Neural Basis of 3D Surface Representation in Primate Visual Cortex

R01: Research Plan

A. Research Aims

The general goal of this project is to characterize the neural mechanisms in the early visual system of the macaque for representing and computing the three-dimensional (3D) surface structure of objects in visual scenes. 3D surface inference and representation is a fundamental problem in vision. Recent neurophysiological studies have shown that neurons in the higher order extrastriate cortices (IT, CIP, MT) are sensitive to large 3D structures and the shape of objects (Bradley et al. 1998; DeAngelis et al. 1998; DeAngelis and Newsome, 1999; Janssen et al. 1999, 2000, 2001; Nguyenkim and DeAngelis, 2003; Tsutsui et al. 2002). There are also several new studies, including ours, that suggest V2 might be involved in the first stage of computing and representing surfaces (Bakin et al. 2000; Lee et al. 2002; von der Heydt et al. 2000, Thomas et al. 2002). However, these studies primarily demonstrate indirect ‘sensitivity’ to surfaces, or tuning to absolute or relative disparity. While quantitative and systematic techniques have been used to study the encoding of binocular disparity by V1 neurons (Ohzawa et al. 1990, 1996, 1997; Anzai et al. 1999a, 1999b, 1999c), similarly rigorous techniques have not been applied to the study of encoding local 3D surface structures in V1 or V2 neurons. Many basic questions remain unanswered: How are arbitrary local 3D surface structures (surface shapes, orientation and discontinuity) represented in the brain? What are the basic neural representational elements? How are they derived from various visual cues and influenced by a contextual surround?

To answer these questions, we propose to investigate the neural activity of early visual cortex (V1, V2) in response to 3D surface shape structures, both locally and in the context of larger scenes. The receptive fields (RFs) of early visual neurons are retinotopically organized and small in size, and thus are ideal candidates for representing local 3D surface structures. Neurons in higher cortical areas (V4, MT, IT) have larger receptive fields and might be useful for representing larger surface patches or more global surface structures. We propose to carry out experiments to systematically characterize the sensitivity of V1 and V2 neurons to surface orientation and other basic surface structures, and to determine their possible 3D receptive field structures which would provide insight to the neural strategies for encoding 3D scenes. We will investigate how these representations are derived and disambiguated in the brain under the influence of various bottom-up visual cues as well as the global image context.

Aim 1: Representation of local 3D surface structures.

Experiments are proposed to characterize the sensitivity of neurons in V1 and V2 to 3D surface structures rendered with dynamic random dot stereograms (DRDS). The basic hypothesis is that neurons in these early visual areas are sensitive or tuned to 3D surface structures, and potentially each has a 3D receptive field for encoding a particular basic shape primitive. We will study neural responses to basic absolute disparity, surface orientations (slant and tilt), and oriented convex and concave structures. Furthermore, we will characterize the 3D receptive fields of these neurons, predicted from efficient coding theory and the statistical analysis of surface variation in natural scenes.

Aim 2: Representation as a function of visual cues.

Here, we propose experiments to determine whether neuronal sensitivity to 3D surface shape is a function of visual cues. The sensitivity of neurons to 3D shapes rendered in shading, texture and stereo disparity cues will be tested both in isolation and in combination. The basic objective is to evaluate whether surface sensitivities based on different cues are carried by the same or different neural channels. We also plan to study the time course of surface shape sensitivity as a function of visual cues as a means to understand the inference processes. We will test the hypothesis that abstract 3D surface structure is inferred and represented early on in the visual areas.

Aim 3: Contextual influence of the surrounding surfaces.

In this aim, we have designed experiments to investigate the influence of global context on the interpretation of local surface cues, and to understand how surfaces are organized in a global context. Surface shape tuning
will be evaluated in several global contexts defined by surrounding surface structures. The first hypothesis to test is that a consistent global context in a large visual scene might help to disambiguate surface interpretation and thus sharpen the surface tuning curve, while inconsistent global context might flatten the tuning curve. The second hypothesis to test is that contextual modulation occurs at the surface or depth levels. Multielectrode recording of neurons whose RFs are located within the same surface or in different surfaces will be used to evaluate the encoding of surface structures by neuronal ensembles.

**Aim 4: Contextual influence of global boundary shapes.**

Experiments are designed to investigate the influence of boundary shape on the interpretation and representation of a shaded surface. The hypothesis to test is that an occlusion border in a larger scene can help to disambiguate local surface interpretation, and that a potential mechanism for accomplishing this is by horizontal propagation of the surface estimate from the border to the interior of the surface.

**B. Background and Significance**

Inferring and representing 3D scene structures is a central problem in vision. A powerful cue to 3D structure is binocular disparity, the difference between the images in the two eyes. Hence, neural tuning to binocular disparity has been intensely studied since the early days of visual neuroscience. Barlow et al. (1967), Pettigrew et al. (1968), and Hubel and Wiesel (1970) were among the first to study the disparity tuning of neurons in the visual cortex (area 17 or V1 in particular) of the anesthetized cat. Later, Poggio and his colleagues (Poggio and Fischer 1977; Poggio et al. 1988) distinguished five classes of disparity-tuned neurons (near, far, zero, tuned excitatory, and tuned inhibitory) in V1, V2 and V3/V3A of the alert macaque. Several groups have found that disparity-tuned cells in extrastriate visual areas are systematically organized, often into “near” and “far” columns (V2, Hubel and Livingstone, 1987; Ts’o et al., 2001; V3, Hubel and Wiesel, 1970; Adams and Zeki 2001; MT, DeAngelis and Newsome, 1999; Bradley et al. 1998; MSTI, Eifuku and Wurtz 1999; V4, Watanabe et al. 2000). However, these studies have not investigated the spatial profile of the relative or absolute depth tuning of these cells, as the stimuli used were typically defined only as different frontoparallel surfaces.

A number of recent neurophysiological experiments began to explore sensitivity of neurons to 3D curved surfaces in the higher order extrastriate cortex (CIP, Sakata et al. 1998, Tsutsui et al. 2002; IT, Janssen et al. 1999, Janssen et al. 2000; V4, Watanabe et al. 2002; MT, Nguyenkim and DeAngelis 2003). Sakata et al. (1998) and Tsutsui et al. (2002) have found that neurons in the caudal intraparietal sulcus (CIPS) are tuned to large 3D surface orientation and 3D curvature defined by random dot stereograms (RDS) and to texture gradients, respectively. Janssen et al. (1999, 2000) have found TE neurons to be sensitive to convex and concave curved surfaces, while Nguyenkim and DeAngelis (2003) found neurons in MT are tuned to the slant and tilt of large surfaces. Functional MRI studies (Tsao et al. 2003) further revealed that a significant number of parietal areas, in particular V3 and V3A, show strongly enhanced activity in response to RDS-defined oscillating checkerboard surfaces. Given that a large number of parietal and inferotemporal cortical regions are sensitive to larger 3D surfaces, it seems likely that basic surface orientation, particularly that of small surface patches, is computed early on in the visual system, before signals are passed on to higher order extrastriate areas.

Our working conjecture is that the early visual areas (particularly V2) are ideal for inferring and representing local 3D surface structures. This is based on several observations. First, almost all of the computational algorithms developed by computer scientists for computing 3D surface shapes based on binocular disparity or optical cues require a high-resolution buffer to represent surface orientation and to carry out the computation. The high-resolution buffer, characterized by small receptive fields of its component elements arranged in a retinotopic coordinate frame, is useful for supporting precise representation and computation of detailed geometrical features, such as surface orientation and surface discontinuities (contours). In a series of experimental and theoretical papers, we have advocated that V1 and possibly the LGN and V2 together can serve as this high-resolution buffer, each with a specialized role, particularly in the representation of surfaces and contours (Lee et al. 1998, Lee and Nguyen 2001, Lee et al. 2002, Lee and Mumford 2003, Deco and Lee 2002). In this framework, surface inference and representation are accomplished through the interaction between the higher order visual areas, which capture global surface context, and the early visual areas, which capture the fine
details of local surface and boundary structures.

Several recent studies have indeed provided strong clues supporting the idea that V2 plays a role in coding 3D surfaces. Bakin et al. (2000) provided some indirect evidence that V2 neurons are sensitive to 3D stereo surfaces presented in Da Vinci stereo. Our group (Lee et al. 2002) found that V2 neurons are preattentively sensitive to pop-out saliency defined by a 3D scene context. This implies that V2 might have an infrastructure with elementary features for encoding 3D surface shapes. The logic is that pop-out is mediated by lateral interaction of the basic representational elements. For example, orientation pop-out in V1 (Knierim and Van Essen 1992) is mediated by the orientation selective neurons. Thus, 3D pop-out in V2 suggests the possible existence of 3D representational elements in V2. von der Heydt et al. (2000) have provided more direct evidence of V2’s role in 3D scene analysis. They found that V2 neurons are responsive to stereoscopic step edges, and that 16% of the cells are sensitive to convex roof edges (manuscript in preparation). More recently, Thomas et al. (2002) showed that some V2 neurons exhibited tuning to relative disparity as defined by center-surround planar surfaces. All of these results suggest to us that V2 might indeed provide an infrastructure for representing and inferring fine 3D surface structures.

