nevertheless, the facilitation of neuronal responses by the embedded contours was strongest only if the animals actively performed the contour

detection task, suggesting an important role of task-specific top-down influences in contour integration at all stages of training. This is strongly supported by another observation that contour-related responses were abolished, not merely attenuated, by anesthesia, a finding that is consistent with the observation that V1 responses associated with figure-ground segregation based on texture differences are absent under anesthesia (Lamme et al., 1998). The removal of contour-related responses under anesthesia suggests that the integration process cannot operate without a measure of top-down control. This is not inconsistent, however, with the idea that local circuits, such as the long range horizontal connections, are involved in contour integration. Rather, we propose that it is the interaction between local circuits and cortical feedback that results in V1 responses to salient contours, and that the operation of local circuits is gated by feedback connections. Supporting evidence for such an interaction is suggested by the findings that higher-order cortical areas, such as the lateral intraparietal area (LIP), are involved in generating saliency maps about visual stimuli according to both stimulus features and behavioral goals (Ipata et al., 2006). The recurrent processing within and across cortical areas may account for

the delay of contour-related responses in V1 and is in agreement with the notion of visual routines (Ullman, 1984) and incremental grouping (Roelfsema, 2006) by which the bottom-up representations of basic stimulus attributes are shaped by processes involving higher-level cognitive influences, resulting in extraction and enhancement of behaviorally relevant information based on the knowledge of specific tasks.

The complete absence of contour-related responses in monkey V1 before training on contour detection seems to be at variance from psychophysical observations that collinear facilitation in perception is present in naive human subjects. It is worth noting that when a naive subject is performing a task as instructed, the appropriate task-specific top-down influences have already taken effect. In fact, immediately after the initial operational training on contour detection and in the first couple of sessions in the early perceptual learning phase, we were already able to find V1 cells showing significant responses to the embedded contours. This is consistent with the psychophysical finding that a proper deployment of attentional resources to

collinear stimuli triggers collinear facilitation effects on perception (Freeman et al., 2001; Shani and Sagi, 2005).

4.4.2 NEURAL MECHANISMS OF PERCEPTUAL LEARNING

Our findings, which support the attention-gated reinforcement learning model (Roelfsema and van Ooyen, 2005), have broader implications for the general neural mechanisms of perceptual learning. Different visual cortical areas are involved in processing different attributes of visual stimuli. For those cortical areas that are intimately involved in processing the information related to the current task, top-down influences specific to the task can dynamically alter the local circuitry to meet the requirements of the perceptual task. During repeated execution of the same task, state changes in cortical circuits occur under task-specific top-down control. Perceptual learning may be a process by which the feedback connections associate with the appropriate intrinsic connections during the encoding of the learned information, and the same set of relationships are utilized during the recall of this information.

4.5 EXPERIMENTAL PROCEDURES

4.5.1 VISUAL STIMULI

Visual stimuli, generated by a VSG2/5 stimulator (Cambridge Research Systems) on a CRT monitor, were composed of white (19.5 cd/m2) line segments on a gray (6.5 cd/m2) background. Each stimulus display contained two circular patterns (Figure 16A), a contour pattern with one, three, five, seven, or nine collinear lines embedded in the center of an array of randomly oriented lines, and a noise pattern without any embedded contour. The noise pattern was identical to the contour pattern if the number of collinear lines in the contour was 1. To construct the stimulus patterns, a circular area of 5.2° in diameter was divided into 13 by 13 compartments, each containing a randomly oriented line segment of 0.2° by 0.05° whose position was slightly jittered within its compartment. This arrangement defined an average line spacing of 0.4° in the stimulus pattern. The orientation and position jitter of each line was re-randomized in each trial. By aligning some adjacent line segments in a collinear, evenly spaced arrangement, a straight contour was generated in the center of the contour

pattern. The center-to-center distance between two adjacent collinear lines was fixed at 0.44° (for details on the geometry of stimulus generation see (Li et al., 2006).

4.5.2 BEHAVIORAL TASKS

A trial began when a lever was pulled by the animal. A 0.08° fixation point (FP) was displayed in the CRT center. Eye positions were sampled at 30 Hz by an infrared tracking system (K. Matsuda et al., 2000, Soc. Neurosci., abstract), which offered a spatial resolution of up to 0.05° (for details see (Li et al., 2006). Within 600 ms after FP presentation the animal was required to fixate within an invisible circular window of 0.5° in radius around the FP. After the animal maintained its fixation for 191 ms, the contour pattern and noise pattern were presented. Subsequent timing and behavioral requirements in a trial were different for different tasks, which are described in detail when each task is introduced in the Results section. The time window allowing the animal to make a response in each trial of a task was 600 ms. Different stimulus conditions in a task were randomized, and each stimulus condition was repeated 10–40 times.

4.5.3 ELECTROPHYSIOLOGICAL RECORDINGS

Recordings in all experiments were confined to the same V1 region for each hemisphere, about 5 by 5 mm in area, which corresponded to eccentricities 3.0° to 4.8°. All recordings were restricted to the first 500 μm from entering the cortex. To further ensure the electrode was in the superficial layers, background spontaneous activity was monitored. Superficial-layer neurons were typically marked by a lack of spontaneous activity (see the population PSTHs before stimulus onset in Figure 17 and Figure 18; also refer to (Snodderly and Gur, 1995; von der Heydt and Peterhans, 1989). Entering from superficial layers into layer 4 is indicated by a remarkable increase in background activity. The RFs were quantitatively mapped using the same methods described elsewhere (Li et al., 2006).

For the experiments that required the animals to perform certain tasks, daily recordings were carried out 4 days a week, and each experimental stage lasted about 1 month for each animal. In the last experiment in which the animals were anesthetized, venous cannulation and tracheal intubation were conducted following an initial induction with ketamine hydrochloride (10

mg/kg). Anesthesia and paralysis were maintained using intravenous infusion of Nembutal (pentobarbital sodium, 1.5 mg/kg/hr) and Norcuron (vecuronium bromide, 0.1 mg/kg/hr). The animals were artificially respired. EKG, expired CO2, and rectal temperature were monitored continuously to ensure that the anesthesia was not very deep. All procedures were conducted in compliance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and under approval of Institutional Animal Care and Use Committee at Rockefeller University.

4.5.4 DATA ANALYSIS

The spike counts within a 400 ms window, beginning from 110 after stimulus onset, was used to calculate the neurometric curve with ROC analysis (Parker and Newsome, 1998; Tolhurst et al., 1983). In brief, the ROC curve for a given number of collinear lines in the contour detection task was calculated by plotting the hit rate (the percentage of spike counts larger than a given decision threshold when the contour pattern was presented to the RF) against the false alarm rate (the percentage of spikes larger than the same threshold when the noise pattern was shown) while

varying the decision threshold from the smallest to the largest spike count. The area under the ROC curve corresponds to the probability that an ideal observer can correctly identify, based on a single spike count, whether the contour or noise pattern was presented for the given contour length, and thus gives a data point on the neurometric curve.

To estimate meaningful error bars and statistical significance of the averaged data, the cell population was resampled by bootstrapping method (Effron and Tibshirani, 1993), and then the spike counts in each trial from a cell were resampled as well (10^3 resampling iterations were used in both cases). This resampling method accounts for the variability both due to the limited number of trials per stimulus display and due to the difference between individual cells.

4.6 ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant EY07968. K. Matsuda developed the infrared eye tracking system. We also thank J. Jones, H. Yamahachi, N. Ramalingam, E. Paul, and S. Hinklein for technical assistance.

5 A NEURAL MODEL OF V1 FOR GAIN CONTROL AND GATING LONG-RANGE HORIZONTAL CONNECTIONS BY TOP-DOWN INFLUENCES

Valentin Piëch, Wu Li, Charles D. Gilbert and George N. Reeke

5.1 SUMMARY

Contour integration and visual pop-out based on texture orientation are important steps for object recognition that are performed in the earliest visual cortical areas. Recent experiments have shown that this processing can be influenced in primary visual cortex by the task that an animal is performing. The network model presented in this paper proposes a simple mechanism by which top-down influences can provide gain control and allow long-range horizontal connections in V1 to be gated via the feedback gain parameter, which controls the amount of local "self-excitation" in the network. We use phase-plane analysis to prove the stability properties of the circuit and to show how changing the self-excitation changes the gain of the

circuit without affecting overall stability. We demonstrate with large-scale simulations that gain control without long-range horizontal connections gives little improvement for naturalistic stimuli, but that a combination of gain control and long-range horizontal connections can lead to non-linear contour integration and the sharpening of edges. We further propose an additional component called the feedback setpoint that allows for exclusive gating of contextual interactions. We then reproduce quantitatively the time course of neuronal responses in V1 for texture masking experiments and pop-out of embedded contours, and account for both the immediate suppression by surround stimuli and the delayed facilitatory component associated with contour saliency, and its dependence on attention. Last, we show that the network behaves very stably in the presence of noise.

5.2 INTRODUCTION

One of the most important steps in object recognition is the ability to distinguish what belongs to the object, and what is part of the background. The borders of most objects consist of smooth edges that only change

direction slowly. The local edge orientation can thus be integrated along the edge following the rule of "good continuation" that has been discovered by the Gestalt school of psychologists (Wertheimer, 1923).

Recent studies in awake behaving primates have shown that the responses of single neurons in primary visual cortex encode on a single trial as much information about the presence of a contour consisting of short line elements embedded in a background of randomly oriented line elements as the behavioral decision of the primate after training (Li et al., 2004).

Moreover, prior to the animals' learning the same contour integration task, the responses did not contain any information about whether the stimulus was present (Li et al., 2008). The information in the response increased in parallel with the animals' performance during learning, the enhanced response after learning disappeared during anesthesia, but was only weakened when performing an irrelevant task in a different part of visual space (Li et al., 2008). This indicated that the long-range horizontal connections are under top-down control.

Previous models of contour integration in V1 (Ernst et al., 2004; Grossberg and Mingolla, 1985a; Li, 1998b; Ullman, 1992; VanRullen et al., 2001; Yen and Finkel, 1998) do not allow for realistic top-down control of long-range horizontal connections, do not tolerate sufficiently strong contextual interactions, or suffer from unrealistic temporal dynamics and stability problems.

We propose a network model that can be shown analytically to be stable, allows elegant integration of top-down influences for gating of long-range horizontal connections, demonstrates powerful non-linear contour integration in large-scale simulations, and reproduces the time course of electrophysiological responses of V1 neurons to contour integration experiments faithfully.

5.3 METHODS

In order to make a mathematical treatment of the network stability feasible, we first examine the dynamics of the simple local circuitry that forms the basis of the full network, and then extend it to include multiple orientations, long-range horizontal connections and feedback.

5.3.1 LOCAL NETWORK EQUATIONS

The simplest local network describing the activity of a single orientation column in V1 in our model consists of an excitatory network node and a divisive-inhibitory network node that are connected to each other; the excitatory node is also connected to itself (see figure 21).

The dynamics are described by the following coupled differential equations:

$$\dot{x} = \frac{1}{\tau_x}\left(-x + \frac{J_{xx}\ f_x(x) + e}{1 + J_{xy}\ f_y(y)}\right)$$

$$\dot{y} = \frac{1}{\tau_y}\left(-y + J_{yx}\ f_x(x) + i\right)$$

In these equations, x, an internal state variable that can be seen as an abstraction of the membrane potential, describes the activity of the excitatory node, and $f_x(x)$ is the response function, i.e. the transformation of the activity x to its response rate. This function can be seen as an abstraction of the membrane voltage to firing transformation that is often characterized with the current-firing relationship of a neuron. The time constant τ_x is the cell time constant, J_{xx} the connection strength of the excitatory network node onto itself and J_{yx} the connection strength of the divisive-inhibitory node onto the excitatory network node (the first index denotes the network node the connection is targeting, and the second one from where it is originating), and e is the external excitatory input to the excitatory network node. Similarly, y denotes the activity of the divisive-inhibitory node, $f_y(y)$ its response rate, τ_y the cell time constant, i the external excitatory input to the divisive-inhibitory node, and J_{yx} the connection strength of the excitatory network node to the divisive-inhibitory one.

The response functions for both the excitatory and divisive-inhibitory network nodes are threshold-linear functions that can be described by the following equations:

$$f_x(x) = \begin{cases} 0, & x < T_x \\ m_x(x - T_x), & x \geq T_x \end{cases}$$

$$f_y(y) = \begin{cases} 0, & y < T_y \\ m_y(y - T_y), & y \geq T_y \end{cases}$$

The slope of the response functions are m_x and m_y respectively, and the thresholds T_x and T_y.

To analyze the network dynamics of the local circuitry, we will resort to phase plane analysis, which depicts the dynamics of the network on a Cartesian plane. The first step is to calculate the so-called nullclines, the paths where the derivatives of the network activities are zero. This is important because at potentially stable solutions to the network dynamics, both derivatives should be 0, and thus, these are points where the nullclines intersect.

The equation for the excitatory nullcline is:

$$\dot{x} = \frac{1}{\tau_x}\left(-x + \frac{J_{xx}\ f_x(x) + e}{1 + J_{xy}\ f_y(y)}\right) = 0$$

$$\Rightarrow J_{xy}\ f_y(y) = \frac{J_{xx}\ f_x(x) + e}{x} - 1$$

There are four different cases that need to be distinguished to fully characterize the excitatory nullcline, as each response function can either be 0 below threshold, or a straight line above threshold. First, if both functions are below threshold, i.e. $f_x(x) = f_y(y) = 0$, the nullcline equation simplifies to:

$$0 = \frac{e}{x} - 1 \ \Rightarrow\ x = e$$

This is a vertical line in the phase plane at the position of the excitatory input to the excitatory node. This is the only case not depicted in Figure 20, but it is also a trivial one in that both responses are silent, and there is always an

intersection between the nullclines because the inhibitory nullcline is a horizontal line below the excitatory threshold (see below).

