

Contextual Modulation in Primary Visual Cortex

Karl Zipser,¹ Victor A. F. Lamme,² and Peter H. Schiller¹

¹The Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and ²Graduate School of Neurosciences, Department of Medical Physics, AMC, University of Amsterdam, Amsterdam, The Netherlands, and The Netherlands Ophthalmic Research Institute, 1100 AC Amsterdam, The Netherlands

We studied extra-receptive field contextual modulation in area V1 of awake, behaving macaque monkeys. Contextual modulation was studied using texture displays in which texture covering the receptive field (RF) was the same in all trials, but the perceptual context of this texture could vary depending on the configuration of extra-RF texture elements. We found robust contextual modulation when disparity, color, luminance, and orientation cues variously defined a textured figure centered on the RF of V1 neurons. We found contextual modulation to have a spatial extent of ~8 to 10° diameter parafoveally. Contextual

modulation correlated with perceptual experience of both binocularly rivalrous texture displays and of displays with a simple example of surface occlusion. We found contextual modulation in V1 to have a characteristic latency of 80–100 msec after stimulus onset, potentially allowing feedback from extrastriate areas to underlie to this effect.

Key words: figure-ground segregation; surface perception; primary visual cortex; awake macaque monkey; single-unit activity; texture; visual perception; context modulation; nonclassical receptive field

Neurophysiological research in primary visual cortex (area V1) has focused primarily on elucidating the characteristics of the receptive fields (RFs) of the neurons in this brain area. The RF of a visual neuron is the restricted region of the visual field in which an appropriate stimulus, such as an oriented bar or a patch of texture, may drive the cell to evoke action-potential responses. Yet the activity of V1 neurons evoked in this manner may be *modulated* by stimuli placed entirely outside the RF (Blakemore and Tobin, 1972; Maffei and Fiorentini, 1976; Nelson and Frost, 1978; Gilbert and Wiesel, 1990; Knierim and van Essen, 1992; Sillito et al., 1995). We call this general phenomenon *extra-RF contextual modulation*. Presumably extra-RF contextual modulation allows neurons to signal some form of comparison between the patterns inside and outside the RF (Allman et al., 1985). But the essential characteristics of extra-RF modulation, and the type of comparison that it may support, remain largely a mystery.

Although not well characterized, the modulatory influence of stimuli placed outside the RF of the V1 neuron constitutes a powerful force in primary visual cortex. A dramatic demonstration of this comes from Lamme (1995), who recorded activity of V1 neurons in awake, behaving monkeys during viewing of textured displays. Lamme used textured stimuli configured such that the RF of a V1 neuron under study received an identical pattern of stimulation from trial to trial. Despite this identical RF stimulation, V1 cells almost always responded more vigorously in trials in

which the orientation, or motion, of the texture pattern on the RF belonged to a circumscribed “figure” (such as the square in Fig. 1a), as compared with trials in which texture was of a homogeneous type across the entire display (Fig. 1b).

Lamme’s experiments suggest that extra-RF contextual modulation constitutes as robust a feature of V1 neural function as the long-studied RF properties of cells in this area, such as orientation tuning (Hubel and Wiesel, 1968). Yet before we may integrate contextual modulation into a comprehensive model of the function of area V1, we must have a better understanding of the basic characteristics of this phenomenon and of the goals that it is designed to accomplish. A key question is whether contextual modulation in area V1 reflects a sophisticated neural correlate of perception or, rather, whether it merely reflects low-level image processing only distantly related to visual awareness. If extra-RF contextual modulation in area V1 closely relates to perception, then this modulation should correlate with perceptual experience under a wide range of stimulus conditions. On the other hand, to the extent that contextual modulation is a low-level phenomenon, it should be relatively easy to dissociate from perceptual experience. We report here results of neurophysiological experiments that we conducted on area V1 of awake, behaving monkeys to attempt to distinguish between these possible functions of contextual modulation.

MATERIALS AND METHODS

Experiments were performed on four male *Macaca mulatta*, each weighing 8–10 kg. Before surgery, monkeys were trained to jump into their primate chairs and were habituated to the laboratory environment. Subsequently, each animal underwent surgical procedures for implantation of a stainless steel cranial post for fixing the position of the head. In the same operation, we implanted the given animal with a scleral coil for monitoring eye position. All surgical procedures were performed using sterile techniques, with monkeys under deep pentobarbital anesthesia; all experimental procedures were performed in accordance with National Institutes of Health guidelines.

After recovery from surgery, monkeys were water-deprived and brought to the laboratory for training. We used a PDP-11/37 computer to

Received Oct. 16, 1995; revised Aug. 28, 1996; accepted Sept. 3, 1996.

This research was supported by a grant from the National Eye Institute to P.H.S., a grant from The Netherlands Organization for Scientific Research to V.A.F.L., and an Office of Naval Research graduate fellowship and a McDonnell-Pew Center for Cognitive Neuroscience at MIT postdoctoral fellowship to K.Z. We thank D. Zipser and numerous other colleagues for valuable discussion. We thank C. J. Doane-Palafox and T. S. Lee for help with some of these experiments, W. M. Slocum for assistance with computer programming, and C. Conner and J. Mendola for reading this manuscript.

Correspondence should be addressed to Dr. Karl Zipser, The Netherlands Ophthalmic Research Institute, P.O. Box 12141, 1100 AC Amsterdam, The Netherlands. Copyright © 1996 Society for Neuroscience 0270-6474/96/167376-14\$05.00/0

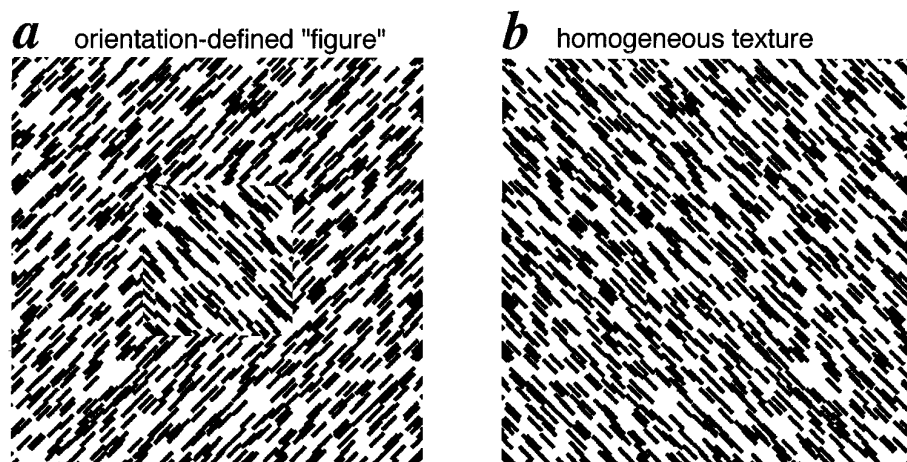


Figure 1. Example texture displays. *a*, Illustration of an orientation-defined figure. *b*, Illustration of a homogeneous texture display. Texture in the center of the orientation-defined figure is identical to texture at the corresponding position in the homogeneous texture display. Typically, the luminance of the gray background was 24 cd/m² and that of the black bars was 6 cd/m², although we saw no evidence that a particular contrast was critical. Furthermore, our results do not seem to depend on the exact texture distribution. Nonetheless, we generally used texture bars 0.5° in length with the pattern illustrated in this figure. The gray between texture bars was the same as the gray that covered the screen in the intertrial period.

regulate and monitor the monkey's behavioral tasks, to collect behavioral and neurophysiological data, and to signal an IBM PC for control of visual stimulation. With its head restrained in the primate chair facing a computer graphics monitor, each monkey was trained to fixate a small luminous spot on the screen and then to make a saccadic eye movement to a luminous target stimulus that appeared in a random position when the fixation spot was extinguished. Analog *x* and *y* eye position signals, measured using the scleral coil (Robinson, 1963), were collected at 200 Hz and digitized with a precision of 0.01° of visual angle. For maintaining fixation and then making the correct saccades, the monkey was rewarded automatically with a drop of apple juice. During training and recording, animals drank a total of 300–500 ml of juice (during 1500 or more trials) per session. Additional rewards of peanuts and fresh fruit were provided once the animals returned to their home cages at the end of the day.

Stimuli were presented on an NEC multisync XL color video display unit, driven by a Number Nine Corporation graphics board with 640 × 480 pixel resolution and a frame rate of 60 Hz. The screen was 32 cm wide and 24 cm high and was viewed at either a 57 or 63 cm distance. In experiments that did not require stereoscopic stimuli, various texture displays covered the entire screen. In experiments that required stereoscopic stimuli, stereo images were displayed side by side on the screen. In this case, all stimuli in each image appeared within a 9 × 9° thin white frame, which remained visible at all times to facilitate fusion of the stimuli. In these experiments, monkeys viewed the screen through a prism haptoscope that allowed the horizontally displaced stereo images to be fused at a comfortable vergence angle.

For human observers (with a separate prism haptoscope for human use), our disparity-defined texture stimuli produced a rich percept of surfaces in depth. Monkey binocular vision is very similar to that of human beings (Bough, 1970; Cowey et al., 1975; Sarmiento, 1975; Miezin et al., 1981; Harwerth et al., 1995; Leopold and Logothetis, 1996), and we presume that with appropriate presentation, the display should have the same richness for monkeys as it does for human observers. A characteristic of binocular image fusion is that sensitivity to binocular disparity is best at the fusion depth (i.e., on the horopter) and declines approximately symmetrically for near and far disparities (Tyler, 1983). In psychophysical tests of our monkeys' ability to detect targets defined through binocular disparity, we found exactly this pattern. Monkeys could effortlessly detect a 0.09° horizontal binocular disparity offset of a textured target from a similarly textured background near the horopter but had decreasing sensitivity to this same offset if target and background appeared at increasingly near or far disparities. This pattern of behavior would not be expected if monkeys failed to fuse the stereo images.

Monkeys initially trained to detect salient orientation-defined texture targets mastered the easier levels of the horizontal disparity texture-target-detection task with no special training. In contrast, when targets were made visible by vertical disparity, monkeys did not transfer easily to this task. Monkeys also could not detect the target defined by binocular disparity when presented with a monocular image. From the combination of these results, it is reasonable to deduce that the monkey's perception of disparity-defined textures in our experiments is similar to that of human observers.

Neurophysiological recording techniques. Neural recordings in awake monkeys were made through a surgically implanted cylindrical stainless

steel electrode chamber (16 mm diameter) overlaying the operculum of area 17. Recording began at least 3 d after surgical implantation of the recording well. Microelectrodes were inserted via the oil-filled, hydraulically closed electrode chamber, through the intact dura, and into occipital cortex. Activity from single cells or clusters of cells was recorded extracellularly with glass-coated platinum-iridium microelectrodes of 0.5–2.0 MΩ impedance (measured at 1000 Hz). The RFs of V1 neurons thus studied were in the lower contralateral visual field with eccentricities between 2 and 6°. To help ensure that our microelectrodes remained in area V1, the RF positions of neurons recorded in each experiment were represented on a graph (maintained for each monkey) that allowed us to observe the orderly retinotopic mapping of the visual field onto striate cortex. Neural recording was principally conducted in superficial cortical layers 2 and 3, judging by microelectrode depth and the characteristic features of deeper input layer 4 (e.g., high spontaneous activity, brisk on and off responses, high degree of monocularly).