One such possible infrastructure is a population of neurons with linear 3D receptive fields. Our application of independent component analysis to 3D range data suggests that 3D Gabor wavelets might be the most efficient elementary basis function for 3D surfaces (see Preliminary Studies). The application of independent component analysis to 2D intensity data has produced simple cell-like receptive fields (also in the form of Gabor wavelets) as the most efficient basis functions for representing intensity images (Olshausen and Field 1996, Bell and Sejnowski 1997). The hypothetical 3D receptive field is similar to the simple cell receptive field, except that the encoded dimension is relative depth rather than luminance intensity. It should be noted that simple cells were considered mainly as oriented edge detectors (Hubel and Wiesel 1972) upon their first discovery. It was later observed that they can be modeled by 2D intensity Gabor wavelets (Marcelja, 1980, Daugman 1985, Lee 1995) and thus form a complete family of basis functions to represent arbitrary intensity images. By the same reasoning, the cells that are found to be sensitive to the polarity of disparity and depth edges (von der Heydt et al. 2000) might also be the ‘simple cells’ in the depth domain. A population of such neurons will be able to represent arbitrary local 3D abstract and invariant surface structures (see Lee 1995), furnishing an infrastructure for the brain to perceive and reason about the three dimensional surfaces in a visual scene.

The 3D receptive field, should it be linear, can be recovered by a spike-triggered average technique with white noise stimuli, or more generally, regression techniques (Romero et al. 2002, Marmarelis 1993, Korenberg 1988, Potetz and Lee, 2003). Anzai et al. (1999a, 1999b) have applied cross-correlation techniques with binary m-sequences to study the relationship between the monocular and binocular RFs of V1 simple and complex cells in anesthetized cats. They found that the binocular interaction RFs of most simple cells are well described as the product of the left and right eye RFs. This multiplicative binocular interaction is important for computing interocular cross-correlation, an algorithm for solving the stereo correspondence problem, but is not appropriate for coding 3D surface. Using the same technique (Anzai et al. 1999b), they also found that the binocular interaction RFs of complex cells exhibit an elongated excitatory subregion along the frontoparallel axis at different binocular disparities. Such a substrate can potentially be the basis of representing planar or curved surface at absolute depth. However, given that they have only examined the binocular interaction field along one dimension (horizontal), the 2D spatial profile of the binocular interaction field is not known.

To understand the basic paradigm of the brain for encoding 3D surfaces and how sensitivity to large surface structures are constructed, it is critical to understand how early visual neurons are tuned to local surface orientation and basic surface structures, as well as the spatial profile of the 3D sensitivity of these neurons. The recent work of Thomas et al. (2002) and von der Heydt et al. (2000) has begun to explore in that direction by studying relative disparity tuning and stereoscopic edge sensitivity in V2. However, sensitivity to local 3D surface has not been directly studied. In this proposal, we plan to study the 3D surface structure tuning of V1 and V2 neurons, and to investigate the possible existence of 3D receptive fields. We will test the neurons’ tuning to basic surface structures that are derived from statistical analysis of natural scenes, and will map the 3D
In addition, there is virtually no direct physiological study that illuminates the cortical mechanisms of surface inference. Most insights to mid-level vision came primarily from psychological observations (Nakayama, Adelson), with little understanding of its neural underpinnings. Thus, in this project, we will use the influence of cues and global context on surface representation as a vehicle to probe at the processes of cortical surface inference. One specific issue is whether cue-integration is weakly coupled, done by different ensembles of cells and then integrated in a later stage (Marr 1982, Bruno and Cutting 1988; Maloney and Landy 1989), or is tightly coupled in a single mechanism (Bulthoff and Mallot 1990, Yuille and Bulthoff 1996). A second issue is whether the influence of global surface and border context is mediated through intracortical horizontal propagation or through recurrent feedback. Through these experiments, we hope to gain further insights to the inference mechanisms in the visual system, particularly in the Bayesian inference framework (Lee and Mumford 2003).

In conclusion, our project addresses several fundamental and unanswered questions concerning how our brain analyzes and represents the 3D surface structures in a visual scene. The experimental findings from this project should advance our understanding of the neural basis of 3D representation and the processes of perceptual inference in the brain. This information is necessary and useful for the development of a new generation of computer vision systems, neural-computer interfaces and other therapeutic technologies for the blind, with long-term clinical significance.

C. Preliminary Studies

This section is organized in two parts. In the first part, we will report on the theoretical and computational investigations we have made in preparation for this project. In the second part, we will describe our technical preparation and experience in V1 and V2 neurophysiology, as well as encouraging results from preliminary experiments that demonstrate the feasibility of the project. The most relevant aim is put in parentheses next to the topic of each preliminary study.

Part 1: Theoretical and Computational Preparation.

Stimulus design from natural scene statistics. (Aim 1, Expt 1) To identify potential efficient codes for representing 3D surface structures, we have collected a database of coregistered intensity and range images (meaning the pixels of the two images correspond to the same point in space) using the Riegl LMS-Z360, the most sophisticated range scanner available. We have performed a series of statistical studies on this database (Potetz and Lee 2003). Applying principal component analysis, we found the natural dimensions for describing surface shape variations in natural scenes. These principal components capture the significant variations of surface shapes in natural environment, and it is reasonable to hypothesize that neurons in the brain will evolve and optimize themselves to become sensitive to these dimensions. Figure 1 shows the first 15 principal components in the natural shape space. These 15 components capture over 95% of the variance in shape space. The principal components of range images resemble a Fourier basis, and can therefore be grouped according to their approximate polynomial order. The first group is the first principal component (top left) – frontoparallel plane. Variation of the absolute depth of the frontoparallel plane has been the basic parameter tested in earlier neurophysiological experiments. The second group represents the second and the third principal components (tilted or slanted planes). The third group represents the fourth to sixth principal components, related to a convex edge or a concave edge at different orientations. Thus, we will use these three types of principal components and their generalization over orientation as stimuli for evaluating the surface shape sensitivity of early cortical neurons.

Conjectures on efficient neural codes for 3D scenes. (Aim 1, Expt 2) Olshausen and Field (1996) and Bell and Sejnowski (1997) have used independent component analysis to explain the emergence of V1 simple cell RFs as an efficient code for capturing correlations in natural luminance images based on the sparse coding principle. What are the efficient codes for range data according such a principle? We applied independent component analysis to the 3D range data and found that the efficient codes for encoding 3D surface information...
turn out to be quite similar to the Gabor wavelets (Lee 1995) which have been used for modeling simple cell RFs. In the 3D case, the receptive field is now coding depth rather than luminance intensity. Examples of the efficient linear 3D codes for representing surface shapes in natural 3D scenes are shown in Figure 2. One of the hypotheses to be tested is that some V1 and V2 neurons might be representing 3D surface structures using these 3D elementary receptive fields.

Figure 1. The first 15 principal components with eigenvalues in descending order from left to right, top to bottom. These 15 components capture over 95% of the variance in shape space.

Figure 2. (a) 3D receptive field predicted from independent component analysis of 3D range data of natural scenes. (b) An idealized 3D Gabor wavelet where the z-variable is depth.

There are other possible coding schemes for representing 3D surfaces. The simplest one is coding disparity at each point. It is also possible that some basis function might be best for shape from shading while another might be best for shape from texture, depending on the relevance of the task and the problem it is trying to solve. Hence, despite our intuition, we entertain a diversity of representational schemes in the cortex, and will adopt a more conservative approach by testing the surface structure tuning of the cells (Aim 1, Expt 1) first before embarking on identifying the 3D receptive field (Aim 1, Expt 2).

Computational algorithms on surface inference (Aim 4) The PI has studied and developed algorithms related to segmentation, surface inference and figure-ground segregation (Lee, Yuille and Mumford 1992, Lee 1995, Zhu, Lee and Yuille, 1995, Yu, Lee and Kanade 2002). Several insights have emerged from these studies. First, many of these algorithms can be modeled by Markov networks, which emphasize constraints embedded in local connections, and the computations are carried out by the propagation of a scene/surface estimate over the network. Second, the processes of boundary inference and surface inference are tightly coupled, yet remain
independent. The boundary inference process will produce a coarse estimate of boundary initially, and this estimate becomes sharper and more refined over time. The dynamics of surface inference in many algorithms (Lee 1995, Grossberg and Mingola 1985, Horn 1985), involves propagation of the surface estimate by a non-linear diffusion process—a diffusion that is stopped by a surface border (Lee 1995). Third, there is a trade-off between the size of the analysis window and the localization power of the boundary detection. The larger the window size, the better one can estimate the texture based on the statistics of the texture elements, but the less capable one is at estimating the boundary (Zhu, Lee and Yuille, 1995). Fourth, such local Markov networks can be generalized to implement many different kinds of rules and constraints, and the computation can be sped up tremendously when implemented in a hierarchy (Yu, Lee and Kanade, 2002). These computational studies have exposed us to a variety of possible algorithms and mechanisms for surface inference and provide a solid theoretical foundation for us to formulate hypotheses and reason about the data we observe.

**Part 2: Experimental and Technical Preparation.**

*Sensitivity of V1 and V2 neurons to 3D surface structures* (Aim 1 and Aim 2). Local 3D codes in the brain have not been systematically investigated. But there are several pieces of recent evidence that strongly indicate the possibility of such a 3D surface code. First is a pair of findings from Rudiger von der Heydt’s group that V2 neurons and some V1 neurons are sensitive to convex roof edges (Qiu, et al. (2000) Soc Neurosci Abstr. 26, 593.2) and stereoscopic edges (von der Heydt et al. 2000). Second, Thomas et al. (2002) found that V2 neurons are sensitive to relative disparity in a center-surround configuration. Such neurons might be coding 3D surface structures defined by the depth of both the center and the background.

**Figure 3.** Population average of a group of V2 neurons that respond better in the later part of their responses to low-contrast shape from shading stimuli than the high-contrast black-and-white stimuli.

