Introduction

This proposal entitled “Cortical mechanisms of visual scene segmentation” (R01 EY017291-01) was submitted in June 2005. Despite general enthusiasm about the importance of the proposed project, the reviewers raised some concerns about the difficulty in achieving the aims (Priority Score: 217, Percentile: 41.9). No concerns were raised about the investigator, environment, or proposed budget. We focus our detailed response below on the reviewer’s major concerns: (1) potential complications in obtaining robust enough data, (2) potential pitfalls in interpretation, and (3) consideration of alternative hypotheses. In response to the reviewers’ concerns, we have substantively changed the text of the entire proposal. Therefore, we do not indicate specific changes in the text itself, but rather discuss them below.

Reviewer 1 Concerns

Reviewer 1’s primary concern was that the figural enhancement signal was too weak to study the influence of more subtle mechanisms of scene segmentation. We believe this was partly due to our insufficient presentation of preliminary data. To address this concern, we have provided a more detailed description and presentation of the original, and additional, pilot data. We devote an entire section in the Preliminary Data to a presentation of responses from 100 neurons tested with figure-ground stimuli comprised of 4 colors, entitled the robustness of the figural enhancement effect. We show enhancement indices for the population. The figural enhancement for luminance- and color-defined figures is much stronger in V1 than for texture defined figures. Furthermore, the enhancement is stronger in V2 versus V1, often on the order of a two-fold difference (Lee et al. 2002; Marcus and Van Essen 2002). This is primarily why our proposal emphasizes color and stereo stimuli. To also improve clarity, we added a series of 1D firing rate vs. space plots (at different times with respect to stimulus onset) with error bars, as suggested by Reviewer 2.

The second major concern raised by Reviewer 1 was that the proposed experiments were not likely to provide a substantial advancement in our understanding of scene segmentation with respect to previous work (e.g., the principle investigator, Lamme, Paradiso, and other colleagues). This was in part due to the first concern—i.e., the figural enhancement is too weak to provide insight about detailed underlying mechanisms. In addition to more clearly showing the figural enhancement is robust enough, we have:

1. Provided better justification for our proposed research by focusing our background on the scene segmentation mechanisms that the proposed experiments are designed to shed light on. The proposed project is motivated and encouraged by past results with respect to the figure-ground enhancement phenomenon, but we believe that there are many critical and important unresolved questions that our experiments can begin to answer.

2. Reorganized the proposal and redesigned most of our experiments, adding two new crucial experiments (Expt 2.3 and Expt 3.2), and simplifying experiments from the original proposal.

3. At the end of each Aim, we summarized the hypotheses being tested and explained why existing data fails to satisfactorily resolve these hypotheses.

The final concern raised by Reviewer 1 is our frequent use of dense spatial sampling method (of RF w/ respect to stimulus location) which requires 600-900 trials to study each single neuron. We devote a section in Preliminary Data to emphasize the importance and feasibility of this method. One major objective of our proposed project is to test specific mechanisms of scene segmentation. The spatiotemporal response profile is critical to reveal evidence supporting mechanisms such as nonlinear diffusion and boundary sharpening, which are functions of space and time. Therefore, we remain committed spatial sampling, but nonetheless reduce the number of experiments that rely on the technique.
Reviewer 2 Concerns

Reviewer 2’s primary concern is our uncritical acceptance that the figural enhancement represents the outcome of image segmentation. Our working hypothesis is that response properties that arise from figure-ground type stimuli are evidence of segmentation mechanisms. A sharp discontinuity in responses between regions (i.e., a border) support the idea that nonlinear diffusion or similar processes are occurring in V1 and V2. Because nonlinear diffusion or similar processes are a necessary step in the majority of computer vision segmentation algorithms, we suggest that this behavior is linked to image segmentation. The enhancement signal itself is only indirect evidence of image segmentation. Image segmentation does not necessarily highlight any particular partitioned region. This signal likely requires an additional process that operates in conjunction with segmentation. So although the enhancement reflects that global segmentation has taken place, it does not explicitly represent the outcome of segmentation.

Reviewer 2 also (as well as Reviewer 1) raises concern with our hypotheses with respect to attention. Specifically, that the figural enhancement is a result of attentional amplification of segmented regions. In the previous proposal and in the revision, we do note this is one of many possible explanations of the figural enhancement. In our revised proposal, we test three potential mechanisms of the enhancement effect: center-surround contrast (Expt 1.2 and Expt 1.3), attention (Expt 3.2) and figure-ground organization (Expt 3.3).

The final major concern by Reviewer 2 is that we limited our hypotheses for segmentation mechanisms to be based on the figural enhancement signal. We did not propose to test alternative hypothesis, specifically, Wolf Singer’s binding-by-synchrony theory. Even though experiments are routinely done with multiple electrodes, we were hesitant to introduce this additional dimension of tests. This is partly due to the additional pitfalls that multi-electrode experiments introduce. In addition, Lamme and Spekreijse (1998) had published a Nature paper explicitly refuting the binding-by-synchrony hypothesis for figure-ground experiments. We were concerned about devoting resources based on a foundation consisting of negative results thus far. However, Lamme and Spekreijse’s experiment was by no means definitive and leaves many questions open. In our revision, we do devote an experiment based on multi-electrode recordings that in part addresses Singer’s hypothesis, but mainly focuses on testing cooperative mechanisms generally related to scene segmentation. Specifically we are testing for cooperative behavior that is analogous to fundamental properties found in graph cut, a very successful image segmentation method used in computer vision. Although in principle this method is similar to feature binding, the correlation analogy is not a label of bound elements, but merely a reflection of cooperative processing. In our pilot experiments for 2-8 neurons responding to color or stereo surfaces thus far, we observe significant stimulus-dependent spike correlation indicating cooperative computations are occurring (see cooperative computation in Preliminary Data). Rather than forcing a two-alternative test of neuronal synchrony versus figural enhancement based on firing rates, we consider each phenomenon (as well as center-surround antagonism) as potential components of scene segmentation.

One additional minor concern raised by Reviewer 2 relates to the phenomenon of color-induced stereopsis (chromostereopsis), and the potential ambiguity in interpreting color-based figural enhancement. We address this issue of a potential bias (blue appears further away relative to other colors, especially red, due increased refraction) in multiple ways. First, we always test each type of stimulus rendered in complementary colors. We include conditions where the figure is green with a red background, and where the figure is red with a green background. In our pilot experiments, figural enhancement effect is observed in both conditions, suggesting that other factors for figure-ground organization (such as shape, convexity, and size) play a substantially stronger role than chromostereopsis.

Reviewer 3 Concerns

Reviewer 3 (as well as Reviewer 2) was concerned about using response latency to derive the flow of information within and across different visual areas. Reviewer 3 suggests that delays in response latencies are not necessarily proof that feedback circuits are involved. Nonetheless, multiple studies do commonly make such interpretations (Bair et al. 2003; Rossi et al. 2001, Smith et al. 2006; Zipser et al. 1996). Inter-areal feedback circuits produce relatively constant delays, whereas intracortical propagation will produce delays that increase with cortical distance. In our pilot experiments, we observed many interesting timing effects and believe that
experiments in Aim 1 will provide insightful data about intra-areal and inter-areal interaction with respect to segmentation mechanisms. We nevertheless agree that response latency differences can only be suggestive and must be carefully interpreted. They do not definitively elucidate hierarchical interaction between V1 and V2. More definitive evidence could come from deactivation experiments which are beyond the scope of the current proposal, but could be a future line of research. We are more concerned with describing cortical scene segmentation mechanisms on an algorithmic level, rather describing specific circuitry. Our intention is not to use response latency data to support or refute feedback. In our view, both feedback and horizontal recurrent processing are likely involved in any contextual computation. We have redirected the description of our experiments testing segmentation mechanisms with a more functional and computational perspective. The stated objectives of our proposal are to evaluate:

1) Nonlinear diffusion and boundary sharpening and their relationship to center-surround contrast and cooperative mechanisms (Aims 1 and 2).

2) The segmentation hypothesis – whether figural enhancement is based on a global segmentation percept or based on a variety of interacting cues (Aim 2).

3) The source of figural enhancement – attention and/or figure-ground organization (Aim 3).

We have attempted to address all of the reviewers’ concerns in our revised proposal, which in the process has become greatly improved. We thank the diligent effort and insightful critiques of the reviewers.
Cortical mechanisms of visual scene segmentation

Research Plan

A. Research Aims

Visual scene segmentation—the partitioning and parsing of a visual image into different coherent parts—is a fundamental process in perceptual organization that facilitates object recognition, scene understanding, and the perception of visible surfaces and shapes. Computational research suggests that robust segmentation cannot be achieved with bottom-up edge detection alone, but rather has to incorporate global contextual constraints about regions, surfaces and figures.

The objective of the proposed research is to test whether mechanisms suggested by computational vision algorithms of scene segmentation are implemented in part in the early visual cortex (V1, V2). Our premise is that (1) the early visual cortical areas are engaged in visual interpretation and perceptual organization, rather than simply extracting or representing simple local features (see Lee and Mumford 2003) and (2) insights from computational vision research can guide neurophysiological study of the visual system from a functional perspective (see Lee and Yuille 2006 for a review). Some common mechanistic themes have emerged from the more successful computer vision segmentation algorithms, namely, nonlinear diffusion, boundary sharpening, and implied rapid and dynamic functional connections. We will explicitly evaluate evidence for these mechanisms.

There are three key findings in the early visual cortex that might reflect segmentation-related mechanisms that are sensitive to global context. These are Singer’s (Gray et al. 1989) synchrony signal, Lamme’s (1995) figural enhancement signal, and von der Heydt’s (Zhou et al. 2000) border-ownership signal. While there have been a number of follow-up studies related to these discoveries over the years, many issues remain unresolved. In particular, figural enhancement signals have been associated with various perceptual and mental constructs – figure-ground segregation, perceptual saliency, scene segmentation, surface information, attention, and awareness. Synchrony has been associated with both feature binding and attention. While all these constructs may be related, they could have distinct meanings and implications for cortical segmentation mechanisms. Furthermore, the relationship among these three signals has not been clarified. In the following specific aims, we propose experiments designed to determine the nature of these signals and how they might reveal the underlying cortical mechanisms of scene segmentation. As a consequence we hope to connect the various observed neural phenomena into a unified framework.

Aim 1. Spatial response profile and center-surround antagonism

The key feature of Lamme’s (1995) figural enhancement signal is that the uniform enhancement of neural activities inside a texture figure diminishes abruptly at the border between the figure and background (see Figure 1). The classic center-surround mechanism for detecting orientation contrast, often modeled as iso-orientation suppression in V1, is thought not to be able to produce such an enhancement profile. Thus, the figural enhancement effect has been attributed to feedback of higher order processes. We propose to monitor the temporal evolution of V1 and V2 neuronal responses when stimulated by different parts of a figure in a background. We will test the hypothesis that neural responses within a figural surface in general, and enhancement responses in particular, exhibit a nonlinear diffusion phenomenon, accompanied by a simultaneous sharpening of the boundary representation – a common theme in modern computational vision algorithms. Experiments are proposed to compare the figural enhancement response and the classic center-surround contrast phenomena in individual neurons to ascertain a possible causal relationship.

Aim 2. Rules and logics of segmentation mechanisms

Traditional surround suppression, implemented via linear models, has coarse spatial resolution. The sharp spatial profile of the figural enhancement effect suggests that a more subtle and sophisticated computation mechanism must be at work. We propose to test the hypothesis that this figural enhancement signal is a reflection of a global segmentation based on surface or region, rather than based on similarity in bottom-up cues. Experiments are proposed to evaluate the importance of a contrast gradient border in the surround, the importance of global percept over local uniformity of cues, and the nonlinear interaction of image cues governing the enhancement responses and functional connectivity among neurons.
**Aim 3. Sources of the figural enhancement effect**

Segmentation mechanisms partition a scene into regions and surfaces, but do not necessarily bestow relative importance to segmented regions. We propose the involvement of an additional process in generating the figural enhancement signal, which has been shown to be correlated with perceptual saliency. Experiments are proposed to evaluate the contribution of attention and figure-ground organization to the figural enhancement effect. Novel stimuli will be used, which allow the dissociation of local border structures, global figural percept and foreground-background relationship between surfaces. The causal relationship between border-ownership signals and figural enhancement signals will be studied by analyzing functional connectivity among neurons.