The next two cases are when only one of $f_x(x)$ or $f_y(y)$ is above threshold. Starting with only $f_y(y)$ being above threshold, the result is:

$$J_{xy}\ m_y\left(y - T_y\right) = \frac{e}{x} - 1$$

$$\Rightarrow y = \frac{e}{J_{xy}\ \ m_y\ x} - \frac{1}{J_{xy}\ \ m_y} + T_y$$

This equation is a hyperbola, which has a vertical asymptote at x = 0. Itcan be seen on the left side of the green dotted line (the excitatory threshold) in figure 22 panel 1Aa, and panels 1-3 Ba. It would be visible in the remaining panels 2Aa and 2Ab if the range of the y-axis in the plots were expanded.

The case where the excitatory response function $f_x(x)$ is above threshold and the inhibitory one below is given by:

$$0 = \frac{J_{xx}\ m_x(x - T_x) + e}{x} - 1$$

$$\Rightarrow x = \frac{J_{xx}\ m_x\ T_X - e}{J_{xx}\ m_x - 1}$$

This is a vertical line. It can be seen in figure 22 panels 1Aa and 2Aa starting at the dashed horizontal red line, which depicts the inhibitory node threshold.

The last case for the excitatory nullcline is where both the excitatory and inhibitory thresholds are exceeded. The corresponding equation is:

$$J_{xy}\ m_y(y - T_y) = \frac{J_{xx}\ m_x(x - T_x) + e}{x} - 1$$

$$\Rightarrow y = \frac{J_{xx}\ m_x(x - T_x) + e - x}{J_{xy}\ m_y\ x} + T_y$$

An interesting property of this equation and therefore of the local network can be seen when x goes to infinity:

$$\lim_{x\to\infty} y = \frac{J_{xx}\ m_x - 1}{J_{xy}\ m_y} + T_y$$

This tells us that when both the excitatory and the inhibitory threshold are exceeded in the network, there is a horizontal asymptote of the excitatory nullcline as x goes to infinity. This asymptote has been depicted as the green dashed line in figure 22 panels Aa and Ba. In figure 22 panels Ba1 – Ba3, one can see that as x grows, the excitatory nullcline approaches this horizontal asymptote; for weak inputs from below (figure 22 panels 1Ba and 2Ba), and for strong inputs from above (figure 22 panel 3Ba).

One last useful insight into the properties of the excitatory nullcline above both thresholds that will be useful later can be gained by calculating its slope:

$$\frac{\partial y}{\partial x} = \frac{\partial}{\partial x}\frac{J_{xx}\ m_x(-T_x) + e}{J_{xy}\ m_y\ x} = \left(\frac{e - J_{xx}\ m_x\ T_x}{J_{xy}\ m_y}\right)\frac{\partial}{\partial x}\frac{1}{x}$$

$$= \left(\frac{e - J_{xx}\ m_x\ T_x}{J_{xy}\ m_y}\right)\left(-\frac{1}{x^2}\right) = \frac{J_{xx}\ m_x\ T_x - e}{J_{xy}\ m_y\ x^2}$$

The important insight here is that the slope is directly proportional to J_{xx}, and can be negative for small J_{xx} (see figure 22 row Aa) and positive for large J_{xx} (see figure 22 panels 1Ba and 2Ba), and goes to 0 for large x (reflected by the horizontal asymptote).

The equation for the inhibitory nullcline is much easier to characterize:

$$\dot{y} = \frac{1}{\tau_y}\left(-y + J_{yx}\ f_x(x) + i\right) = 0$$

$$\Rightarrow y = J_{yx}\ f_x(x) + i$$

The equation for the inhibitory nullcline below the excitatory threshold is a horizontal line with the equation:

$$y = i$$

Above threshold, the equation is:

$$y = J_{yx}\ m_x(x - T_x) + i$$

which is a straight line with the slope $J_{yx}\, m_x$. Since both J_{yx} and m_x, the connection strength of the excitatory node to the divisive-inhibitory node and the response function slope, are always positive values, the line has a positive slope. Correspondingly, in figure 22, all panels showing the phase plane analysis (1Aa to 3Aa and 1Ba through 3Ba) show the inhibitory nullcline as a horizontal line to the left of the dotted green line, the excitatory neuron threshold, and a straight line with a positive slope above it.

5.3.2 STEADY-STATE SOLUTION OF THE SYMMETRIC NETWORK

While we explicitly simulate excitatory and divisive-inhibitory nodes in our simulations, it is possible analytically to derive the steady-state solution by letting the time constant τ_y of the divisive-inhibitory node approach 0, i.e. by

having the divisive-inhibitory current react instantaneously. It is easy to see that this has no influence on the position of the nullcline intersection, but only on the dynamics until reaching the steady state. The divisive-inhibitory node network equation then can be replaced by the equation for the inhibitory nullcline, and substituted into the equation for the excitatory network node:

$$\dot{x} = \frac{1}{\tau_x}\left(-x + \frac{J_{xx}\ f_x(x) + e}{1 + J_{xy}\ f_y(y)}\right) = \frac{1}{\tau_x}\left(-x + \frac{J_{xx}\ f_x(x) + e}{1 + J_{xy}\ f_y\left(J_{yx}\ f_x(x) + i\right)}\right) = 0$$

The solution we are most interested in is the one where both $f_x(x)$ and $f_y(y)$ are above threshold; the one where either or both are below threshold are trivial to derive. The steady-state equation is:

$$x = \frac{J_{xx} m_x(x - T_x) + e}{1 + J_{xy}\ m_y\left(J_{yx}\ m_x(x - T_x) + i - T_y\right)}$$

$$\Rightarrow J_{yx}\, J_{xy}\, m_x\, m_y\, x^2 + \left(J_{xy}\, m_y\left(i - T_y\right) + 1 - J_{yx}\, J_{xy}\, m_x\, m_y\, T_x - J_{xx} m_x\right)x - e + J_{xx} m_x\, T_x = 0$$

This is a quadratic equation and can be easily solved with the standard formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

with

$$a = J_{yx}\, J_{xy}\, m_x\, m_y$$

$$b = J_{xy}\, m_y\left(i - T_y\right) + 1 - J_{yx}\, J_{xy}\, m_x\, m_y - J_{xx} m_x$$

$$c = J_{xx}\, m_x\, T_x - e$$

Although there are two solutions, the left of the two is below threshold for the parameters we chose. But even if one were to choose the parameters differently, for geometric reasons (and this can also be easily demonstrated

by calculating the eigenvalues of the system at each point), the right one of the two is a stable equilibrium, while the left one is not.

The numerical value of y at the steady state can now be easily calculated with help of the equation for the inhibitory nullcline.

5.3.3 FEEDBACK GAIN AND SETPOINT

In order to help understand how this network could be controlled by top-down influences, we introduce two variables we call the feedback gain and the feedback setpoint.

The feedback gain is just the connection strength of the excitatory node to itself, and is the only variable that is hypothesized to be under top-down control:

$$FB_{gain} = J_{xx}$$

As the feedback gain is increased, so is the slope of the excitatory nullcline as can be seen in figure 22 comparing rows A (small feedback gain) with rows B (large feedback gain). The result is that any change in the excitatory input e to the excitatory network node or any change in the excitatory input i to the divisive-inhibitory node will lead to a larger change in the activities of both network nodes as the intersection of the nullclines moves a larger distance along the inhibitory or excitatory nullcline respectively since the nullcline intersection angle is smaller for larger feedback gain values (see figure 22 rows Aa and Ba).

The feedback setpoint is defined as the ratio of the connection strength of the excitatory node to itself to the connection strength of the divisive-inhibitory node to the excitatory node:

$$FB_{setpoint} = \frac{J_{xx}}{J_{xy}}$$

In the current simulations, the feedback setpoint is defined to be either a constant value (simulations in figure 24 row C), or controlled in a bottom-up

fashion by the input so that the network activity given by the bottom-up input is not changed by changing the feedback gain in the absence of long-range connectivity (as depicted by the gray arrows in figure 21 B; simulations in figure 24 row D). While this may seem difficult to achieve, it actually only requires that the feedback setpoint be linearly proportional to the bottom-up input. This can be shown by solving for x the equation for the steady-state solution we just derived, and then calculating the resulting y value from the equation for the y nullcline, imposing the condition that it be constant while changing J_{xx}. We omit this result for the general case, as without substituting in the numeric network parameters, it will give an unreadable formula.

For the network parameters that we chose, the numeric value for the input e that will not be changed by changing the feedback gain is identical to the feedback setpoint. This can also be seen in figure 24 row D, where both the feedback setpoint and the input are chosen to be 2, and hence, changing the feedback gain only changes the local slope of the excitatory input and hence the input sensitivity to additional inputs, but not the nullcline intersection.

It is important to understand that changing both the feedback gain and setpoint does not imply literally changing the synaptic strength between the neurons in primary visual cortex as it is done in our network, but likely manipulating the effective coupling strength in a more subtle way.

5.3.4 LONG RANGE CONNECTIVITY

Now that the local network is mathematically characterized, it can be extended by adding long-range connections that will connect network nodes over longer horizontal distances.

To do so, we define that at each location in space, there will be K neurons each tuned to a different orientation θ spaced apart by π/K. Further definitions are that, in subscripts, index 1 indicates the local neuron, while 2 refers to the neuron it is connecting to. The angle between the optimal orientation a neuron is tuned to and the line connecting two neurons at different spatial locations in cortical coordinates is ϑ_1 for the local neuron and ϑ_2 for the "remote" one, each being limited to the range of $-\pi/2$ to $\pi/2$.

For notational brevity, it is useful to define an angular tuning function that is based on the cosine and has as arguments an orientation, an optimal orientation and a full-width-at-half-maximum (FWHM) tuning width. This function will be 1 when the input angle is identical to the optimal angle, 0.5 when it is ½ tuning width away from the optimal angle to either side, and 0 if it is one width or more away:

$$cos_{fwhm}(\varphi, \varphi_{opt}, \varphi_{fwhm})$$

$$= \begin{cases} \frac{1}{2}\cos\left(\pi \frac{\varphi - \varphi_{opt}}{\varphi_{fwhm}}\right) + \frac{1}{2}, & |\varphi - \varphi_{opt}| < \varphi_{fwhm} \\ 0, & |\varphi - \varphi_{opt}| \geq \varphi_{fwhm} \end{cases}$$

With these definitions, the formula for the long-range excitatory-to-excitatory connections becomes:

$$L_{xx} = l_{xx} \exp\left(-\frac{\Delta x^2 + \Delta y^2}{d_\sigma{}^2}\right) cos_{fwhm}\left(-\vartheta_2, \vartheta_1, \frac{\pi}{2}\right) cos_{fwhm}\left(\frac{|\vartheta_1| + |\vartheta_2|}{2}, 0, \frac{\pi}{2}\right)$$

In short, there is a Gaussian decay with distance, the optimal connection strength is achieved for cocircular geometry, and there is a penalty for curvature (with the optimal configuration being collinearity), times a factor l_{xx} that determines overall synaptic connection strength. The chosen form for the curvature penalty (and the rest of the function) additionally ensures that the connectivity is symmetric, i.e. that the connection from neuron 1 to neuron 2 is identical to the one from neuron 2 to neuron 1.

The excitatory to divisive-inhibitory network node long-range kernel strength is almost identical:

$$L_{yx} = l_{yx} \exp\left(-\frac{\Delta x^2 + \Delta y^2}{d_\sigma{}^2}\right) cos_{fwhm}(-\vartheta_2, \vartheta_1, \pi /2)\, cos_{fwhm}\left(\frac{|\vartheta_1| + |\vartheta_2|}{2}, \frac{\pi}{2}, \frac{\pi}{2}\right)$$

The only difference is that the optimal configuration now is a parallel line and not a collinear one.

5.3.5 FURTHER DEFINITIONS FOR THE FULL NETWORK

There are a number of further definitions that were used for the simulations of the full network.

First, the optimal orientation of the input at every location in space and its strength had to be determined. This was done as in Sigman et al. 2002 by using steerable filters (Freeman and Adelson, 1991). Only the optimal orientation was used as input to the network, convolved with a tuning function, i.e. the cos_{fwhm} function, to allow for the network nodes to have a realistic input tuning.

Because of this realistic input tuning, and the fact that the long-range horizontal connections are also tuned around an optimal orientation, the connections from the local divisive-inhibitory neuron to the local excitatory node do not just connect to the excitatory node with the same orientation: They also connect to the nodes at the same spatial location with different optimal orientations with same tuning width as the bottom-up input tuning.

5.3.6 PARAMETER CHOICE

The parameters in the model were chosen to be as simple as possible while still taking into account the quantitative relationships found from experimental data.

The input tuning width θ_{In} and the divisive-inhibitory to excitatory local tuning width θ_{Jxy}, were chosen to be $\pi/4 = 45°$, and K = 12 different orientations are explicitly represented so that there are 12 excitatory and 12 divisive-inhibitory nodes at each location in space with a tuning separation of $\pi/12 = 15°$.

The excitatory and divisive-inhibitory node thresholds T_x and T_y were chosen to be 1, and the response function slopes m_x and m_y were chosen as 1 and 2 respectively. The reason for choosing the later response function slope m_y as 2 is that the current-firing relationship of inhibitory neurons is 6 times as steep as the one for excitatory neurons (e.g. see Novak et al. 2003), but the divisive-inhibitory network nodes would probably be a mixed pool of

excitatory and inhibitory neurons to perform divisive inhibition, and likely contain more excitatory neurons, so a compromise value of 2 was chosen. Note that as long as J_{yx} is chosen to be $1/m_y$ the steady-state firing rate of both excitatory and divisive-inhibitory network nodes will be very similar.

The cell time constant τ_x was defined as 1, and since both excitatory and inhibitory neurons have similar cell time constants *in vivo* in V1 (e.g. Nowak et al. 2003: regular spiking cells 10.4 ± 3.5 ms and 7.6 ± 4.2 ms), and the divisive-inhibitory pool would likely contain both types of neurons, τ_y was set to 1 as well.