Within 3 weeks of insertion of the electrode chamber, the dura mater hardened and became covered with an epithelium up to 6 mm thick. Such tissue barriers caused difficulty with recording, because microelectrodes tended to break before entering the cortex and, more importantly, because moving the microelectrode through these tissues could cause displacement of the brain. We found the latter effect to be highly deleterious to the expression of extra-RF contextual modulation, perhaps because the physical displacement generally depressed neural activity or perhaps because it specifically compressed feedback fibers in layer 1. We took three measures to counter this problem. First, the supra-dural epithelium was thinned through gentle aspiration (performed with the monkey under ketamine anesthesia). Second, we interspersed week-long breaks from recording between each week of experimentation, because we found that this kept the dura from hardening to such an extent that recording became difficult. Third, to avoid brain displacement, we moved the microelectrode through the supra-dural epithelium and the dura with the following pattern: a quick advance of about 10 μm, followed by a brief pause, followed by another advance, etc. In this way, we avoided building mechanical pressure on the brain. The average rate at which we lowered the microelectrodes was ~1 cm per hour.

Plotting of RFs. To plot the extent of the RF of a V1 neuron under study, we moved computer graphics-generated bars of variable size and orientation over the neighborhood of the RF as the monkey fixated. We initially drew RF boundaries by hand with felt-tip markers on an auxiliary stimulus monitor while we simultaneously watched the moving bar stimulus and monitored the evoked neural activity with an audio amplifier. After this, we tested our estimate of RF dimensions by flashing bars and textures inside and outside this area. We confirmed the reliability of our RF plotting techniques by flashing texture stimuli in a region surrounding the measured RF while leaving the RF unstimulated. Whereas neurons responded vigorously to direct RF stimulation, stimulation with surrounding texture evoked at best an extremely weak response (see Results, Fig. 2*d*). Our RF plotting techniques thus were adequate to allow us to isolate extra-RF stimulation from direct RF stimulation.

Texture experiments. We studied each V1 neuron with static, flashed texture displays that contained the same stimulus pattern in the region over the RF from trial to trial. Texture over the RF consisted of black bars on a gray background; the gray between texture bars was the same as

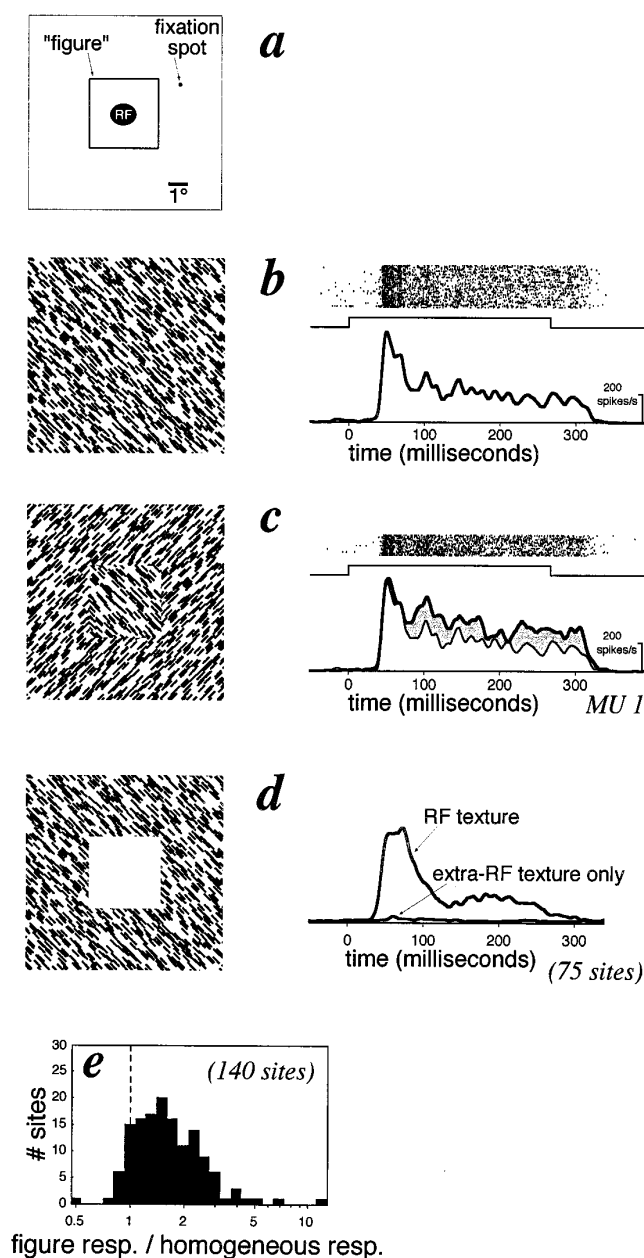


Figure 2. Extra-RF contextual modulation for orientation-defined texture figures. *a*, Configuration of the fixation spot, figure, and RF. The RF is completely enclosed by the figure contour. *b*, Illustration of the response of a multiunit site to stimulation with the homogeneously textured display flashed on a gray background for 267 msec. *c*, Illustration of the response of the same site when a 3.6° wide orientation-defined figure was flashed on in randomly interleaved trials. The initial response is nearly identical, but the tonic phase of the response is elevated in this condition compared with *b*. The response profile for the homogeneous texture display is shown in composite for comparison (fine-line waveform), and gray shading highlights the positive difference in response. *d*, Comparison of the average responses of all 75 recording sites for which we have quantitative data for stimulation both with RF texture and with extra-RF texture alone. Extra-RF texture alone gave at best an extremely weak response. *e*, Histogram of extra-RF contextual modulation ratios for the orientation display. This ratio is defined by the average response to the figure display divided by average response to the homogeneous texture display. Average response rates were measured in the interval of 100–250 msec after stimulus onset (thereby ignoring the initial transient response). Ratio values > 1.0 indicate larger responses to figure displays. Single- and multiunit sites were qualitatively and quantitatively similar (see text for details).

the gray that covered the screen in the intertrial period. In some trials, the display appeared as a homogeneously textured field (e.g., Fig. 1*b*). In other randomly interleaved trials, the display appeared to have a textured figure (e.g., Fig. 1*a*) centered on and completely covering the RF. Although various visual cues were used in our experiments to segment the texture figures from their backgrounds, texture within the figure was identical to that in the corresponding region of the homogeneous texture display. Details concerning particular texture displays are presented in the accompanying figure legends.

We used two types of homogeneous texture display in our experiments. The first type was a true homogeneously texture display, as illustrated in Figure 1*b*. We also used a pseudo-homogeneous texture display constructed, for example, by pairing a textured figure with a background texture of the same orientation. The line terminations formed by the figure contour in the pseudo-homogeneous display served as a control against the possibility that similar line terminations in other displays could be the source of the extra-RF contextual modulation that we investigated. In practice, differences between these two types of texture display are only visible under careful foveal inspection. With control experiments on 53 multiunit recording sites, we found that V1 neurons generally produced indistinguishable responses to the two types of homogeneous texture display when the RF is placed well within the "figure" contour. The median ratio of response to true- and pseudo-homogeneous texture displays was 1.01. Furthermore, responses to the two display types were significantly different in only 13% of the 53 sites ($p < 0.05$, two-sided t test), and these differences were small. For simplicity, we will ignore the distinction between the true- and the pseudo-homogeneous texture displays in the remainder of this report.

The temporal progression of a behavioral trial for most of our texture experiments was as follows. At the beginning of a trial, a fixation spot appeared on the gray monitor screen, and the monkey foveated this spot. Approximately 200 msec after foveation of the spot occurred, a texture display appeared on the screen for a fixed interval (e.g., 250 msec in some experiments), after which the screen returned to the prestimulus gray. Approximately 200 msec after the texture offset, the fixation spot was extinguished, and a target spot appeared in a random position around the fixation spot. The monkey was rewarded with a drop of apple juice for maintaining stable fixation throughout the trial and then making a saccade to this target. In an alternative experimental paradigm, the monkey was required to saccade to a texture-defined stimulus (either over the RF or in the opposite hemifield) after the extinguishing of the fixation spot. Operationally, stable fixation meant that the monkey's eye position remained within a fixation window (not visible in the stimulus display) that was centered on the fixation spot. The fixation window size varied from $1^\circ \times 1^\circ$ to $0.3^\circ \times 0.3^\circ$; the typical value was $0.5^\circ \times 0.5^\circ$.

Given that the results in this study are based on comparison of neural responses in trials in which the texture display was either homogeneous or contained a salient textured figure, it is of considerable importance to determine whether the presence of the figure in the flashed texture display could subtly influence eye movements that might, in turn, alter neural responses. We addressed this topic quantitatively by selecting recordings in which neural responses showed strong modulation depending on whether the texture display was of the homogeneous type or contained a texture-defined figure in randomly interleaved trials. For each trial, the mean and variance in both x and y eye position was measured during the texture display interval. The distributions of these mean and variance measures were indistinguishable for the homogeneous and nonhomogeneous texture displays; separate χ^2 tests for x and y values fail to reject the null hypothesis that the content of the texture display has no influence on mean or variance of eye position during fixation. From these results [which agree with an analysis by Lamme (1995)], we conclude that our observations of modulation of neural activity described here are not an artifact of eye movements.

Data collection and analysis. Neural spike data were collected using either hardware and software from a Brainwave Systems Corporation data collection setup or a simple two-level spike amplitude discriminator. Data files containing spike, event, and eye position information were saved on an IBM PC (486) in binary form and converted to ASCII for analysis on UNIX and Macintosh computer systems. Data analysis was conducted using a combination of our own C++ analysis routines and commercially available software (i.e., Mathematica and MATLAB).

RESULTS

Here, we present the results of neural recordings in area V1 in six hemispheres of four awake, behaving rhesus monkeys. Our quan-

titative data consist of findings from experiments on 118 isolated V1 neurons and 228 multiunit sites (in which inseparable signals from two or more cells were recorded simultaneously). As we will describe in reference to Figure 2, single- and multiunit sites behaved similarly in our experiments. Thus, we will not generally be concerned with the distinction between single- and multiunit sites except where the cue receptivity of individual neurons is of interest. We recorded principally in superficial layers 2 and 3. The V1 cells that we studied had RFs in the lower, contralateral visual field with eccentricities ranging from 2 to 6° of visual angle.

We use the expression *extra-RF contextual modulation* (or “contextual modulation” for short) to describe how a neuron’s response to direct RF stimulation may be influenced by patterns appearing entirely outside the RF. The technique common to our experiments on V1 contextual modulation consists of measuring the response of a given V1 neuron or multiunit site to a homogeneous texture display (e.g., Fig. 1*b*) and using this as a standard against which to compare the responses of the same cell or multisite to various test displays containing an identical texture pattern over the RF and different patterns outside the RF area. For example, Figure 1*a* shows a textured display containing a square “figure” region that segments from the background through the 90° difference in orientation of texture elements between these two regions. In our experiments, we positioned the figure so that it was centered on and completely covered the RF of V1 neurons under study (e.g., Fig. 2*a*). In the absence of any sort of extra-RF contextual modulation, V1 neurons would respond identically to these displays.