**B. Background and Significance:**

Scene segmentation is an important task in visual inference that requires high spatial resolution to represent region surfaces and boundaries. Early visual areas such as V1 and V2 with neurons with small receptive fields (RF) retinotopically organized are suitable for supporting such a computation. There is ample neurophysiological evidence implicating V1 in the detection and representation of oriented edges defined by luminance, color and texture contrast (Hubel & Wiesel 1978, Connor et al. 1997; Lee et al. 1998, Grosof et al. 1993; Sheth et al. 1996; Kapadia et al. 2000; Lee & Nguyen 2001). Horizontal collaterals can facilitate linking oriented segments into contours (Ts’o et al. 1986; Ts’o & Gilbert 1988; Gilbert et al. 1996; Angelucci et al. 2002). V2 neurons, meanwhile, appear to encode contours of a more abstract nature, with a greater percentage of the neurons responding to subjective contours (von der Heydt et al. 1984; von der Heydt and Peterhans 1989; Peterhans and von der Heydt 1989).

**Computational algorithms of scene segmentation**

Computational vision research, however, suggests that mechanisms of local edge detection by themselves are insufficient to produce robust scene segmentation. To date, there are three classes of computational algorithms that have been shown to be relatively successful in scene segmentation: (1) Markov random field (e.g. Geman and Geman 1984, Lee, Mumford and Yuille 1992), (2) graph-cut (e.g. Shi and Malik 1998), and (3) region competition (Zhu, Lee and Yuille 1995, Zhu and Yuille 1996, Tü and Zhu 2002, Tu et al. 2005).

In Markov random fields (Geman & Geman 1984, Blake & Zisserman 1987, Koch et al. 1989, Lee et al. 1992, see Lee and Yuille 2006 for review), surface inference and boundary detection are interactive and concurrent. Surface inference uses a diffusion mechanism to propagate information from locations of strongest surface evidence—e.g., cue contrast borders (color, texture, luminance)—to all locations within the surface. This process can propagate global contextual information about surfaces to improve local boundary detection. The predictions of this mechanism are: (1) spreading and ‘homogenization’ of surface signals within each region, (2) gradual accentuation of surface discontinuities, (3) gradual sharpening of boundaries, (4) blocking of propagating surface signals across boundaries (Lee 1995).

The second class of algorithms, called graph cut (Shi and Malik 1998, Sharon et al. 2001), incorporates global scene information by specifying the affinity between pairs of image pixels as a function of the input image. The procedure looks for reasonable “cuts” that partition the scene into compact and coherent regions based on the affinity matrix. Graph cut methods utilize a richer set of global constraints (defined by affinities or connections) than those implemented by Markov models. The connections can extend spatially indefinitely, although spatial proximity is typically an important factor in defining pixel pair affinity. Connection strengths are also a function of similarities in basic cues, which can include Gestalt grouping rules. The implementation of graph cut algorithms still requires propagation of constraints to reach a global solution (Tolliver and Miller 2005). Therefore, graph cutting may also have a property similar to diffusion as described in Markov random fields.

Graph cut procedures can also be viewed from the perspective of region grouping or feature binding. Therefore, it corresponds closely with one particular neurally motivated theory: the binding-by-synchrony theory (von der Malsburg 1981). Neurons that respond to features of one object fire their spikes at the same time (synchronize), but neurons responding to features of different objects do not. The synchronization of spikes among neurons within an object parallels the affinity or connection strengths between pixels within a region in graph cutting.

The third class of algorithms, called region competition, can be viewed as an extension of the first two classes described above. Region competition allows more advanced priors and higher order models of regions...
and objects to be specified. This procedure is implemented within an hierarchy to guide the segmentation process using recurrent constraints (Zhu et al. 1995, Tu et al. 2005). The priors at higher levels of the hierarchy are used to explain lower level representations (Mumford 1992). The first two classes of algorithms therefore can be seen as bottom-up mechanisms to generate good proposals which can then be evaluated and by the high level models.

All three classes of algorithms emphasize the interaction between surface inference (or region grouping) with boundary detection, and to various extents incorporate cooperative recurrent computations (see Lee and Yuille 2006, for a review). The surface grouping process propagates global information so that boundary detection no longer depends solely on local edge detection. Therefore, if these algorithms are predictive of cortical mechanisms of scene segmentation, neural responses in early visual cortex should exhibit the mechanistic features embodied by these algorithms. Cortical responses over time should demonstrate nonlinear diffusion (e.g., filling in; Sasaki and Watanabe 2004), boundary sharpening, and boundary blocking of surface responses. In addition, there should be changes in functional connectivity among cortical neurons based on cue-based and global segmentation rules. The primary objective of the proposed research is to evaluate evidence of these common mechanistic features to build a solid foundation for further investigation of the neural mechanisms responsible for generating them.

**Neural representations of scene segmentation**

In addition to the neurophysiological work on responses to edges and edge continuation as discussed earlier, there is a wealth of neurophysiological findings on how global segmentation could be represented in the visual cortex. There is empirical evidence that supports von der Malsburg’s binding-by-synchrony (Eckhorn et al. 1988; Gray et al. 1989; Singer and Gray 1995; Castelo-Branco et al. 2000), but there is also evidence refuting it in its broadest interpretation (Lamme and Spekreijse 1998, Thiele and Stoner 2003; Roelfsema et al. 2004; Palanca and DeAngelis 2005). Currently, the validity of the evidence is still highly debated (Shadlen and Movshon 1999; Singer 1999).

Most supporting evidence comes from dynamic coherent contours (Gray et al. 1989, Samonds et al. 2006a). Findings of synchrony are often accompanied by observation of stimulus-dependent gamma oscillation (Castelo-Branco et al. 2000, Kreiter and Singer 1996; Samonds and Bonds 2005) as well as attention (Fries et al. 2001). It seems that there should be a natural and immediate relationship between these response phenomena (synchrony, gamma oscillation) and figure-ground enhancement phenomenon (but see Lamme and Spekreijse 1998). The mutual interaction of neurons within a transitive network could produce short-term increases in coupling strengths between neurons within the same surface, which could naturally lead to synchrony and oscillation. This could serve as a very effective mechanism and/or amplification for associating ('binding') elements belonging to the same surface.

Lamme (1995) proposed an alternative hypothesis based on his observation that neurons respond more when their RFs are located inside a compact “figure” than when they are located in a “background”. The local RF stimuli for figure and background conditions were identical. The enhancement responses of the neurons at different locations within the figure were found to be roughly uniform (even when the figure is 4-6 times larger than the receptive field). Lamme found that the enhancement abruptly terminated at the border or surface discontinuity. He suggested that this enhancement signal can be used to label a “global segmentation representation”. This behavior was consistent among a variety of cues (Zipser et al. 1996) including luminance (Lee et al. 1998, Rossi et al. 2001) and more complex figures embedded in a field of distractors (Lee et al. 2002). In addition, the enhancement is correlated with the perceptual saliency of an object based on the behavioral performance of monkeys (Lee et al. 2002). Although the enhancement signal in texture figures was subsequently found to be rather weak in V1 (roughly 15%) in later studies (Lee et al. 1998, Rossi et al. 2001), it has been used as a springboard for a series of important studies (Zipser et al. 1996, Lamme and Spekreijse 1998, Super et al. 2001) arguing for a role of this signal in figure-ground segregation (Lamme 1995), surface representation (Zipser et al. 1996), texture segmentation (Lamme and Spekreijse 1998), object-based attention (Roeselfema et al. 1998), working memory (Super et al. 2001), and visual awareness (Super et al. 2003). In addition, this signal has been found to be much more robust in more salient cues than texture such as luminance and color (Lee et al. 1998; Rossi et al. 2001; see Preliminary Results below).

The fact that this signal has been associated with so many perceptual and mental constructs raises many questions. While they are related, they are not equivalent. Which one of these constructs is necessary and suf-
Figure 1: Population averaged responses to texture figures in a contrasting background. Both texture images above were presented at spatially varying locations relative to the neuron’s receptive field, with the image shifted horizontally after each presentation. The cells effectively sampled the image at 0.5° intervals. Neural responses to both images were spatio-temporally aligned and summed to give the plot on the right, which shows an initially uniform response yielding to a response with figural enhancement about 80 ms after stimulus onset. The X-axis indicates the location of the RF relative to the image, with 0 at the center of the figure, whose sides were 4° wide. From Lee et al. (1998).

ficient for generating the enhancement signal has not been definitively tested. In addition, while much work has been devoted to establishing that the enhancement arises from extrastriate cortex and depends on awareness and consciousness (Super et al. 2003), no experiments have been carried out to elucidate the mechanisms underlying its generation, beyond suggesting that feedback is involved. The connection of this phenomenon with the center-surround contrast mechanisms well known in V1 and other visual areas has not been carefully considered, beyond dismissing the latter as a null hypothesis. The second objective of the proposed research is clarify the the functional/perceptual status of this signal and to establish the extent that this signal depends on global segmentation and figure-ground organization, and to investigate the various possible mechanisms underlying its generation.

It should be recognized that scene segmentation only implies the partitioning of a scene into distinct regions, and does not by itself bestow importance (relative saliency) to the segments. The enhancement signal would require an additional computation. What is this additional process? Lamme (1995) only suggested feedback may play a role in generating this signal but proposed no specific mechanisms. Here we will discuss three major proposals, which are by no means exhaustive. They serve as a starting point of our investigation.

The first proposal, usually considered as a null hypothesis, is that the enhancement is generated by bottom-up automatic center-surround contrast mechanisms. This contrast mechanism can work in various cue domains. For example, the ON-center-OFF-surround mechanism in retinal ganglion cells operate on luminance. In V1, it has been shown to operate in the orientation domain—i.e., iso-orientation suppression (Maffei and Fiorentini 1976, Li and Li 1994; Knierim and Van Essen 1992). This center-surround interaction is thought to be limited in spatial extent, and was viewed as a orientation contrast detector. Computationally, lateral inhibition can also be conceptualized to serve the function of ‘normalization’ (e.g. Schwartz and Simoncelli 2001). Recent evidence suggests the spatial extent of the surround influence might be more extensive (Cavanaugh et al. 2002) based on grating stimuli, and based on luminance contrast modulation (MacEvoy et al. 1998, Lee et al. 1998, Rossi et al. 2001). However, all these center-surround studies were not explicitly motivated by considerations of scene segmentation mechanisms, and were not done in conjunction with an assessment of the spatiotemporal response profile of the neurons to the image. The working model of center-surround antagonism is lateral inhibition, which is essentially a linear filtering operation. However, much evidence suggests it to be nonlinear and potentially more complex (Bair et al 2003, Webb et al. 2005). Instead of dismissing the classical center-surround contrast mechanism as a null hypothesis (Lamme 1995), we propose to embrace it as a possible contributing mechanism. Experiments in Aim 1 are designed to establish a relationship between the surround contrast phenomena and the spatial response profile that is indicative of global segmentation. Ex-
periments in Aim 2 are designed to demonstrate that the center-surround contrast mechanism is not linear, but sensitive to subtle image structure in the surround, which is responsible for bringing out segmentation-based responses.

The second proposal is that this enhancement might be a highlighting signal generated by a bias signal from the extrastriate cortex that serves to color the region and facilitate a robust segmentation (Lee et al. 1998, Lee et al. 2002). This highlighting signal could emerge from interaction between low level representations and higher order models as suggested by region competition (Lee et al. 1998, Tu et al. 2005), This might also be related to attentional mechanisms in the visual cortex, which are modeled as biased competition (Desimone & Duncan 1995, Deco and Lee 2004). Even though monkeys are performing a fixation task in Aims 1 and 2, it is possible the salient bottom-up contrast representation of the figure stimulus attracts and engages attentional mechanisms reflexively. Attention-based feedback enhancement can further facilitate processing of the features in the salient region represented in V1. Attentional selection on visual processing has been intensely studied in extrastriate cortex (Moran & Desimone 1985; Luck et al. 1997; Maunsell 1995; Reynolds et al. 2000), and has been observed in V1 particularly when the scene has many competing elements (Motter 1993, Lee et al. 2002, Ito & Gilbert 1999; Hupe et al. 1998; Roelfsema et al. 1998; McAdams & Reid 2005). To what extent attention is necessary and involved automatically in the segmentation process (e.g., Boynton 2005) will be tested in Aim 3.