The default feedback gain was set to $FB_{gain} = 1$, and therefore $J_{xx} = FB_{gain} = 1$. The default feedback setpoint was set to $FB_{setpoint} = 2$, and as a result, $J_{xy} = FB_{gain} / FB_{setpoint} = 0.5$. The divisive-inhibitory to excitatory connection strength was $J_{yx} = 1/m_y = 0.5$ so that the response rate of excitatory and divisive-inhibitory network nodes was identical (see figure 26).

The excitatory input i to the divisive-inhibitory network nodes was set to the divisive-inhibitory threshold $i = T_y = 1$, and for simplicity, the divisive inhibitory network nodes do not receive any further input. The excitatory input e to the excitatory nodes was $e = 0.75 * T_x = 0.75$ plus the input stimulus, which was the strength of the maximum orientation with a tuning FWHM of $\pi/4 = 45°$, and scaled from 0 to 2. This insured that for the simulations with a constant feedback setpoint $FB_{setpoint} = 2$, only the strongest inputs were a attenuated ($e_{max} = 2.75$) while most contours were amplified. The long-range connectivity scaling constants were both chosen for all simulations as $l_{xx} = l_{yx} = 0.1$ to ensure that at a low feedback gain of $FB_{gain} = 0.4$, the long-range connectivity had little influence on the simulations, while at a high feedback gain $FB_{gain} = 2.5$, it had a profound one. Choosing the two constants identical ensured agreement with the experimental data that there is no clear orientation preference in any cortical direction.

Lastly, the noise values for figure 27 were chosen to be additive noise with a Gaussian amplitude distribution of $\sigma_n = 0.1$ for figure 27B and $\sigma_n = 0.2$ for

figure 27C, both with an exponential distribution of 0.1 cell time constants in the time domain. The first set of values is numerically the same as in Li 1998, but since the input is scaled from 0 to 4 in the Li model, and 0 to 2 in ours, it is effectively twice as large. The second distribution with $\sigma_n = 0.2$ is twice as large as our first one, and given that the threshold is 0.25 away from the background input, about as large as possible without driving a lot of spontaneous background activity.

5.3.7 IMPLEMENTATION

The phase plane analysis was implemented in Mathematica (Wolfram Research, Champaign, IL), and allows for interactive real-time changing of the parameters to explore the network parameters very quickly.

The large-scale simulations have been implemented in Matlab (The Mathworks, Natick, MA), and accelerated via the use of 3D-Fourier transforms for convolution. Additionally, the Jacket toolbox (Accelereyes, Atlanta, GA) has been used, which allows offloading the simulation onto the

GPU, and accelerating the simulations by another order of magnitude. This allows calculating the large-scale simulation in Figure 24 with more than 1 million network nodes (220 x 220 x 24 network nodes) in under two minutes.

5.4 RESULTS

The basic circuit of the network is shown in panel A of Figure 21: It consists of an excitatory network node and a divisive-inhibitory network node. These are connected to each other with connection strength J_{yx} for the excitatory connection from the excitatory node to the divisive-inhibitory one, the connection strength J_{xy} for the divisive-inhibitory connection from the divisive-inhibitory to the excitatory node, and finally, the excitatory network node is connected to itself with connection strength J_{xx} (see Methods for details).

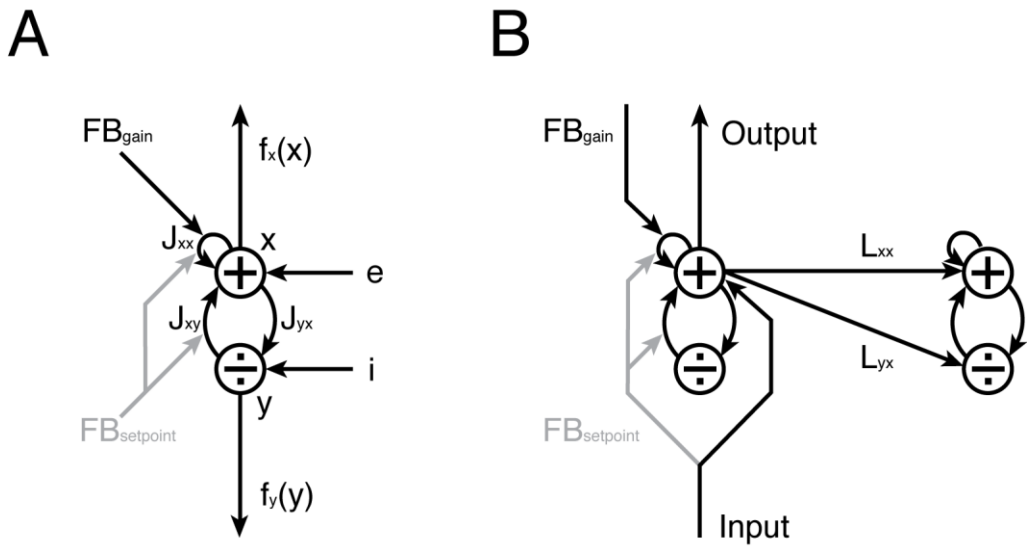

Figure 21. Schematic of network circuit.

A: Local circuit consisting of an excitatory (+) and a divisive-inhibitory (÷) network node. The excitatory to excitatory node connection strength is J_{xx}, the divisive-inhibitory to excitatory one Jxy, and the excitatory to divisive-inhibitory one J_{yx}. The excitatory input to the excitatory node is e, and the excitatory input to to divisive-inhibitory node is i. The activities x for the excitatory and y for the divisive-inhibitory node are converted into the response rate $f_x(x)$ and $f_y(y)$ respectively. The feedback gain FB_{gain} directly changes J_{xx}. Finally, the feedback setpoint FBsetpoint influences the ratio between J_{xx} and J_{xy}. This circuit is used for the phase plane analysis in figure 22.

B: Local circuit and long-range horizontal connections. The local excitatory network node connects to other excitatory network nodes at different spatial positions with connections strength L_{xx}, and to other divisive-inhibitory nodes with connection strength L_{yx}. The connection patterns of L_{xx} and L_{yx} can be seen in figure 23 for a network node responding optimally to a vertically oriented bar. The feedback gain is controlled via top-down influences, and the feedback setpoint is either constant (see figure 24 rows B and C) or proportional to the bottom-up input (see figure 24 row D). This circuit was used for the simulations in figures 24 through 27.

Additionally, there is an excitatory input e to the excitatory network node and an excitatory input i to the divisive-inhibitory network node. These external inputs, together with internal circuit connections, drive the activities x and y of the excitatory and divisive-inhibitory network node respectively. The node activities x and y, an abstraction of the membrane potential, are then translated into the response rates $f_x(x)$ and $f_y(y)$, each with a threshold-linear response function (see Methods for details).

To make the network circuit more flexible, we introduce two new variables that manipulate two of the connection strength parameters in the network in the following way: The feedback gain FB_{gain} directly changes the excitatory self-connection strength J_{xx}, and the feedback setpoint $FB_{setpoint}$ changes the ratio of excitatory self-connection strength to the divisive-inhibitory connection strength J_{xx}/J_{xy} (see Methods and Figure 21). Note that this does not increase the number of free parameters in the network, but is just a reformulation that makes the effect of manipulating the two network parameters more obvious.

This simple connection scheme makes the circuit very stable yet computationally surprisingly powerful. This can be most easily demonstrated with a standard technique of dynamic systems analysis, the phase plane analysis (see Methods and Figure 22 Panels Aa and Ba).

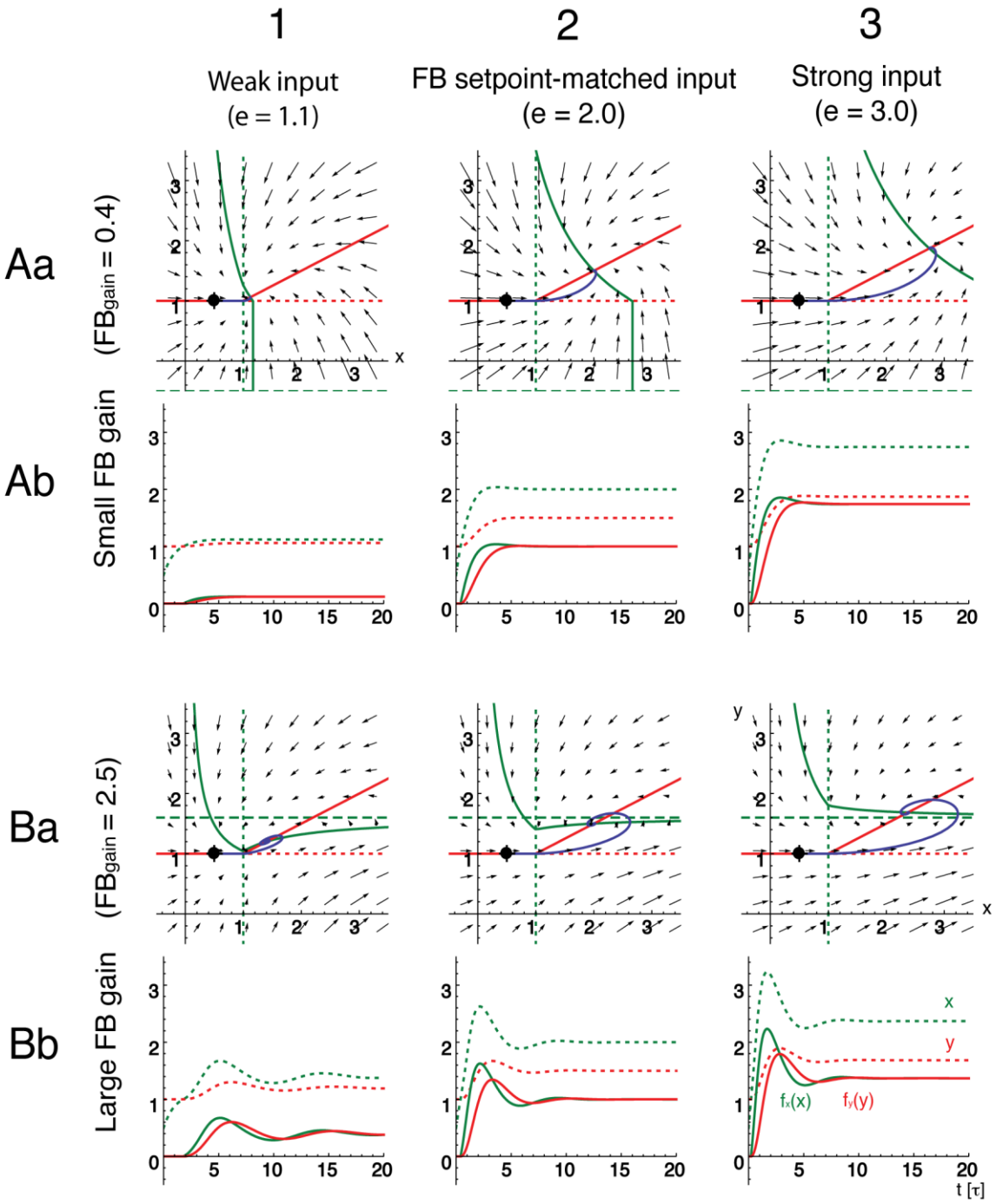

Figure 22. Phase plane analysis and activity and response time courses for low and high feedback gain at low, medium and high input strength.

Column 1: Network behavior to a weak, supra-threshold input. The input is enhanced for a higher feedback gain.

Column 2: Network behavior to an input whose strength is matched to the feedback setpoint. The steady-state activities and responses are independent of the feedback gain.

Column 3: Network behavior to an input higher than the match to the feedback setpoint. The input is suppressed to higher feedback gain.

Rows A: Small feedback gain (FB_{gain} = 0.4) leading to little amplification of weak inputs, and little suppression of strong inputs

Rows B: Large feedback gain (FB_{gain} = 2.5) leading to large amplification of weak inputs, and stronger suppression of strong inputs

Subpanels a: Phase plane analysis. The green solid line depicts the excitatory nullcline, and the solid red line the inhibitory nullcline. The excitatory threshold is depicted by a vertical dotted green line, and the inhibitory threshold by a horizontal dotted red line. The asymptote of the excitatory nullcline is depicted by the horizontal dashed green line. The solid black circle with a cross indicates the initial conditions in the phase plane, and the solid blue line depicts the trajectory of the network activities in the phase plane.

Subpanels b: Activity and response time courses. The dotted green and red lines depict the excitatory and inhibitory activities (an abstraction comparable to the membrane voltage), and the solid green and red lines the excitatory and inhibitory response rate.

In this analysis, the vector field representing the derivatives dx/dt and dy/dy of the activities x and y from the coupled differential equations are plotted in the Cartesian plane (depicted by the little arrows in Figure 22 rows Aa and Ba), and the nullclines, the places where these derivatives are zero for the excitatory and the divisive-inhibitory network node, are added (green and

red sold lines in Figure 22 Panels Aa and Ba). From any initial condition (or network state), one can find the evolution of the network dynamics by integrating the equations with small time steps, which is the equivalent of moving along the directions of the little arrows (the blue line in Figure 22 panels Aa and Ba). If there is a stable solution that can be reached by integrating the network equations, it must be in a place where the derivatives are zero, or equivalently, where the two nullclines intersect.

The great stability of this circuit can be appreciated without understanding the mathematical details by considering the shape of the nullclines, which are guaranteed to intersect, and even more, in such a way that at least one intersection is stable (see Methods for details): The divisive-inhibitory nullcline (solid red line in Figure 22 panels Aa and Ba) is always a horizontal line to the left of the excitatory node threshold (vertical dotted green line in Figure 22 panels Aa and Ba), and then a straight line with a positive slope to the right. The excitatory nullcline (solid green line in Figure 22 panels Aa and Ba), on the other hand, has a vertical asymptote at $x = 0$ (Figure 22 panels Aa and Ba), becomes a vertical line below the divisive-

inhibitory node threshold (horizontal dotted red line, Figure 22 panels 1Aa and 2Aa), and has a horizontal asymptote for sufficiently large values of $J_{xx} = FB_{gain}$ [1] (see Methods, dashed green line in Figure 22 panels Ba).