Figure 2, *b* and *c*, compares the response activity of one V1 multiunit site to the homogeneous texture and to the orientation-defined figure displays. As a monkey foveated the fixation spot on a gray computer monitor screen, a given texture display appeared for 267 msec. The V1 multiunit site showed little activity for the uniform gray display but responded to the appearance of the homogeneous texture display with a vigorous burst of action potentials (Fig. 2*b*). After this initial burst, the cells’ responses declined to a lower maintained discharge rate. When we stimulated this site with the orientation-defined texture figure (width 3.6°) in randomly interleaved trials, we recorded different results (Fig. 2*c*). Although the neurons responded to the onset of the figure display with nearly the same burst of activity as to the homogeneous texture display, the response rates diverged ~80 msec subsequent to texture onset. Despite the fact that texture within the RF was identical to that for the homogeneous texture display, the orientation-defined figure display thereafter caused the cells of the multiunit site to maintain a significantly ($p < 0.05$, one-sided *t* test) more vigorous response rate than did the homogeneous texture display (as is indicated by the *gray shading* of response profile in Fig. 2*c*). Extra-RF texture alone did not appreciably activate the V1 neurons (Fig. 2*d*).

The difference in responses of the V1 multiunit site for the homogeneous texture display and the orientation-defined figure display is an example of extra-RF contextual modulation. We quantify this contextual modulation by calculating a ratio, the average response rate for the test display (in this case, the orientation-defined figure) divided by the average response to the homogeneous texture display. Because contextual modulation typically evolves only after the initial transient response, throughout this study we will only consider activity 100–250 msec after stimulus onset in our ratio metric. Applying this ratio measure to a large sample of V1 recordings ($n = 92$ single-unit and 48 multiunit sites) with RFs centered in either a square or disc-

shaped orientation-defined figure of width 2.7–4°, we arrive at the histogram in Figure 2*e*. For each cell or multiunit site, we chose the orientation of RF texture best suited for the cell. These data replicate the observation by Lamme (1995) that V1 neurons with remarkable consistency respond more vigorously when their RFs are within an orientation-defined figure than when over a homogeneously textured background (i.e., most entries in the histogram are above the ratio value 1.0). Single-unit and multiunit sites were qualitatively and quantitatively similar in behavior. The median contextual modulation ratio for the 92 single-unit sites was 1.61, whereas for the 48 multiunit sites it was 1.53. Furthermore, the hypothesis of independence between the distributions of contextual modulation ratios for single- and multiunit sites was rejected by a χ^2 test. Forty-five percent of the single units and 57% of the multiunit sites showed significantly greater response rates to the orientation-defined display as compared with the homogeneous texture display ($p < 0.05$, one-sided *t* test).

The basic pattern of neural response that we have described above was observed whether the experimental subjects were required merely to passively fixate (the normal condition) or to make saccades to texture figures; thus, we replicated Lamme’s result (1995). It is therefore unlikely that the results we report are merely an indirect result of modulation by visual attention, because the effects do not appear to depend on the behavioral task being performed by the monkey subjects.

Do diverse visual cues evoke extra-RF modulation?

Lamme’s original experiments (1995) showed that both orientation- and motion-defined figures may evoke contextual modulation in V1. If extra-RF contextual modulation is closely related to our perception of figure/ground segregation, then this modulation should indeed be evoked by the same broad range of cues that support image segmentation. In this section, we specifically address the question: what is the range of static visual cues that evoke extra-RF contextual modulation in V1 neurons? The different cues that we use to delineate a texture figure from the background texture are illustrated in the left column of Figure 3.

Binocular disparity

We illustrate a rendition of a textured disc segmented from the background through binocular disparity cues in Figure 3*b*. The disc appears to float above a textured background. The disc texture over the RF duplicates that in the corresponding region of the homogeneous texture field. No previous study has investigated the potential for binocular disparity cues to evoke extra-RF contextual modulation.

Color, luminance

In Figure 3, *c* and *d*, we illustrate disc displays in which either color or luminance act as cues for segmenting the disc from background texture. Although previous studies have investigated effects of color on extra-RF contextual modulation in primate extrastriate cortex (Zeki, 1973; Schein and Desimone, 1990), pure color and luminance cues have not been tested previously in this manner in primate area V1.

Orientation

We also included an orientation-defined disc in the set of stimuli (Fig. 3*e*).

Combination of cues

Figure 3*f* illustrates a rendition of the combination disc display, in which orientation, disparity, color, and luminance all serve to offset the disc from the texture background.

Disc alone

Another way to visualize the texture disc is through the complete lack of background texture. In Figure 3g, we illustrate a display of this type, called the “disc-alone” condition. The texture disc in this case is identical to that in other displays. In trials in which the

disc-alone condition appeared, the area around the disc remained a uniform gray.

We show the response activity of one isolated V1 neuron (cell *a*) to these displays in the right column of Figure 3. For each of the disc displays, this cell gave essentially the same response: after a burst of activity at texture onset, the cell exhibited a robust rate of activity for each disc, well above the response level for the homogeneous texture display. The magnitude of the contextual modulation for the cell in Figure 3 was very similar for the various disc-defining cues (a topic to be addressed below).

We studied a total of 64 V1 neurons using the textured displays described in Figures 3, the disc in each case being centered on the RF. We focused exclusively on single-unit responses for this experiment, because the response selectivity of individual neurons for the various cues is of interest, and multiunit data would cloud this issue. For most isolated cells, we used discs 3.6° in diameter ($n = 40$). For the remaining isolated cells, we used smaller discs, although never discs $< 2.7^\circ$ in diameter (which is well above RF size). For each cell, we chose the orientation of RF texture best suited for the cell. Aside from these manipulations, the same texture displays were used for each experiment. Thus, beyond varying orientation, we did not attempt to “optimize” the RF texture for each cell. Indeed, optimizing RF texture does not appear critical for evoking contextual modulation (Lamme, 1995). The criterion for selecting a cell for experimentation was that it gave clear responses to at least one of the texture displays; this was the case with approximately one-third of the neurons that we isolated. In general, we did not attempt to classify cells as simple or complex, although it is likely that most cells in the sample are of the complex type, because these are more responsive to the flashed random texture patterns (De Valois and De Valois, 1988).

For each of the 64 isolated V1 neurons thus tested, we calculated extra-RF modulation ratios for each disc display (i.e., disc

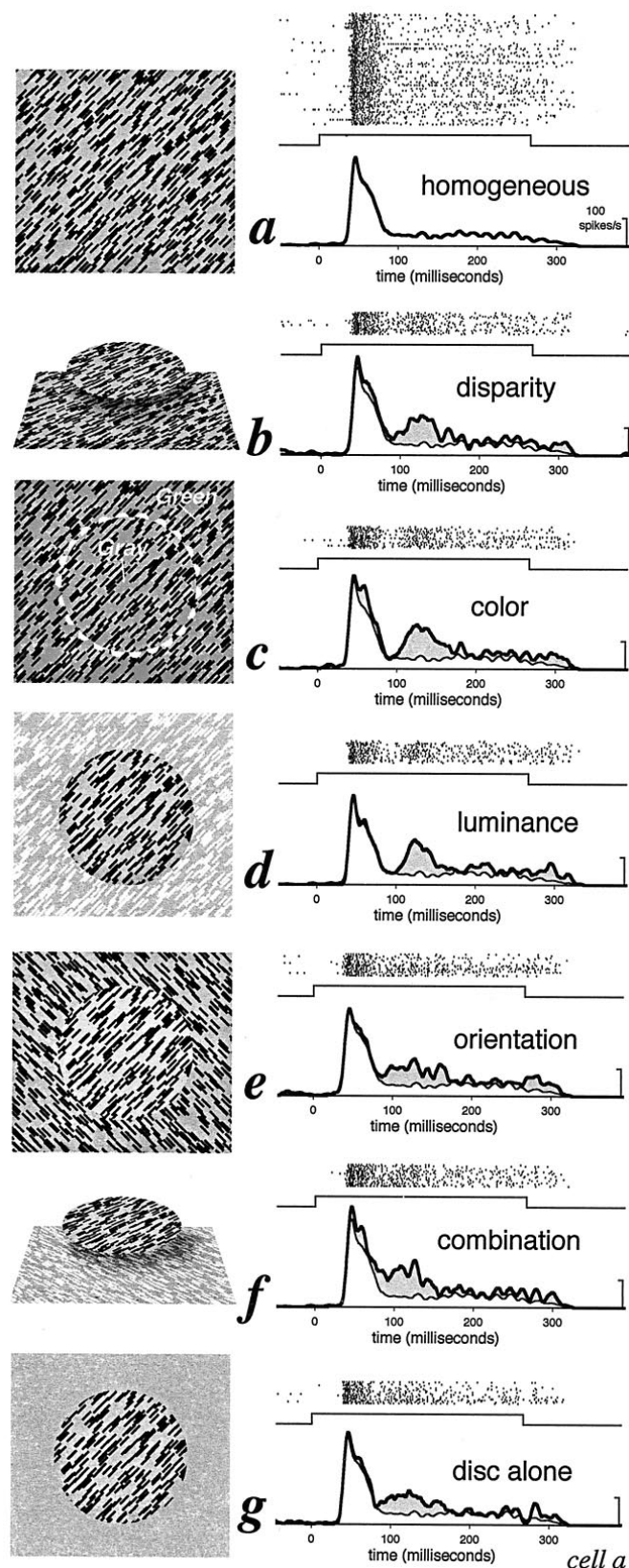


Figure 3. Extra-RF contextual modulation for diverse figure-defining cues. The responses of one isolated parafoveal V1 neuron (cell *a*) are illustrated in this figure; quantitative description of this cell's responses appears in Figure 5. *a*, Illustration of the responses to the homogeneous texture display. In each of the following conditions (*b–f*), the pattern and disparity of RF texture were identical to that in the corresponding region of the homogeneous texture display. *b*, Illustration of the response when the RF is centered in a texture figure 3.6° wide, defined by binocular disparity. The disc figure appeared at zero disparity, the background texture at 0.14° far disparity. The neuron's initial response to this display was nearly the same as to the homogeneous texture display. Yet after the initial response, the disparity-defined disc evoked significantly more vigorous responses. *c*, Illustration of the response when the disc figure was defined by chrominance cues; in this condition, the space between texture elements in the background was a green [CIE coordinates (x, y) = (0.344, 0.486)] equiluminant to the gray between texture elements in the disc [gray CIE coordinates (x, y) = (0.333, 0.333)], as confirmed with measurement by chrominance and luminance meters. We chose green (as opposed to, say, red or blue) solely to minimize chromatic aberration. The color-defined disc evoked a response very similar to the disparity-defined disc. *d*, Illustration of the response when the disc figure was defined by luminance cues; luminance of bars outside the disc was 43 cd/m², with the gray background the normal 24 cd/m² and the black bars the normal 6 cd/m². Again, the response of the V1 neuron was very similar to that for the disparity-defined disc. *e*, Illustration of the response to an orientation-defined disc. The cell here also showed elevated activity after the initial response as compared with the homogeneous texture display. *f*, Illustration of the response to a disc defined by each of the four preceding cues. The response magnitude for this “combination” display is not significantly different from that for discs defined by the four constituent cues. *g*, Illustration of the response for the disc-alone condition, in which the region outside the disc remained a constant gray throughout the trial. The response magnitude for this condition was not significantly different from the preceding five conditions.