The third proposed mechanism for the figural enhancement is the propagation of border signals (Lee et al. 1998). These signals could propagate both inward and outward from the borders. In the case of a square or disk, the signals inside the figure might become more concentrated because each point inside the surface is closer to all the boundary points than each point outside. Thus propagation of border signals only depends on segmentation. However the propagation of border signals could also depend on figure-ground organization. Border signals might only propagate inward toward the surface that ‘owns’ the border. This would require that some level of border-ownership has been established at this early stage of visual processing (e.g., Zhou et al. 2000). Earlier experiments always used squares and disks as the figure in stimuli. The findings therefore cannot be used to distinguish these two possibilities. Experiments in Aims 1 and 2 evaluate the propagation hypothesis, and experiments in Aim 3 are designed to distinguish whether the effect depends solely on local border structures (e.g., enclosure) or on figure-ground organization (i.e., border ownership).

**Representation of surfaces and regions**

We have thus far used the concepts "surface" and "region" interchangeably, but they are distinct technically. Surface, strictly speaking, should have a 3D meaning or connotation, or what Marr (1982) called the 2.5D sketch, while image regions could arise from albedo or texture difference within a single surface. While a number of studies (Rossi and Paradiso 1999, MacEvoy et al. 1998) suggest that V1 neurons are correlated with brightness perception, which is an aspect of surface perception, there is actually no direct evidence that V1 is involved in surface representation. Because all experiments on figural enhancement signals and neural synchrony are based on 2D image cues (except one experiment in Zipser et al. 1996), it is difficult to distinguish whether the neural phenomena observed in V1 depend on 2D image cue segmentation or 3D surface segmentation. To resolve this issue more definitively, we propose to carry out spatial sampling experiment on 3D images defined by random dot stereograms in Aim 1 (Experiment 1.3), as well as to investigate whether figures defined by a non-uniform cue can induce the segmentation effect in Aim 2 (Experiment 2.2).

On the other hand, there is stronger evidence that implicate V2 in surface computation and representation. First, Von der Heydt et al. (2000) found that V2 neurons and very few V1 neurons are sensitive to stereo edges and exhibit border-ownership signals, which is clearly a property of surface representation. Second, Bakin et al. (2000) observed that V2 neurons but not V1 are sensitive to a da Vinci stereo effect, a stereo perception induced by the interpretation that there is a surface occluding another. Third, Cumming et al. (1999) and Thomas et al. (2002) showed that V2 but not V1 neurons are sensitive to relative disparity, which could be a substrate for surface discontinuity detection. Fourth, the perception of some color and brightness illusions such as the Craik-O’Brian-Cornsweet illusion (Hung et al. 2001, Roe et al. 2005) can be observed primarily in V2 but not in V1. Interestingly, Lee et al. (2002) found that the enhancement response for a shape from shading oddball in a cluttered scene appeared in V2 preattentively, but not in V1. But then, only after the monkeys were trained to detect the oddball stimulus, the signals began to appear in V1. This could be taken to suggest that V2 might be the primary site for processing these shape from shading interpretations and performing...
the pop-out segmentation based on surface representation. Then V2 feeds back to V1 to color the segmented object.

We conjecture that segmentation computation embodied in V1 likely operates at the cue level. Segmentation computation in V2 is likely based on surface representation. Significant interaction of V2 with the image representation in V1 is expected. Therefore, it is important to investigate the figural enhancement signals and the functional connectivity in both visual areas (V1 and V2) to evaluate this conjecture. The proposed research will provide valuable new information to understand the distinction in the functional roles of V1 and V2 in scene segmentation.

**Summary and Significance**

The proposed study is motivated by insights provided by computer vision research on the possible neural mechanisms for scene segmentation. The first objective of the proposed research is to evaluate the common mechanistic features that have been found in several effective segmentation algorithms, namely, (1) contrast-dependent nonlinear diffusion, (2) interactions between surface and boundary computations, and (3) dynamic functional connectivity between nodes in a network as a function of input images. Experiments in Aims 1 and 2 are designed to pursue this objective.

The experiments in this proposal will utilize several major neural phenomena that have been proposed to be related to scene segmentation: figural enhancement, spike synchrony, and border-ownership enhancement. Although there are many studies on these phenomena because of their potential significance and implications, these studies have not established definitively the extent these signals really depend on global segmentation, attention, and figure-ground organization. The second objective of the proposed research is to determine the functional status of these signals by dissociating various physical and perceptual factors. Experiments in Aim 3 are designed to pursue this objective.

These three neural phenomena have been seen as competing and often conflicting hypotheses. In addition, classic center-surround contrast is often considered as a null hypothesis or dismissed as irrelevant. Our hypothesis is that these four phenomena are connected to one another, each reflecting an aspect of scene segmentation. The third objective of the proposed research is to elucidate the possible relationships among center-surround contrast mechanisms, figural enhancement effect, neural synchrony, and border-ownership signals in the context of region and surface segmentation.

This proposal contains a number of novel experiments which are designed (1) to evaluate neural basis of computational vision algorithms, (2) to understand the precise functional roles of the neural phenomena, and (3) to probe at the connections between the various neural phenomena. The experimental outcome will answer a number of long-standing questions on the relationship between the neural signals and the different perceptual processes, and will help clarify the role of V1 and V2 in scene segmentation. This study is an important step toward a unified understanding of cortical scene segmentation. The theories and mechanisms underlying visual grouping and scene segmentation have very close analogies to other cognitive modalities such as language, reasoning, and categorization. A better understanding of the neural mechanisms of scene segmentation therefore could provide important insights to cognitive functions beyond visual perception. Knowledge gained from the proposed research can also have practical significance for developing scientifically-based diagnostic and rehabilitation programs for the visually impaired, and for advancing the technology of robust computer vision systems.

**C. Preliminary Results**

We have carried out a series of pilot experiments to establish the feasibility of our study, and to address the following specific concerns.

1. **Robustness of the figural enhancement representation.**

   The figural enhancement effect for texture figures typically manifests as an 10-15 percent increase in firing rate for a 4° diameter disk (for RF located between 2-4 degrees in eccentricity) (Lee et al. 1998). There is some question whether the signals are strong enough to carry out the proposed studies of their mechanisms. Earlier research (e.g. Lamme 1995, Zipser et al. 1996, Lee et al. 1998, Marcus Van Essen 2002, Rossi et al. 2002) typically repeated each condition for 10 or 15 trials. This is often sufficient for the effect to reach statistical significance for multi-unit activities (MUA), but not for single unit recording. However, the figural enhancement effect is
considerably stronger for luminance figures (Lee et al. 1998, Rossi et al. 2001) and color figures (Kelly and Lee 2005, Liu et al. 2005). Figure 2 shows the PSTH of a V1 neuron’s response when its receptive field (about 0.7 degrees in spatial extent) is inside a 4 degree square color figure (row 1) and in the same color background (row 2) for the four colors.

Figure 2: The raster plots and peri-stimulus-time histogram (PSTH) of the response of a neuron in the figure/ground conditions for four isoluminant colors. The red figure and red ground conditions are illustrated, with the RF indicated by the ellipse. The stimulus was presented for 450 msec, starting at 0 msec in the plot, while the monkey fixated. Each condition was repeated 30 times.

The difference between the responses in the two rows along each column is the so-called “figural enhancement effect”. It can be observed that along each column, the initial activity of the neuron was the same. A stronger response emerged in the figure conditions at about 70 ms after stimulus onset. Many V1 and V2 neurons respond well only to the borders but not inside the figure. In our pilot study, among 100 neurons that did respond inside the figural surface (when tested with color stimuli), 6% of the cells showed significant enhancement for one color, 16% for two, 27% for three, 51% for all four colors.

To demonstrate the robust strength of the enhancement effect in color stimuli obtained with 20-30 trials per condition, we show in Figure 3 scatter plots of the figural enhancement indices (F-G)/(F+G) for color figure-ground stimuli, where F is the response of the cell inside a uniform color figure, computed within the 100-400 msec window after stimulus onset, and G is the response of the cell in the same time window to the background of the same color, outside a figure of complementary color. The mean figural enhancement indices in this population of V1 neurons are quite strong, equal to 0.26, 0.36, 0.24, 0.38 for red, green, yellow and blue figures respectively. As a point of reference, the mean figural enhancement index in V1 for texture stimuli (Lee et al. 1998) is about 0.1, which was sufficiently strong to be used in a number of important studies (e.g Super et al. 2001, Lamme and Spekreijse 1998). Although the magnitude of the enhancement can vary depending on stimulus cue dimension, its prevalence and strength suggest that it can be suitably leveraged in our proposed

Figure 3: Distribution of the figural enhancement indices for the four colors. Filled circles indicate single unit recordings, open circles indicate multi-unit.
Figure 4: Spatial sampling applied to a pair of complementary stimuli: a red figure in a green background, and a green figure in a red background. Ellipses indicate the RF location of the neuron. On the right is the spatial profile of the neuron’s response to the red figure in green background (red line) and to the green figure in red background (black line) in different 30-msec time windows after stimulus onset. The error bars are the standard error of the spike counts within each 30 ms time window at each spatial location. Location 0 is where the receptive field of the cell overlaps the center of the square. Thus, -2 and 2 are the locations of the borders.

2. Feasibility of the spatial sampling paradigm.

Spatial sampling (Experiment 1.3) to estimate how a neuron responds to different parts of the image is important for testing our hypotheses about mechanisms of segmentation. In order to obtain the spatiotemporal profiles of a neuron’s response to a pair of complementary stimuli as proposed, over 600 trials from a single neuron in the course of 1 hour in each recording session. While labor intensive, this technique is feasible as we can routinely maintain a single unit for over 2000 trials or over two hours of recording. We have used this dense sampling paradigm in earlier studies (Lee et al. 1998, Lee and Nguyen 2001, Liu et al. 2005). For the proposed study, we will use multi-electrode recording techniques to increase the yield of each recording session. This will have the added benefit of providing multi-electrode data to analyze how neurons work together to encode boundaries and surfaces.

Figure 4 shows an example of a V1 neuron’s spatiotemporal response profile to a pair of complementary red/green figure-ground stimuli (with a 4° color square in a contrasting background), obtained using spatial sampling. This neuron responded strongly to the onset of the uniform red stimulus in its receptive field, but not to the onset of the uniform green stimulus (see the responses in the 30-60 msec window). However, over time the neuron started to respond more strongly inside the green figure, overtaking the response to the red background. The response inside the green figure appeared to propagate in from the border, exhibiting a gradual ‘filling-in’ of response. At the end, the response inside both figures is stronger than the response to the background.

Sometimes, the figural enhancement response might not be strong enough to overcome the color tuning of the neurons. In that case, the responses to the complementary stimulus pairs can be added together at each spatial location at each point in time, as shown in Figure ??c. A positive figural enhancement response is marked by an elevated response profile inside the figure (in the region of offset between -2° and 2°) relative to the background (from 2° to 4°), as shown in this cell. It is worth noting that the spatiotemporal profile of this neuron (and of most neurons) shows a strong response to the border of the figure. For other cells with stronger color tuning, the figural enhancement response might not overcome the color tuning effect. See Lee and Yuille (2006) for other examples.

Finally, it can be observed that in the later part of its response, the neuron began to respond better to the right border (at the +2 location) than the left border (at the -2 location) for stimuli, regardless of the polarity of the color contrast. This effect is consistent across a large window of response (150-400 msec), and is consistent within each of the 30 msec running windows shown in Figure 5. This might be an observation of what von der Heydt and colleagues (Zhou et al. 2000) called the border-ownership signal. In this particular V1 cell, the border-ownership signal emerged after the figural enhancement response, suggesting that there might be some interaction between these two signals, a subject we will pursue further in Aim 3.

This example demonstrates that the spatial sampling paradigm is feasible and capable of providing a
Earlier studies on the figural enhancement effect and border-ownership effect were based on square figures, which do not allow the dissociation of the effect due to the local convexity of border structures from the effect of a global figural percept. For the proposed study in Aim 3 to more rigorously test the dependence of these two signals on figure-ground organization and to explore their plausible causal relationship, we devised the nut-and-bolt stimuli (Figure 7). These are rendered in color and as random dot stereograms to allow the

Figure 5: 3D plot of the neuron’s spatiotemporal responses to the red figure condition (a), green figure condition (b) and their combination (c). Location 0 is where the receptive field of the cell overlaps the center of the square. Thus, -2 and 2 are the locations of the borders.