When starting from an initial network state (an example is depicted by the black circle with a cross in Figure 22 panels Aa and Ba), the network state follows the vector field in the phase plane until it reaches the stable equilibrium point where the two nullclines intersect (solid blue line in Figure 22 panels Aa and Ba).

The time course of the trajectory in the phase plane can be seen in Figure 22 panels Ab and Bb: The dotted green and red lines are the time courses of the activities x and y, and the solid green and red lines are the response rates $f_x(x)$ and $f_y(y)$ of the two network nodes. It is important to understand the difference between the activities and the response rates. The former is an

[1] To be precise, there is always a horizontal asymptote, but the excitatory nullcline will become vertical as soon as it is below the inhibitory node threshold, so that the asymptote will only be ever relevant if it is higher than the inhibitory node threshold (i.e. if the horizontal dashed green line is above the horizontal dotted red line). It is obvious that this does not have a negative impact on stability.

abstraction of the membrane potential, while the latter is an abstraction of the firing rate. One can see that as soon as the activities cross the threshold, whose numerical values are 1 in all simulations, the response rates appear in the plot, or equivalently, the neuron starts to fire (Figure 22 panels Ab and Bb).

We can now use our understanding of the phase plane to understand the network behavior for different feedback gains to different input strengths.

At a low feedback gain, here $FB_{gain} = 0.4$, the slope of the excitatory nullcline for suprathreshold inputs is always negative (Figure 22 panels Aa, see Methods for mathematical details). A weak input e = 1.1 gives a steady-state activity of x =1.12 and output of $f_x(x) = 0.12$ (Figure 22 panel Aa1 and Ab1), a feedback setpoint-matched input of e = 2.0 gives a steady-state activity of x = 2.0 and output $f_x(x) = 1.0$ (Figure 22 panel Aa2 and Ab2), and a strong input of e = 3.0 gives a steady-state activity of x = 2.74 and output $f_x(x) = 1.74$ (Figure 22 panel Aa3 and Ab3). This shows that at low feedback

gain, a weak input is essentially not amplified, a feedback setpoint-matched input is unchanged, and a strong input is nonetheless somewhat attenuated.

At a high feedback gain of FB_{gain} = 2.5, the shape of the excitatory nullcline is quite different in that it approaches a horizontal asymptote for all input strengths (note how the solid green line in Figure 22 panels Ba approaches the horizontal dashed green line). A weak input e = 1.1 gives a steady-state activity of x =1.4 and output of $f_x(x) = 0.4$ (Figure 22 panel Ba1 and Bb1), a feedback setpoint-matched input of e = 2.0 still gives a steady-state activity of x = 2.0 and output $f_x(x) = 1.0$ (Figure 22 panel Ba2 and Bb2), and a strong input of e = 3.0 gives a steady-state activity of x = 2.36 and output $f_x(x) = 1.36$ (Figure 22 panel Ba3 and Bb3). This shows that at high feedback gain, a weak input is strongly amplified, a feedback setpoint-matched input is unchanged, and a strong input is more strongly attenuated.

To summarize, at low feedback gain, the network leaves the input virtually unchanged except for attenuating the strongest inputs, while at high feedback gain, weak inputs are strongly amplified and strong inputs are

attenuated more strongly. A second important realization is that a feedback setpoint-matched input has the same steady-state output independent of the feedback gain, even though the slope of the excitatory nullcline is much more positive (compare Figure 22 panels A2 with panels B2). This means that only additional inputs Δe or Δi would be more strongly amplified, such as those arriving via long range horizontal connections (see Figure 24 row D).

One last intuition that can be gained from the phase-plane analysis (and proven mathematically, see Methods) is that because there is always a horizontal asymptote no matter how strong the "self-excitation" J_{xx} is (only the location of the asymptote shifts upwards with increasing J_{xx}), the network activities will always be bounded even for the case of many connected nodes because an increase in their activity is in the "engineering worst-case scenario" similar to a corresponding increase in J_{xx}. This property is very useful when connecting many network circuits together, as many other recurrent network models have to introduce global normalization, a saturating response function, strong oscillations or only allow weak

recurrent connections to keep the network activities from growing without bounds under some conditions.

Now that we have established the network dynamics of the local circuit and explored the functions of feedback gain and feedback setpoint, we can extend this local circuit to become a network of the superficial layers of primary visual cortex including long-range horizontal connections.

Each orientation column is modeled by this local circuit, which responds optimally to a stimulus at its optimal orientation, and at each spatial location, K = 12 orientation columns form a hypercolumn so that circuits representing different optimal orientations are separated by $\pi/K = 15°$ (see Methods for details). As depicted in Figure 21 panel B, we add excitatory connections L_{xx} from each local excitatory network node to other excitatory network nodes and excitatory connections L_{yx} from the excitatory network node to the divisive-inhibitory network node, depending on the geometric relationship between the stimuli they optimally respond to. This relationship is shown in

Figure 23 in panel A for an orientation column that responds optimally to a vertical stimulus (see Methods section for the mathematical description).

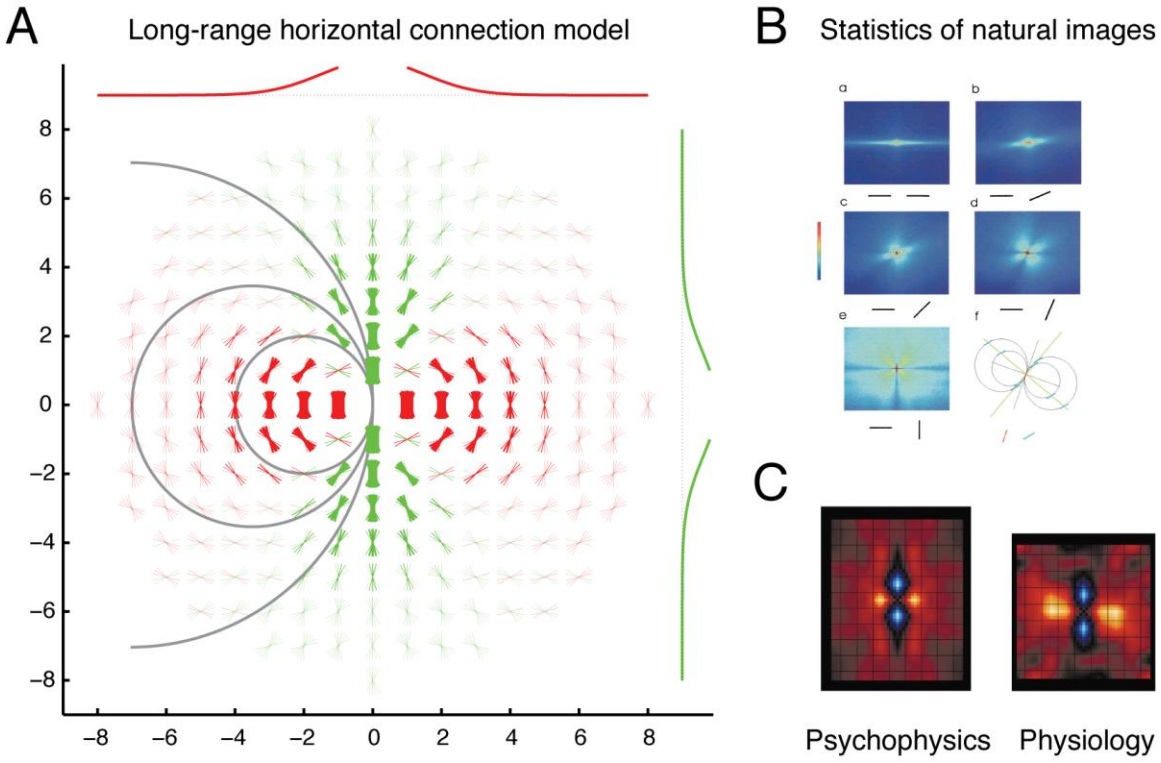

Figure 23. Long range connectivity, statistics of natural images, and psychophysical and physiological evidence. A: Long range connectivity for an excitatory node with the vertical orientation preference. The green oriented bars represent connections to other excitatory nodes, while the red ones represent connections to divisive-inhibitory network nodes These are reflected in the contextual interactions surrounding the RF (Kapadia et al. 2000; see panel C). The orientation of the bar depicts the orientation of the node it is connecting to, while both the thickness and the intensity of the color represent the connection strength. The green graph to the right of the connection representation and the red one above it additionally show the connection density decay, which is modeled by an Gaussian with a decay constant of three hypercolumn distances (i.e. the density at ~2.25 - 2.4 mm is 1/e = 36.7%, in agreement with data in Stettler et al. 2002) . Also note that both excitatory and inhibitory connections follow the rule of cocircularity (and collinearity, which is a special case of cocircularity; see panel B,

Sigman et al. 2001) for the optimal connection strength at a given spatial position, as can be seen by comparing the gray circles with the bars.

B: Statistics of natural images: Spatial dependence of the histogram of cooccurring pairs for different geometrical configurations. (a) The probability of finding a pair of iso-oriented segments as a function of their relative position; a pair of segments at relative orientation of 22.5° (b), 45° (c), 67.5° (d), or 90° (e). (f) Cocircularity solution for a particular example of two segments. The solutions to the problem of cocircularity are two orthogonal lines, whose main have values $(\psi + \varphi) / 2$ or $(\psi + \varphi + \pi) / 2$. For the example given, φ (red segment) = 20°, ψ (blue segment) = 40°, and the two solutions (green lines) are 30° and 120° (all angles from the vertical axis). Reproduced from Sigman et al. 2001.

C: Psychophysical and physiological results for excitatory and inhibitory interactions of vertical bar elements as a function of spatial arrangement . Two-dimensional maps obtained from psychophysical studies and from neural recordings are shown side-by-side for comparison. Both maps were obtained with high contrast stimuli. Psychophysical map: 5° flanks, fovea, subject MK. Physiological map: context map, average of 12 recording sites, 50% contrast. The basic features of the maps are quite similar. Reproduced from Kapadia et al. 2000.

The overall decay of the kernel strength with distance was well-fitted with a Gaussian distribution with a decay constant of three hypercolumn distances (i.e. the density at ~2.25 - 2.4 mm is 1/e = 36.7%, in agreement with the labeling with GFP by recombinant adenovirus in (Stettler et al., 2002)). Within this envelope, the strongest connection strength between two elements is for a collinear or cocircular configuration, in agreement with the statistics of natural images (compare Figure 23 panel A with panel B reproduced from (Sigman et al., 2001)). Last, there is a penalty for increasing curvature, and a maximal connection strength of excitatory to

excitatory connections for collinear segments, while the maximal connection strength for excitatory to divisive-inhibitory network nodes is for parallel flanking elements (see Methods for details). This ensures that there is an overall preference for connecting orientation columns of similar orientation preference (Stettler et al., 2002; Ts'o et al., 1986), and gives very good agreement with the psychophysical and physiological measures of the contextual interactions between elements of similar orientation (compare Figure 23 panel A with panel C reproduced from (Kapadia et al., 2000)).

Now that the long-range connectivity is established, we can demonstrate the effectiveness of the network in responding to a naturalistic stimulus, namely the most commonly used test image in computer science of a girl known as Lena wearing a hat (Figure 24 panel A1).

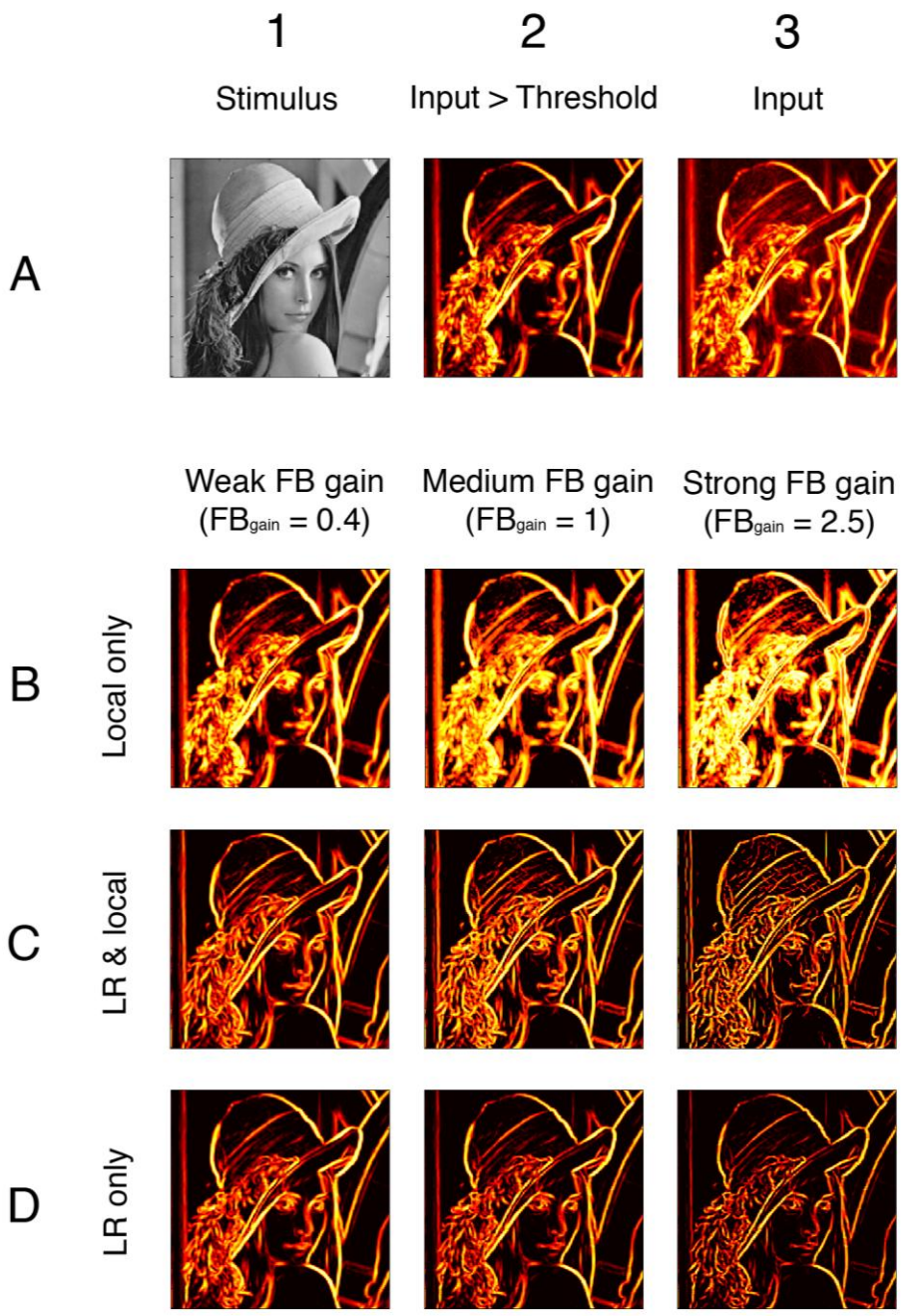

Figure 24. Large-scale simulation with naturalistic stimuli at different feedback gain levels.