response/homogeneous display response). The ratio measure is independent of absolute neural response rate. In Figure 4, we show histograms of these modulation ratios pooled by disc type. The data show that for the great majority of neurons, each of these disc displays evoked greater responses than did the homogeneous display, (i.e., most values in the histogram fall above the extra-RF modulation ratio value 1.0). The median modulation ratios and the percentage of cells significantly modulated for each disc display are as follows: for disparity-defined discs, the median modulation ratio was 1.67, and 50% of cells responded significantly more vigorously to the figure than to the homogeneous texture display ($p < 0.05$, one-sided t test); for color, the values were 1.74 and 52%; for luminance, 1.44 and 34%; for orientation, 1.69 and 52%. The extra-RF contextual modulation ratio values for the combination display (1.73 median modulation ratio and 48% of cells showing significant modulation) were similar to those for the other disc displays. This is an interesting result, because we might expect that extra-RF modulation arising in response to a display in which a number of potent cues segment the disc would reflect a summation of effects from individual cues and thus be substantially greater than extra-RF modulation evoked by any individual cue. Our data show that this is not the case. Finally, for the disc-alone condition, the median modulation ratio was 1.45, with 37% of cells significantly modulated.

We show examples of isolated V1 cells with a range of cue receptivity in Figure 5. In this figure, we only consider the five disc types used on all 64 cells (i.e., we exclude the disc-alone condition). In the top of the figure, we show responses rates for two cells (including cell *a* from Fig. 3) that each had very similarly positively modulated response rates for each of five disc displays (i.e., disparity-, color-, luminance-, orientation-, and combination-defined discs). In the bottom of the figure, we show responses rates from two other cells that displayed cue-dependent contextual modulation (i.e., discs defined by different cues yielded highly dissimilar responses). To quantify the cue-dependence of contextual modulation for a given cell, we defined a cue-variance index (*CVI*), which is simply the SD of average disc responses in excess of the homogeneous display response, divided by the homogeneous display response. A large value of *CVI* for a given cell indicates strong cue-dependence of contextual modulation, whereas a cell with a *CVI* of zero would have the same response to each disc display.

To classify cells according to the cue selectivity of their contextual modulation, we adopted conservative criteria for describing “cue-invariant” behavior. These were (1) that a given cell had significantly greater responses ($p < 0.05$, one-sided t test) to each of the five common disc displays compared with the homogeneous texture display; and (2) that the cell’s *CVI* was ≤ 0.25 . This definition is necessarily somewhat arbitrary, because the distribution of *CVI* values is essentially continuous, with no clearly separate modes that could be used to segregate cells. Nonetheless, the cutoff value we chose serves to select only those cells that intuitively appear to respond equivalently to the various cues, and the additional criteria of multiple significance tests ensure that this appearance is unlikely to be by chance. Twelve percent ($n = 8$) of the 64 isolated cells tested thus were classified as cue-invariant, whereas 27% of the cells ($n = 17$) were not significantly modulated by any disc display, and the remaining 61% of cells ($n = 39$) showed some significant contextual modulation without meeting the full criteria for “cue-invariance.”

One simple explanation for the invariance in response to disc displays is that the neurons reach some saturating level of activa-

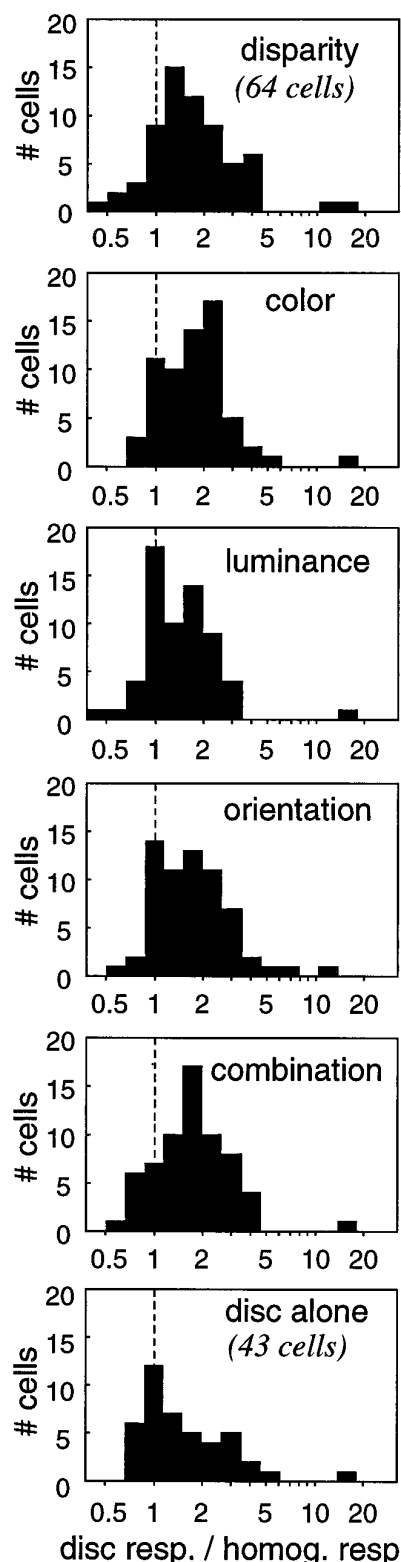


Figure 4. Single-unit extra-RF modulation ratios for diverse cues. Activity measures, as always, are from 100 to 250 msec after stimulus onset. The first five histograms compile modulation ratios for various figure-defining cues for all 64 cells tested with each of the following: the disparity-, color-, luminance-, orientation-, and combination-defined figures. The last histogram compiles modulation ratios for the 43 neurons tested with the disc-alone condition that were also tested with the preceding five disc displays. The form of each distribution is similar (i.e., most cells have ratio values > 1.0). See text for details.

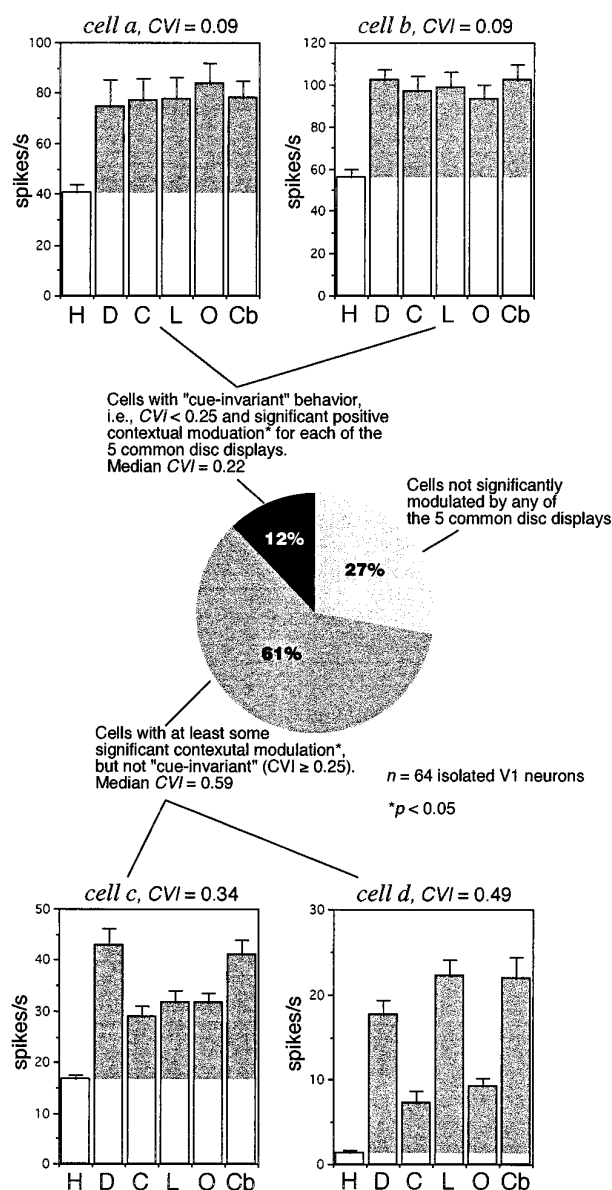


Figure 5. Variation of cue receptivity among single units for diverse cues. This figure deals with the 64 isolated V1 neurons tested with the homogeneous texture display (*H*) and the five common disc displays: disparity (*D*), color (*C*), luminance (*L*), orientation (*O*), and combination (*Cb*). We define a cue average variance index (*CVI*) as the standard deviation of a cell's responses to disc displays in excess of the response to the homogeneous texture display, normalized by the response to the homogeneous texture display. For cell *a* (the same cell as in Fig. 3), with activity levels shown as a bar chart in the upper left of the figure, this corresponds to the SD of the heights of the gray portions of the response bars divided by the height of the leftmost bar (the homogeneous display response). We conservatively define a neuron to be cue-invariant in extra-RF contextual modulation if it has a *CVI* < 0.25 and shows significantly greater response to each of the five disc displays as compared with the homogeneous texture display (*p* < 0.05 for one-sided *t* test for each disc). The center of the figure shows a pie chart that divides the cells into three classes: cells not significantly modulated by any of the five disc displays, cells that are cue-invariant, and cells with significant modulation that fall short of the cue-invariant classification. At the top and bottom of the figure are example cells.

tion that causes the response for each disc display to converge at the same activity level. We can counter this argument by simply noting that the neurons in fact did not reach saturating levels of

activity during stimulation with the normal texture displays. For example, the most cue-invariant isolated V1 neurons in our sample (cell *b* in Fig. 5) had an overall vigorous response for disc displays but could be driven to a response level 63% larger by using a different RF stimulus (for this cell, monocular texture stimulation in the right eye) (data not shown). Observations such as these make it very unlikely that the cue-invariance of extra-RF contextual modulation arises from simple saturation in the response of cells from which we recorded.

In summary, in this section, we showed that within the population of V1 neurons, robust extra-RF modulation exists for each of the diverse cues that we tested. These results suggest that extra-RF modulation serves a function that generalizes across visual cues. If widespread extra-RF modulation had existed for only a subset of the disc displays (say, those defined by orientation and luminance but not those defined by color or disparity), this phenomenon could at best serve only a restricted role tied to particular visual cues (such as orientation or luminance analysis). Instead, our results suggest that contextual modulation serves an integrative function across diverse cues. This means that cues traditionally considered separate subjects of study, such as color and binocular disparity, are linked in the sense that extra-RF contextual modulation in V1 commonly uses both. Although it has been suggested that different visual cues (such as color and binocular disparity) are processed independently by separate anatomical modules in the visual system (Livingstone and Hubel, 1987, 1988), our results show that many V1 neurons treat these cues interchangeably, at least in terms of contextual modulation.