Figure 6: The behaviors of two cells during the asynchronous update paradigm. The entire screen was turned blue at -350 msec, followed by the appearance of a yellow background surrounding the RF at 0 msec. The blue disk made visible was 6 degree in diameter. (a) shows a neuron that had a transient response to the blue screen onset, but gave a stronger and more sustained response to the surround update. (c) shows a neuron that exhibited little response to the blue screen onset, but responded in a sustained manner to the appearance of the yellow background outside the receptive field. (b) and (d) are the responses of the two neurons when the surround update did not take place.

wealth of information and insight. This would not be observable if only the responses of the neurons at the center of the figure are studied, as in most previous studies.

3. Effectiveness of the asynchronous update paradigm in isolating the surround effect.

In Figures 4 and 5 we observed a diffusion or “filling-in” effect in the green figure but not the red figure. One possible reason is that when an entire stimulus is presented on the screen simultaneously, the response to direct RF stimulation (the appearance of red color in the RF) overlaps the response due to the surround contrast. To test the hypothesis that the color enhancement signal propagates in from the border, we use an asynchronous update paradigm to dissociate the two effects. Here, we present the RF stimulus first and then, at a later time, change the surround. Figure 6 shows how two cells responded when first the entire screen turned blue, and then later (at 0 msec in the PSTH plot), the surround color was changed to yellow, making visible a blue disk 6° in diameter (in a yellow background) over the receptive field (less than 0.7° in size). Remarkably, many cells did not responded to the uniform color RF stimulation, and those which did tended to give a transient response. On the other hand, the appearance of the surround contrast elicited a strong and robust response in the cells. Figure 6 shows the responses of two typical cells. In this experiment, the surround update occurred at 0 msec, and the change in RF stimulation occurred at -350 msec. The differences between (a) and (b), or between (c) and (d) are due to the surround stimulus. The responses to the surround update are often strong and observable, and it is feasible to estimate this timing of this response onset quite accurately using standard methods (Bair et al. 2002).

4. Figural enhancement effect in nut-and-bolt stimuli

Earlier studies on the figural enhancement effect and border-ownership effect were based on square figures, which do not allow the dissociation of the effect due to the local convexity of border structures from the effect of a global figural percept. For the proposed study in Aim 3 to more rigorously test the dependence of these two signals on figure-ground organization and to explore their plausible causal relationship, we devised the nut-and-bolt stimuli (Figure 7). These are rendered in color and as random dot stereograms to allow the
dissociation of many figure-ground factors. Since the figural enhancement effect has never been demonstrated in such stimuli before, we performed a pilot study in 13 V1 neurons and 8 V2 neurons. Spatial sampling was applied to examine their responses to different locations of the nut and bolt complementary pairs. The results for V1 and V2 neurons are shown in Figure 7 for location 2, where the stimuli are identical within a 9° diameter window for the nut and bolt conditions. The average response for V2 neurons demonstrated a robust difference between the “nut” (figure) and “bolt” (ground) conditions. V1 neurons showed a small but inconsistent effect. The enhancement effect in the population of all V1 and V2 neurons is shown in the pair of histograms to the right of Figure 7. Some V1 neurons show significant enhancement (shaded bars), and the average effect is slightly greater than one. In V2, more neurons demonstrated significant enhancement.

While this pattern of response could still arise from the neuron’s sensitivity to the border structures or stimulus very far away from the receptive field, this positive result provides a reasonable starting point for the proposed experiment. Advanced tests will be performed to determine whether the responses are controlled by the physical surface ordering relationship or the perceptual choice of the monkey.

5. Cooperative computation in 3D surfaces.

To investigate scene segmentation as a surface-partitioning process, it is important to use 3D stereo stimuli (dynamic random dot stereogram) in our study (Aim 1-3). To investigate cooperative computation among neurons as a mechanism for generating segmentation, it is important to study the functional connectivity of the neurons within the surfaces and between surfaces (Aim 2). While the methods for studying stereo sensitivity and the functional connectivity are by now standard, we have not published papers on these issues. To demonstrate our ability in carrying out stereo experiments and correlation analysis, we performed a pilot experiment to investigate stimulus-dependent correlation among neurons in response to a DRDS-defined 3D surface to evaluate evidence of cooperative computation within the surface (Samonds et al. 2006b). This is a first step to evaluate whether the dynamic change in functional connectivity can serve as a mechanism in segmentation as proposed in Experiment 2.3.

A planar surface is projected over the receptive fields of a pair of disparity-tuned neurons. The surface is displayed at eleven different depth planes using dynamic random dot stereograms for 1 second, repeated sixty times. For each individual neuron, the disparity tuning curve was computed based on firing rates. We compared the firing rate tuning to correlation between two neurons with different RF locations. Out of 26 pairs of disparity tuned neurons, 5 pairs exhibited significant correlation peaks (> 3 SD, and area under peak \( p < 0.05 \) bootstrapped). Peaks had similar heights and bandwidths as correlation measured to low contrast gratings in anesthetized macaques (Kohn and Smith, 2005), which is expected for 25% sparse DRDS. For higher contrast stimuli, we expect higher and narrower peaks (Kohn and Smith, 2005).

For higher contrast stimuli, we expect higher and narrower peaks (Kohn and Smith, 2005). Pairs functionally connected all had similar disparity tuning (within 0.1°). When we examined the correlation coefficient versus time (area under the peak), interestingly, robust disparity tuning based on the correlation (green) emerged earlier than the disparity tuning based on firing rates (see Figure 8, right column). This suggests disparity tuned neurons in V1 are coupled together and their mutual interaction might spatially integrate disparity information to improve the global depth estimate for a continuous plane. The functional connectivity and firing rate appear to interact as evidenced by the sharpening of the firing rate tuning curves over time (e.g., Ringach et al. 1997; Menz and Freeman 2003). This finding is interesting because it provides evidence of cooperative computation within a surface. Examining the temporal evolution of firing rates and functional connectivity in different image contexts allows us to gain insight about the underlying cooperative mechanisms for scene segmentation.
Figure 8: The left column shows two examples of pairs of disparity tuning curves (error bars: standard error). Each pair was selected from 4 electrodes recorded simultaneously (inset: RF plots; dotted circle represents DRDS aperture). The center column shows corresponding cross-correlation histograms (CCH; Aertsen et al. 1989) for preferred (dark blue) and non-preferred (light blue) disparities for each pair of cells. The right column shows that correlation coefficients (green; integrated under the CCH from -25 to 25 ms) were disparity dependent (error bars: 95% confidence) soon after stimulus onset when firing rate shows minimal selectivity for disparity. Firing rates gradually became more selective for disparity.

and surface inference.

D. Research Design and Methods

In the following sections, we will first describe the general methods employed in all of the experiments. We then describe three sets of experiments that address, respectively, the three specific aims, and we will conclude with brief comments on the schedule and on critical details of methodology.

Recording Methods: In this project, we will use the standard technique of single-unit recording on awake monkeys with single or multiple electrodes. The general technique of recording, surgical installation of recording chambers, and other general issues on experimental techniques are standard and will be detailed in the Appendix. A protocol covering this study has been approved by the Institutional Animal Care and Use Committee of Carnegie Mellon University, in accord with Public Health Service guidelines for the care and use of laboratory animals.

Recordings will be made transdurally with epoxy-coated tungsten electrodes through a surgically implanted well overlying the operculum of area V1. Neurons will be isolated based on spike multi-dimensional feature clustering methods (utilizing spike height, width, and shape) available in our recording systems, Tucker-Davis’ Open Explorer and Cyberkinetics’ Cerebus; see also Shoham et al. 2003). The eye position signals will be measured using the scleral search coil technique and sampled at 200 Hz during the experimental sessions. V1 and V2 neurons will be recorded from the same wells. Some V2 neurons will be recorded from the narrow surface strip next to the V1/V2 border, where a transition from V1 to V2 is marked by a reversal in the progression of the receptive fields toward and away from the mid-line in the visuotopic map, accompanied by an increase in their size. Other V2 neurons will be reached by advancing the electrodes through V1 into the posterior bank of the lunate sulcus. The transition into V2 will be accompanied by a sudden shift in receptive field location and by an increase in receptive field size. The exact locations of the recording will be further confirmed by structural functional magnetic resonance imaging in our facility.

In each recording session, the receptive field of the cell will first be measured using the minimum response field method (Barlow et al. 1967) or the grating summation method (Cavanaugh et al. 2002). The receptive fields of neurons at the site of these wells will be between 0° to 5° in eccentricity. At those eccentricities, V1
Asynchronous

| Monkey fixates | Target color appears | Background changes color | Target disappears, monkey makes a saccade to a dot |

Synchronous

| Monkey fixates | Target and background appear | Target disappears, monkey makes a saccade to a dot |

Figure 9: Asynchronous and synchronous update paradigms.

behavioral field sizes typically range from 0.5° to 0.8° in diameter, and V2 receptive field size ranges from 0.8° to 2.0°. The layer of the cells will be estimated based on the depth of penetration during slow withdrawal of the electrodes upon the completion of the recording session. Although most of the experiments are designed for single-electrode recording of a single neuron, we will also employ multiple electrode recording techniques to record from 4-8 neurons in each recording session.

**Behavioral Tasks.** The monkeys will be trained to perform three behavioral tasks for the various experiments – fixation, distraction, and figural report. In the *fixation* task, the monkey fixates on a small dot while stimuli are presented on the computer monitor for 400-600 msec, depending on the stimulus presentation paradigms discussed below. When the stimulus presentation is completed, the screen turns gray, and the fixation dot moves to another location on the screen. The monkey is required to make a saccade to the new fixation dot location to get a juice reward. This final step serves solely to keep the monkey alert and awake, and to prevent it from completely ignoring events occurring on the computer display. In the *distraction* task used only in Aim 3 to evaluate the role of attention, the monkeys are required to engage in a demanding task that is designed to draw attention away from the receptive fields of the neurons being tested. In the *figural report* task used in Aim 3, the monkeys are required to report the shape of the figure they perceive in the stimulus display. The specifics of these two tasks are described in detail in Aim 3.

**Stimulus Design.** Visual stimuli used in the study contain simple and complex shapes (squares, disks, nuts and bolts) rendered in different isolated surface cues, color, shading, stereo, texture, brightness. This project will primarily focus on color, shading and stereo cues, but texture and brightness cues will also be studied in some experiments for comparison. The color cues are isoluminant and iso cone-contrast. The RGB settings of these color stimuli will be chosen according to the specifications in De Valois et al. (1997) (see also Wachtler et al. 2003). The stereo cues are rendered in dynamic random dot stereograms (DRDS) (Poggio et al. 1988; Cummings & Parker 1999; Thomas et al. 2002) presented at 8-12 frames per second. Stereo goggles with liquid crystal shutters (CrystalEyes PC; Stereographic Corp., San Rafael, CA), which alternatively transmit left- and right-eye views of the display monitor at a monocular frequency of 60 Hz. Texture stimuli similar to earlier studies (Lamme 1995) will be used, rendered with oriented line segments. A variety of stimuli will be used to address the different issues, and are discussed in detail in each specific Aim.

**Presentation Paradigm.** Two major paradigms for presenting stimuli will be used (see Figure 9). In the *synchronous update* paradigm, the entire test display will be presented simultaneously on the screen. The intertrial screen is gray. In the *asynchronous update* paradigm, designed to dissociate the RF stimulation effect from the surround stimulation effect, the figure and the background are presented at different times. There are three possible configurations. In the *figure* configuration, the entire screen is first changed from gray to one particular cue (e.g. red), then after 300 msec, the stimulus outside the receptive field is changed to a different cue (e.g. green), making a red figure visible over the receptive field location. In the *ground* configuration, the entire screen is first updated to one cue, and then 300 msec later, a figure of a different cue appears at a location at least 6° away from the receptive field. In the *no change* configuration, the second update never materializes. In all three conditions, the RF stimulus is not changed in the second update.