Third, our study on shape from shading (Lee et al. 2003) demonstrated that V2 neurons, as well as some neurons in V1, are sensitive to saliency defined by 3D shape from shading but not from simple 2D black-and-white patterns. This provides a solid foundation for the contextual influence experiments in Aim 3. In addition, we have also observed that many V2 neurons show better discrimination for convex and concave stimuli than the high contrast black-and-white WA/WB stimuli. Figure 3 illustrates this finding, showing a population of V2 neurons whose later responses to the convex shape from shading stimulus (LA) are stronger than their responses to the high-contrast WA stimulus. These findings are suggestive of the idea that V2 neurons might provide an infrastructure of 3D elementary functions for representing 3D shapes. Lastly, a number of studies have begun to show both the dorsal and ventral extrastriate areas (Nguyenkim and DeAngelis 2003, Tsutsui et al 2002) are sensitive to large 3D surface planes. Hence, it is highly likely that early visual neurons are participating in the inference and coding of local surface orientation to support further integration and inference in higher visual areas.
In addition, we have carried out preliminary experiments to test the basic 3D surface orientation tuning of V2 neurons using dynamic random dot stereograms, with encouraging preliminary results. Stereo goggles with liquid crystal shutters (CrystalEyes PC; Stereographic Corp., San Rafael, CA) are used to alternatively transmit left- and right-eye views of the display monitor at a monocular frequency of 60 Hz. We have trained the monkeys to do a target detection task, and confirmed that the monkeys can perceive a binocular stereo-disparity defined target that was not perceptible based on monocular cues alone. We measured the responses of several V2 neurons to 3D sine wave gratings, different absolute disparities, and planar 3D surface orientation. Two of the cells seemed to show a tuning curve to different planar surface orientations, particularly in the later part of their responses (Figure 4).

**Figure 4.** Two V2 neurons from a preliminary experiment that show sensitivity tuning to DRDS-defined surface orientations in the later part of their responses (200-600 ms).

_Neural responses to surfaces and boundary_ (Aim 2). The interpretation of surface structures is intimately related to the perception and computation of surface boundaries. It stands to reason that the two representations should stand side-by-side in the early visual cortex. Even though virtually no direct or conclusive study has been carried out on surface representation, there have been a number of studies investigating the neural representation and computation of surface borders. The PI has recorded from V1, V2 and V4 neurons to study the neural dynamics of surface boundary formation. He has investigated how V1 and V2 neurons respond to surfaces and contours that are defined by different kinds of cues, including texture (Lee et al. 1998), color (Yan and Lee 2000), luminance (Lee et al. 1998), shading (Lee et al. in preparation), and illusory surface and contour (Lee and Nguyen 2001). These experiments gave the PI experience and insight into the basic neural responses to various visual features and more abstract perceptual constructs, such as the saliency (figure-ground highlighting signals) of surfaces and sensitivity to edges defined by different kinds of visual cues. In particular, he has found that responses to texture boundaries (Lee et al. 1998) and illusory boundaries (Lee and Nguyen 2001) can be observed in V1 neurons. The typical paradigm in these earlier experiments involves positioning the receptive fields of neurons at different parts of the stimulus. This allows us to study the spatiotemporal dynamics of neurons in response to different parts of a larger scene. This is a paradigm that we will employ in the large scene contextual modulation study proposed in Aims 3 and 4. We have seen many V1 and V2 neurons that are sensitive to surface boundaries defined by different cues (color, texture, shading, illusory contour), and also exhibit surface saliency to the different cues. However, we have not directly tested and settled the issue of whether the representation of the boundaries by individual neurons is cue-specific or cue-invariant. Aim 2 seeks to systematically investigate this issue in the context of surface representation.

_Contextual modulation in visual processing_ (Aims 3 and 4). The interpretation and representation of a local surface depends on global context. This is most evident in the example shown in Figure 16. Understanding how
the surround context can affect a neuron’s interpretation and representation of the 3D scene is therefore key to understanding the information being carried by the neuron. Given that the possible surround configurations are almost infinite, in Aim 3 of the proposal we will study contextual surround influence using a relatively standard center-surround paradigm that the PI has used in a different context. The PI has used texture stimuli (Lee et al. 1998) and shape from shading stimuli (Lee et al. 2002) in a pop-out paradigm to demonstrate how higher order information such as 3D shapes can affect the neural activity in V1 and V2. Figure 5 shows that V1 and V2 neurons respond differently to a stimulus when it is surrounded by its opposite stimulus than when it is surrounded by a similar stimulus. The former condition is similar to the “contrast” condition, while the latter is similar to the “consistent” condition in the Aim 3 experiment. This paradigm is useful to provide a first order measure of the sensitivity of the neuron to higher order contextual modulation. Experiment 1 in Aim 3 essentially is a generalization of the earlier experiment (Lee et al. 2002) to the stereoscopic 3D domain to investigate interaction within and across surfaces and depth planes.

**Figure 5.** Contextual modulation in visual processing. We found that 3D information can affect the saliency of a local representation in V1 and V2 (Lee et al. 2002). The temporal PSTH shown on the right indicates that at 100 ms after stimulus onset, the neurons respond significantly more strongly to the oddball condition than to the uniform condition. This enhancement was found to be inversely proportional to the reaction time of the monkey in performing oddball detection, suggesting that it might be a signal for perceptual pop-out saliency. The figures shown are the convex oddball pop-out condition and the uniform condition. (Lee et al. 2002).

**Receptive field identification using naturalistic stimuli** (Aim 1). The technique for recovering linear and nonlinear receptive fields of neurons in awake monkeys is important to Experiment 4 in Aim 1. Most of the existing work on receptive fields using white-noise spike-triggered averages has been done in anesthetized monkeys where it is possible to collect an enormous amount of data from each neuron. The awake behaving monkey preparation imposes a severe constraint on the amount of data that can be collected from each cell. We have worked on two techniques to overcome this problem. The first is experimentation with the chronically implanted Utah/Bionic multielectrode arrays. The second is the development of a rapid receptive field identification technique using naturalistic stimuli. In order to map the receptive field quantitatively in awake monkeys, we have developed, implemented and tested a system identification method for rapidly identifying linear and nonlinear terms in a receptive field in awake behaving monkeys. The method is a regression method that basically seeks the best function that maps the input stimulus to the response of the neurons. It is based on the methods originally developed by Marmarelis (1993) and Korenberg (1988). By using naturalistic stimuli that stimulate the cell more effectively and efficiently, we managed to recover the first order Volterra kernel (linear receptive field) and the second order Volterra kernel of the temporal transfer function of the neuron in 300 trials. These kernels are surprisingly accurate in predicting the cells’ response to time varying luminance signals within the receptive field (see Figure 6) (Romero and Lee, 1999, Romero et al. 2001).

We have also performed an initial investigation using a simulation to verify if a first order depth kernel can be obtained from shape samples drawn from the natural shape space. In this study, we present 500 surfaces,
generated at random from the space spanned by the 15 principal components of the surface shape variation in the natural environment. We assume a linear 3D receptive field. The inner product of the RF and the surface is then passed through a Poisson process to generate a spike train for 600 ms. Figure 7 shows the result from the regression technique, thus proving the technical feasibility of our approach.

Figure 6. The first and second Volterra kernels recovered with naturalistic stimuli (left column) can be used to predict V1 neural responses to novel stimuli accurately in our experiments (Romero and Lee 1999, Romero, Yu, Afshar and Lee 2002).

Figure 7. Simulation to demonstrate how regression can help us to recover the 3D receptive field. The top left panel is the original RF and the top right panel is the reconstructed RF. The bottom plot shows the principal component coefficients of the original RF (in blue) and the reconstructed RF (in green).

Multielectrode recording (Aims 1-4). We have conducted multi-electrode recording studies in V1, using custom manufactured mini-drives. In one study (Yan and Lee 2000), we routinely recorded from 5-8 neurons at the same time using the DataWave Experimental Workbench system. We have also experimented with 100-electrode array implants (manufactured by Cyberkinetics, Inc, formerly Bionic Technologies, Inc). We have had one successful implant, with 5 active electrodes from which we were able to record for weeks. We now have acquired a 128-channel Cerebus system that would allow us to record from up to 128 channels simultaneously. Most of the experiments in this proposal will rely on a dual Alpha-Omega 16 channel microdrive system which provides 2 groups of 8 independently drivable channels that will allow us to probe at two independent cortical locations sites simultaneously. Multiple electrode recording is important in this project for several reasons. First, it will increase the yield per recording session. Second, it allows us to study the relative timing of the onset of surface sensitivity across cortical locations more precisely. Third, it allows us to investigate possible ensemble coding strategies of surfaces and layers.

D. Research Design and Methods

General Techniques. Single-unit recording will be performed on V1 and V2 neurons in awake, behaving macaque monkeys. Awake monkeys will be used for several reasons. First, we believe that surface computation
and representation is fundamentally a perceptual inference process. While it is possible some inference can be done unconsciously and automatically by the neural hardware, it has been found that many neural correlates of perceptual inference are only present in awake monkeys (Logothetis, NK 1999), and that feedback is activated most strongly when perceptual scrutiny is involved (e.g. Lee et al. 2002, Hochstein and Ahissar 2002). Since our working hypothesis is that surface inference is an interactive process involving feedforward, horizontal and feedback computations, it is important to carry out these experiments in an animal that can be trained to perform a surface inference task in a normal operating state and can be trained to tell us what they see. All the recent experiments that demonstrate sensitivity to surface structures in higher cortical areas were also carried out in awake monkeys for these very reasons. See Section F for detailed considerations on the use of vertebrate animals.