Spatial extent of extra-RF modulation

Complementary to the question of what cues evoke contextual modulation is the question: how large is the spatial extent of this phenomenon? We measured this by varying disc diameter from trial to trial, while keeping the RF centered. Figure 6*a* shows sample responses from one V1 multiunit site tested with the homogeneous texture display, whereas Figure 6*b* illustrates the entire diameter-tuning curve for the same multiunit site. The magnitude of contextual modulation declines with increasing disc diameter and vanishes at ~10° diameter.

We studied 33 single- and 51 multiunit sites in experiments with variable sized discs. We used only orientation (*n* = 65), color (*n* = 5), or luminance (*n* = 14) cues for this part of the study, so that the entire monitor screen (32 × 24° in dimensions) could be covered with texture. Single and multiunit sites had similar characteristics. Figure 6*c* illustrates the median contextual modulation ratio for all 84 sites, measured at each disc diameter. This smooth, monotonically falling spatial tuning function reaches the level of the homogeneous texture background at ~10° diameter. Only at the smallest disc diameter (1.8°) did we occasionally find significant deviations from this pattern (perhaps reflecting an interaction between the disc contour and the RFs of neurons in these cases). In Figure 6*d*, we graph the fraction of sites with significant contextual modulation (*p* < 0.05 for one-sided *t* test) as a function of disc diameter. For discs with diameter up to ~8°, the proportion of sites showing significant modulation is greater than that expected by chance.

Contextual modulation with binocularly rivalrous displays

Up to this point, we have dealt with displays in which inhomogeneity in texture outside the RF is correlated with the expression of contextual modulation. In this and the following section, we treat

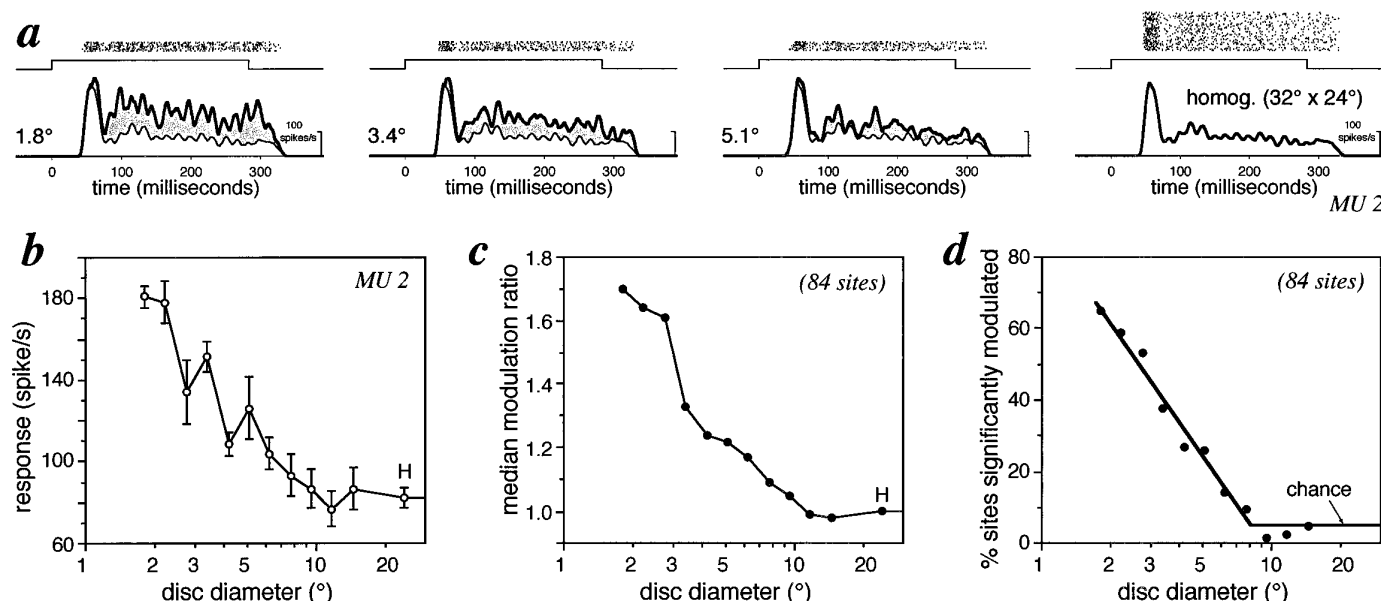


Figure 6. Spatial extent of extra-RF contextual modulation tested with discs of variable diameter. *a*, Illustration of some sample responses from one multiunit site tested with variable diameter discs defined by luminance. *b*, Illustration of the entire diameter-response function for the same multiunit site. Response rates fall off essentially monotonically with disc diameter. *c*, Illustration of the median extra-RF modulation ratio as a function of disc diameter for all 84 sites tested ($n = 65$ tested with orientation-defined discs, 14 with luminance-defined discs, and 5 with color-defined discs). Extra-RF modulation declines monotonically with disc diameter, reaching the value 1.0 at $\sim 10^\circ$ diameter. *d*, Illustration of the fraction of sites with significant modulation ($p < 0.05$, one-sided t test) as a function of disc diameter. The fraction of modulated sites reaches chance level at $\sim 8^\circ$ diameter.

displays where this simple link is broken; in other words, we study test texture displays that are not homogeneous but nonetheless fail to evoke contextual modulation or, equivalently in our terminology, evoke the same response as a homogeneously textured display. The first such texture displays that we will describe involves the use of binocular rivalry. Examples of rivalrous texture displays used in our study are illustrated in Figure 7*a*. Each row of this figure shows the images presented to left and right eyes and an *approximate* representation of the cyclopean percept obtained when these images are fused. In our experiments, monkeys viewed pairs of texture displays through a haptoscope.

The first row of Figure 7*a* illustrates the case in which homogeneous texture is presented to each eye, but the texture orientation differs by 90° between eyes (case 1). The cyclopean percept is of a fairly homogeneous texture field combining texture elements from both eyes. The second row illustrates the case in which one eye views a homogeneous texture field while the other views a field containing an orientation-defined figure (case 2). The stable cyclopean percept here is of a clearly delineated square texture surface with rivalrous texture patterns surrounded by a nonrivalrous background. The third row shows the case in which an orientation-defined figure appears to both eyes, but the orientation of texture at corresponding points in the display differs by 90° between the eyes (case 3). As has been observed previously with closely related displays (Kolb and Braun, 1995), the *cyclopean percept* in this case is surprisingly homogeneous. Some pieces of contour are visible in the fused display, but the overall sense of figure/ground segregation seen in the monocular images is clearly lost. Note that the texture in the central region of the displays is the same in all three cases. (One consequence of maintaining the same rivalrous texture over the RF from trial to trial is that in the cases in which no figure is perceived, the background texture is also rivalrous. Although it seems unlikely that this fact in itself is

the basis for the results we describe below, future experiments should test this explicitly.)

We recorded from 40 multiunit and 6 single-unit sites in area V1 while presenting displays like those in Figure 7*a* to awake, fixating monkeys. Displays were configured such that the RF of a V1 neuron under study (or the aggregate RF of a group of cells) fell completely within the square region of the display that sometimes appeared as a figure. In this manner, the RF was stimulated with *exactly the same texture pattern* from trial to trial, whereas texture entirely outside the RF could vary, as seen in Figure 7*a*.

The responses recorded with rivalrous displays for one V1 multiunit site are illustrated in Figure 7*b*. Texture was flashed on a gray background for 200 msec as a monkey foveated the fixation spot. Cells at this site showed almost no activity for the uniform gray display but responded to the appearance of the case 1 texture display with a vigorous burst of action potentials. After the initial response, activity decayed to a reduced level for the remainder of the texture display interval. Using case 2 texture displays in randomly interleaved trials, we recorded dramatically different results. Although the cells initially responded to the onset of the case 2 display in the same way as in the previous case, the subsequent sustained activity level was far greater. This extra-RF contextual modulation occurred whether the orientation-defined figure appeared in the left or the right eye. However, when the orientation-defined figure appeared in *both* eyes (case 3), the response profile was virtually identical to that for case 1.

Figure 7*c* illustrates results from a separate multiunit site. This site showed strong ocular dominance for the right eye. Contextual modulation in case 2 displays occurred predominantly for the condition with the figure in the right eye. Still, when rivalrous figures appeared in *both eyes* (case 3), the response again was the same as for case 1, despite the fact that the right eye stimulus was