**Aim 1. Spatial response profile and center-surround antagonism**

**Overview:** The experiment in this aim has two primary objectives: (1) elucidate the relationship between a neuron’s spatial response to figure-ground images and its center surround contrast mechanisms; and (2) char-
Experiment 1.1: Surround effects across cues

Rationale. This test serves two purposes. First, it is a screening test to assess the strength of the surround effect as a function of different surface cues. The responses are then used to select appropriate cues for further experiments on individual neurons. Second, as an experiment unto itself, it is designed to assess the relationship between the surround effect and the tuning properties of the neurons, and to measure the degree of cue-invariance of the surround effect in V1 and V2.

Experimental Design and Procedure The monkey will perform the fixation task while stimuli are presented in the asynchronous update paradigm (see Figure 9 and Presentation Paradigm) in the figure configuration. That is, the entire screen is changed first to a cue; then, 300 msec later, the surround stimulus outside the receptive field of the neuron is updated to another cue, making a figure visible over the receptive field. The figure will be a 3° diameter disk, a size selected to be larger than the RF sizes of V1 and V2 neurons at the recorded eccentricities (3-4 times larger than the V1 RFs and 1.5-2 times larger than the V2 RFs). Figure 9 shows an example of a red cue stimulating the receptive field of the cell, with the surround then changing from red to green to make a red figure visible over the receptive field.

In this experiment, stimuli defined in color, brightness, texture, and stereo cues (as shown in columns 1-4 of Figure 10) will be tested. Each stimulus (after the surround update) contains a figure in a background of contrasting cue. Four colors (red, green, yellow, blue), four texture orientations (horizontal, vertical, and 45° obliques), two luminances (black and white), and three disparities (near, zero, far) will be tested. In the disparity case, the surround can be in front of or behind the center disk after the second update, resulting in 6 conditions. As a control, additional black and white ring stimuli of the same diameter will be presented in the second update. Each of the 17 total conditions will be repeated 10-20 times. When this experiment is used only as a screening test for other experiments, it is possible to test only a subset to select a ‘reasonable’ cue, not necessarily the best cue.

Data Analysis and Expected Outcomes Two questions motivate the initial data analysis. 1. Should a recorded neuron be studied in subsequent experiments? 2. Which cue dimension should be chosen for testing this neuron in the subsequent experiments of this and other Aims?

Figure 6 demonstrates the responses of two cells obtained with this experiment’s asynchronous update paradigm. A distinct neural response occurs after the surround update, which, because the V1 RF is typically much less than 1° wide and the disk is 3° wide, can be attributed to the appearance of the surround contrast and not direct cue stimulation on the receptive field. As a control, meanwhile, it is important that the ring stimuli do not elicit a significant response in the neurons. Our pilot data suggest this is usually the case.

We will use the responses 200-300 msec after the first update as the ground condition (G) and the responses 70-170 msec after the second update as the figure condition (F) for calculating the standard figural enhance-
Data will be further analyzed to evaluate the magnitude and onset timing of the surround effect as a function of isolated cue parameters in both visual areas (V1 and V2), using standard methods (e.g. DiCarlo and Maunsell 2005, Lee and Nguyen 2001, Bair et al. 2002). One expected outcome would be that V2 neurons demonstrate the surround effect in a more cue invariant way, while V1 neurons’ surround effect might depend on the neurons’ selectivities for the cue parameters. Another expected outcome would be a difference in the onset timing of the surround effect across cues, and across the two visual areas. These timing differences might be informative, to some extent, about the hierarchical order of the mechanisms as a function of cues, and whether recurrent interaction is chiefly responsible for processing each cue in the two areas.

**Pitfalls and solutions.** A key concern with awake monkey experiments is the potential eye movement or jitter triggered by the abrupt onset of the surround contrast. This concern will be addressed by close monitoring and analysis of the eye movement data. In addition, the ring stimulus, which should elicit no significant response, serves as a control to ensure that the RF is not large enough or moving around enough to receive direct bottom-up stimulation. The use of dynamic random dot stereograms for generating stereo stimuli might cause additional eye movement artifacts, as the convergence or divergence of the eyes can be triggered by the disparity signals (Motter & Poggio 1984, 1990). To address this concern, the fixation spot will be surrounded by a small gray disk at the zero-disparity plane to minimize distortion in fixation.

A second concern is that the appearance of the figure by an abrupt change of the surround can attract top-down attention reflexively, which could be an intensifier or even the sole source of the surround response. We will address this concern in two ways. First, in tests 1.2 and 1.3, we will use the synchronous update paradigm, which should reduce the effect of attention capture. However, one cannot rule out the chance that reflexive attention is involved even in that scenario, and for this reason we will carry out an experiment in Aim 3 that uses a distraction task to assess the influence of attention on the surround signal.

**Experiment 1.2: Center-surround contrast response**

**Rationale.** This experiment is designed to assess the spatial extent of the surround influence on a neuron in a chosen cue dimension (e.g. color or disparity). The objective is to elucidate the relationship between the classical center-surround contrast phenomenon and the spatiotemporal response profile of the neurons to figure-ground images, the subject of study in Experiment 1.3.

**Experimental Design and Procedure.** The monkey performs a fixation task while the stimulus is presented using the synchronous update paradigm (Figure 9), which presents an image statically on the screen for 400 msec. Stimuli (see Figure 10) with a disk figure rendered in one cue (e.g. red) and the background rendered in the complementary cue (e.g. green) will be presented. The disk will be centered on the receptive field of the neuron, and will be defined in two complementary cues (e.g. red vs green) with several diameters: 0.5°, 3°, 4°, 5°, 7°, 9°, 12°, and full screen. The “ground” (G) response will be computed from the full screen stimulus presentation.

**Data Analysis and Expected Outcome.** Data will first be analyzed to determine the spatial extent of the surround effect. We will compute the figural enhancement effect (within 100-400 msec window post-stimulus onset) as a function of disk diameter. When this effect or the raw spike counts are plotted against the disk diameter, the expected outcome is that the surround effect of most of the neurons will decrease with an increase in disk diameter, but the effect on some cells will increase and then decrease, or decrease and then increase, showing the variety of patterns observed in earlier studies (Li and Li 1992, Hawken et al. 2001, Johnson et al. 2001, MacEvoy and Paradiso 2001). We will group these cells into different categories according to the results of this experiment and examine how each group behaves in response to simple figure-ground images in the next experiment.

Data can be further analyzed to assess the onset timing of the enhancement responses by comparing the temporal response of the neuron to disk figures of each particular diameter against its response to the uniform screen condition. Earlier studies suggest that the onset time increases with disk size for luminance stimuli (Rossi et al. 2001) and color stimuli (Liu et al. 2005) but remains unchanged for texture stimuli of different sizes (Zipser et al. 1996). Information about onset timing of responses in stereo stimuli, as well as other cues in V2, would be new, along with comparative information on how the different cues are processed by the...
**Figure 11:** Spatial sampling paradigm. By shifting the figure relative to the RF location (shown as an ellipse), the neuron’s response to different parts of the image can be monitored over time.

different areas.

**Pitfalls and Solutions.** The onset time estimate based on the comparison between two responses is potentially less accurate than the onset time based on a single response, such as what is obtained by the asynchronous update paradigm. By design, both Experiments 1.1 and 1.2 will test disks $3^\circ$ in diameter, which will allow us to calibrate the onset time estimate from the two paradigms.

**Experiment 1.3: Spatiotemporal responses to figure-ground images**

**Rationale.** Experiment 1.3 is designed to assess the spatiotemporal response of the neurons to figure-ground stimuli with two objectives: (1) to evaluate whether the nonlinear diffusion effect and the boundary sharpening effect implicated by computational mechanisms of segmentation could be observed in V1 and V2’s neural activities; (2) to evaluate, together with Experiment 1.2’s findings, the mechanistic relationship between the spatiotemporal response and the center-surround contrast responses.

**Experimental Design and Procedure.** The monkey will perform the fixation task and the stimulus will be presented with the synchronous update paradigm as in Experiment 1.2. A spatial sampling technique will be applied to study the spatiotemporal responses of V1 and V2 neurons to the images depicted in Figure 10: in each trial, a particular part of the image is presented to the receptive field of the cell (Figure 11). Each stimulus contains a single $4^\circ \times 4^\circ$ disk in a contrasting background, as shown in Figure 10. For each neuron, the pair of complementary cues tested in Experiment 1.2 will be used in this experiment.

Over successive trials, the figural disk is placed at 15 different locations relative to the center of the neuron’s receptive field. The spatial sampling is denser ($0.25^\circ$ interval) near the border and coarser away from the border. From each monkey, a small set of neurons, if exhibiting a strong texture surround effect, will be tested with the texture stimuli to calibrate with earlier studies (Lamme 1995, Lee et al. 1998).

For the stereo stimuli, a pair of figure and window stimuli will be tested. For the figure stimulus, the figure is at optimal disparity, while the background is behind it ($-0.3^\circ$ disparity past the optimal disparity). For the window stimulus, the foreground is at optimal disparity, the window is in front, and the monkey fixates at a dot at zero-disparity.

**Data Analysis and Expected Outcomes** The spatial sampling strategy in this experiment allows us to monitor the temporal evolution of the neurons’ responses to each part of the stimuli, allowing us to gain a glimpse of the time course of the underlying computational algorithm. Two questions motivate the initial data analysis. (1) Do the boundary responses sharpen over time? (2) Do the responses in two abutting surfaces become smoother within the surfaces but more sharply different across the region border? These effects, which reflect predictions from computational vision algorithms, are illustrated in Figure 12.

We will call the cells that respond only to borders ‘boundary cells’, and cells that respond inside the figure ‘surface cells’. For the boundary cells, we will evaluate evidence for the boundary contraction effect (Lee 1995) by analyzing the spatial width of their spatial response profiles across the border at different time windows after stimulus onset. The prediction is that the response profile is characterized by a peak at the exact location of a boundary and decays exponentially with distance away from the border. One quantitative measure is to fit the spatial profile with a curve, and use the half-height width of this curve as a measure of the width of the boundary representation in each time window. In one pilot study, we found that the width thus measured exhibit a significant contraction (see Lee and Yuille 2006).

For the surface cells, we will examine evidence for the nonlinear diffusion effect (Grossberg and Mingolla 1985, Koch et al. 1986, Lee 1996). This effect could be manifested in two ways: first, the responses within each surface/region will become more smooth; second, the difference in responses between two surfaces will become accentuated over time at the surface/region discontinuity. The observation shown in Figure 5 and
Figure 12: (a) The boundary cell’s response is predicted to contract spatially around the exact location of the border. The activity at each location represents the response of a boundary cell at that location. The curve represents the responses of a population of identical boundary processors distributed over space, or equivalently the spatial response profile of a boundary unit to different parts of an image with a luminance edge. Over time (as $p$ decreases), the spatial response envelope becomes narrower and narrower (Lee 1995). (b) The surface cell’s initial response $R_i$ is a filtered version of the input $d_i$, which evolves over time by a nonlinear diffusion process to the spatial response profile possibilities $R_f$, which is marked by an abrupt discontinuity in responses between the two regions.

Figure 4 is encouraging evidence. With finer spatial sampling near the border, we will be able to quantify the development of the sharpness of the surface response discontinuity by using the slope of the line that best fits the ‘response cliff’.

The nonlinear smoothing effect cannot be achieved by simple convergence of input via a Gaussian kernel, or linear surround suppression, which will typically exhibit a spatial profile that is smooth across the border (c.f. $R_i$ in Figure 12). We will estimate the optimal suppressive kernel modeled as a linear filter that is consistent with the data obtained in Experiment 1.2 and then proceed to prove that such a kernel cannot be used to predict the nonlinear diffusion effect as exemplified by $R_f$. Existing results on the abrupt change in responses across surface discontinuities are encouraging evidence, but since Experiments 1.2 and 1.3 have never been carried out together in the same neurons, it has not been demonstrated that the spatial response profile cannot be explained by simple lateral inhibition.

Will will use spatial sampling data, in conjunction with the size-dependent onset time estimated in experiment 1.2, to estimate to what extent horizontal connections and feedback are responsible for mediating the enhancement (Bair et al. 2002; Smith et al. 2006; Rossi et al. 2001, Zipser et al. 1996).

The temporal evolution in V1 and V2 of the spatial response profile of neuronal responses to the different images could potentially be very informative with regard to the interaction dynamics in those areas. One could imagine contextual surface information is computed and propagated in V2 and then fed back to V1 to interact with the boundary representation there, which can in turn block the propagation processes in both V1 and V2. Information regarding when fine and coarse information can be observed in V1 and V2 at different points in time could help to infer the interactive process, although definitive result on cortical interaction can only be answered with reversible deactivation studies, a possible future line of research.

*Pitfalls and Solutions.* One concern about the spatial sampling paradigm is that it requires a large number of trials to sample a neuron’s response. Another concern is whether the spatiotemporal response profile will be precise or robust enough to give us information about the propagation mechanisms or hierarchical interaction. We have addressed both issues in detail in the Preliminary Data section.

**Summary and Significance of Aim 1 experiments**

Experiments in Aim 1 are designed to evaluate the figural enhancement effect, or surround influence, across cues (Expt 1.1), across figure sizes (Expt 1.2), and across space (Expt 1.3) in individual neurons. While related experiments have been attempted previously, our specific hypotheses and research questions must be addressed by new efforts.

Our first hypothesis is that the responses of neurons to images exhibit the nonlinear diffusion and boundary sharpening phenomena suggested by computational vision algorithms. Testing this hypothesis (Expt 1.3) requires denser spatial sampling of responses around the border and coarser sampling of responses within surfaces for both ‘boundary’ and ‘surface’ neurons, which has not been done before.

Our second hypothesis is that the center-surround contrast mechanism (Expt 1.2) of a neuron is sufficient (although possibly still necessary) to explain the spatiotemporal response profile of the neuron (Expt 1.3) that...
is symptomatic of nonlinear diffusion. Testing this hypothesis requires Expt 1.2 and Expt 1.3 to be carried out together in the same neuron, which has also not been achieved.

Our third hypothesis is that the surround effect is due to an interaction between cell tuning and input stimuli. This requires knowledge of the tuning of the neurons in the tested cue dimension. An earlier experiment (Zipser et al. 1996) on cue-invariance did not test the cues in isolation (texture always is the carrier, modulated by the cues) and did not test the tuning curve of the neurons for color or disparity. Hence, the rules behind why some neurons show enhancement for some cues but not others are not known, and how cue-invariance can be accomplished is a mystery. Expt 1.1 is designed in part to address these issues.

Furthermore, all Lamme and colleagues’ earlier experiments on figural enhancement effect were exclusively done in V1. The proposed study will study V2 as well, with the understanding that the responses in V2 will be more sensitive to surface representation. By itself, extending these classical studies to V2 will present opportunities for qualitatively new advances.

**Aim 2: Rules and logics of segmentation mechanisms**

*Overview* Experiments in Aim 1 are designed to demonstrate that the linear center-surround mechanism might explain some enhancement at the center of figures, but cannot explain the nonlinear diffusion effect, as exemplified by the sharp response discontinuity at the borders of surfaces. This suggests that the center-surround mechanism should be sensitive to more subtle image structures in the surround, in a nonlinear fashion. Our proposal is that part of the underlying function of this surround mechanism is to implement segmentation. Hence, we predict that the enhancement effect should depend on factors governing perceptual segmentation, a hypothesis we test in the experiments of this aim. Experiment 2.1 tests the dependence of the enhancement effect on contrast gradients and intervening contours. Experiment 2.2 tests whether the enhancement signal is based on region or surface segmentation, rather than simply on cue segmentation. Experiment 2.3 records neurons simultaneously to evaluate cooperative mechanisms for mediating the segmentation effect and to further test the synchrony hypothesis of segmentation.

**Experiment 2.1 Sensitivity to contrast gradients and intervening contours**

*Rationale.* The nonlinear diffusion mechanism for segmentation stipulates that the propagation of surface information is sensitive to cues that are evidence of surface discontinuity. Specifically, the diffusion will be slowed down at locations where the cue changes rapidly, or the cue contrast gradient is high. The prediction inspired by this mechanism is that (1) the enhancement signal will disappear when when the contrast border is blurred in the surround, even though most of the pixel values on the display are identical, (2) enhancement can be produced by combining weak surface discontinuity cues together, even though each by itself might not be sufficient to produce the response.

*Experimental Design and Procedure.* To test these predictions, we will test the following five types of stimuli: a background, a disk, a blurred ‘disk’, a ring, and a blurred disk with a ring, all shown in Figure 13. The ring stimulus has been found not to elicit an enhancement response, even though it will strongly excite neurons in the surround because it can be seen as an object on a uniform background. When combined with the blurred disk (type 5), however, the ring’s border evidence and the blurred disk’s contrast evidence can combine synergistically to create a segmented figure percept.

The monkeys will perform a fixation task while stimuli are presented in the simultaneous update paradigm. Two complementary foreground/background cues will be tested each time, resulting in 10 conditions. Similar stimuli can be created for texture, brightness, disparity, and shading. The main design constraint is that the cue contrast between the foreground and the background should not be too strong. Only one cue dimension from color, brightness, and stereo will be tested, and the preferred cue for each recording session will be determined based on the cell’s response in the screening test (Experiment 1.1).

If time permits, spatial sampling will be used to obtain the cell’s spatiotemporal responses to the type 3, 4, and 5 stimuli using the sampling positions as described in Experiment 1.3.

*Data Analysis and Expected Outcome.*

Figure 13 illustrates the expected magnitude of the enhancement responses as predicted by the nonlinear diffusion type of segmentation algorithms. That is, it should be strongest when there is a high cue gradient border in its surround (type 2), and attenuated greatly (as not predicted by the linear filter model) when the contrast gradient is reduced, i.e. the border becomes fuzzy (type 3). The ring stimulus (type 4) is not expected
to elicit much enhancement if the background is uniform because there the ring is really perceived as a ring. On the other hand, when the ring is superimposed on the blurred disk (type 5), we would expect to see the enhancement become much stronger than the sum of what arises from types 3 and 4. The ring can activate the boundary response, which can then prevent diffusion from crossing the border. This would in turn result in a nonlinear diffusion effect—the disk will appear to be more uniform and homogeneous than the raw stimulus input would have suggested. If spatial sampling is applied to type 5 stimuli, we would predict that the responses within the ring will become more homogeneous over time.

A positive result will add strong credence to the nonlinear diffusion mechanism hypothesis. In our pilot experiment similar to Expt 1.1, we have found that an enhancement response can be elicited by type 2 stimuli (disk) but not by type 4 stimuli (ring). The predictions about stimuli types 3 and 5 remain to be tested. A negative result will show that type 3 stimuli elicit as much response as type 2 stimuli, since the pixel values around the receptive field are roughly equal for the two stimuli (only the border is blurred in one case). Such a result will indicate that the surround effect is simply mediating linear lateral inhibition.

**Pitfalls and Solutions.** In order to have a blurred border, the contrast between the RF stimulus and the background stimulus has to be relatively weak. This lower contrast might reduce the strength of the enhancement response, weakening the signal to noise ratio of the effect. To address this issue, we tested the enhancement effect in figures with a range of cue contrasts and found that enhancement can be observed reliably in disk stimuli with moderately low cue contrast, as long as the segmentation percept is strong.

**Experiment 2.2 Global segmentation with nonuniform surface cues**

**Rationale.** In all figure-ground experiments to date, figures in the stimuli are only rendered with homogeneous cues, such as texture at a single orientation, uniform color, or uniform disparity. In natural scenes, however, the surface of an object in 3D space often exhibits gradients in luminance, texture, and stereo disparity, and yet such surfaces and regions are readily segmentable perceptually. The objective of this experiment is to test whether the figural enhancement reflects a more abstract global segmentation that accounts for gradients in surface cue.

**Experimental procedure.** This experiment will be carried out in conjunction with Experiment 1.1 and will begin with the same asynchronous update paradigm in the figural configuration. Stimuli will include disk figures defined by a gradient in one of the cues. Figure 14 shows luminance gradients (shading) and stereo stimuli to be tested. Each test will contain 4 conditions: two where the same continuous gradient covers the whole screen and remains static, and two that begin with whole-screen gradients and then update to the figural configuration by reversing the gradient in the surround outside of a disk centered on the neuron’s receptive field.

For cells that show a significant enhancement effect, the same techniques used in Experiment 1.1 and Experiment 1.3 will be used to measure their spatiotemporal and center-surround responses to the stimuli.

**Data Analysis and Expected Outcome.** This experiment is a simple extension of Experiment 1.1 and will be analyzed in much the same way. The no-change configuration is tested so that comparison with the figural configuration will reveal the figural enhancement or surround effect in those trials. If enhancement is ob-
Figure 14: Figure-ground stimuli with figure and ground rendered in luminance or disparity gradients. Note that disparity images are rendered graphically for easy visualization. In the experiment they will be rendered with DRDS.

erved, it would suggest enhancement signals reflect a global segmentation that is both independent of cue uniformity and more consistent with the image’s 3D surface segmentation. Based on our conjecture that V2 is responsible for surface segmentation and V1 is responsible for cue segmentation (see Background), we expect V2 will exhibit a strong enhancement effect while V1 exhibits no effect or some weak effect during the later part of the responses by virtue of feedback. We also expect that spatial sampling will yield a uniform enhanced response within the figure at the V2 level but not necessarily in V1 as V1 neurons might be more cue-dependent.

Pitfalls and Solutions. We have not tested these stimuli specifically. However, we did previously demonstrate, in the context of visual pop-out (Lee et al. 2002), that V1 and V2 neurons do experience enhancement for shape from shading stimuli. These are very similar to the luminance gradient stimuli but smaller (1° and 2° in diameter). Since strong enhancement can be observed in luminance and color figures, we are confident of a positive outcome at least in V2.

Experiment 2.3. Mechanisms of cooperative computation

Rationale Nonlinear diffusion-based segmentation involves boundary blocking and surface grouping. These mechanisms can be implemented in a locally connected recurrent network or via feedback from higher levels of visual processing. Cooperative computations within such interactive networks would suggest an abundance of correlated spiking activity (i.e., synchrony). Because these computations segment regions, we would predict greater synchrony among neurons belonging to a segmented region and less synchrony between neurons belonging to different regions/surfaces. We propose that stimulus-based changes in functional connectivity can be a mechanism by which nonlinear diffusion for segmentation is implemented. Such mechanisms are indeed successfully utilized in computer vision for segmentation in graph-cut algorithms (Shi and Malik 1998, Tolliver and Miller 2005). The objective of this aim is to investigate this hypothesis, as well as to test the related ‘binding-by-synchrony’ theory with respect to segmentation (Singer and Gray 1995).

Experimental procedure To test for cooperative computations, we will use stimuli tested in Experiments 2.1 and 2.2. To quantify cooperation among neurons, we will record from multiple neurons (2-8) simultaneously and measure ”effective connectivity” between pairs of neurons (Aertsen et al. 1989). Effective connectivity starts with the joint-PSTH, which is corrected for correlation expected by chance (stimulus-locked) and then normalized for firing rate. Finally, a cross-correlation histogram (CCH) is integrated from this resulting histogram. We define effective, or functional, connectivity as the area under the half-height full bandwidth of the cross-correlation peak at or near 0 ms lag time (e.g., see Figure 8). This quantity is equivalent to Pearson’s correlation coefficient for correlated or synchronous spikes within the temporal window of integration. We will examine peaks that exceed at least 3 SD above mean (Menz and Freeman 2004) and will bootstrap correlation coefficient estimates to derive confidence intervals (Ventura et al. 2005a).

For the five types of stimuli in Experiment 2.1 (Figure 13), we will study neural responses for pairs of neurons when the image is placed in three possible locations with respect to their RFs: (1) both RFs inside the figure (in-in), (2) both outside the figure (out-out), and (3) one inside and one outside the figure (in-out) (example locations for type 2 stimuli are shown in Figure 15). We will test if stronger effective connections are observed in the in-in or out-out configurations and if weaker effective connections are observed in the in-out configuration.
Figure 15: RF placement configurations for disk, blurred disk, ring, and blurred disk with ring.

Figure 16: The two neurons are experiencing similar luminance in their local receptive fields in A and C, and in B and D, but a very different context. The expected outcome is that the neurons will synchronize (i.e. be functionally connected) in the in-in condition (B and C) but not in the in-out condition (A, D).