Recordings will be made transdurally with epoxy-coated tungsten electrodes through a surgically implanted well (recording chamber) overlying the operculum of area V1. Eye movements of both eyes will be monitored using implanted scleral search coils and sampled at the rate of 200 Hz. 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 to V2 will be accompanied by a sudden shift in the locations of the neurons’ receptive field 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. Based on our experience, the receptive fields of the neurons at the site of these wells would be between 0° to 7° in eccentricity.

In most experiments, stimuli are presented in the form of DRDS (dynamic random dot stereogram) (Poggio et al. 1988). In a DRDS, a movie of 60 frames per second will be projected to each eye. Each frame is rendered with a different white noise random dot stereogram (RDS), even though the underlying surface structures could be static. The purpose of the DRDS presentation is to effectively average out the chance response due to different random dot visual stimuli (see Poggio et al. 1988 ). Each stimulus is presented within a 5 degree window aperture in a zero-disparity DRDS background. In most experiments, the monkey has to perform an oddball detection task. This is used to ensure stereo fusion and perception of the stereo images. It will also allow us to correlate neural responses to perceptual behavior. For stereoscopic presentation, we have used 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. We will explore other non-goggle stereoscopic presentation technologies as well.

The random dot stereograms used in the experiments are generated in two steps. First, the disparity at each point on the surface is computed from its distance from the viewer. Second, this disparity is used to displace the pixels of a white-noise image so that the disparity between the altered image and the original white-noise image matches the disparity required by the desired 3D surface. In performing this second step, we make the assumption that the variation of depth along the surface is small compared with its absolute depth (57 cm). This allows us to use an orthographic camera model. This approximation can become noticeable when generating images that contain sharp disparity gradients. The symptom is that occlusion cues are the same in both eyes, whereas in natural images, occlusion cues change as the viewpoint moves slightly. Disparity gradients can sometimes introduce texture gradients in the monocular image, but the effect is not strong enough to result in a significant deviation of the cell's response.

While some of these experiments can be done with single-unit recording, simultaneous recording of many neurons from many sites in parallel is desirable, partly for increasing the yield per experiment, but more importantly for discovering potential additional information on 3D surface encoding based on the ensemble correlated responses of the neurons. Multielectrode recording is particularly crucial in Aim 3 and 4 where the interaction of neurons within a surface or across a surface will be evaluated. Therefore, we will acquire the Alpha-Omega multielectrode (double MT) system to conduct the experiments. This device allows us to place two groups of electrodes independently at two cortical sites.
**General Experimental Procedure:** When recording from a cell, we will first test each neuron’s receptive field location and basic tuning properties (such as orientation, length and size tuning) using standard moving bar and flashing bar methods (Hubel and Wiesel 1962, 1968). V1 neurons will be classified as simple cells and complex cells using luminance sine wave gratings (Skottun et al. 1991). V2 neurons will be tested to evaluate their basic orientation tuning, equal-luminance color tuning and planar disparity tuning. After the standard tuning measures are obtained, neurons will be tested with drifting 3D sinusoidal gratings to rapidly assess the spatial frequency and orientation tuning of the neurons to 3D stimuli (Figure 8). We will test eight orientations and three spatial frequencies, with a constant temporal frequency of 6 Hz or lower. The spatial frequency sensitivity is useful for deciding the proper spatial scale of surface structure stimuli. After these rapid basic assessments of the neuron, we will conduct the planned experiments. In most experiments, unless specified otherwise, the monkeys are required to perform a perceptual task (oddball detection). This ensures that they actually fuse and perceive 3D structures in the dynamic random-dot stereograms. The monkeys will be trained to perform these perceptual tasks to achieve 90 percent correct performance rate prior to any recording experiment.

**Figure 8.** Examples of 3D sine wave of two different orientations (to be rendered in DRDS in actual experiment) for rapid assessment a neuron’s sensitivity to 3D structures different orientations and spatial frequencies.

In all experiments, the size of the surface structure stimulus will be scaled according to the size of the optical receptive field and the estimated optimal 3D spatial frequency tuning of the cell. In general, the scale of the surface structure will be slightly larger than the minimum responsive receptive field of the neuron. The receptive fields of V1 neurons in the areas we record will typically range from 0.5–1.5° in diameter, and those of V2 neurons will range from 1–2° in diameter. Hence, the diameter of the shape stimulus will be set to 1° or 1.5° for V1, and 1.5° or 3° for V2, except in the experiment that explicitly tests scale-sensitivity. The shape stimulus in Aim 1 will be rendered using a DRDS. In other Aims, the shape stimulus will be rendered in DRDS, shading or other visual cues. The underlying shape is static in each presentation. The surrounding stimulus will be a DRDS of zero-disparity. Experiments will be carried out on two monkeys, with at least 100 neurons from each area of each monkey for each experiment. Multiple electrode recording techniques will be used when allowed by the paradigm.

**Aim 1: Representation of local 3D surface structures: receptive field.**

**Rationale.** The basic objective of this series of experiments is to systematically characterize the neuronal tuning to 3D surface shapes that have been derived in a principled manner. The hypothesis we will test is that early visual neurons will show tuning to basic surface shape dimensions, in particular local surface orientation (slant and tilt), and convexity and concavity. These basic shape dimensions are derived from principal component analysis of 3D shape variations in natural environments in our preliminary computational study and thus constitute the logical dimensions to test neuronal surface shape sensitivity. A secondary hypothesis to investigate is that some early visual neurons might have linear 3D receptive fields, derived as the independent components of the 3D range data in natural scenes (Figure 2). We will perform two experiments to test these
hypotheses. The first experiment examines the tuning of neurons to several fundamental shape dimensions, in particular the first three types of principal components in natural shape variations. The second experiment attempts to estimate the 3D receptive field of the neuron using a ridge regression technique (Ryan 1997, Potetz and Lee 2003) with shape samples drawn from the natural shape space.

**Experiment 1** (Tuning to basic 3D surface structures)

In this experiment, we will evaluate the tunings of V1 and V2 neurons to basic surface structures. From our computational analysis of natural 3D scenes, we found the first three major types of surface structures can be characterized as frontoparallel planes (type 1), slanted oriented planes (type 2) and oriented convexity/concavity (type 3), as shown in Figure 9. We will test these first three types of shape components systematically. First, we will test the cell’s response to seven disparity/depth planes relative to the fixation plane (zero-disparity) to obtain the basic disparity tuning curve. Second, we will test the slanted plane rotated in eight orientations, with a 45 degree slope, to obtain the tuning to surface orientation. Third, we will test the oriented convexity and concavity in eight orientations, with a total of 16 conditions. These 31 stimuli will be presented in a block, and in the form of DRDS stimuli (dynamic random dot stereograms). Each stimulus is presented within a 5 degree aperture window in a zero-disparity RDS background.

![Figure 9](image)

**Figure 9.** First row: type 2 stimuli – slanted plane at different surface orientations and slopes. Four examples of one particular slope with center at zero-disparity are shown. Second row: type 3 stimuli – oriented convex stimuli, and isotropic convex and concave stimuli. The receptive field of the neuron will be placed at the center of the 5 degree diameter window.

In each trial, the monkey fixates during 500 ms of DRDS presentation. Three static surface shapes rendered in DRDS will be presented at three locations, one centered at the receptive field. The monkey has to detect the odd shape, which is never in the receptive field location, and make a saccadic eye movement to it. The chance response rate is 50% (as the monkey should quickly realize that the target seldom occurs in the RF location). The experimental paradigm is shown in Figure 10.

For type 2 stimuli and type 3 stimuli, if the stimuli are not centered accurately on the receptive field, the neuron’s sensitivity to absolute disparity could manifest as a tuning to surface orientation or curve surfaces. Hence additional controls are needed.

If tuning to type 2 stimuli is observed, we will test the surface orientation tuning with surfaces at two different depths (-4 cm and +4 cm) relative to the fixation plane (an example of one orientation is shown in Figure 11, first row) with a total of 16 additional conditions. The background plane is always zero-disparity. This test therefore also evaluates whether the surface orientation tuning is independent of absolute depth. If time permits, we will also test the surface orientation of the slanted planes at two additional slopes, at the depth that produces the best surface orientation tuning, with a total of 16 additional conditions.

If a tuning to type 3 stimuli is observed, we will test the curved surface orientation tuning with surfaces at two different depths (-4 cm and +4 cm relative to fixation plane). The background plane in this case is of
different disparity, while the convex shape relative to the background plane is invariant. This test therefore also evaluates whether the curved surface orientation tuning is independent of absolute depth. If time permits, we will carry out another test to evaluate whether the tuning is independent of scale by testing the curved surface orientation tuning (at the best depth) with the stimulus scaled to two other different sizes (see Figure 11, second row).

![Figure 10](image)

**Figure 10.** A typical stimulus display (the actual display will be a dynamic random-dot stereogram, but the 3D underlying shape is shown). In the experiment, the monkey will be shown three movies at three locations simultaneously, with one covering the cell’s RF. His task is to make a saccade to the location of the odd movie when the fixation spot disappears.

![Figure 11](image)

**Figure 11.** Effect of depth and scale: First row: examples of type 2 stimuli at different depths. Second row: examples of type 3 stimuli at different scales. Actual stimuli are rendered in DRDS.