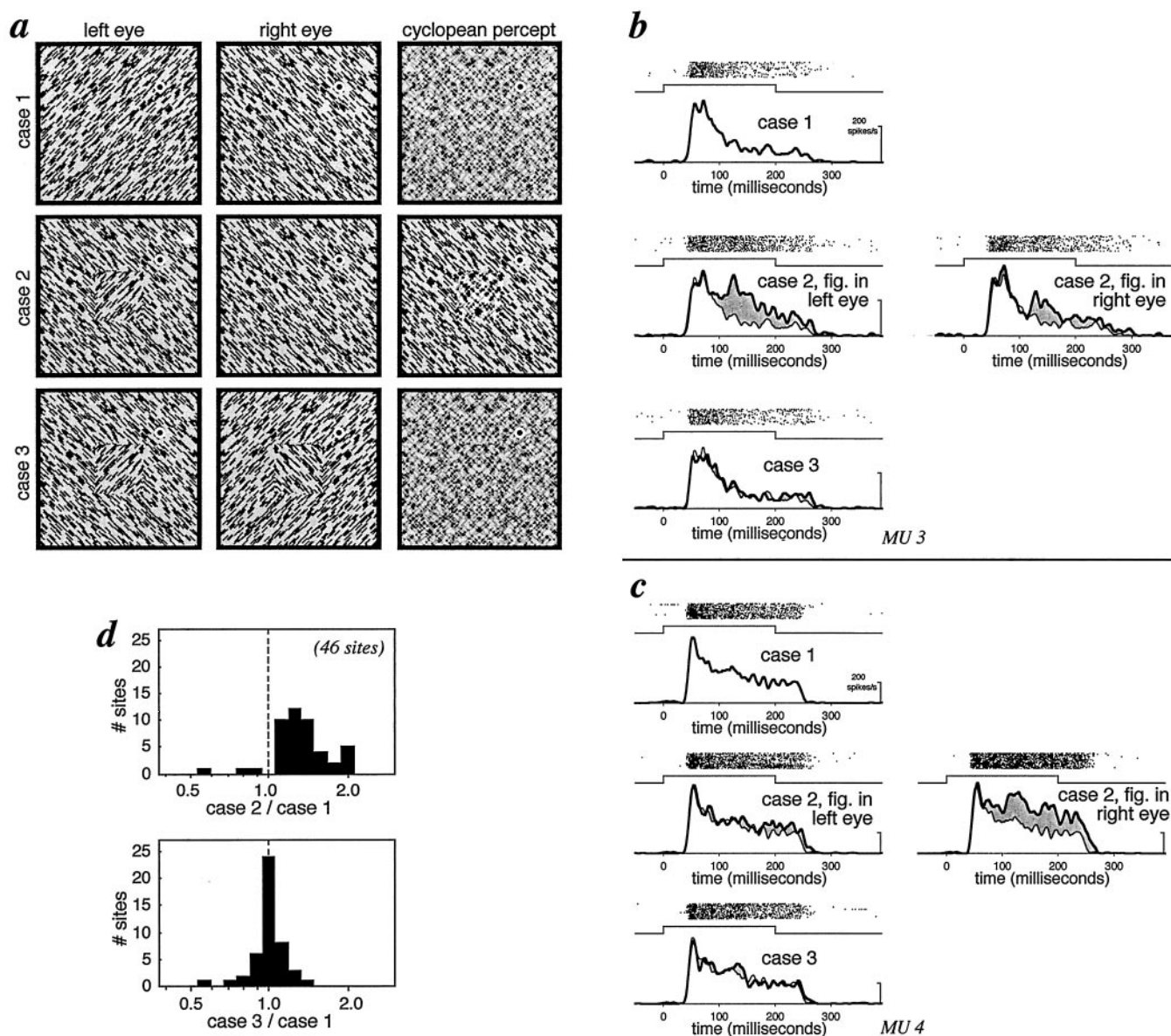


Figure 7. Extra-RF contextual modulation and binocular rivalry. *a*, Illustration representing examples of three types of binocularly rivalrous displays (case 1, case 2, and case 3) for each illustrating left- and right-eye images and an approximate representation of the cyclopean percept. See text for complete description. *b*, Illustration of the response of one multiunit site to these displays. *c*, Illustration of another multiunit example. *d*, Histograms of extra-RF modulation ratios for case 2/case 1 and case 3/case 1, with case 1 filling the role of the homogeneously textured display. See text for details.

identical to that in the case 2 condition that produced strong modulation.

The results across our sample of 46 V1 sites were remarkably consistent with those shown in Figure 7, *b* and *c*. We again quantify the results by calculating a ratio, the response rate to case 2 or case 3 displays divided by the response rate to the case 1 display. The top of Figure 7*d* illustrates a histogram of extra-RF context modulation ratios for case 2, with the conditions of the figure in either the left or the right eye averaged. As with the examples above, the average responses to case 2 were typically greater than to case 1; ratio values fall consistently above 1.0 (the median value is 1.45; 76% of sites had significantly greater responses to at least one of the case 2 displays as compared with the case 1 display, $p < 0.025$ in one-sided t test for figure in each eye). The bottom of Figure 7*d* illustrates a histogram of extra-RF

context modulation ratios for case 3. As with the examples above, responses to case 3 were typically the same as to case 1; ratio values cluster tightly about 1.0 (the median value is 1.01; only 2% of sites showed activity significantly greater than for the case 1 display, $p < 0.05$ in one-sided t test). Thus, displays that generate a cyclopean percept of a homogeneously textured field evoke the same level of neural activity (given identical RF stimulation) as a truly homogeneous texture field, *even though the monocular images may contain clearly defined figures*.

An important question is, how do neural responses correlated with perception of rivalrous displays relate to the ocular dominance characteristics of individual V1 cells? The data in our rivalry experiment (predominantly recordings of multiple-unit activity in superficial layers of striate cortex) do not contain a sufficiently large proportion of sites with strong ocular bias to

establish quantitative relationships on this point. Nonetheless, it is noteworthy that in the examples that we do have of sites whose receptive fields were predominantly activated by stimulation in one eye (e.g., Fig. 7c), there is a clear interaction of contextual stimuli across the eyes.

In summary, we have seen with the results in Figures 2–6 that a change in the global perceptual nature of an image can substantially alter the firing rate of V1 neurons the RFs for which cannot detect the change in stimulus. However, with our experiments using binocularly rivalrous stimuli in Figure 7, case 3, we demonstrate that a large change in the image stimulus that has *little or no perceptual consequence* (because of rivalry) does not alter the firing rate of V1 neurons.

Contextual modulation and perceived distal structure

One possible interpretation of the results thus far is that contextual modulation better reflects the perceived structure of the stimulus (e.g., figure vs ground or figure size) than it reflects the particular cues (such as disparity or color) that delineate this structure. Our purpose for this section lies in studying more directly how extra-RF modulation relates to the perceived distal structure of our stimulus displays. Our approach is to vary the perceived distal structure of the display region containing the RF, while at the same time keeping RF texture stimulation the same from trial to trial. A key display that allows us to do this is illustrated in Figure 8a. The display appears as a homogeneously textured field, with the modification that we can manipulate the perceived depth of a band of texture surrounding the RF by varying binocular disparity cues (i.e., the band of texture between the *white dashed lines* in Fig. 8a; dashed lines are not in the actual display). It is in our opinion a reasonable assumption that our monkeys perceived the various manipulations of this display as do human (see Materials and Methods); however, we cannot offer proof here of this assumption.

In the case in which we cause the texture band surrounding the RF to have the same binocular disparity as the other regions of the display, we simply generate a standard homogeneous texture display. In the top of Figure 8b, we illustrate this display. In Figure 8, *c (top)* and *d (top)*, we illustrate the average response profiles of two V1 multiunit sites to stimulation with this display. Each showed an initial vigorous burst of activity in response to texture onset, followed by a much diminished response rate for the remainder of the texture display interval.

Moat

We could alter the perceived distal structure of the display by causing the band of texture surrounding the RF to appear farther away in depth from the remaining area of texture (typically through 0.14° uncrossed horizontal disparity). We refer to this receded region as a “moat,” illustrated in the center of Figure 8b. As seen from the illustration, with establishment of the moat, the RF no longer appears positioned on a large textured field but rather appears to be positioned on a small square surface isolated from the textured background by the moat. In the experiment, moat depth was only apparent through binocular disparity cues, although we provide some shading cues to depth in Figure 8b for schematic purposes.

In Figure 8, *c (center)* and *d (center)*, we illustrate the resulting average response profiles of the two multiunit sites. The initial response of the multiunit sites to the moat display was nearly identical to the response to the homogeneous texture display. However, for both sites the response rates diverged ~ 100 msec

after texture onset, with the moat display causing the cells at each site to maintain a more vigorous response rate than did the homogeneous texture display (gray shading of the response profiles.) Thus, we see that the moat display evoked extra-RF modulation of the same nature as we have seen with the various tests that we have already described in previous sections of this paper.

Frame

We could also modify the display in Figure 8b in a different way by having the texture band surrounding the RF appear nearer in depth than the remaining area of texture (through 0.14° crossed horizontal disparity). In this case, the perceived distal structure (Fig. 8b, *bottom*) is completely different from the moat display. In the frame display, the RF appears positioned not on a small textured surface but on a large textured surface *continuous* with the textured background, as though a narrow textured “frame” were merely floating above and partially occluding the homogeneous texture display. In Figure 8, *c (bottom)* and *d (bottom)*, we illustrate the average response rates of each multiunit site to the frame display. The results stand in strong contrast to the response to the moat display, because the multiunit responses to the frame display either closely follow those to the homogeneous texture display or are even less vigorous.

Remarkably, this asymmetry of effect for the moat display compared with the frame display was highly consistent among the 14 single- and 132 multiunit sites that we studied with these stimuli. We demonstrate this in Figure 8e, which illustrates histograms of extra-RF contextual modulation ratios for these recording sites. In the top of the histogram, we show the values for moat response/homogeneous response. Extra-RF modulation ratio values in this case fall consistently above 1.0, indicating that neural responses for the moat display generally exceeded those for the homogeneous texture display (the median value is 1.68; 63% of sites showed responses for the moat display significantly greater than to the homogeneous texture display, $p < 0.05$ in one-sided t test). In the bottom, we show the ratio values for frame response/homogeneous response. In contrast to the moat case, here the extra-RF modulation ratio values cluster near or below 1.0, indicating that for the frame display, neurons responded in a manner similar to or weaker than that to the homogeneous texture display (the median value being 0.75; 37% of sites showed responses to the frame display significantly *less* than to the homogeneous texture display, whereas only 2% of sites showed significantly greater activity, $p < 0.05$ in one-sided t tests). The square region inside the moat or frame was between 2 and 3.6° for different recording sites. Control experiments at each recording site showed that cells did not respond to the extra-RF texture band alone, or gave at best extremely weak responses, regardless of whether it appeared at near (frame), far (moat), or zero disparities (data not shown).

Perturbations in moat and frame displays that retained the essential character of their perceived distal structure evoked qualitatively similar results to those just described. For example, the asymmetry in effect of moat and frame displays for evocation of extra-RF modulation did not depend on having the displays centered at zero disparity (the standard case, e.g., multiple-unit site 5 in Fig. 8c) but was equally evident when we moved texture displays back in depth relative to the fixation spot (e.g., multiple-unit site 6 in Fig. 8d). Furthermore, we could vary the magnitude of the moat and frame disparities to larger or smaller values than our $\pm 0.14^\circ$ standard without qualitatively altering the basic moat/frame modulation asymmetry (data not shown).