A1: The raw test stimulus, a black-and-white version of the most commonly used test image for image processing.

A2: The input going into the network that would be able to elicit a supra-threshold response in the absence of local and long-range connections. Only the strength at the optimal orientation is shown.

A3: The input going into the network including subthreshold activity.

Row B: The network output at the optimal orientation in the absence of long-range connections with a feedback setpoint of 2. Increasing the feedback gain results in an amplification of weak inputs as in figure 22 panels 1Ba and 1Bb, and a suppression of the strongest inputs as in panels 3Ba and 3Bb.

Row C: The network output at the optimal orientation with long-range connections at a feedback setpoint of 2. Increasing the feedback gain results both in an amplification of the weakest inputs and directional contour integration and "sharpening" of the object contours through the long-range connectivity.

Row D: The network output at the optimal orientation with long-range connections at a feedback setpoint that is matched to the input in such a way that changing the feedback gain has no effect on the activity and responses driven by bottom-up inputs (see figure 22 panels 2Ba and 2Bb). The result is that only directional contour integration and "sharpening" of the object contours is taking place, without a direct amplification of weak inputs. In this way, long-range horizontal connections can be gated exclusively.

Figure 24 panel A1 depicts the raw image. The network itself never directly sees this image; instead it receives as input the results of calculating the local orientation strength with help of a quadrature pair of steerable filters, mimicking the processing of the visual inputs by complex cells (see Methods for details). The strength at optimal orientation is shown in Figure 24 panel A3 (for details, see Methods section). If this input is presented to the network with all local and long-range horizontal connections J_{xx}, J_{xy}, J_{yx}, L_{xx}

and L_{yx} set to zero, the network output is that shown in Figure 24 panel A2, which is just a thresholded version of the input.

Each of the next three rows (Figure 24 rows B, C, and D) shows a different network configuration, while each column depicts each network configuration with a weak, an intermediate, and a strong feedback gain.

The first row (Figure 24 row B) has no long-range horizontal connections, i.e. $L_{xx} = L_{yx} = 0$. The result is that as the feedback gain is increased, weak inputs are amplified, and some of the strongest ones are attenuated, but without taking any information from the other orientation columns into account. This is equivalent to a non-linear scaling of the inputs. Not too surprisingly, the effect of this transformation is not very useful; it is just a non-linear compression of the local orientation contrast.

The second row (Figure 24 row C) is a simulation of the network with a constant feedback setpoint $FB_{setpoint} = 2$ (and long-range connectivity restored). This means that the feedback setpoint is not adjusted, and thus

would not require any mechanisms to control it. The network circuit would thus be the one in Figure 21 panel B without the gray connections. As a result, the gain of the local circuit is higher for weak inputs, and leads to both an amplification of weak bottom-up inputs and a more pronounced effect of long-range horizontal connections for weak contours. At the highest feedback gain level in (Figure 24 panel C3), one can see both how strong the weakest contours have become because of the non-linear contour integration, and how much they have been sharpened at the same time, almost to the point of a sketch).

The last row (Figure 24 row D) is a simulation of the network in which the feedback setpoint is matched to the input strength at every location in a bottom-up manner by the gray connections in Figure 21 panel B. The result is that increasing the feedback gain only affects contextual interactions mediated by long-range horizontal connections and has no influence on bottom-up inputs. This is possible because when the feedback setpoint is matched to the input, the intersection of the nullclines is independent of the feedback gain (compare Figure 22 panels Aa2 and Ab2 with panels Ba2 and

Bb2 and see Methods; with the chosen parameters, the numerical values for the feedback setpoint are identical to the input values to be matched). At the highest feedback gain level in (Figure 24 panel D3), one can see that while the contours have been "sharpened" and "cleaned up" by the non-linear contour integration, the analog aspect of the input has not been lost as in Figure 24 panel C3). This represents "exclusive" gating of the long-range horizontal connections and contextual interactions because in the absence of contextual stimuli, the feedback gain has no effect on the stimuli in the "classical" receptive field.

Next, we use the model quantitatively to reproduce the time course of electrophysiological recordings of contextual interactions in primary visual cortex of awake behaving primates with the same network parameters. Since the stimuli consist of oriented bars with equal contrast, there is no qualitative difference between the version of the model with a constant feedback setpoint, and the one that adjusts it to match the input. We thus performed the following simulations on the model with a constant feedback setpoint of $FB_{setpoint} = 2$.

The first set of experiments compared the human psychophysical performance in an orientation discrimination task in the presence of masks consisting of oriented bars with eletrophysiological recordings from superficial layer neurons in primary visual cortex of awake behaving primates that were performing an irrelevant task (Li et al., 2000) while being shown the stimuli with the corresponding masks, and found good agreement between the two. Figure 25 panel A1 shows the time course of the population response to a single bar presented without a mask, with a mask consisting of orthogonal bars, with a mask of randomly oriented bars, and with a mask consisting of parallel bars. Figure 25 panel A2 depicts the peristimulus time histogram of a single example cell to the same stimuli, which shows larger response differences to different masks than the population, but with the same ordering of the response strengths.

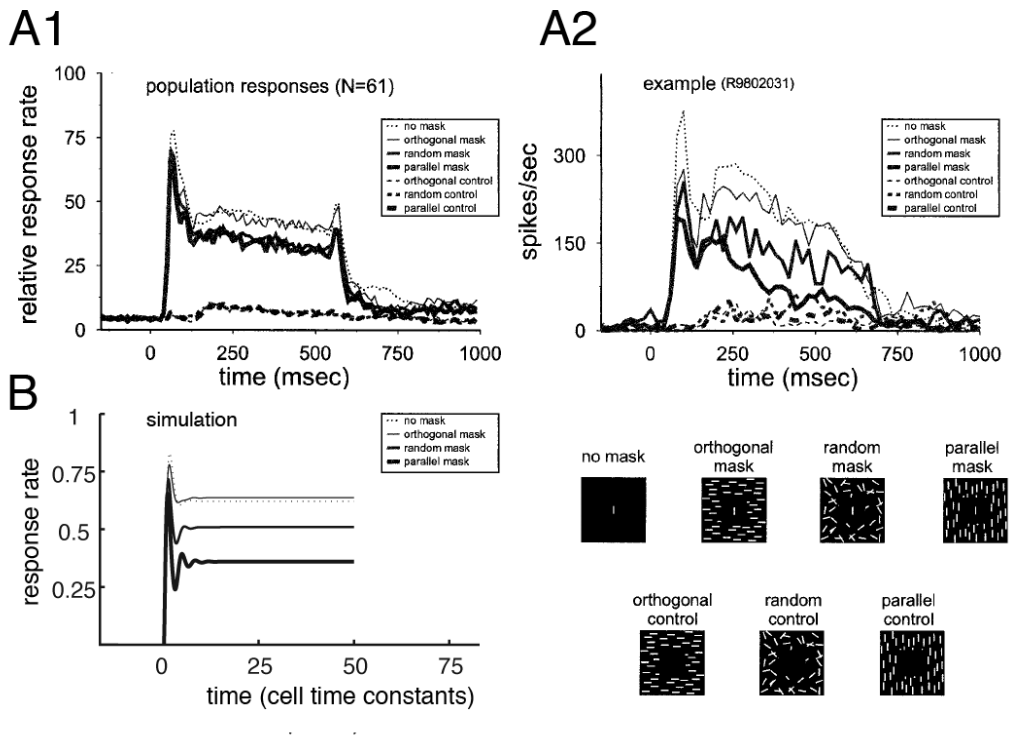

Figure 25. Neural and simulated response time courses to textured masks in the surround.

A1: Population responses from all recording sites (PSTHs averaged for all the recording sites for each bin, n = 61). Before averaging across all recording sites, the PSTHs from each recording site under all test conditions were normalized in such a way that the peak of the PSTH for no-mask condition was always 100. Binwidth here is 10 ms. Reproduced from Li et al. 2000.

A2: Peristimulus time histograms (PSTHs) from 1 typical recording site under different masking conditions. Neuronal responses to the surround lines alone are also shown here as controls. Time 0 indicates stimulus onset, 500 ms indicates stimulus offset. A binwidth of 20 ms was used to construct the PSTHs. Reproduced from Li et al. 2000.

B: Network simulation at the default feedback gain of 1 and a feedback setpoint of 2. For the simulation, the 8 positions immediately surrounding the test stimulus did not receive any input, similar to the electrophysiological experiments. Note that the responses show a similar time course as the experimental data, with the orthogonal mask giving almost the same response strength as the bar alone except at the stimulus onset, and that the masks consisting of randomly oriented and parallel bars give markedly lower responses in the same order as the neural

data. While the population effect seems weaker then the simulation result, the example cell shows a stronger effect. The simulation could be adjusted to be more similar to either data set by decreasing or increasing the feedback gain.

In Figure 25 panel B, we have simulated the experiment on a 32 x 32 element grid at the default feedback gain of 1 and a feedback setpoint of 2. For the simulation, the 8 positions immediately surrounding the test stimulus did not receive any input, similar to the electrophysiological experiments. Note that the responses show a similar time course as the experimental data Figure 25 panels A1 and A2, with the orthogonal mask giving almost the same response strength as the bar alone except at the stimulus onset, and that the masks consisting of randomly oriented and parallel bars give markedly lower responses in the same order as the neural data. While the population effect seems weaker than the simulation result, the example cell shows a stronger effect. The simulation could be adjusted to be more similar to either data set by decreasing or increasing the feedback gain.

The next simulation reproduces the results of contour integration and perceptual learning of pop-out of embedded collinear contours in a patch of randomly oriented bars (Li et al., 2006, 2008). In brief, awake behaving

primates were trained to detect embedded collinear contours consisting of 1 through 9 bars in a patch of randomly oriented bars, while electrophysiological recordings were made from neurons in the superficial layers of primary visual cortex. Before learning, the contours did not elicit strong responses, but after learning there was more than a twofold increase of the steady-state response. The size of the response increase depended on the actual task being done and the effect of learning completely disappeared under anesthesia (Li et al., 2006, 2008).

These effects are simulated by proposing an envelope of spatial attention that is mediated by top-down influences, and has the shape of a Gaussian of the same standard deviation as the long-range horizontal connections convolved with the stimulus, but without orientation specificity. The reason for this choice is that the spatial extent of feedback from V2 to V1 has the same spatial extent as the long-range horizontal connections, but lacks orientation-specificity (Stettler et al., 2002). This attentional envelope is hypothesized to change the feedback gain in our model from the default value of 1 to a peak value of approximately 3 at the center of the location

where the contour would appear, and is shown in Figure 26 panel A1 together with the bottom-up input with an embedded 9-bar stimulus. The same attentional envelope is used for all simulations since the animal did not know how many collinear elements the figure on the next trial would contain.

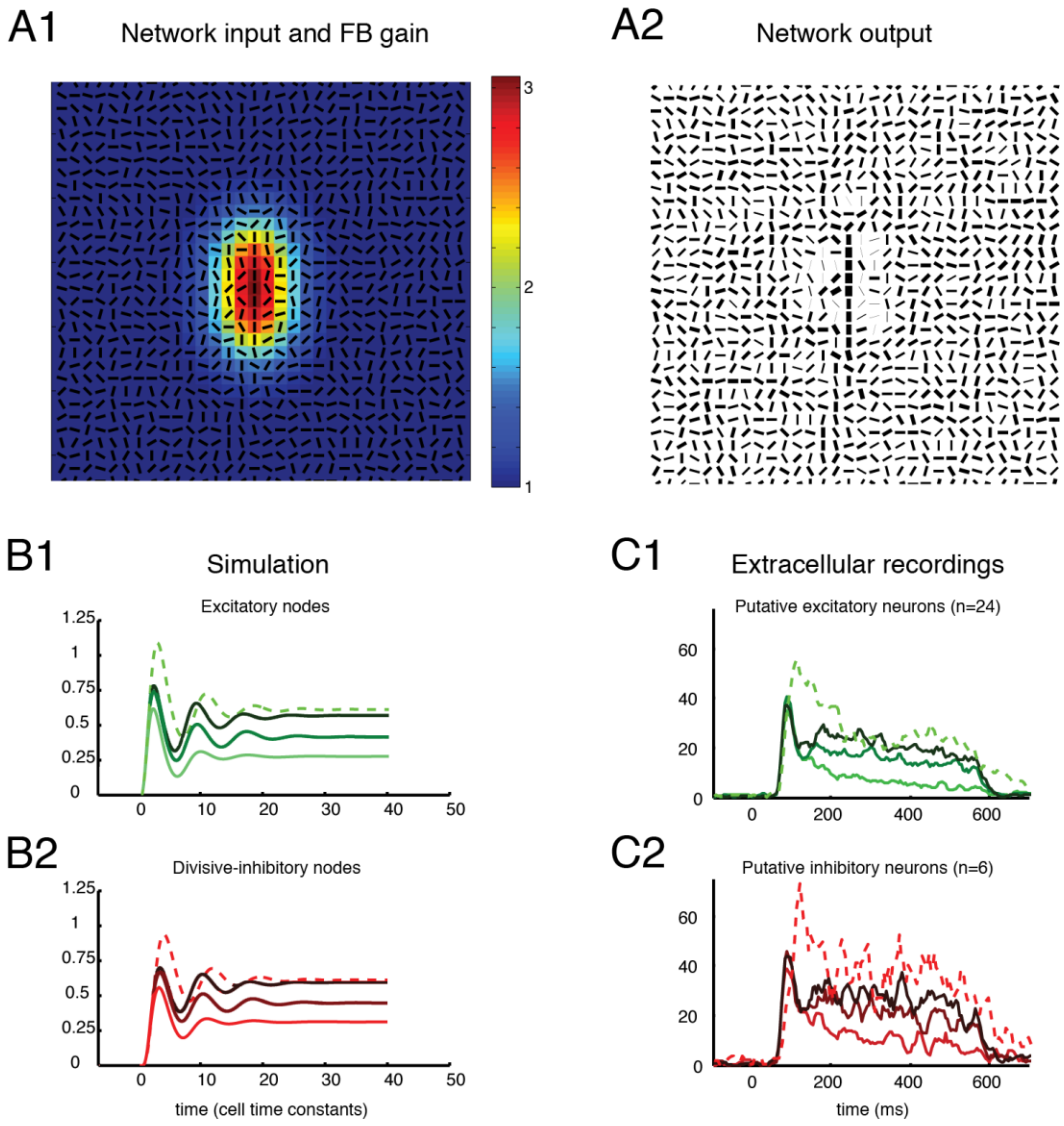

Figure 26. Pop-out of collinear elements embedded in a random background.