**Experiment 2** (Mapping of the 3D receptive field)

This experiment seeks to test the hypothesis of 3D receptive field of the neuron by evaluating the strength and validity of the first order Volterra kernel (Rieke et al. 1997, Romero et al. 2001). The stimuli we used will be shape samples drawn randomly from the distribution of shapes that approximates the natural environment. The distribution is spanned by the 15 principal components shown in Figure 1. The stimuli drawn therefore are relatively smooth and natural looking, with a frequency of occurrence that approximates the natural occurrence of the 3D surface structures. One static shape will be presented for 500 ms each trial using DRDS. Typically, 1000-1500 shapes will be tested. The ridge regression technique will be used to estimate the 3D receptive field by regressing the mean spike counts of the neuron in response to each shape against the shape itself (see Figure
As in Experiment 1, three DRDS stimuli will appear on the screen, and the monkey has to make a saccade to the odd stimulus.

**Data Analysis and Expected Outcome:**

**Sensitivity to 3D surface structures.** Experiment 1 seeks to obtain the neuron’s tuning curves to three basic types of surface structures: frontoparallel planes, slant planes, and oriented convex and concave surfaces. We expect to find a certain percentage of V1 complex cells and V2 neurons to exhibit tuning to 3D surface structures rendered in DRDS. In this experiment, each condition will be repeated at least 20 times, and the mean spike count within a time window post-stimulus onset will be used to compute the surface orientation tuning curve. One example of such a tuning curve is shown in Figure 4 in the Preliminary Studies. Tuning responses will be evaluated across different time windows to assess the temporal onset and evolution of the shape sensitivity tuning. The statistical significance of the tuning responses will be checked using T-test and ANOVA.

For type 1 stimuli, should a cell be found to be tuned to the disparity stimulus, we expect to see the five classes of disparity tuning curves described by Poggio et al. (1988), and that some of these tuning curves can be modeled by 1D Gabor filters (Zhu and Qing 1996, Bridge et al. 2001). For type 2 stimuli, we must be careful not to misinterpret absolute disparity sensitivity as relative disparity sensitivity. As the orientation of the sloped plane changes, the depth of the center for the stimulus remains constant. However, at more eccentric locations, the depth of that location changes as the plane orientation changes. Thus, a cell that is sensitive only to absolute disparity may still respond differentially to sloped planes of different orientations if its receptive field is not fully aligned with the stimulus. We need to control for this possibility by comparing the orientation tuning curves at different absolute depths (see the top row of Figure 11). If the variance in neural response across different absolute depths is greater than the variance in response due to surface orientation, then the cell may not encode relative depth. ANOVA tests will help us to determine if a cell is actually sensitive to relative disparity. For example, figure 12c illustrates a case where the receptive field is misaligned relative to the center of the stimulus and the sensitivity of the neuron to absolute disparities is manifested as a surface orientation tuning curve. The artifactual nature of such tuning is discovered only by comparing the tuning responses at three different depths. In general, we expect to find more depth-invariant surface tuning among V2 neurons, and more depth-sensitive surface tuning among V1 neurons.

![Figure 12](image-url)

**Figure 12.** Three examples of possible surface tuning curves measured at three different depths. (a) positive depth-invariant surface orientation tuning. (b) positive depth-sensitive surface orientation tuning. (c) artifactual surface orientation tuning due simply to sensitivity to absolute local disparity. Red curve: response to near surface slant; blue curve: response to surface slant centered at zero-disparity; green curve: response to far surface slant.

**3D receptive field.** The second experiment tests the hypothesis of the 3D receptive field by estimating the first order kernel of the neuron in response to 3D input using a regression technique. In the regression analysis, the independent variable X is the input shape, and the dependent variable Y is the mean spike count of the neuron in response to that shape. Given the shape is not white noise, the cross-correlation will needed to be multiplied by the inverse of the covariance matrix in the regression, i.e. the kernel is $H = (XX^T)^{-1}XY$. 

0 45 90 135 180 225 270 315
Slant surface orientation

0 45 90 135 180 225 270 315
Slant surface orientation

0 45 90 135 180 225 270 315
Slant surface orientation
If the kernels recovered tend to be quite noisy outside the receptive field, a ridge regression technique, $H = (XX^T + \lambda I)^{-1}XXY$ (see Ryan 1997), can be used to smooth out the irrelevant noisy signal. Figure 7 illustrates how a 3D receptive field can be recovered accurately using this technique with 500 shape samples. We have used similar techniques to recover the temporal kernels of V1 neurons (Romero and Lee, 1999; Romero et al. 2001) and to find conjoint kernels in both optical and range space (Potetz and Lee, 2003). Using the spike count in different windows post-stimulus onset, we can also estimate the temporal evolution of the 3D receptive field. To verify the validity of the receptive field, we will use this recovered 3D receptive field to predict the response of the neuron to a novel 3D shape rendered in DRDS. The percentage of variance explained can be used as a measure of the validity of the kernel. We are successful in recovering the temporal kernel of V1 complex cells and in using that kernel to predict the response of a neuron to a stimulus movie (see Figure 6). If the neuron indeed has a 3D receptive field, we should be able to recover it and evaluate it against the linear 3D receptive fields predicted by independent components of 3D scenes (see Figure 2). These functions are said to form the most efficient and compact codes for representing arbitrary 3D surface shapes.

**Neural correlate of behavior.** The monkey’s performance in the oddball detection task will be evaluated both in terms of accuracy and reaction time. These behavioral measures will be compared with the onset of surface sensitivity tuning. We expect the reaction time to be proportional to the onset of the surface tuning, with a delay. Furthermore, when comparing the responses of the correct trials and the incorrect trials in experiment 2, we may find the shape tuning curve to be less significant in the incorrect trials. The neural data in the correct trials is expected to contain more shape information than those in the wrong trials. The amount of shape information in neural activity should be directly proportional to the performance accuracy of the monkey.

**Pitfalls and Solutions:**

*Attention and eye movement effect.* In the experiment, the monkey is required to fixate and fuse a pair of red dots (one for each eye) and detect the odd stimulus and make a saccade to it. Eye movement toward the RF is known to be associated with an additional attentional enhancement effect (Judge et al. 1980). To remove this additional effect due to eye movement, the oddball will always appear in one of the non-receptive field locations, so the monkey will always make saccade away from the RF. Another artifact could be introduced by vergence eye movements induced by changes in the disparity of visual cues at the fovea or at the surround. This as well as microsaccades can cause a large variation in the responses of early visual neurons with small receptive fields (Leopold and Logothetis 1998). Therefore, it is important to monitor the eye movements of both eyes carefully during each trial. Small eye movement can potentially blurred the 3D receptive field that is recovered, though fixation has been shown to be rather stable (Read and Cumming 2003). This might be less important for V2 neurons. From our experience in estimating V1 kernels, the eye movement effect seem to be less important than we anticipated. The recovered kernels can be used to predict the neural response remarkably well (see Figure 6).

*Sensitivity to visual cues.* V1 and V2 neurons are sensitive to small differences in visual cues in the random dot stereogram. These differences could produce a significant variation in the responses to the different conditions. To ensure the difference in neural responses is due to the difference in the underlying 3D surface shape, rather than the random visual features in the RDS, a large number of RDS will be presented in the form of a dynamic random dot stereogram movie. With 60 frames per second, the chance responses due to random noise patterns should be averaged out to produce a more robust measure of surface shape tuning. To further safeguard against this artifact, we will, for some neurons that show sharp tuning, repeat the experiment with one of the monkeys’ eyes covered. The surface tuning curve should vanish with monocular stimulation.

*Perception of stereo images.* Our effortless perception of the DRDS does not guarantee that the monkey will see the stereo image. To have some confidence that the monkeys actually fuse and perceive the stereo images, we will ask the monkeys to perform a odd-stimulus detection task that should require them to perceive the underlying 3D shapes in all three locations in each trial. We found closing one of our eyes will render it impossible for us to perform the task. Nevertheless, we need to be vigilant about potential alternative tricks the monkeys might use to perform the task without seeing the 3D shape in stereo.

*Variation due to position.* The stimuli will be centered on the receptive field of the neurons. Misalignment
between the receptive field and the stimulus can potentially produce an artifactual surface orientation tuning due to the neuron’s sensitivity to absolute depth. Testing the surface orientation tuning curve at different absolute depths will help to detect such a problem and rule out the artifacts.

**Linearity of the kernel.** The linearity of the 3D kernel is a hypothesis to be tested rather than an assumption. Experiment 1 is independent of this hypothesis. Experiment 2 is designed to test it. There is no guarantee that V1 or V2 neurons could be characterized in terms of a linear 3D receptive field despite its simple theoretical appeal.

**Amount of data required to recover kernels.** Regression analysis in noisy situations requires a certain amount of data to estimate the input-output relationship (receptive field) of the neuron. The amount of data needed depends on the amount of noise. It is therefore not feasible to carry out Experiments 1 and 2 in the same neuron in most cases. We have designed the experiments to be modular and separable. Given the labor intensive nature of the recording of each cell, it is important to employ multiple-electrode recording techniques to improve the yield of each recording session.

**Aim 2: Representation as a function of visual cues.**

**Rationale.** 3D surface structures are abstract and have to be inferred indirectly from various cues available to our visual systems. These cues include binocular disparity, texture, shading and motion. Combining multiple bottom-up cues is an effective way of disambiguating surface interpretation and representation (compare Figure 13b, c and d). To study the role of cues in surface inference, we propose experiments to determine whether the surface tuning/representation in the early visual areas is cue-specific, with multiple surface representations, or cue-invariant, with a unitary invariant surface representation. If the representation is found to be defined by multiple cues, we will study the contribution of each of the cues in isolation and in combination.

**Experimental Procedure**

The basic experimental paradigm is similar to Experiment 1 in Aim 1. Three stimuli are presented on the screen for 500 ms while the monkey fixates. The monkey makes a saccade to the odd stimulus when the fixation spot disappears to complete the task.