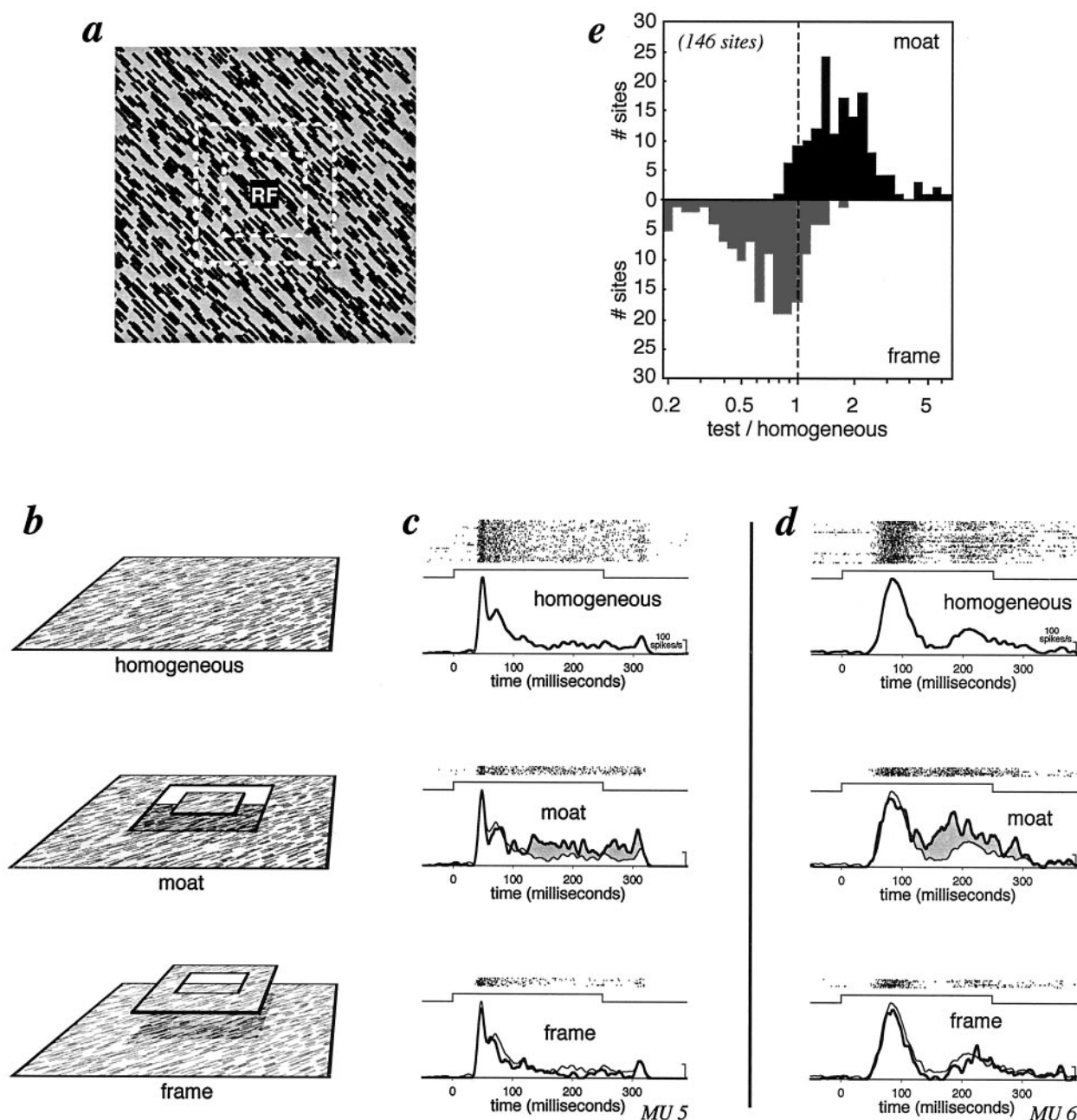


Figure 8. Extra-RF contextual modulation and perceived distal structure. *a*, Configuration of a texture display in which a band of texture surrounding the RF may vary in apparent depth through binocular disparity cues. The default texture was typically at zero disparity, although we used far background disparity as the default in some experiments. *b*, Illustration of how this display may be configured to appear as a homogeneous texture display, moat display, or frame display. Typically, the disparity offset of moat and frame was $\pm 0.14^\circ$, although this value was not critical. Monkey binocular vision is similar to that of man (see Materials and Methods), and we assume that our monkey subjects perceive these displays as do human observers. *c*, Illustration of the responses of a multiunit site to the displays in *b*. For this experiment, RF texture was always at zero disparity. *d*, Illustration of responses of another multiunit site. For the data shown, RF texture was at 0.14° far disparity (with moat and frame moved back accordingly to preserve the relative depth arrangements). The contextual modulation pattern is the same as in *c*. The site in *d* also gave the same pattern of response when RF texture was at zero disparity (data not shown). *e*, Extra-RF contextual modulation ratios for 146 sites for moat (*top*) and frame (*bottom*) displays. See text for details.

In summary, when the RFs of V1 neurons appear to rest on a large flat textured surface (i.e., the homogeneous texture display), cells consistently give a small response, even when this surface is partially occluded by a frame. However, when the RFs of V1 neurons appear within a smaller “figure” surface surrounded by a moat, consistent contextual modulation is evoked.

It seems natural to ask whether the moat/frame asymmetry stems from some asymmetry in the RF disparity tuning of cells in

our sample. In fact, we did not find any overall bias of single- or multiunit sites for a particular RF disparity tuning. In other words, the normal results for presentation of moat and frame displays may be elicited from cells that prefer either near or far disparity stimuli (data not shown). Analogous dissociations have been observed by Lamme (1995), wherein extra-RF contextual modulation evoked by orientation cues has no correlation with the sharpness of orientation tuning of individual V1 RFs; further-

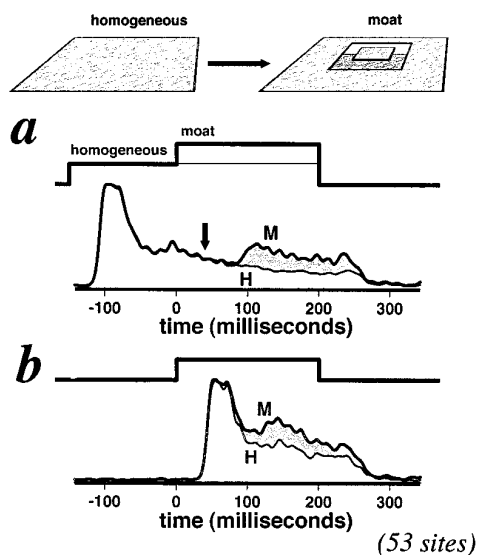


Figure 9. Characteristic delay of extra-RF contextual modulation. In a two-step texture presentation procedure, we initially present the homogeneous texture display and then 150 msec later, change to the moat display by manipulating only extra-RF texture. As the RF is entirely within the moat-defined figure, it receives static RF texture stimulation whether or not the moat appears. In *a*, we compare average response profiles of 53 sites for the two-step moat presentation (trace *M*) and simple long-duration homogeneous texture (trace *H*). The responses are identical until ~80 msec after the moat appears, after which the neural response rebounds for the moat condition. In *b*, we show the results of the analogous one-step moat experiment performed on the same 53 sites in randomly interleaved trials. Despite the different time course of RF stimulation, the timing of extra-RF contextual modulation is the same.

more, Lamme also found that direction selectivity of V1 RFs was uncorrelated with contextual modulation evoked by motion cues. Taken together, these data suggest an overall dissociation between specific types of RF tuning and the extra-RF contextual modulation received by V1 neurons.

Temporal characteristics of V1 contextual modulation

A striking trend in the results that we have collected is the delay in the expression of extra-RF contextual modulation in V1. This delay is important in the discussion of whether contextual modulation reflects perceptual experience, because the delay could allow complex and lengthy neural computations to contribute to the expression of this phenomenon. But is this delay indeed a characteristic of contextual modulation, or is it an artifact tied in some trivial way to the recent history of RF stimulation? For example, is the delay of contextual modulation a mere artifact of saturation in neural response at texture onset?

To show that the delay in the onset of extra-RF modulation is a characteristic feature of the phenomenon itself and not merely a simple side effect of the recent history of RF stimulation, we need to show that this delay is independent of the time at which the RF itself was first stimulated. We test this by using a two-step procedure in which we first present a homogeneous texture display (thereby generating the initial burst of neural activity) and then subsequently modifying *only* the extra-RF stimulus. We can contrast these results to the response recorded when the homogeneous texture display remains unchanged throughout the entire period. In Figure 9*a*, we illustrate results of an experiment of this type performed on 53 V1 multiunit sites. In the first step of texture presentation, the homogeneous display appears for 150 msec. In

the second step, a narrow band of texture surrounding but outside the RF is replaced with texture of farther binocular disparity. The result is that the display in this second step contains a figure region surrounded by a gap or moat. The average neural response for this two-step condition is illustrated by trace *M* and is compared with the response to a long-duration homogeneous display (trace *H*). We see that after the initial burst of activity, the response rate settles into a steady state of activity. However, between 80 and 100 msec after the display changed to the moat-defined figure configuration, the response rate *rebounds* to a more elevated level of activity (indicated by the gray shading of the response profile). The vertical arrow indicates the time at which the cells would have started to respond had the texture within the RF itself been modified in the second step of the two-step condition. Note that the average response at this point in time is in fact identical to the average response to the static homogeneous texture display. Interestingly, the delay of modulation in the *two-step* presentation (highlighted in gray) is the same as for the modulation evoked by a normal *one-step* presentation of moat versus homogeneous displays (Fig. 9*b*, showing average data from the same sites collected in randomly interleaved trials).

Also in interleaved trials, we included a two-step presentation similar to that in Figure 9*a*, except that the texture band added in the second step was of the same disparity as the homogeneous texture; thus, despite a texture change between steps one and two as the band was added, both steps had the same steady-state appearance of a homogeneous surface. Unlike the two-step presentation in which the moat was added in the second step, this procedure yielded no consistent effect: responses were statistically indistinguishable from those for the static homogeneous texture presentation in 87% of recording sites ($p > 0.05$ for two-sided t test), and for the remaining sites, there was no bias toward increased or decreased response (data not shown).

The results in this section are important, because they indicate that extra-RF modulation need not be triggered by an initial burst of activity. Rather, the results show that extra-RF modulation may be triggered even when neurons have achieved a steady state of firing from constant RF stimulation. They suggest that extra-RF contextual modulation is a neural process distinct from the normal RF functioning of a V1 neuron, because in contrast to the delay in expressing extra-RF modulation, V1 neurons display their tuning specificity for visual stimuli with their first action potential responses to visual stimulation (Celebrini et al., 1993).

It has been suggested that the delay in expression of contextual modulation in our texture experiments is a phenomenon related to the delay in neural response that can be observed with low-luminance contrast stimuli (Geisler and Albrecht, 1992). This speculation is based on the assumption that our texture figures in some sense have low "effective" contrast analogous to the low-luminance contrast. However, this assumption fits neither with phenomenological observations of our actual displays (i.e., figures do not appear to be "low contrast" on the monitor screen) nor with behavioral data (i.e., monkeys consistently are able to initiate eye movements to texture figures with short latencies in the range of 120–150 msec), but for true low-contrast luminance stimuli, the latency may be twice as long (Schiller, 1993).

DISCUSSION

Given the images impinging on the retinae, the visual system must model the three-dimensional structures of the distal world. Distal world structure cannot be found through image-filtering alone, however, because the structures of the distal world modeled so

richly through our perception do not in fact exist in the retinal images (Kanizsa, 1979; Marr, 1982). Rather, distal structure must be *inferred* from their reflected traces of contour and texture in the retinal images (Nakayama and Shimojo, 1992; Adelson, 1993; Anderson and Julesz, 1995). Moreover, because we have a relatively fixed vantage point on any scene at any given moment, the visual system must also make inferences about forms not directly visible, such as the manner in which surfaces continue beneath occluding structures (Nakayama et al., 1989; Kovacs and Julesz, 1994; Rensink and Enns, 1995). For the visual system to accomplish these tasks, it must employ sophisticated mechanisms for translating retinal images into models of the three-dimensional structures in the distal world.

The function of area V1 has long appeared far removed from these concerns. The RFs of V1 neurons are well described as spatially localized filters, jointly tuned for orientation and spatial frequency (Schiller et al., 1976a,b; Movshon et al., 1978; De Valois et al., 1982a,b). The “lines” or “edges” that may stimulate these cells (Hubel and Wiesel, 1968) do so not because they form the contours of surfaces or objects in any perceptual context. Rather, they do so merely because the cells are tuned for the specific two-dimensional spatial frequency content of these stimuli, *regardless of their perceptual context* (De Valois et al., 1979). The tuning characteristics of V1 RFs for color and binocular disparity are likewise well described as simple filters that have no direct connection to the perceptual interpretation of distal world structure (Lennie et al., 1990; DeAngelis et al., 1991).