A1: Network input consisting of 9 vertical collinear elements embedded in a background of randomly oriented elements. The feedback gain is a uniform background of 1 and a Gaussian kernel with the same width as the long-range horizontal connections convolved with a 9 element vertical line at the location of the embedded contour stimulus with a strength chosen to give a maximum feedback gain of approximately 3. This simulates the allocation of spatial attention to the stimulus location. All simulations in this figure use the same feedback gain distribution even for shorter 1 and 5 bar contours to simulate the animal expecting a 1 through 9 bar length

contour on an individual trial. The feedback gain is not orientation-specific, in agreement with experimental data that the feedback connections from V2 lack orientation specificity (Stettler et al. 2002).

A2: Network output at the optimal orientation averaged over 50 cell time constants. Note the pop-out of the embedded contour and the suppression of the surround random bars.

B and C: The dashed line indicates the response to a single bar without background. The lightest lines depict the response to single bar with a background consisting of random lines, the darker ones those of an embedded 5 bar stimulus, and the darkest ones those to the 9 bar stimulus.

B: Network output for excitatory (B1) and divisive-inhibitory (B2) network node at the optimal orientation at the center of the stimulus. Both the excitatory and divisive-inhibitory nodes show very similar time courses as a result of the slope of the inhibitory nullcline of the local network (see figure 22 panels Aa and Ba). Note the agreement with the extracellular recordings in panels C1 and C2.

C: Time course for putative excitatory (C1) and inhibitory neurons (C2) identified by the shape of the waveform from extracellular recordings. Note the similarity for both types of neurons except for the seemingly larger variability caused by the much smaller number of inhibitory (n=6) compared to excitatory neurons (n=24). Post-hoc analysis of the data of MA in Li et al. 2006.

Figure 26 panel A2 shows the network output averaged over 50 cell time constants by displaying the thickness of the oriented bars proportional to the average output. The pop-out of the embedded 9-bar stimulus is very obvious due to the high feedback gain at its location. Also note the relatively small but noticeable differences between elements forming smooth contours and their surround bars in the areas outside of the attentional spotlight: At least

the fast component of perceptual learning might consist of learning where to allocate spatial attention based on small initial differences in saliency. Next, Figure 26 panels B show the time course in response to a single bar without background (dashed line), and 1, 5, and 9 bars embedded in a random background (light, darker, and darkest line), in green in B1 for the center excitatory network node, and in red in B2 for the center divisive-inhibitory network node.

The simulation is in close agreement with the time course of 24 putative excitatory neurons in Figure 26 panel C1 and 6 putative inhibitory neurons in Figure 26 panel C2 that have been identified in a post-hoc analysis of the spike wave forms recorded in those experiments. It is striking how similar the time courses of both putative excitatory and inhibitory neurons are. We can fully account for this in our model by choosing the connection strengths of the local excitatory nodes to the divisive-inhibitory nodes as the inverse of the slope of the response function of the divisive-inhibitory neuron, i.e. $J_{yx} = 1 / m_y$ (see Methods and Figure 22).

One other similarity between the simulation and the neuronal data is that there is an almost immediate difference in both between presenting the bar alone or embedded in a background (Figure 26 panels B and C – compare the dashed with the solid lines), but very little difference between the 1, 5, and 9 bar conditions. The simulation does show a small difference in the beginning, but given that it does not take conduction and synaptic transmission delays into account, the difference would be even smaller in a biophysically more realistic model, so it is surprising how good the agreement is nonetheless.

The final Figure 27 demonstrates the robustness of the model in the presence of noise. Despite the fact that the model is recurrent and does not exhibit strong oscillations, one can see that even at the highest noise level simulated (which effectively has a 4 times larger amplitude than in a comparable model, such as in (Li, 1998b)), the model starts to fail gracefully and does not generate artificial contours etc., but rather looses a bit more detail (e.g. compare the weak contours in the top quarter hat in Figure 27 panels A1, B1 and C1). Even more, notice that the onset response is much less susceptible

to noise in Figure 27 panels B2 and C2. This seems to be in agreement with experimental data (compare the onset response of the neuronal data in Figure 26 panels C1 and C2 with the later part of the response), and would be useful for rapid decision making essential for survival.

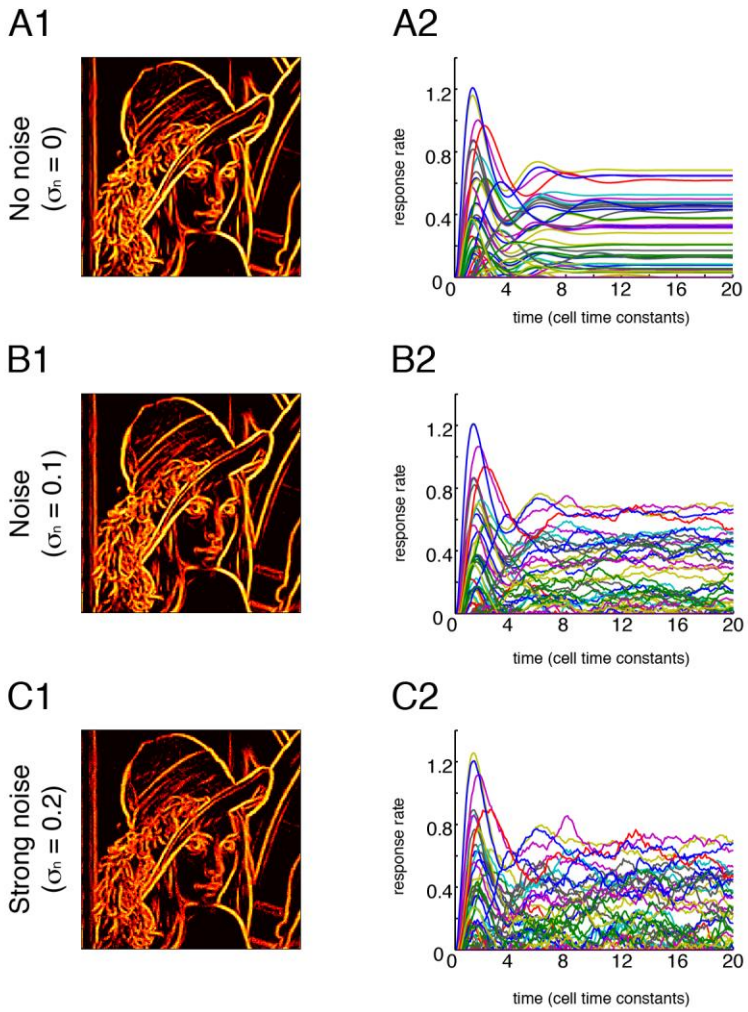

Figure 27. Robustness of the network to noisy inputs.

Column 1: Average output of the optimal orientation over the simulation duration of 20 cell time constants.

Column 2: Response time course of every 40th network node in x and y direction.

Row A: The network output at the optimal orientation in the absence of noise, as in the simulations in figure 23.

Row B: The network output at the optimal orientation in the presence of Gaussian noise with an amplitude of sn = 0.1 as in Li 1998. Note that the input in our simulations is scaled from 0 to 2 and not 0 to 4, so that the noise here is twice as strong as in that model.

Row C: The network response at the optimal orientation when the Gaussian noise amplitude is doubled to sn = 0.2. Increasing the noise much beyond this point would make a subset of network nodes reach threshold regularly since the background input is 0.75, and the threshold 1.

One additional observation is that the onset response seems to be less affected by noise then the later steady state, in agreement with experimental data (see figures 24C1 and C2)

5.5 DISCUSSION

5.5.1 INTRINSIC LOCAL CONNECTIVITY AND LONG-RANGE HORIZONTAL CONNECTIONS

Recurrently connected excitatory and inhibitory neurons make up an essential "basic circuit" of many models of cortical function and brain circuits in general (Shepherd, 1990). This would seem to suggest that it should be possible to find at least two different classes of neurons whose behavior can be distinguished. Yet in experiments studying pop-out of embedded contours, the behavior of all cells in V1 seems very homogeneous in that after learning, all cells show facilitation for more salient embedded

contours (Li et al., 2004, 2006), and even more, all cells show a similar suppression to the contextual background.

This can be relatively easily achieved for local interactions by the recurrent coupling of excitatory and inhibitory neurons, or in our case, the excitatory and divisive-inhibitory network nodes, by choosing the connection strength from the excitatory to the inhibitory (divisive-inhibitory in our case) network nodes to be the inverse of the slope of the inhibitory response function, $J_{yx} = 1/m_y$ (see Methods).

But in order to preserve this relationship even when including long-range influences that are strong enough to change the firing rates more than twofold in the presence of strong inhibition, it is important not only that the inhibitory nullcline have a positive slope, but that the excitatory nullcline does as well. This can be achieved by connecting the excitatory network node recurrently to itself (Miller, 2008; Tsodyks et al., 1997). This is also an essential component of our model, but with two differences: First, the strength of this connection is directly under control of top-down influences,

and can thereby control the gain of the circuit so that contour integration and pop-out are under the control of (spatial) attention. Second, using a divisive-inhibitory instead of a subtractive-inhibitory network node greatly improves the stability of the network by introducing a horizontal asymptote for the excitatory nullcline. This ensures that the network activity cannot accidentally grow without bounds for some inputs.

5.5.2 TOP-DOWN INFLUENCES

Previous work from our laboratory has shown that neurons in the superficial layers of V1 can adjust their properties to encode more information pertaining to a task when an animal is performing the actual task compared to an irrelevant one (Li et al., 2004, 2006), and that during the perceptual learning phase, the information relevant to the task in the neuronal responses increases in parallel with learning (Li et al., 2008).

In our model, this is simulated by increasing the feedback gain at the spatial location where the stimulus is expected (see Figure 26). This view is

compatible with either space- or feature-based attention (McAdams and Maunsell, 1999; Treue and Martinez Trujillo, 1999), but also with object-based attention (Roelfsema et al., 1998), insofar as the embedded contour is viewed as an object that will give rise to an attentional envelope via feedback connections. Indeed, differentiating between these modes of attention may not be necessary as long as they all lead to a similar attentional envelope.

One extension of this model that still has to be explored is to allow feedback to gate different sets of long-range horizontal connections for different tasks, and to allow perceptual learning for the set that is engaged in the task.

5.5.3 TIMING OF CONTEXT-RELATED RESPONSES

Placing stimuli outside of the classical receptive field often leads to a suppression of neural responses in V1 (Bair et al., 2003; Bishop et al., 1973; Knierim and van Essen, 1992; Li et al., 2000; Nothdurft et al., 1999; Walker et al., 2000). Moreover, this contextual inhibition is very quick, often

lowering the peak of the initial onset response (Bair et al., 2003; Knierim and van Essen, 1992; Li et al., 2000; Nothdurft et al., 1999). On the other hand, recent studies of the neuronal correlates of the pop-out of embedded contours in a complex background in V1 (Li et al., 2004, 2006) have shown an additional strong facilitatory component that is proportional to the saliency of the embedded contour, and that is delayed by 100-150 ms. Previous studies investigating the influence of higher-order, contextually dependent properties have shown similar delays (Bauer and Heinze, 2002; Kinoshita and Komatsu, 2001; Lamme, 1995; Lee et al., 2002; Li et al., 2000, 2001; Roelfsema et al., 1998; Rossi et al., 2001; Super et al., 2001; Zipser et al., 1996).

This has given rise to a debate on the origin of the delayed facilitatory component: It has been argued that the delay must be due to feedback (Lamme, 1995; Lee et al., 2002; Li et al., 2001; Zipser et al., 1996), though feedback may actually be associated with an early inhibitory component (Bair et al., 2003). A different view is that perhaps the relatively slow

conduction velocity of the unmyelinated axons of the long-range horizontal connections may be responsible for the delay (Bringuier et al., 1999).

Our network has reproduced both effects quantitatively without the necessity for a delay due to feedback or the slow propagation velocity of the axons: The immediate surround inhibition can been seen in the suppression of the initial peak in Figure 26 due to the presence of the surrounding random bars (compare dashed line in panels B with solid lines), while the strong facilitatory component does not start to develop until after approximately 5 cell time constants, and takes about 10 cell time constants to fully develop, because of the delay that the network exhibits before reaching equilibrium. This time would be somewhat shorter than the 100-150 ms that has been reported in vivo (Li et al., 2006): Assuming a cell time constant of 10.4 ms (Nowak et al., 2003), the facilitation would start around 60 ms. But there are factors that may increase this duration in vivo: The current implementation of the model does not include any synaptic delays or slow conduction velocity of the long-range horizontal connections (Bringuier et al., 1999),

nor does it take into account any other delays that could result from the biophysical implementation of the circuit in vivo.