For the stimuli used in Experiment 2.2, we will study pairs of neurons in three locations for complementary images (see Figure 16). Unlike Experiment 2.1 stimuli, however, these stimuli are exactly the same with respect to the two neurons’ RFs in the in-in and in-out conditions. We predict the neurons will have stronger effective connections in the in-in configuration versus the in-out configuration based on the conjecture that the contrast border in between the two RFs interrupts the functional connectivity between the neurons.

To ensure that we measure statistically significant correlation between spike trains, we will repeat each condition for at least 40-60 trials, which we have determined to be sufficient (e.g., see Figure 8).

Data Analysis and Expected Outcome

If cooperative computations play a role in segmentation, functional connectivity between neurons should be high within the same region and low across different regions. This should be true even for gradient images (Figure 16 where local stimuli for the two RFs are the same).

We expect the results in Experiment 2.3 to complement the firing-rate based nonlinear diffusion mechanism results from Experiments 2.1 and 2.2. Changes in functional connectivity could be driven by bottom-up cues (similarity in the input the two neurons receive), as well as top-down modulation or feedback (higher order information or priors – e.g., attention). Both bottom-up and top-down factors likely influence firing rates in a similar manner. The effective connectivity measurement provides the additional advantage of allowing us to examine the temporal dynamics of functional connectivity (Aertsen et al. 1989). Therefore, we can study the timing of both changes in firing rates and functional connectivity, which in turn permits an examination of the interaction between both response properties. Simultaneous increases in firing rates should enhance functional connectivity by increasing stimulus-locked correlations and making integration in the cortical network more effective. At the same time stronger functional connections likely amplify firing rates and also contribute to more reliable integration (Alonzo et al. 1996). Such cooperative interactions have been observed in V1 for contour integration (Gray et al. 1989; Kapadia et al. 1995; Samonds et al. 2006a), and preliminary results suggest similar interactive mechanisms may exist for disparity-based surface integration (Samonds et al. 2006b).

For the first class of stimuli (Figure 13 top row), when the figural enhancement response is weak, we expect the in-out correlation will be strong. When the figural enhancement response is strong at the center of the disk, the in-out correlation should be weak. Therefore, we predict functional connectivity will be stronger in the in-out condition for type 1, type 3, and type 4 stimuli than for type 2 and type 5 stimuli (see Figure 13 bottom row).

For the second class of stimuli (Figure 16), we expect the correlation to be stronger in the in-in condition in
one stimulus than for the in-out condition in the complementary stimulus. Overall, the correlation in the in-in conditions (B,C) should be stronger than the corresponding in-out conditions (A,D).

**Pitfalls and solutions**

Recording from multiple channels simultaneously introduces several potential technical and analysis problems. We have both Cyberkinetics’ 128-channel Cerebus Data Acquisition system and Tucker-Davis’ 16-channel Pentusa Data Acquisition system to minimize many technical issues (i.e., isolating channels and maximizing signal-to-noise). We limit our analysis to high signal-to-noise multi-units and use spike sorting (Shoham et al. 2003) to remove noise and artifact, and to isolate robust single units. We examine interspike intervals, auto-correlation histograms, and response properties to ensure that each unit is well isolated. We are especially careful to prevent any cross-talk at each level of the recording system. Only examine units on separate electrodes were examined. Preliminary cross-correlation results are consistent with results in anesthetized macaques (Kohn and Smith 2005) and cats (Samonds et al. 2004, 2006a).

Interpretations of cross-correlation results must also be done carefully. Although we use the term effective or functional connectivity, anatomical connections are only one of many possible explanations of cross-correlation peaks. We can only be certain that significant correlation peaks imply that responses are not independent (Aertsen et al. 1989; Brody 1998). However, we can compare our CCHs to previously published data (e.g., Kohn and Smith, Samonds et al. 2004, 2006a) that have been systematically documented with respect to intracortical connectivity. We also examine response properties for correlation resulting from slow covariation of firing rates (Brody 1998, 1999a; Gerstein and Kirkland 2001) and collaborate with colleagues in Carnegie Mellon’s Statistics Department on dealing with cross-correlation interpretation and analysis (Kass et al. 2005; Ventura et al. 2005a,b).

**Summary and significance of results of Aim 2**

Experiments in Aim 2 are designed to test detailed predictions of computational segmentation algorithms. Expt 2.1 will clarify interactions between boundary and surface signals, particularly blocking boundary signals. Expt 2.2 allows us to investigate how strongly segmentation signals are governed by bottom-up cues and global perception of surfaces or regions. Expt 2.3 is designed to investigate the interaction among neurons in mediating these phenomena.

Our first hypothesis is that center-surround interactions in neurons are not linear but are influenced by subtle differences in image structures in the surround. They depends on strong discontinuities in the surround. However weak cues can generate segmentation-related neural responses via nonlinear diffusion. The outcome of Experiment 2.1 will advance our functional understanding of center-surround properties of V1 and V2 neurons in the context of scene segmentation.

Our second hypothesis is that segmentation-related responses in V1, and more likely in V2, are not restricted to figures with uniform cues. Segmentation responses can arise from more abstractly defined figures. To test this hypothesis we will use non-homogeneous cues, such as opposing gradients, to define figures. All earlier experiments on segmentation stimuli are based on figures defined by uniform cues. Shading cues have been used in pop-out tasks (Lee et al. 2002) with encouraging results. Experiment 2.2 can be considered a more explicit test on whether segmentation signals in V1 and V2 arise from the existence of regions or surfaces or local cues.

Our third hypothesis is that segmentation-related signals and nonlinear diffusion emerge from cooperative computations among neurons. This hypothesis requires simultaneously recording from pairs of neurons. Lamme and Spekreijse (1998) have carried out similar experiments using type 2 stimuli (rendered in texture) and concluded that neural synchrony does not reflect binding local features into segregated regions. They alternatively suggest that segmentation is signaled by figural enhancement. Our goal is different in that we are not simply trying to test whether synchrony is a binding ‘label’ for a segmented region per se, but rather to characterize changes in functional connectivity with respect to the underlying mechanisms for generating segmentation. We will nonetheless be able to conduct a test of the binding-by-synchrony hypothesis as a consequence of our paradigm (particularly with the gradient cue-based figure). The use of a broad range of stimuli will provide a rich set of data to evaluate our hypothesis on neural interactions.

**Aim 3: Sources of the figural enhancement effect**

**Overview.**
While the figural enhancement signal could be a reflection of global segmentation, the computational vision notion of segmentation is mainly concerned with partitioning scenes into regions or surfaces and is not required to identify or highlight regions’ importance or salience. The enhancement of neural responses within figures suggests a possible involvement of further, higher order processes dedicated to these tasks. In the Background section (B), we have proposed three possible mechanisms for generating the enhancement signals. In Aim 1, we investigate the contributions of the center-surround contrast mechanism. In Aim 3, we will investigate the contributions of attentional selection (Expt 3.1) and figure-ground organization (Expt 3.2).

**Experiment 3.1: The contribution of attention**

**Rationale.** Is attention necessary for the generation of the figural enhancement (Triesman and Kanwisher 1998)? In all the fixation experiments, the figure could trigger attentional mechanisms reflexively. To evaluate whether attention plays a role in modulating the enhancement processes, we will measure and compare the responses of neurons inside and on the border of a figure while the monkey performs both a simple fixation task and a distraction task designed to draw attention away from the receptive field (Chun and Marois 2002).

**Experimental Design and Procedure.** The monkeys will perform a fixation task and a distraction task in each recording session (Figure 17). During both tasks, the actual receptive field stimuli are identical, consisting of a $3^\circ$ figure on a contrasting background, and do not interfere with the fixation or distraction stimuli associated with the monkeys’ behavior. These figure stimuli are presented simultaneously using the figure update paradigm, with the receptive field either at their centers (figure condition), on their borders (border condition), or away from them altogether (ground condition). Stimuli employing two different levels of cue contrasts will be examined, both maximum contrast as in Figure 10, and weak contrast as shown in Figure 17. Each cue contrast will furthermore be shown with two complementary cue pairings, such as light figure/dark ground and light ground/dark figure for luminance. Altogether, there are a total of 12 conditions (2 contrasts, 2 cues, 3 locations).

During experimental sessions, experiment 1.1 will first be performed to select the optimal cue. Next, in the fixation task, no dots will appear at the top of the screen, and the subject is only asked to make a saccade to a target dot. When recording from each cell, the fixation task will be performed first. The distraction experiment will then be performed only if a significant figural enhancement effect is observed in the fixation task.

In the distraction task, the monkey has to carefully attend the five small dots during the 400 ms stimulus presentation. The subject’s task is to detect a transient (34 ms) dimming of one of the dots and make a saccade to that dot within 180 ms. Catch trials will be introduced in which none of the dots are dimmed. For these trials, the monkey has to fixate for 400 ms without making any saccade in order to receive a reward. If he fails either to make a saccade in the dimmed trial or fails to remain fixating in the catch trial, he will not be rewarded and will instead be penalized by a one-second timeout.

40 trials will be collected for each of 12 conditions in the distraction task versus 20 trials per condition in the fixation task. This is because the distraction task terminates at variable times, and at least 20 trials with trial durations of at least 200 msec will be required for an analysis of both.

**Data Analysis and Expected Outcome**

Data will be analyzed to address whether distraction attenuates the figural enhancement signal. We will sort the correct trials in the distraction experiment according to the duration of the trial before the monkey’s saccade to the correct target. We will take the trials over 200 msec in duration before saccade onset and compute the figural enhancement index based on the spikes within the 80-200 msec window, in both the figure condition and the ground condition. This index will be compared with the figural enhancement index obtained from the same cell during the fixation task.

If attention plays a role in the enhancement response, we expect the figural enhancement signal for weak contrast stimuli to be attenuated in the distraction task. The enhancement signal for strong contrast stimuli...
may or may not suffer because a salient enough figure may be able to capture sufficient residual attention to produce figural enhancement. Attenuation across all stimuli may be observed during the distraction condition, which may be obscured in the figural enhancement ratio index. Thus, raw responses of the neurons in the 12 conditions (2 contrast, 2 cues, and 3 locations) will be carefully compared for the two tasks. If attention plays no role in the enhancement response, then we expect little or no difference between the responses in the two tasks.

Pitfalls and Solutions.

Whether the distraction task is demanding enough is a serious concern. Marcus and Van Essen (2002) detected only a marginally significant effect of attention in V2 for their illusory figure stimuli. In their experiment, however, the monkey had to alternate in the same recording session between a task requiring attention to the RF stimulus and another task that involves detecting shapes at another location away from the receptive field. It might be difficult for the monkeys to shut off their attention to the RF stimulus when they are performing the other task. Furthermore, there is nothing in the behavioral design that would prevent the monkeys from occasionally shifting their attention to the RF stimulus even while engaged in another task. Our distraction task is more demanding in time and concentration specifically to prevent such occasional attentional shifts. Furthermore, we do not require the monkeys to do a task that demands attention to the RF stimulus and may cause involuntary attention at its location. To further ensure the monkeys have no opportunities to shift attention, we will make the task more difficult by reducing the time allowed for reaction so that the monkey can achieve no more than 85% correct trials at best.

Experiment 3.2: The role of figure-ground organization

Overview. Local convexity such as the border of a square is a factor in favor of figure perception and by itself is sufficient to produce enhancement (Rossi et al. 2001). However, the figure-ground organization is a global percept. To what extent does the figure surface enhancement response depend on local border convexity versus global context? Experiment 3.2 is designed to answer three pertinent questions. (1) Does the enhancement response depend on global context of figure-ground organization or local convexity of borders? (2) Does the enhancement response depend on the figural percept or the foreground-background ordering relationship of surfaces? (3) Does the enhancement result from the propagation of border signals within the surface that owns the border?

Experimental Design and Procedure. To dissociate the influence of the local convexity structures of the border from the global figural percept, we employ the “nut” and the “bolt” stimuli shown in Figure 18. The stimuli are constructed by trimming parts of a $9^\circ \times 9^\circ$ square. Of particular interest are the responses of neurons with their RFs at the two depicted locations in the image pairs. In both locations, the general principle of the experiment is evident: the local image structures are constant but the global figure varies, allowing us to directly address the first question.