The problem of *synthesizing*, from V1 RF filter information, a perceptual model of distal world structure (as, for example, in reconstructing the three-dimensional form of physical surfaces) has traditionally been assumed to occur only at later stages of visual processing than striate cortex. This view has appeared sensible both because the filter description of V1 seemed conceptually complete (De Valois and De Valois, 1988), and because there was little compelling evidence that V1 neurons could be doing anything qualitatively different from simple image filtering. The extra-RF contextual modulation recently described by Lamme (1995) and in the present paper stands in strong contrast to the filter properties of the V1 RFs themselves, however. We have seen that contextual modulation is a phenomenon of broad spatial scope (Fig. 6), which is nonetheless sensitive to fairly small-scale perturbations of the stimulus display (Fig. 8). It may be evoked by a wide variety of visual cues (Fig. 4), and under certain stimulus conditions, can respond invariantly to individual cues or to cues in combination (Fig. 5). Yet under other conditions, strong image features will fail to evoke extra-RF modulation and may even block its effects (Fig. 7). Although these observations in no way challenge theories of the functional role of the RF itself, the complexity and apparent flexibility of extra-RF contextual modulation in V1 clearly does challenge the view that the role of the V1 neuron is *solely* to filter local regions of images in a cue-specific manner.

There is, of course, a great difference between saying on the one hand that V1 neurons do more than simple RF filtering, and on the other that V1 in fact participates in perceptual modeling of distal world structure. Although the former point is now indisputable, the latter calls for debate. Our approach has been to observe whether contextual modulation recorded in monkey V1 consistently follows our perception of textured displays. To an astounding degree, this has been the case. The results on binocular rivalry, cue combination and invariance, and moat and frame, considered together with Lamme's previous data (1995), in our opinion call

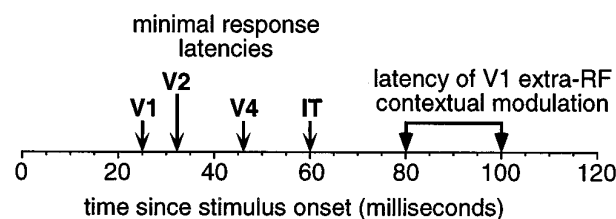


Figure 10. Comparison of response latencies of various visual areas and extra-RF contextual modulation in area V1. Extra-RF contextual modulation in V1 occurs well after the extrastriate cortical areas are initially activated. If we assume that feedforward and feedback pathways have the same time delay (Domenici et al., 1995), we see that the V1-inferotemporal cortex (IT) latency would permit IT signals to reenter V1 before extra-RF contextual modulation is fully expressed at the poststimulus onset time $t = 100$ msec.

for serious consideration of the hypothesis that area V1 has access to and participates in formation of perceptual interpretation of the visual scene.

As is illustrated in Figure 10, the various primate extrastriate cortical areas are all activated before the appearance of contextual modulation that we observed in V1 (Maunsell, 1987; Maunsell and Gibson, 1992; Miller and Desimone, 1993). The large spatial scope and the complex RFs of extrastriate neurons (Peterhans, 1989; Albright, 1992; Snowden et al., 1992; von der Heydt and Marcar et al., 1995), coupled with the delay in expression of modulation in V1, should allow feedback signals from extrastriate areas to support the neural events that we have observed. In contrast, it is not clear from our current understanding of circuitry *within* V1 (Lund, 1988; Gilbert and Wiesel, 1989; McGuire et al., 1991; Kapadia et al., 1995) that lateral connections in striate cortex could underlie our results. One possible interpretation of our data is that visual processing involves a series of temporally discrete steps in which information is initially fed forward through V1, is further processed in extrastriate cortex, and then returns to V1 after an interval of ~ 50 msec. Such delayed feedback would permit V1 access to perceptual interpretations of the visual scene, such as surface or figure/ground representations of the distal world, ascribed previously only to extrastriate cortical areas.

REFERENCES

- Adelson EH (1993) Perceptual organization and the judgment of brightness. *Science* 262:2042–2044.
- Albright TD (1992) Form-cue invariant motion processing in primate visual cortex. *Science* 255:1141–1143.
- Allman J, Miezin F, McGuinness E (1985) Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu Rev Neurosci* 8:407–430.
- Anderson BL, Julesz B (1995) A theoretical analysis of illusory contour formation in stereopsis. *Psychol Rev* 102:705–743.
- Blakemore C, Tobin EA (1972) Lateral inhibition between orientation detectors in the cat's visual cortex. *Exp Brain Res* 15:439–440.
- Bough EW (1970) Stereoscopic vision in the macaque monkey: a behavioural demonstration. *Nature* 225:42–44.
- Celebrini S, Thorpe S, Trotter Y, Imbert M (1993) Dynamics of orientation coding in area V1 of the awake primate. *Vis Neurosci* 10:811–825.
- Cowey A, Parkinson AM, Warnick L (1975) Global stereopsis in rhesus monkeys. *Q J Exp Psychol* 27:93–109.
- DeAngelis GC, Ohzawa I, Freeman RD (1991) Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature* 352:156–159.
- De Valois KK, De Valois RL, Yund EW (1979) Responses of striate cortex cells to grating and checkerboard patterns. *J Physiol (Lond)* 291:483–505.

- De Valois RL, Yund EW, Hepler N (1982a) The orientation and direction selectivity of cells in macaque visual cortex. *Vis Res* 22:431–544.
- De Valois RL, Albrecht DG, Thorell LG (1982b) Spatial frequency selectivity of cells in macaque visual cortex. *Vis Res* 22:545–559.
- De Valois RL, De Valois KK (1988) *Spatial vision*. New York: Oxford UP.
- Domenici L, Harding GW, Burkhalter A (1995) Patterns of synaptic activity in forward and feedback pathways within rat visual cortex. *J Neurophysiol* 74:2649–2664.
- Geisler WS, Albrecht DG (1992) Cortical neurons: isolation of contrast gain control. *Vision Res* 32:1409–1410.
- Gilbert CD, Wiesel TN (1989) Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *J Neurosci* 9:2432–2442.
- Gilbert CD, Wiesel TN (1990) The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision Res* 30:1689–1701.
- Harwerth RS, Smith 3rd EL, Siderov J (1995) Behavioral studies of local stereopsis and disparity vergence in monkeys. *Vision Res* 35:1755–1770.
- Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol (Lond)* 195:215–243.
- Kanizsa G (1979) *Organization in vision. Essays in gestalt perception*. New York: Praeger.
- Kapadia MK, Ito M, Gilbert CD, Westheimer G (1995) Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron* 15:843–856.
- Knierim JJ, van Essen DC (1992) Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J Neurophysiol* 67:961–980.
- Kolb FC, Braun J (1995) Blindsight in normal observers. *Nature* 377:336–338.
- Kovacs I, Julesz B (1994) Perceptual sensitivity maps within globally defined visual shapes. *Nature* 370:644–646.
- Lamme VAF (1995) The neurophysiology of figure-ground segregation in primary visual cortex. *J Neurosci* 15:1605–1615.
- Lennie P, Krauskopf J, Sclar G (1990) Chromatic mechanisms in striate cortex of macaque. *J Neurosci* 10:649–669.
- Leopold DA, Logothetis NK (1996) Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 379:549–553.
- Livingstone MS, Hubel DH (1987) Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. [Review] *J Neurosci* 7:3416–3468.
- Livingstone MS, Hubel D (1988) Segregation of form, color, movement, and depth: anatomy, physiology, and perception. [Review] *Science* 240:740–749.
- Lund JS (1988) Anatomical organization of macaque monkey striate visual cortex. [Review] *Annu Rev Neurosci* 11:253–288.
- Maffei L, Fiorentini A (1976) The unresponsive regions of visual cortical receptive fields. *Vision Res* 16:1131–1139.
- Marcus VL, Xiao DK, Raiguel SE, Maes H, Orban GA (1995) Processing of kinetically defined boundaries in the cortical motion area MT of the macaque monkey. *J Neurophysiol* 74:1258–1270.
- Marr D (1982) *Vision*. San Francisco: Freeman.
- Maunsell JHR (1987) Physiological evidence for two visual subsystems. In: *Matters of intelligence* (Vaina L, ed), pp 59–87. Dordrecht, Holland: Reidel.
- Maunsell JHR, Gibson JR (1992) Visual response latencies in striate cortex of the macaque monkey. *J Neurophysiol* 68:1332–1344.
- McGuire BA, Gilbert CD, Rivlin PK, Wiesel TN (1991) Targets of horizontal connections in macaque primary visual cortex. *J Comp Neurol* 305:370–392.
- Miezin FM, Myerson J, Julesz B, Allman JM (1981) Evoked potentials to dynamic random-dot correlograms in monkey and man: a test for cyclopean perception. *Vision Res* 21:177–179.
- Miller EK, Desimone R (1993) Activity of neurons in anterior inferior temporal cortex during a short term memory task. *J Neurosci* 13:1460–1478.
- Movshon JA, Thompson ID, Tolhurst DJ (1978) Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J Physiol (Lond)* 283:53–77.
- Nakayama K, Shimojo S (1992) Experiencing and perceiving visual surfaces. *Science* 257:1357–1363.
- Nakayama K, Shimojo S, Silverman GH (1989) Stereoscopic depth: its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception* 18:55–68.
- Nelson JJ, Frost BJ (1978) Orientation-selective inhibition from beyond the classic visual receptive field. *Brain Res* 139:359–365.
- Rensink RA, Enns JT (1995) Preemption effects in visual search: evidence for low-level grouping. *Psychol Rev* 102:101–130.
- Robinson DA (1963) A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Trans Biomed Eng* 101:131.
- Sarmiento RF (1975) The stereoacuity of macaque monkey. *Vision Res* 15:493–498.
- Schein SJ, Desimone R (1990) Spectral properties of V4 neurons in the macaque. *J Neurosci* 10:3369–3389.
- Schiller PH (1993) The effects of V4 and middle temporal (MT) area lesions on visual performance in the rhesus monkey. *Vis Neurosci* 10:717–746.
- Schiller PH, Finlay BL, Volman SF (1976a) Quantitative studies of single-cell properties in monkey striate cortex. II. Orientation specificity and ocular dominance. *J Neurophysiol* 39:1320–1333.
- Schiller PH, Finlay BL, Volman SF (1976b) Quantitative studies of single-cell properties in monkey striate cortex. III. Spatial frequency. *J Neurophysiol* 39:1334–1351.
- Sillito AM, Grieve KL, Jones HE, Cudeiro J, Davis J (1995) Visual cortical mechanisms detecting focal orientation discontinuities. *Nature* 378:492–496.
- Snowden RJ, Treue S, Andersen RA (1992) The response of neurons in areas V1 and MT of the alert rhesus monkey to moving random dot patterns. *Exp Brain Res* 88:389–400.
- Tyler CW (1983) Sensory processing of binocular disparity. In: *Vergence eye movements: basic and clinical aspects* (Schor CM, Ciuffreda KJ, eds). Boston: Butterworth.
- von der Heydt R, Peterhans E (1989) Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. *J Neurosci* 9:1731–48.
- Zeki SM (1973) Colour coding in rhesus monkey prestriate cortex. *Brain Res* 53:422–427.