**Experiment 1:** Cue-specificity.

To answer whether surface structure tuning is cue-specific or cue-invariant, after testing the basic properties of the neurons (as described in the General Experimental Procedures) we will test the sensitivity tuning of the neurons to the two different types of surface structures as described in Aim 1: planar surface orientation (type 2 stimuli), and oriented curved surfaces (type 3 stimuli) rendered either in optical cues (texture, shading) and/or stereo cues (DRDS). There are 9 conditions that comprise the type 2 stimuli: 1 frontoparallel plane and 8 slanted planes of different surface orientations; and 18 conditions for the type 3 stimuli: 1 spherical convexity, 1 spherical concavity, 8 oriented convex surfaces and 8 oriented concave surfaces. These 27 surface structures will be rendered in the two different cues, optical (shading and texture) cues (Figure 13d), and dynamic random dot stereograms (Figure 13e), with a total of 54 conditions. For each underlying shape, multiple texture renderings will be presented, 5 images per trial, each for 150 ms. With 20 trials, 100 different texture patterns will be presented and their responses will be averaged to minimize the variance due to random visual features.

**Experiment 2:** Cue-interaction.

If the optical cues (shading and texture) produce sharp surface structure tuning curves in Experiment 1, whether or not the cell exhibits a tuning curve to DRDS-defined surfaces, we will proceed to ask two questions: First, could the representation be further decomposed into separate texture and shading channels? Second, would introducing additional binocular disparity based on the texture and shading features further sharpen the representation? It is notable that binocular disparity defined by texture and shading features and binocular disparity defined by random dot stereograms can be carried by separate channels (Roe and Ts’o 1995, Ts’o et al. 2001, von der Heydt et al. 2000).

To answer these questions, we will choose the shape variation dimension (type 2, type 3, or convex versus concave) to which the cell exhibits surface structure sensitivity tuning and use at most 4 parametric conditions
to sample the surface structure tuning curve. We will render the underlying shapes in the following different ways: (1) shape from shading (Figure 13b), (2) shape from texture (Figure 13c), (3) shape from shading and texture (Figure 13d), (4) shape from texture and stereo, (5) shape from shading and stereo, (6) shape from consistent shading, texture and stereo (Figure 13f) and (7) shading-texture cues in conflict with the stereo cues. For this last case, an example would be shading-texture indicating a convex surface while stereo information indicates a concave surface.

If the neurons only respond to the DRDS-defined surface structures, but not to the texture-shading defined stimuli, the remaining question is whether the cells exhibit surface tuning to the shape defined from consistent shading, texture and stereo, where the stereo is defined by disparity in shading and/or texture features rather than RDS, i.e. renderings (4) and (5). If the result is positive, then we will test rendering (6) and (7) to further study cue interaction.

![Image of surface structures rendered in different visual cues](image_url)

**Figure 13.** 3D surface structures rendered in different visual cues. (a) underlying 3D surface; (b) shading cue; (c) texture cue; (d) shading and texture cues; (e) pure stereo cue; (f) stereo, shading and texture cues combined. All stereo images are arranged for cross-fusing.

**Data Analysis and Expected Outcome.**

The images in Figure 13 illustrate the powerful effect of cue interaction in surface inference. The interpretation of shape from texture or shading alone is quite ambiguous, but becomes compellingly clear when multiple cues are combined. Hence, we expect cooperation among cues will help to disambiguate surface interpretation, thus sharpening the surface sensitivity tuning curves. Competition, on the other hand, will provide conflicting evidence that causes ambiguity in surface interpretation. This will likely result in a flattening of the surface tuning curves and emergence of multiple peaks. Furthermore, the temporal development of the tuning curves for the different cues might have different time courses. Optical cues are expected to act more quickly but less reliably, while random-dot stereo cues are expected to act in a slower but more robust fashion. When stereo cues and optical cues are combined, one might observe a rapid emergence of a shape sensitivity tuning curve which would sharpen over time. When stereo cues and optical cues appear in conflict, the tuning curve might be flattened or even experience a reversal as time evolves. To evaluate these predictions, the tuning curves (similar to those in Aim 1, Experiment 1) will be computed from average spike counts in different time windows for each type of cue rendering. Peri-stimulus time histograms (PSTH) for each condition will also be computed, smoothed with a 10 ms time window, and compared against the other conditions to assess the temporal evolution of shape sensitivity for each cue rendering.
The data from these experiments will address a long-standing debate on whether surface inferences based on different cues are weakly coupled, done in different channels in early visual areas and combined in a later stage (Marr 1982, Bruno and Cutting 1988; Dosher et al. 1986, Maloney and Landy 1989) or tightly coupled in a unified inference process (Yuille and Bulthoff 1996). If the inference based on different cues is strongly coupled in a single inference process, we would expect to see many cells to show consistent surface tuning curves for the different cues, particularly in the later stage of their responses. It is however likely that we may also find the shape tuning curves of some neurons to be specific to only one or two visual cues. There are at least two possible scenarios from which such an observation may arise. First, the cue-specific neuron performs inference within its module and then the results are integrated by the cue-invariant neuron as in a weakly coupled system (Marr 1982). In this case, one might see some causal relationship between the two types of neurons. This can potentially be determined by a cross-correlation study (Nicolelis 1998) if we record from the two types of cells simultaneously. Second, the inference can be strongly coupled but the mechanisms are distributed. In this case, all the cells compute surface information in parallel, it might be difficult to discern the difference in timing or causal relationship between the different classes of neurons.

**Pitfalls and Solutions.** In addition to pitfalls similar to those in Aim 1, there are some additional concerns specific to the cue-invariant experiments.

**Too few V1 and V2 neurons exhibit surface tunings.** Recent literature (Thomas et al. 2002, Von der Heydt et al. 2000, Lee et al. 2002; Bakin et al. 2002, Anzai et al. 1999a,b,c) has provided many clues and indirect evidence that V1 and V2 neurons are sensitive to 3D surface structures and not just absolute disparity. In our preliminary experiments, we have found some V2 cells demonstrated sensitivity to surface orientation defined in RDS or DRDS (see Figure 4). Furthermore, recent studies in the extrastriate cortex indicate 3D surface sensitivity exists in many dorsal and ventral extrastriate cortical areas. Therefore, we are optimistic both on empirical and theoretical grounds. In the event that only a small percentage (< 20 percent) of V1 and V2 neurons exhibit surface sensitivity tuning to prevent Aim 2 (cue-integration in surface inference) from being effectively carried out, we will consider performing Aim 2 experiments V3/V3a as well. V3/V3A is easily accessed through the same recording well, and has been shown to have activity in response to surface stimuli in some fMRI studies (Tsao et al. 2003) and is known to harbor disparity columns (Adams and Zeki 2001). The PI has not recorded systematically from V3/V3A before, but his next door colleague, Dr. Carol Colby, has worked in these areas (Nakamura and Colby 2000, 2002). Therefore, it should not be difficult to extend the study to V3/V3A if necessary. Our current plan however is to study V1 and V2 first and V3/V3A in the future.

**Shading Stimuli.** The interpretation of shading strongly depends on an assumption of lighting direction. Even though there is a intrinsic bias in our visual system to assume lighting from above, such a prior is not completely dominating. If the tuning curves depending on shape from shading turn out to be random, we will consider training the monkeys for 20 sessions to view stereo 3D shapes shaded with lighting from above, to establish a stronger lighting from above prior. In order to dissociate the learning effect and default processing, we would need to perform the recording experiments in two stages: pre-training and post-training.

**Size of the receptive field.** V1 and V2 receptive fields tend to be small, and this limits the number of texture elements that can be placed in the receptive field. Perception of the texture surface likely involves horizontal integration of information of neurons covering different parts of visual space. Because horizontal interaction takes time, it is not completely certain that optical cues would necessarily be faster than stereo cues in defining a surface percept. The dynamics of temporal responses of surface shape sensitivity are expected to be more complex than we described above.

**Effect of texture elements** The smaller receptive fields of early visual neurons make them very sensitive to the precise placement of the texture elements. Coupled with microsaccades, this can produce a significant variance in the response of neurons that might not be associated with surface structures. One solution is to present many random texture renderings of the same underlying surfaces so that the response due to local texture feature differences can be averaged out, in very much the same way as in DRDS.

**Aim 3: Contextual influence of the surrounding surfaces.**

**Rationale.** The inference of 3D surface structures from bottom-up visual cues is often ambiguous in
the absence of global contextual information. To understand how surface interpretation is constrained and disambiguated by surround context, and how a local surface is represented when situated in a larger scene, we propose to study the activity of V1 and V2 neurons in response to a local DRDS-defined or shading-defined surface in several different global contexts of surrounding surfaces.

**Experimental Procedure.**

**Experiment 1 (Influence of the smooth surfaces in the surround)**

In this experiment, we will assess the DRDS-defined surface shape tuning in the context of surrounding surfaces (type 2 stimuli). The hypothesis to test is that a consistent background surface will sharpen the surface orientation tuning curve of the neuron because the consistent surrounding information can be integrated to improve the surface inference. On the other hand, a contrasting surround could recruit stronger saliency/attentional signals but distort and flatten the surface orientation tuning curves. We will test the surface orientation tuning in four conditions: when the center plane is embedded in the surround plane (consistent condition), when the center plane is orthogonal to the surround plane (contrast condition), when the center plane is parallel to the surrounding plane (same surface orientation), but situated at different depths, either in front or behind the background plane (see Figure 14). The stimulus will be defined in DRDS and is arranged in a center-surround organization, with the center diameter equal to 3° visual angle and the surround spanning the full screen. The monkey will perform only the fixation task. There are therefore 36 (9 x 4) conditions (8 surface orientations + 1 flat plane, 4 conditions) for the basic experiment. We will use the Alpha-Omega multielectrode drive to record from two groups of neurons simultaneously. One group of electrodes will record from cells whose RFs are inside the center surface, while a second group of electrodes will monitor cells whose RFs are located in the surrounding plane in all the four conditions. If time permits, we will repeat this experiment with the center stimulus’ diameter increased to 5° visual angle to test the hypothesis that the surface orientation tuning curves are sharpened with an increase in surface area up to a certain extent.