5.5.4 EXTENT OF CONTEXTUAL INTERACTIONS

A recent study measuring the extent of functional connectivity in vivo has found that at lower stimulus contrast, local field potentials had large amplitudes and traveled over larger distances in primary visual cortex than at higher contrast (Nauhaus et al., 2009).

Our model would provide an explanation and a mechanism for this finding assuming that the feedback setpoint is either constant or adjusted to a smaller degree than matching it to the input: Under these circumstances, the gain both for bottom-up and contextual influences would be larger than at higher contrast, and as a consequence, lead to larger field potential amplitudes and further traveling waves. Even more, our model predicts that this may be also under top-down control and may be altered in a task-dependent manner by adjusting the feedback gain.

5.5.5 RELATIONSHIP TO PREVIOUS MODELS

The components of our model are central to many other well-established models of primary visual cortex: Divisive inhibition has been used to explain many of the properties of the classical receptive field (Carandini and Heeger, 1994; Heeger, 1992); a network of coupled excitatory and inhibitory node pairs with similar-shaped long-range horizontal connection kernels (Li, 1998a, 1999a; Li, 1998b, 1999b, 2000, 2001) provided the most complete explanation of pop-out and texture segmentation in primary visual cortex; and an inhibition-stabilized self-excitatory network to explain data from intracellular recordings in surround inhibition experiments (Miller, 2008) has been proposed.

But it is only the combination of divisive inhibition with self-excitation under top-down control that allows for gain control and gating of long-range horizontal connections, is guaranteed to have bounded network activities, and avoids the need to have network parameters that give rise to all-or-none oscillations that need to be contained by a saturating response function and non-linear normalization currents. To our knowledge, no other model can

gate long-range horizontal connections exclusively with a simple network mechanism based on top-down feedback, and allows for stable responses despite strong recurrent connections over a large range of feedback gain.

5.5.6 SUMMARY

The network presented in this paper proposes a simple mechanism by which top-down influences can provide gain control and allow long-range horizontal connections to be gated via a feedback gain parameter that controls the amount of local "self-excitation" in the network. We use phase-plane analysis to prove the superior stability properties of the circuit and how changing the self-excitation changes the gain of the circuit without affecting overall stability. We show with large-scale simulations that gain control without long-range horizontal connections gives little improvement for naturalistic stimuli, but that a combination of gain control and long-range horizontal connections can lead to nonlinear contour integration and sharpening of edges. We further propose an additional component called the feedback setpoint that allows for exclusive gating of contextual interactions. We then reproduce quantitatively the time course of neuronal responses in

V1 for texture masking experiments and pop-out of embedded contours, and account for both the immediate suppression by surround stimuli and the delayed facilitatory component associated with contour saliency, and its dependence on attention. Last, we show that the network behaves very stably in the presence of noise.

5.6 ACKNOWLEDGEMENTS

This work was supported by The Neurosciences Research Foundation and the National Eye Institute. We thank J. McManus for categorizing excitatory and inhibitory neurons based on the action potential wave form.

6 APPENDIX

7 BIBLIOGRAPHY

Adini, Y., Sagi, D., and Tsodyks, M. (1997). Excitatory-inhibitory network in the visual cortex: Psychophysical evidence. Proc. Natl. Acad. Sci. U. S. A. *94*, 10426-10431.

Adini, Y., Sagi, D., and Tsodyks, M. (2002). Context-enabled learning in the human visual system. Nature *415*, 790-793.

Ahissar, M., and Hochstein, S. (1993). Attentional control of early perceptual learning. Proc Natl Acad Sci U S A *90*, 5718-5722.

Bair, W., Cavanaugh, J.R., and Movshon, J.A. (2003). Time course and time-distance relationships for surround suppression in macaque V1 neurons. J. Neurosci. *23*, 7690-7701.

Ball, K., and Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. Science *218*, 697-698.

Bauer, R., and Heinze, S. (2002). Contour integration in striate cortex - Classic cell responses or cooperative selection? Exp. Brain Res. *147*, 145-152.

Bennett, R.G., and Westheimer, G. (1991). The effect of training on visual alignment discrimination and grating resolution. Percept Psychophys *49*, 541-546.

Bishop, P.O., Coombs, J.S., and Henry, G.H. (1973). Receptive fields of simple cells in the cat striate cortex. J Physiol *231*, 31-60.

Blakemore, C., and Tobin, E.A. (1972). Lateral inhibition between orientation detectors in the cat's visual cortex. Exp Brain Res *15*, 439-440.

Bonneh, Y., and Sagi, D. (1998). Effects of spatial configuration on contrast detection. Vision Res *38*, 3541-3553.

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

Braun, J. (1998). Vision and attention: the role of training. Nature *393*, 424-425.

Bressloff, P.C., and Cowan, J.D. (2002). An amplitude equation approach to contextual effects in visual cortex. Neural Comput *14*, 493-525.

Bringuier, V., Chavane, F., Glaeser, L., and Fregnac, Y. (1999). Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. Science *283*, 695-699.

Carandini, M., and Heeger, D.J. (1994). Summation and division by neurons in primate visual cortex. Science *264*, 1333-1336.

Crist, R.E., Kapadia, M.K., Westheimer, G., and Gilbert, C.D. (1997). Perceptual learning of spatial localization: Specificity for orientation, position, and context. J. Neurophysiol. *78*, 2889-2894.

Crist, R.E., Li, W., and Gilbert, C.D. (2001). Learning to see: experience and attention in primary visual cortex. Nat. Neurosci. *4*, 519-525.

DeAngelis, G.C., Freeman, R.D., and Ohzawa, I. (1994). Length and width tuning of neurons in the cat's primary visual cortex. J Neurophysiol *71*, 347-374.

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. Annu Rev Neurosci *18*, 193-222.

Dragoi, V., and Sur, M. (2000). Dynamic properties of recurrent inhibition in primary visual cortex: contrast and orientation dependence of contextual effects. J Neurophysiol *83*, 1019-1030.

Dresp, B. (1993). Bright lines and edges facilitate the detection of small light targets. Spat Vis *7*, 213-225.

Effron, B., and Tibshirani, R.J. (1993). An Introduction to the Bootstrap (New York: Chapmann and Hall).

Ermentrout, G.B., and Cowan, J.D. (1979). A mathematical theory of visual hallucination patterns. Biol Cybern *34*, 137-150.

Ernst, U.A., Mandon, S., Pawelzik, K.R., and Kreiter, A.K. (2004). How ideal do macaque monkeys integrate contours? Neurocomputing *58-60*, 971-977.

Fahle, M. (1997). Specificity of learning curvature, orientation, and vernier discriminations. Vision Res *37*, 1885-1895.

Fahle, M., and Edelman, S. (1993). Long-term learning in vernier acuity: effects of stimulus orientation, range and of feedback. Vision Res *33*, 397-412.

Fahle, M., and Morgan, M. (1996). No transfer of perceptual learning between similar stimuli in the same retinal position. Curr Biol *6*, 292-297.

Fendick, M., and Westheimer, G. (1983). Effects of practice and the separation of test targets on foveal and peripheral stereoacuity. Vision Res *23*, 145-150.

Field, D.J., Hayes, A., and Hess, R.F. (1993). Contour Integration by the Human Visual-System - Evidence for a Local Association Field. Vision Res. *33*, 173-193.

Finney, D.J. (1952). Probit Analysis - a Statistical Treatment of the Sigmoid Response Curve (Cambridge: Cambridge University Press).

Fiorentini, A., and Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. Nature *287*, 43-44.

Freeman, E., Sagi, D., and Driver, J. (2001). Lateral interactions between targets and flankers in low-level vision depend on attention to the flankers. Nat Neurosci *4*, 1032-1036.

Freeman, W.T., and Adelson, E.H. (1991). The Design and Use of Steerable Filters. Ieee Transactions on Pattern Analysis and Machine Intelligence *13*, 891-906.

Geisler, W.S., Perry, J.S., Super, B.J., and Gallogly, D.P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. Vision Res *41*, 711-724.

Ghose, G.M., Yang, T.M., and Maunsell, J.H.R. (2002). Physiological correlates of perceptual learning in monkey V1 and V2. J. Neurophysiol. *87*, 1867-1888.

Gibson, E.J. (1963). Perceptual learning. Annu Rev Psychol *14*, 29-56.

Gilbert, C.D. (1998). Adult cortical dynamics. Physiol. Rev. *78*, 467-485.

Gilbert, C.D., Sigman, M., and Crist, R.E. (2001). The neural basis of perceptual learning. Neuron *31*, 681-697.

Gilbert, C.D., and Wiesel, T.N. (1979). Morphology and Intra-Cortical Projections of Functionally Characterized Neurons in the Cat Visual-Cortex. Nature *280*, 120-125.

Gilbert, C.D., and Wiesel, T.N. (1983). Clustered Intrinsic Connections in Cat Visual-Cortex. J. Neurosci. *3*, 1116-1133.

Gilbert, C.D., and Wiesel, T.N. (1989). Columnar Specificity of Intrinsic Horizontal and Corticocortical Connections in Cat Visual-Cortex. J. Neurosci. *9*, 2432-2442.

Gilbert, C.D., and Wiesel, T.N. (1990). The Influence of Contextual Stimuli on the Orientation Selectivity of Cells in Primary Visual-Cortex of the Cat. Vision Res. *30*, 1689-1701.

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

Grossberg, S., and Mingolla, E. (1985a). Neural Dynamics of Form Perception - Boundary Completion, Illusory Figures, and Neon Color Spreading. Psychol Rev *92*, 173-211.

Grossberg, S., and Mingolla, E. (1985b). Neural Dynamics of Perceptual Grouping - Textures, Boundaries, and Emergent Segmentations. Perception & Psychophysics *38*, 141-171.

Haenny, P.E., and Schiller, P.H. (1988). State dependent activity in monkey visual cortex. I. Single cell activity in V1 and V4 on visual tasks. Exp Brain Res *69*, 225-244.

Heeger, D.J. (1992). Normalization of cell responses in cat striate cortex. Vis Neurosci *9*, 181-197.

Herzog, M.H., and Fahle, M. (1997). The role of feedback in learning a Vernier discrimination task. Vision Res. *37*, 2133-2141.

Herzog, M.H., and Fahle, M. (1998). Modeling perceptual learning: difficulties and how they can be overcome. Biol. Cybern. *78*, 107-117.

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

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

Hubel, D.H., and Wiesel, T.N. (1965). Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas (18 and 19) of the Cat. J Neurophysiol *28*, 229-289.

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

Hupe, J.M., James, A.C., Girard, P., and Bullier, J. (2001a). Response modulations by static texture surround in area V1 of the macaque monkey do not depend on feedback connections from V2. J Neurophysiol *85*, 146-163.

Hupe, J.M., James, A.C., Girard, P., Lomber, S.G., Payne, B.R., and Bullier, J. (2001b). Feedback connections act on the early part of the responses in monkey visual cortex. J Neurophysiol *85*, 134-145.

Ipata, A.E., Gee, A.L., Gottlieb, J., Bisley, J.W., and Goldberg, M.E. (2006). LIP responses to a popout stimulus are reduced if it is overtly ignored. Nat Neurosci *9*, 1071-1076.

Ito, M., and Gilbert, C.D. (1999). Attention modulates contextual influences in the primary visual cortex of alert monkeys. Neuron *22*, 593-604.

Ito, M., Westheimer, G., and Gilbert, C.D. (1998). Attention and perceptual learning modulate contextual influences on visual perception. Neuron *20*, 1191-1197.

Kapadia, M.K., Ito, M., Gilbert, C.D., and 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., and Gilbert, C.D. (1999). Dynamics of spatial summation in primary visual cortex of alert monkeys. Proc. Natl. Acad. Sci. U. S. A. *96*, 12073-12078.

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

Karni, A., and Sagi, D. (1991). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. Proc Natl Acad Sci U S A *88*, 4966-4970.

Kastner, S., and Ungerleider, L.G. (2000). Mechanisms of visual attention in the human cortex. Annu Rev Neurosci *23*, 315-341.

Kinoshita, M., and Komatsu, H. (2001). Neural representation of the luminance and brightness of a uniform surface in the macaque primary visual cortex. J Neurophysiol *86*, 2559-2570.

Knierim, J.J., and van Essen, D.C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. J Neurophysiol *67*, 961-980.

Kourtzi, Z., Betts, L.R., Sarkheil, P., and Welchman, A.E. (2005). Distributed neural plasticity for shape learning in the human visual cortex. Plos Biology *3*, 1317-1327.

Kourtzi, Z., Tolias, A.S., Altmann, C.F., Augath, M., and Logothetis, N.K. (2003). Integration of local features into global shapes: Monkey and human fMRI studies. Neuron *37*, 333-346.

Kovacs, I., Kozma, P., Feher, A., and Benedek, G. (1999). Late maturation of visual spatial integration in humans. Proc. Natl. Acad. Sci. U. S. A. *96*, 12204-12209.

Lamme, V.A.F. (1995). The Neurophysiology of Figure Ground Segregation in Primary Visual-Cortex. J. Neurosci. *15*, 1605-1615.

Lamme, V.A.F., Zipser, K., and Spekreijse, H. (1998). Figure-ground activity in primary visual cortex is suppressed by anesthesia. Proc. Natl. Acad. Sci. U. S. A. *95*, 3263-3268.

Lee, T.S., Yang, C.F., Romero, R.D., and Mumford, D. (2002). Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. Nat. Neurosci. *5*, 589-597.

Li, C.Y., and Li, W. (1994). Extensive integration field beyond the classical receptive field of cat's striate cortical neurons--classification and tuning properties. Vision Res *34*, 2337-2355.

Li, W., and Gilbert, C.D. (2002). Global contour saliency and local colinear interactions. J. Neurophysiol. *88*, 2846-2856.