At location 1, the neuron’s RF experiences the same local color-contrast edge with both stimulus types; in fact, the stimulus is identical within a $6^\circ$ window. Globally, however, one observes that the neuron is on the border of the “nut” stimulus in the left panel and the “bolt” stimulus in the right panel. At location 2, the local stimuli for both conditions are identical within a $9^\circ$ diameter aperture. In the nut condition, however, it is in the background, while in the bolt condition, it is inside the bolt figure. In the color nut-and-bolt experiment, these two stimuli and their horizontally mirrored counterparts will be tested, resulting in 4 possible configurations.

We will also test the nut and bolt stimuli rendered with dynamic random dot stereograms (see Stimulus Design for details). Figure 18 shows a set of 4 configurations to test. In both configurations A and B, the nut is considered the ‘figure’, but in A it is an object in relief, while in B it is an indented shape. In both configurations C and D, the bolt is considered the ‘figure’, but in C it is an object, while in D it is an indented shape. The stimuli are designed to dissociate the ‘figure’ percept and the foreground-background ordering relationship of the surfaces, which allows us to address the second question. Since the nut and bolt can face either left or right, there will be a total of 8 configurations. The stimuli are designed such that RF locations on the right in each panel (location 2) all receive exactly the same local DRDS stimulus at the preferred disparity of the neuron across the different configurations.

To investigate whether the enhancement arises from border signals propagating within the surfaces that own them, we will test the color and the stereo stimuli by simultaneously recording from a pair of neurons,
Figure 18: The nut (left) and the bolt (right) stimuli in color (top row) and in stereo (bottom row, defined in dynamic random dot stereograms).

one with the RF at the edge of the figure and the other with RF at the center of the image. For this test to be effective for the specific stimuli shown, the RF on the left in each panel (location 1) should belong to a border-ownership neuron and RF on the right (location 2) should belong to a surface neuron that exhibits the figural enhancement response.

In this experiment, the monkeys have to perform a figural shape recognition task. The stimulus will be presented for 400 msec in the synchronous update paradigm, while the monkeys fixates. Then the stimulus will disappear, followed by the appearance of four choice stimuli (2 nuts and 2 bolts pointing left or right, rendered in outline) at the four corners of the screen. The monkey has to make a saccade within 200 msec to the correct choice to get a reward.

**Data Analysis and Expected Outcomes**

To address the first question, we will compare the responses at location 2 between the color nut and bolt conditions in the trials that the monkey makes a correct response. If the enhancement signal depends on the global figural percept rather than the local convexity of contrast borders, the bolt condition (as shown in Figure 18) is expected to elicit a stronger response than the nut condition even though the local stimuli within the 9° diameter window are exactly the same. Our pilot data provide strong encouraging evidence in support of this hypothesis (see Figure 7).

To address the second question, we will compare the responses at location 2 in the four DRDS stereo nut and bolt configurations. If the figural enhancement depends on the figural percept as reported by the figural shape seen by the monkeys, rather than the strict foreground/background order of the surfaces, we would expect to see the responses at location 2 in configurations C and D to be stronger than those in A and B. On the other hand, if the enhancement is determined by the foreground/background order of the surfaces, rather than the figural choice, then the responses at location 2 in configurations B and C will be stronger than those in A and D.

When tested with a border-ownership neuron at the edge location, a left-border neuron should respond more in the specific depicted bolt condition than in the depicted nut condition, while the right-border neuron should behave in the opposite way (see Figure 18). For the stereo stimuli, the situation is more complex: if the border-ownership signal depends on which figure is selected, then responses at location 1 in configurations C and D will be stronger than A and B for the left-border neuron, and vice versa for the right border neuron. On the other hand, if the border-ownership signal depends on foreground/background surface order relationships, then responses at location 1 in configurations B and C will be stronger than A and D for the left-border.
neuron, and vice versa for the right-border neuron.

To answer the third question, we will perform correlation analysis between the activities of the two neurons. For the color stimuli, the expected outcome is follows: when the neuron at the edge location (1) is a right-border neuron, the two neurons should synchronize (i.e. show stronger functional connectivity) in the bolt condition but not in the nut condition; when the neuron at location 1 is a left border neuron, the neurons should not synchronize in either the nut or the bolt conditions. For the stereo stimuli, similar reasoning can be applied. If the neuron at the edge is a left-border neuron, the two neurons should synchronize in conditions B and C but not in conditions A and D. If the neuron at the edge is a right-border neuron, the two neurons should not synchronize in any of the configurations.

**Pitfalls and Solutions.** Most of the pitfalls have been covered by earlier discussion on stereo experiments and synchrony analysis. The new challenge here is the training of the monkeys to report the global figure shape it perceives. It is important to ensure the monkeys base their report on the global figural shape they perceive, rather than some local features in the stimuli. We will train them to do the color task first. During the training process, we will vary the figure stimuli in various ways, rotating them, deleting parts, rendering with different cues, or only outline, to make sure they are not discriminating the shapes by detecting some local features.

**Summary and significance of Results in Aim 3**

Experiments in Aim 3 are designed to further evaluate the possible mechanisms of figural enhancement, specifically attention (Expt 3.1) and figure-ground organization (Expt 3.2).

Our first hypothesis to test is that the enhancement effect is a type of attentional signal constrained by the segmentation boundaries in V1. This attention is triggered reflexively and automatically by salient figures in the stimuli. Earlier experiments (Lamme 1995, Marcus and Van Essen 2002, Lee et al. 2002) have tested this hypothesis, but with mixed or negative results, and the responses at the border have also not been tested. The proposed experiment will test this hypothesis more rigorously by making the distraction task more demanding and by avoiding attracting the monkeys to the RF stimuli in alternating trials. A negative result will serve as a necessary control; a positive result, particularly one accompanied by an observation of response attenuation at the border as well as at the center of the disk, would provide an important piece of evidence in support of a role for attention and feedback in segmentation. Attention is not expected to be the primary source of the signals, but rather a potentiating signal that interacts with the computation already taking place in the early visual areas.

The second hypothesis to test is that the enhancement effect is a product of figure-ground organization and emerges from the propagation of the border signals in the surfaces that own the borders. For this reason it should have close ties to the border ownership signals themselves. Earlier figure-ground experiments typically used squares or disks, leaving the possibility that local convexity of contrast borders might be the determining factors in both the signal enhancement effect and the border ownership effect. Our experiment with the nut and bolt stimuli will provide definitive evidence that the border-ownership signals and the figural enhancement signals depend on figure-ground organization and not on local convexity structures of the borders. In addition, it will dissociate the roles of perceptual figural choice of the subject, which is more abstract, from the physical order of surfaces in all these signals.

**Overview and Schedule**

Funding support for two postdocs or one postdoc/one graduate student and partial support for the PI are requested each year. The experiments described in Aim 1 are partly a continuation of our pilot study and will begin immediately. The complete set of experiments will be carried out by a postdoc or graduate student in the first two years on two monkeys (both V1 and V2 neurons), and the basic results will be reported within three years. Experiments 2.1, 2.2, and 3.1 will begin in the second year and continue through years three and four. These will be carried out by the same postdoc/graduate student as an extension of Aim 1. Experiments 2.3 and 3.2 will begin in the third year and will be carried out by the second postdoc. Due to the requirement of multielectrode recordings and a stereo setup, this will continue through years four and five and will involve two additional monkeys. The PI will be involved in all aspects of this project. Four monkeys will be used in this project.
E. Human Subjects

None.

F. Vertebrate Animals

1. Proposed use of the animals. Rhesus monkeys (Macaca mulatta) will be used for this research. The experiments will involve recording from single neurons in the behaving monkey for approximately four hours per day for 4-5 days per week. Four adult monkeys (male or female) 4-12 kg and 2-12 years of age will be used. In each animal a cranial pedestal and scleral search coils will be implanted surgically at the beginning of the experimental period and craniotomy will be performed to permit placement of a recording chamber. The full experiment will span at least one year in each monkey. Each experiment will terminate with euthanasia and histological reconstruction of the recording sites.

2. Justification of animal use, choice of species, and numbers used. Rhesus monkeys are superior to any other common experimental species in their ability to master sophisticated visuomotor tasks. This is manifest in four ways: (1) they are able to learn sophisticated contingencies as in the case of memory-guided saccades; (2) they commonly emit thousands rather than hundreds of operant eye movements during a session; (3) they are able to move their eyes over a large range; (4) their gaze is steady during fixation. Other species of monkey, such as Macaca fascicularis, have not been found to share these advantages and are not widely used in cognitive and visuomotor experiments. We estimate that four individuals will be required for collection of adequate data. The necessity for using multiple monkeys arises several factors including the following: (1) there is a time limit of around one year for the chamber and cranial implant to remain in good condition; (2) for statistical reliability, each test must be carried out on many neurons in each area, and data must be collected from more than one individual.

3. Veterinary care. Monkeys in the PPL colony are cared for by a full-time AALAS-accredited animal care technician. All aspects of housing and care are fully compatible with applicable USDA regulations and are in accordance with the NIH Guide for the Care and Use of Laboratory Animals. Veterinary oversight is provided by the University of Pittsburgh Division of Laboratory Animal Resources.

4. Procedures to minimize discomfort, distress, pain and injury. Three situations exist where the animal might be subject to these conditions: (1) during surgery; (2) during restraint for handling or routine testing; (3) during training and experimental recording sessions.

5. Surgery. The initial surgery for installation of the cranial implant and eye coils is not a major surgery in the sense that no mesenchymal barrier is penetrated. Nevertheless, it is carried out under full sterile precautions during continuous maintenance anesthesia by inhalation (Isoflurane, approx. 1.5%). An analgesic (Butorphanol, approx. 0.05 mg/kg) is administered immediately postsurgically and at intervals thereafter as appropriate.

6. Restraint for handling or routine testing. For brief restraint, e.g., for giving an injection, the squeeze cage will be used. For prolonged restraint, e.g., during medical tests and treatment, we will use ketamine HCl (10 mg/kg, i.m.). Monkeys will be transferred between cage and chair by use of a pole-and-collar arrangement as recommended in the NIH Guide for Care and Use of Laboratory Animals.

7. Restraint during training and experimental recording sessions. During training and recording, monkeys sit in chairs that provide rigid head restraint but permit considerable movement of the rest of the body. Monkeys can and do repeatedly change their posture to achieve comfort. The height of the floor of the chair is adjusted to accommodate the dimensions of the individual monkey. Each monkey will be confined to the chair only during the course of its daily experiment. The head will be fixed to the chair only after the chair is in position in the experimental area.

8. Care of chronic implants. Each animal’s implanted recording cylinder will be inspected and cleaned at least three times per week. A bacteriostatic solution will be placed in the cylinder before resealing. The skin margin at the edge of the implant will be inspected on the same schedule and cleaned if appropriate.

9. Control of food and fluid intake. It is central to most experiments in this series that monkeys perform tasks for reward (water or dilute fruit juice) and therefore is necessary that their intake outside the experimental sessions be supervised (Desimone & Duncan 1995). Each monkey’s weight and intake will be monitored on a regular basis. In general, monkeys will be allowed to work for fluid until satiated or will be given sup-
plemental water at the end of the session. On average, during days when there is no session, animals will be given at least as much water as they have consumed on average during preceding days when they worked to satiation.

10. Animal care philosophy. The monkeys used in these experiments will be required to perform visuo-motor tasks that demand extended effort and concentration. It is essential that they be in excellent health and that the laboratory environment not induce stress. This is in the interest of the animal, the investigator, the scientific community, and society, which ultimately will reap the benefit in basic and applied knowledge.

11. Euthanasia. At the end of the experimental period, each monkey will be killed so that electrode penetrations can be reconstructed histologically. Euthanasia will be carried out by the following procedure. First, a high dose of ketamine HCl will be given (15-20 mg/kg, i.m.) to induce dissociation and analgesia. This will be followed by an overdose of sodium pentobarbitol (200-250 mg/kg, i.v.). The monkey will then be perfused intracardially with saline followed by 10% formalin to fix the brain. This method is consistent with recommendations of the Panel on Euthanasia of the American Veterinary Medical Association.
G. References


Tolliver D, Miller GL (2005) Graph partitioning by spectral rounding: applications in image segmentation and clustering. Submitted to ECCV.