**Figure 14.** The four conditions to be tested for each oriented surface plane, shown in 1D, with the rough prediction of the surface orientation tuning curves on the left. The actual stimuli in the experiments will be rendered in 3D dynamic random dot stereograms. The number in cm is the distance away from the viewer relative to the fixation (zero-disparity) plane. The bottom row shows the 3D landscape rendering of the four conditions. In all cases, the stimulus over the neuron’s receptive field (i.e., the center plane) is identical.

**Experiment 2 (Interactions of surface elements across depth planes)**
In this experiment, we will investigate the response of the neurons to a curved-surface object in the receptive field with many curved-surface objects in the surround. The hypothesis to test is that the action of the surround is operated at the levels of the surface or depth plane. The secondary hypothesis is that the convex-concave tuning will sharpen in a consistent background and both the convex and concave signals will be enhanced in non-consistent background as a signal of saliency. While Experiment 1 investigates the surround effect in the context of smoothed and continuous surfaces, this experiment investigates the surround interaction among objects within and across depth planes. Two types of stimuli will be used to cover both possible stereo channels (one based on DRDS and one based on forms) in V2. The first type is a DRDS-defined surface with many convexities and concavities embedded in it. The second type is an array of stereo- and shading-defined convex spheres and concavity floating in the void. Figure 15 shows the rendering of the stereo-shading stimuli of a convex sphere on the RF of the neuron with four kinds of surround configurations: consistent, contrast, in-front and behind. In total, there are 16 conditions (convex+concave shading+RDS 4 configuration). For the second type of stimulus, we need to add additional controls, White-Above and White-Below (WA/WB) patterns placed in the four depth configurations (8 conditions).

**Figure 15.** (a) The four conditions (shown in 1D) to be tested for an convex surface on the RF of the neuron. (b) contrast condition to be rendered in DRDS. (c) The four conditions rendered as four pairs of stereo-shading images. (d) 2D patterns arranged at the same depth, and 2D patterns arranged in different depths. All stereo images are arranged for cross-fusing.
**Data Analysis and Expected Outcome.**

For experiment 1, we will compute the surface orientation tuning curves using the same method as in Aim 1 based on the cell’s response to the center stimulus in the four surround configurations. If the local representation is completely independent of the surround, then the tuning curves would be invariant across the four configurations. However, we do predict considerable contextual modulation. One possible scenario of expected outcomes is shown in Figure 14. A consistent background should produce the sharpest tuning curve because it makes more consistent information available to help the neuron in estimating the surface orientation (Figure 14, red curve). A contrasting background will increase the activity of the neuron because contrast increases saliency. However, the inconsistent surround likely will distort or flatten the surface orientation tuning curves (Figure 14, black curve). When the center plane is in front of (pop-out) or behind (pop-in) the surrounding plane of the same surface orientation, the surface tuning curve will be equally as sharp as the consistent condition (or nearly so), but will be sharper than the inconsistent one. When the center stimulus is increased from 3° to 5° in diameter, the tuning curve will be sharper but the magnitude of response will be less.

The pop-in and pop-out configuration might be accompanied by an additional effect: the plane in front might be considered as a figure, associated with enhancement, while the plane behind could be considered as background and associated with suppression (Zipser et al. 1996). However, these predictions are still controversial (Rossi et al. 2001)

Experiment 2 evaluates the hypothesis that surface elements within the same depth plane exhibit stronger interactions. The prediction is similar to that of Experiment 1, i.e. a consistent background will enhance the neuron’s discriminative sensitivity between convex and concave shapes, but reduce the magnitude (saliency) of the raw response to the RF stimulus, while a contrasting background will decrease convex-concave sensitivity but increase the magnitude of the raw response. Objects that are in front or behind the depth plane of other objects will not be inhibited (hence the stronger raw response) but their convex-concave tuning curves would not be compromised either. The WA/WB stimuli are particularly important in this context. Based on our earlier experiments, the WA/WB tuning is expected to be weak in V2, and should not show difference between the contrast and consistent conditions. However, since the pop-out and pop-in conditions position the 2D flat stimulus (imagine it as a column sticking out) at a different depth plane than the surround, we would expect to see pop-out enhancement (relative to the consistent condition) for these stimuli!

To evaluate possible ensemble neural codes (Shadlen and Movshon 1999, Singer 1999, Hopfield 1995, Salinas and Sejnowski 2001), we will analyze the correlation between neurons whose RFs are situated in the same surface, as well as between neurons whose RFs are situated in different surfaces. The consistent, pop-out and pop-in slant surface conditions provide a particularly interesting opportunity to evaluate the possibility that correlated activity or synchrony might serve to “bind” surfaces together. If synchrony or correlated firing is used for binding or signaling the same surface, we expect cells with similar surface tuning covering the center and the surround will synchronize in the consistent condition, even though they are experiencing different absolute disparities. On the other hand, the same cells in the pop-out (in-front) and the pop-in (behind) conditions should not synchronize, even if they experience the same absolute disparities. Recording from 8 neurons at both the center and the surround locations in the conditions tested in Experiment 1 will be useful for evaluating this hypothesis. The same technique can be used to evaluate ensemble coding in Experiment 2, except the hypothesis becomes whether synchrony can be detected in neurons covering objects in the same depth plane despite being absent when they are in different depth planes.

**Pitfall and Solutions:**

The single-unit recording portion of both experiments is relatively simple and, based on our earlier study (Lee et al. 2002), is likely to furnish positive and interesting results. In addition, it will furnish novel information on the center-surround interaction at the surface or depth plane level. The multielectrode recording part of the experiments is challenging both in execution and in analysis. In execution, the difficulty is in finding a pair of neurons at two locations that is consistent in surface orientation coding. The Alpha-Omega multielectrode...
drive allows us to sample 8 neurons per location, which will increase the likelihood of success. There is a rich
literature of neural ensemble synchrony and its analysis (Kreiter and Singer, 1992, 1996, Usrey and Reid 1999,
Mehta et al. 2002, Nicolelis 1998), which we will not go into here in detail. In short, the PI has a long-standing
collaboration with statistician Rob Kass at Carnegie Mellon who has been working on some of the more diffi-
cult statistical issues concerning the analysis of synchrony. An important feature of our experimental paradigm
is that the existence of synchrony will be deemed meaningful on one condition only if there is an absence of
synchrony in the control condition.

Aim 4: Contextual influence of the global boundary shapes.

Rationale. An important contextual constraint for determining surface perception is the shape of the oc-
cclusion boundary of the surface. The stimuli in Figure 16 are characterized by the same luminance sine wave
grating. Yet, as the boundary shape varies, very different interpretations emerge of the surface orientation at
each point as well as the illumination directions. This is because shading is an ambiguous cue. The shape of
the boundary provides a powerful constraint to disambiguate the surface interpretation. In this Aim, we seek
to understand how boundary shape at a distance can exert an influence on the interpretation and representation
of surfaces. The hypothesis to test is that surface estimates are first determined at the occlusion border and
propagate into the surface interior.

Figure 16. Illustration of the influence of border shape on surface and lighting interpretation. The lumin-
nance sine wave grating is identical in all the images. As the boundary shape varies parametrically, the
interpretation of the illumination direction is changed accordingly, causing changes in the interpretation of
surface orientation associated with each luminance pattern.

Experimental Procedure.

Experiment 1 (Influence of boundary shape)

Here we test the hypothesis that boundary shape remote from the receptive field can exert an influence on
the surface interpretation. In this experiment, after preliminary testing on each neuron’s basic edge orientation
tuning and surface orientation tuning, we will test its response to the two stimuli in Figure 17 by shifting the
two images so that the receptive field is situated at the 6 labeled locations. The locations are labeled 1, 2, 3 and
1', 2', 3', respectively. Location 1 and location 1' have the same luminance pattern in monocular display, as
do 2 and 2', 3 and 3'. The responses of the neuron to these corresponding locations are to be compared. Note
that the labeled locations in these two particular stimuli can test the neurons only in three surface orientations,
as shown in the blue arrows of Figure 17. In order to cover all eight surface orientations, these two stimuli
are rotated in 90° steps to generate three other pairs of images. Four pairs of stimuli, each with six locations,
results in 24 conditions to be tested.

The receptive field of the neuron will be centered between the two bounding borders as shown and its edge
should be at least 3° of visual angle away from any boundary border. A cautionary note is that the images
are bistable monocularly, i.e. each can be interpreted in two different ways with different assumptions on
lighting direction. When rendered in binocular stereo pairs, the 3D surface structures in the images become
unambiguous. To be sure what surface structure the monkeys are seeing, we will train the monkeys for 20
sessions to view the 3D structures using a binocular display, in which the monkeys will be asked to push or
pull a lever depending on which 3D structure they perceive. During recording experiments the monkeys will
be shown the eight images only in binocular and monocular display, and the monkey has to push or pull the
lever to indicate what 3D structures they are seeing.

As a control to ensure that the observed effect is due primarily to surface representation, and not directly a
result of the boundary curve near the RF, we will measure the neurons’ responses when their RFs are located
at the black circle locations. In those locations, the RFs are seeing the same boundary shape above them as
when they are in the red circle locations, yet in the absence of shading information and a supporting contour,
y they are not perceived as being on the surface, and should not exhibit a sensitivity to surface shape.