Li, W., Piech, V., and Gilbert, C.D. (2004). Perceptual learning and top-down influences in primary visual cortex. Nat Neurosci *7*, 651-657.

Li, W., Piech, V., and Gilbert, C.D. (2006). Contour saliency in primary visual cortex. Neuron *50*, 951-962.

Li, W., Piech, V., and Gilbert, C.D. (2008). Learning to link visual contours. Neuron *57*, 442-451.

Li, W., Thier, P., and Wehrhahn, C. (2000). Contextual influence on orientation discrimination of humans and responses of neurons in V1 of alert monkeys. J. Neurophysiol. *83*, 941-954.

Li, W., Thier, P., and Wehrhahn, C. (2001). Neuronal responses from beyond the classic receptive field in VI of alert monkeys. Exp. Brain Res. *139*, 359-371.

Li, Z. (1998a). Pop-out, feature conjunctions, and asymmetries in visual search in a model of intracortical interactions in V1. Perception *27*, 1488-1489.

Li, Z. (1999a). A model of various eases and asymmetries in visual search by V1 mechanisms. Invest. Ophthalmol. Vis. Sci. *40*, S345-S345.

Li, Z.P. (1998b). A neural model of contour integration in the primary visual cortex. Neural Comput *10*, 903-940.

Li, Z.P. (1999b). Visual segmentation by contextual influences via intra-cortical interactions in the primary visual cortex. Network-Comp Neural *10*, 187-212.

Li, Z.P. (2000). Pre-attentive segmentation in the primary visual cortex. Spatial Vision *13*, 25-50.

Li, Z.P. (2001). Computational design and nonlinear dynamics of a recurrent network model of the primary visual cortex. Neural Comput *13*, 1749-1780.

Maffei, L., and Fiorentini, A. (1976). The unresponsive regions of visual cortical receptive fields. Vision Res *16*, 1131-1139.

Martinez, L.M., and Alonso, J.M. (2001). Construction of complex receptive fields in cat primary visual cortex. Neuron *32*, 515-525.

Martinez, L.M., and Alonso, J.M. (2003). Complex receptive fields in primary visual cortex. Neuroscientist *9*, 317-331.

Maunsell, J.H., and Cook, E.P. (2002). The role of attention in visual processing. Philos Trans R Soc Lond B Biol Sci *357*, 1063-1072.

McAdams, C.J., and Maunsell, J.H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. J Neurosci *19*, 431-441.

McKee, S.P., and Westheimer, G. (1978). Improvement in vernier acuity with practice. Percept Psychophys *24*, 258-262.

Miller, K.D. (2008). Feedforward and recurrent processing in cat primary visual cortex. pp. Talk at the Physics and Biology Departement 1/22/2008, Rockefeller University, New York.

Motter, B.C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. J Neurophysiol *70*, 909-919.

Nauhaus, I., Busse, L., Carandini, M., and Ringach, D.L. (2009). Stimulus contrast modulates functional connectivity in visual cortex. Nat Neurosci *12*, 70-76.

Nelson, J.I., and Frost, B.J. (1978). Orientation-selective inhibition from beyond the classic visual receptive field. Brain Res *139*, 359-365.

Nothdurft, H.C., Gallant, J.L., and Van Essen, D.C. (1999). Response modulation by texture surround in primate area V1: Correlates of "popout" under anesthesia. Visual Neurosci *16*, 15-34.

Nowak, L.G., Azouz, R., Sanchez-Vives, M.V., Gray, C.M., and McCormick, D.A. (2003). Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses. J Neurophysiol *89*, 1541-1566.

Panzeri, S., and Treves, A. (1996). Analytical estimates of limited sampling biases in different information measures. Network-Comp Neural *7*, 87-107.

Parker, A.J., and Newsome, W.T. (1998). Sense and the single neuron: Probing the physiology of perception. Annu. Rev. Neurosci. *21*, 227-277.

Piech, V., Li, W., Gilbert, C.D., and Reeke, G.N. (2009). A neural model of V1 for gain control and gating long-range horizontal connections by top-down influences. Manuscript in preparation.

Poggio, G.E. (1995). Mechanisms of stereopsis in monkey visual cortex. Cereb Cortex *5*, 193-204.

Poggio, T., Fahle, M., and Edelman, S. (1992). Fast Perceptual-Learning in Visual Hyperacuity. Science *256*, 1018-1021.

Polat, U., and Bonneh, Y. (2000). Collinear interactions and contour integration. Spat Vis *13*, 393-401.

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

Polat, U., and Sagi, D. (1993). Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. Vision Res *33*, 993-999.

Polat, U., and Sagi, D. (1994a). The architecture of perceptual spatial interactions. Vision Res *34*, 73-78.

Polat, U., and Sagi, D. (1994b). Spatial interactions in human vision: from near to far via experience-dependent cascades of connections. Proc Natl Acad Sci U S A *91*, 1206-1209.

Posner, M.I., and Gilbert, C.D. (1999). Attention and primary visual cortex. Proc. Natl. Acad. Sci. U. S. A. *96*, 2585-2587.

Raizada, R.D.S., and Grossberg, S. (2003). Towards a theory of the laminar architecture of cerebral cortex: Computational clues from the visual system. Cereb. Cortex *13*, 100-113.

Ramachandran, V.S., and Braddick, O. (1973). Orientation-specific learning in stereopsis. Perception *2*, 371-376.

Rao, R.P., and 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.

Recanzone, G.H., Merzenich, M.M., and Jenkins, W.M. (1992a). Frequency discrimination training engaging a restricted skin surface results in an emergence of a cutaneous response zone in cortical area 3a. J Neurophysiol *67*, 1057-1070.

Recanzone, G.H., Merzenich, M.M., Jenkins, W.M., Grajski, K.A., and Dinse, H.R. (1992b). Topographic reorganization of the hand representation in cortical area 3b owl monkeys trained in a frequency-discrimination task. J Neurophysiol *67*, 1031-1056.

Recanzone, G.H., Schreiner, C.E., and Merzenich, M.M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. J Neurosci *13*, 87-103.

Rockland, K.S., and Lund, J.S. (1982). Widespread periodic intrinsic connections in the tree shrew visual cortex. Science *215*, 1532-1534.

Rockland, K.S., Lund, J.S., and Humphrey, A.L. (1982). Anatomical binding of intrinsic connections in striate cortex of tree shrews (Tupaia glis). J Comp Neurol *209*, 41-58.

Roelfsema, P.R. (2006). Cortical algorithms for perceptual grouping. Annu Rev Neurosci *29*, 203-227.

Roelfsema, P.R., Lamme, V.A., and Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. Nature *395*, 376-381.

Roelfsema, P.R., Lamme, V.A., and Spekreijse, H. (2004). Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. Nat Neurosci *7*, 982-991.

Roelfsema, P.R., and van Ooyen, A. (2005). Attention-gated reinforcement learning of internal representations for classification. Neural Comput *17*, 2176-2214.

Rossi, A.F., Desimone, R., and Ungerleider, L.G. (2001). Contextual modulation in primary visual cortex of macaques. J Neurosci *21*, 1698-1709.

Rossi, A.F., Rittenhouse, C.D., and Paradiso, M.A. (1996). The representation of brightness in primary visual cortex. Science *273*, 1104-1107.

Saarinen, J., and Levi, D.M. (1995). Perceptual learning in vernier acuity: what is learned? Vision Res *35*, 519-527.

Saffell, T., and Matthews, N. (2003). Task-specific perceptual learning on speed and direction discrimination. Vision Res *43*, 1365-1374.

Schmidt, K.E., Goebel, R., Lowel, S., and Singer, W. (1997). The perceptual grouping criterion of colinearity is reflected by anisotropies of connections in the primary visual cortex. Eur J Neurosci *9*, 1083-1089.

Schoups, A., Vogels, R., Qian, N., and Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. Nature *412*, 549-553.

Schwartz, O., and Simoncelli, E.P. (1999). Accounting for surround suppression in V1 neurons using a statistically optimized normalization model. Invest. Ophthalmol. Vis. Sci. *40*, S641-S641.

Series, P., Georges, S., Lorenceau, J., and Fregnac, Y. (2001). A network view of the structure of center/surround modulations of V1 receptive field properties in visual and cortical spaces. Neurocomputing *38*, 881-888.

Series, P., Georges, S., Lorenceau, J., and Fregnac, Y. (2002). Orientation dependent modulation of apparent speed: a model based on the dynamics of feed-forward and horizontal connectivity in V1 cortex. Vision Res. *42*, 2781-2797.

Shani, R., and Sagi, D. (2005). Eccentricity effects on lateral interactions. Vision Res *45*, 2009-2024.

Shannon, C.E. (1948). A mathematical theory of communication. Bell Syst. Tech. J. *27*, 379-423.

Shepherd, G.M. (1990). Synaptic Organization of the Brain, 3rd edn (Oxford University Press).

Shiu, L.P., and Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. Percept Psychophys *52*, 582-588.

Sigman, M., Cecchi, G.A., Gilbert, C.D., and Magnasco, M.O. (2001). On a common circle: Natural scenes and Gestalt rules. Proc. Natl. Acad. Sci. U. S. A. *98*, 1935-1940.

Sigman, M., and Gilbert, C.D. (2000). Learning to find a shape. Nat. Neurosci. *3*, 264-269.

Sigman, M., Pan, H., Yang, Y.H., Stern, E., Silbersweig, D., and Gilbert, C.D. (2005). Top-down reorganization of activity in the visual pathway after learning a shape identification task. Neuron *46*, 823-835.

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

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? Neuron *24*, 49-65.

Snodderly, D.M., and Gur, M. (1995). Organization of striate cortex of alert, trained monkeys (Macaca fascicularis): ongoing activity, stimulus selectivity, and widths of receptive field activating regions. J Neurophysiol *74*, 2100-2125.

Somers, D.C., Todorov, E.V., Siapas, A.G., Toth, L.J., Kim, D.S., and Sur, M. (1998). A local circuit approach to understanding integration of long-range inputs in primary visual cortex. Cereb. Cortex *8*, 204-217.

Stemmler, M., Usher, M., and Niebur, E. (1995). Lateral interactions in primary visual cortex: a model bridging physiology and psychophysics. Science *269*, 1877-1880.

Stetter, M., Bartsch, H., and Obermayer, K. (2000). A mean-field model for orientation tuning, contrast saturation, and contextual effects in the primary visual cortex. Biol Cybern *82*, 291-304.

Stettler, D.D., Das, A., Bennett, J., and Gilbert, C.D. (2002). Lateral connectivity and contextual interactions in macaque primary visual cortex. Neuron *36*, 739-750.

Super, H., Spekreijse, H., and Lamme, V.A.F. (2001). Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). Nat. Neurosci. *4*, 304-310.

Tolhurst, D.J., Movshon, J.A., and Dean, A.F. (1983). The Statistical Reliability of Signals in Single Neurons in Cat and Monkey Visual-Cortex. Vision Res. *23*, 775-785.

Treisman, A., Vieira, A., and Hayes, A. (1992). Automaticity and preattentive processing. Am J Psychol *105*, 341-362.

Treue, S. (2001). Neural correlates of attention in primate visual cortex. Trends Neurosci *24*, 295-300.

Treue, S., and Martinez Trujillo, J.C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. Nature *399*, 575-579.

Ts'o, D.Y., Gilbert, C.D., and Wiesel, T.N. (1986). Relationships between Horizontal Interactions and Functional Architecture in Cat Striate Cortex as Revealed by Cross-Correlation Analysis. J. Neurosci. *6*, 1160-1170.

Tsodyks, M.V., Skaggs, W.E., Sejnowski, T.J., and McNaughton, B.L. (1997). Paradoxical effects of external modulation of inhibitory interneurons. J Neurosci *17*, 4382-4388.

Ullman, S. (1984). Visual routines. Cognition *18*, 97-159.

Ullman, S. (1992). Low-level aspects of segmentation and recognition. Philos Trans R Soc Lond B Biol Sci *337*, 371-378; discussion 379.

VanRullen, R., Delorme, A., and Thorpe, S.J. (2001). Feed-forward contour integration in primary visual cortex based on asynchronous spike propagation. Neurocomputing *38*, 1003-1009.

Vogels, R., and Orban, G.A. (1985). The effect of practice on the oblique effect in line orientation judgments. Vision Res *25*, 1679-1687.

von der Heydt, R., and Peterhans, E. (1989). Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. J Neurosci *9*, 1731-1748.

von der Heydt, R., Peterhans, E., and Baumgartner, G. (1984). Illusory Contours and Cortical Neuron Responses. Science *224*, 1260-1262.

Walker, G.A., Ohzawa, I., and Freeman, R.D. (2000). Suppression outside the classical cortical receptive field. Vis Neurosci *17*, 369-379.

Wang, Q., Cavanagh, P., and Green, M. (1994). Familiarity and pop-out in visual search. Percept Psychophys *56*, 495-500.

Watanabe, T., Nanez, J.E., and Sasaki, Y. (2001). Perceptual learning without perception. Nature *413*, 844-848.

Watanabe, T., Nanez, J.E., Sr., Koyama, S., Mukai, I., Liederman, J., and Sasaki, Y. (2002). Greater plasticity in lower-level than higher-level visual motion processing in a passive perceptual learning task. Nat Neurosci *5*, 1003-1009.

Weinberger, N.M., and Bakin, J.S. (1998). Learning-induced physiological memory in adult primary auditory cortex: receptive fields plasticity, model, and mechanisms. Audiol Neurootol *3*, 145-167.

Wertheimer, M. (1923). Untersuchung zur Lehre von der Gestalt. Psychologische Forschung *4*, 301-350.

Yen, S.C., and Finkel, L.H. (1998). Extraction of perceptually salient contours by striate cortical networks. Vision Res. *38*, 719-741.

Zhaoping, L., Dayan, P., and Herzog, M.H. (2002). Positional variation in perceptual learning: quadratic ideal observation, recurrent pre-processing, and transfer. Perception *31*, 94-94.

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