Figure 17. Illustration of the influence of border shape on surface representation. Red circles are placement
locations of the receptive field of the neuron. Location 1 and location 1’ have the same luminance pattern,
but different surface orientation interpretation as indicated by the blue arrows. A neuron that is tuned to a
surface plane slanted upward (up arrow) will respond better at location 2 than at location 2’, and better at
location 1’ than location 1 etc.

**Experiment 2** (Constraint propagation mechanisms)

The hypothesis to test is that the surface estimate originates at the occlusion border and propagates into the
surface interior. In addition, neurons covering the same surface may exhibit synchrony or correlated activity. If
the results of Experiment 1 are positive, i.e. that a surface orientation sensitive cell shows a systematic shift in
surface interpretation with the change in boundary shape, and if time permits, we will proceed to evaluate the
second hypothesis. This will be to test if the surface constraint is mediated by horizontal propagation. We will
insert a second group of eight electrodes to another cortical site in the same visual area. The stimuli chosen
are constrained to two sets (a stimulus and its 180° rotation). For example, for a neuron sensitive to a plane
slanting up (up arrow), we can only choose the two stimuli shown in Figure 17 or their 180° rotations. This
provides a severe constraint on where we can put the second group of electrodes. They have to go to a cortical
location where the receptive fields are spatially aligned in a vertical column with the first cell. Assuming we
have mapped out the retinotopic maps of V1 and V2, this is not difficult. Among the neurons recorded by the
second group of electrodes, the analysis will be easier if at least one of them has the same surface orientation
tuning as the positive neuron in the first group.

Once the two neurons or two sets of neurons are in place, we can record from them as we shift the images
vertically or horizontally in successive trials. Note the image is stationary during each trial for 500 ms. At
least three horizontal shifts (across columns) and four vertical shifts (across rows) will be examined (as shown
in Figure 18). The three vertical shifts are labeled as 1, 2, 3 and 1’, 2’, and 3’ as in Experiment 1. The four
configurations due to vertical shifts are shown: (1) the RFs of one group of neurons (group A) are placed in the
blue circle, while the RFs of the second group are placed in the red circle; (2) group A is in the red circle and
group B is in green circle; (3) group A is in the green circle and group B is in the brown circle; (4) group A is
in the brown circle and group B is below the brown circle. In configuration 1, group A is nearer a border than
group B, while in configuration 2, group B is nearer a border than group A. In both of these configurations, the
neurons share the same surface, while in configuration 3, group A is in the surface, while group B is outside
the surface in the background. In configuration 4, both groups of neurons are in the shapeless background.
Comparing these four horizontal shifts are important as they provide crucial information about the relative timing of the shape sensitivity onset, as well as possible ensemble coding for surfaces. Given that synchrony detection and latency detection both require a larger number of trials, we will try to obtain at least 50 trials per condition. To avoid losing data due to losing a cell in the middle of the recording, we will record from four shifts along one vertical column at a time.

Figure 18. Locations of the electrode group placement in Experiment 2. The two groups of electrodes will be placed in the same vertical column (relative to this rotated version of the stimuli) in the following three configurations: (1) group A’s RFs in the blue circle, group B’s RFs in the red circle; (2) group A’s RFs in the red circle, group B’s RFs in the green circle; (3) group A’s RFs in the green circle, group B’s RFs in the brown circle. The experiment will repeated for at least three different luminance patterns, i.e. the three vertical columns, 1, 2, 3 or 1’, 2’ and 3’.

Data Analysis and Expected Outcome.

If the boundary shape can indeed disambiguate and change surface interpretation and representation in the early visual areas, the change in border configuration from image A to image B should produce a shift in the neural response that is predictable from the surface orientation tuning of the neuron. Consider the stimuli in Figure 17, where location 1 and location 2’ can be seen to slant in a 45° direction, while location 1’, location 3’ and location 2 can be seen to slant to 90° (up arrow), and location 3 be seen to slant to 135°. Suppose a cell prefers surface that is oriented to 90°, then we expect it will respond better in location 2 of image A than in location 2’ of image B, and that it will respond better in locations 1’ and 3’ than it will in location 1 and 3. Given that we will rotate the images to cover all eight surface orientations, we should be able to detect such a change in response to the appropriate stimulus. It is possible that a neuron is tuned to both surface orientation and luminance gradient, so that it is only sensible to compare the response at each pair of corresponding locations, but not across locations. Comparison across locations with different luminance patterns can be made using some relative measure such as $R_1/R_{1'}$.

The prediction of experiment 2 is that the relative onset time of evidence of surface interpretation should emerge earlier when the RFs are closer to the border, and later when they are further away. To evaluate this prediction, we will compare the onset time of the surface shape sensitivity of the neurons on a trial by trial basis in the different conditions. Given neurons might have different response speeds, we can only compare relative delay rather than absolute delay between onsets. We will compute the delay as the difference in time to the onset of sensitivity between neurons in group A and B. The horizontal propagation hypothesis predicts the propagation time from border to interior is given by $(D2 - D1)/2$, where $D1$ is the onset time of neuron A minus the onset time of neuron B in configuration 1, and $D2$ is the onset time of A minus the onset time of neuron B in configuration 2. We will compute the correlated firing of neurons A and B (by cross-correlation normalized by shuffling) in configurations 1 and 2, and compare that to the correlated firing of the two neurons in configuration 3. If synchrony or correlated firing is associated with neurons being on the same surface, then synchrony should be stronger in configurations 1 and 2 than in configuration 3. However, this synchrony test is valid only for for the configurations along columns 1 and 1’ where the luminance is roughly the same between
the surface and the background. For the result to be solid, simultaneous recording of RFs covering different luminance patterns within the surface will also need to be evaluated. Synchrony detection and latency detection both require a large number of trials to be accurate. For this reason, we will try to obtain at least 50 trials per condition. We will record four conditions along one column before recording another another column in case the experiment is prematurely terminated for a variety of reasons.

**Pitfalls and Solutions:**

*Distant border exerts no effect.* There could be no change in neural response due to the change in border configuration far away from the receptive field. The lack of contextual influence at that particular spatial scale in a particular area might suggest that the representation in the early visual areas might be less abstract than we predicted. Earlier experiments in binocular rivalry, illusory contour and figure-ground segregation have suggested that a certain percentage of neurons (10% in V1, 20% in V2, and 40% in V4) are modulated by perception. Therefore, it is quite reasonable to expect to see a similar percentage of neurons in these areas to exhibit a neural correlate of surface perception that is modulated by distant contours. In the event of a negative result, we can turn our study to V3/V3a.

*Boundary directly stimulating the RF.* The experiment depends heavily on our estimate of the receptive field size. When the RF is under-estimated, the border could be too close to the receptive field, and the effect might not be considered “contextual”. In this case the surface tuning effect might be due to direct boundary stimulation. Configuration 3 in Experiment 1 and the black circle location in Experiment 1 are designed to rule out the contour alone as sufficient to produce the effect.

*Estimate of latency and synchrony.* Statistically it is difficult to estimate the exact timing of a certain signal, particularly on a trial by trial basis. One solution is to repeat the same condition many times (e.g., 50 or more) and use the control condition to build a model of instantaneous firing rate. The onset time is then estimated as the moment when the spike train becomes significantly different from what is predicted from the instantaneous firing rate in the control condition (Ventura and Kass). Estimation of significant synchrony must take chance synchrony into account. A standard approach is to normalize the cross-correlogram by a cross-correlogram that is produced by shuffling the spike trains across different trials, which predicts chance cross-correlation (cites).

*Ambiguous surface interpretation.* The surface interpretation of the stimuli is ambiguous in monocular display. It is a potential problem that the monkeys can see the display in two different ways. This can be solved by training the monkeys and asking them to perform a task to tell us what they perceive, which would allow us to more confidently determine the neural correlate of the monkey’s perception in a more precise, trial by trial basis.

**Time Table**

<table>
<thead>
<tr>
<th>Aims/Expt</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aim 1, Expt 1</td>
<td>train/record</td>
<td>record</td>
<td>analysis/paper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aim 1, Expt 2</td>
<td>train/record</td>
<td>record</td>
<td>analysis/paper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aim 2, Expt 1</td>
<td></td>
<td>record</td>
<td>record</td>
<td>analysis/paper</td>
<td></td>
</tr>
<tr>
<td>Aim 2, Expt 2</td>
<td></td>
<td>record</td>
<td>record</td>
<td>analysis/paper</td>
<td></td>
</tr>
<tr>
<td>Aim 3, Expt 1</td>
<td>train/record</td>
<td>train/record</td>
<td>analysis/paper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aim 3, Expt 2</td>
<td>train/record</td>
<td>train/record</td>
<td>analysis/paper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aim 4, Expt 1</td>
<td></td>
<td>record</td>
<td>record</td>
<td>analysis/paper</td>
<td></td>
</tr>
<tr>
<td>Aim 4, Expt 2</td>
<td></td>
<td>record</td>
<td>record</td>
<td>analysis/paper</td>
<td></td>
</tr>
</tbody>
</table>

Three monkeys will be used for Aims 1 and 2, and an additional three monkeys will be used for Aims 3 and 4. This includes year 5, in which occasional recording and data collection would still be required.
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. Six 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 ten 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 et al., ’92). 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 supplemental 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 visuomotor 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 sacrificed 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 for the brain. This method is consistent with recommendations of the Panel on Euthanasia of the American Veterinary Medical Association.
REFERENCES:


