

# Neural Models of Binocular Depth Perception

S.R. LEHKY,\* A. POUGET,† AND T.J. SEJNOWSKI†

\*Laboratory of Neuropsychology, National Institute of Mental Health, Bethesda, Maryland 20892;

†Computational Neurobiology Laboratory, The Salk Institute, San Diego, California 92138;  
and University of California, San Diego, California 92093

We have known since Wheatstone (1838) that disparities between images presented to the two eyes induce a strong sensation of depth. More recent experiments with random-dot stereograms have shown that disparity is a sufficient cue for stereopsis (Julesz 1960, 1971). Disparity-tuned neurons in visual cortex were first demonstrated in the cat (Barlow et al. 1967; Nikara et al. 1968; Pettigrew et al. 1968) and later in the macaque monkey (Hubel and Wiesel 1970; Poggio and Fischer 1977; Poggio and Poggio 1984). Similar disparity mechanisms probably exist in the human visual cortex.

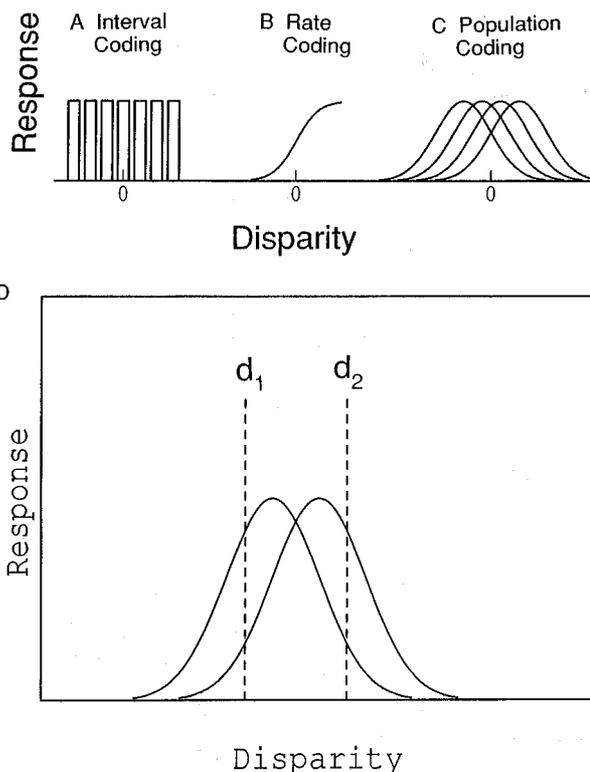
We have developed a model for the representation of stereo disparity by using a population of neurons that is based on tuning curves similar in shape to those measured physiologically (Lehky and Sejnowski 1990). The model predicts depth discrimination thresholds that agree with human psychophysical data only when the population size representing disparity in a small patch of visual field was in the range of about 20–200 units. This population model of disparity coding at a single spatial location was extended to include lateral interactions, as suggested by psychophysical data on stereo interpolation (Westheimer 1986a).

Disparity is a measure of depth relative to the plane of fixation. Additional sources of information are needed to estimate the distance to the point of fixation, such as that provided by eye vergence, vertical disparity, and accommodation. We have developed a simple model that combines disparity information in a distributed representation and vergence information to compute the absolute depth of objects from the observer (Pouget and Sejnowski 1990). Although these are models of binocular vision, a number of the ideas presented here generalize to the representations of other sensory cues.

## Representations of Disparity

In a local representation, disparity is unambiguously represented by the activity of a single neuron. An example of this is shown in Figure 1A, where the value of disparity is indicated by which neuron fires. To cover the entire range of disparities with high resolution, there must be a large number of such narrowly tuned, nonoverlapping units. This form of local representation is called interval encoding and has been used almost universally in models of stereopsis (Marr and Poggio 1976; Mayhew and Frisby 1981). A second form of

local representation is rate encoding (Fig. 1B). Here, a single unit codes all disparity values by its firing rate. As disparity increases, the firing rate of the unit increases monotonically. Examples of models that have used a rate-coded representation of depth are Julesz's dipole model (Julesz 1971) and the model of Marr and Poggio (1979). A third type of encoding, which we use here, is a distributed representation, or population



**Figure 1.** Three methods of encoding disparity and an example of discrimination in a population code. (A) Interval coding: A separate unit is dedicated for each disparity. (B) Rate encoding: Disparity is encoded by the firing rate of a single neuron. (C) Population coding: Disparity is encoded in the pattern of activity in a population having broad, overlapping disparity tuning curves. (D) Population code based on two idealized disparity tuning curves. As stimulus disparity is changed (e.g., from  $d_1$  to  $d_2$ ), the response of one unit goes up and other goes down, as indicated by the intersection of the dashed lines with the tuning curves. The changes in activities of all units in the population are combined to see if the total change is statistically significant relative to the noise in the units. If so, then the change in disparity is considered perceptually discriminable.

code, in which disparity is encoded by the pattern of activity within a population of neural units, each broadly tuned to disparity and extensively overlapped with each other (Fig. 1C). In a distributed representation, the activity of a single unit is ambiguous.

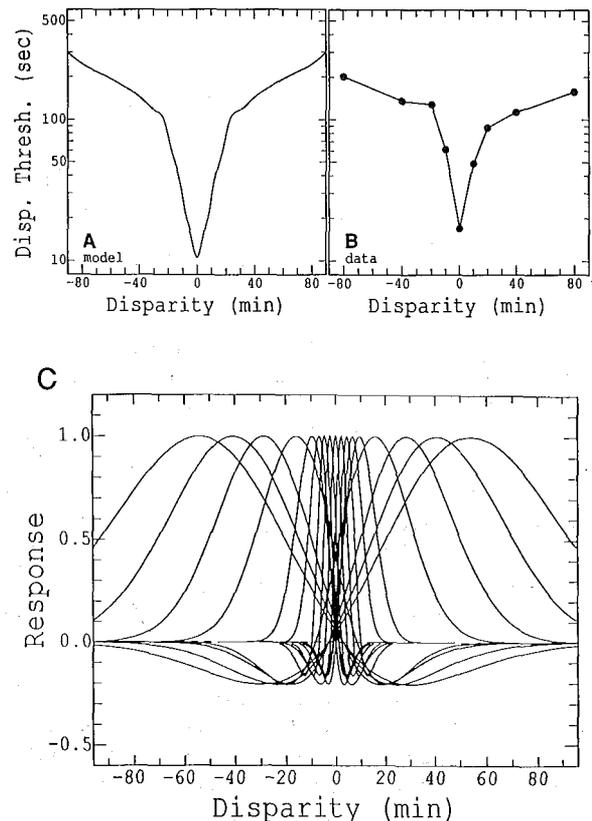
### Neurophysiological Data

Poggio et al. (1985, 1988) have reported that neurons in monkey visual cortex are grouped into three classes on the basis of their responses to disparity: (1) the "near" neurons, broadly tuned for crossed disparities, (2) the "far" neurons, broadly tuned for uncrossed disparities, and (3) the "tuned" neurons, narrowly tuned for disparities close to zero. The tuned neurons have an average bandwidth of  $0.085^\circ$ , and their peaks are almost entirely restricted to the range  $\pm 0.1^\circ$ . Near neurons are excited by crossed disparities and inhibited by uncrossed disparities, whereas the opposite holds true for far neurons. In both cases, the response curves have their steepest slope near zero disparity, as they go from excitation to inhibition. The excitatory peaks for far and near neurons are on average at about  $\pm 0.2^\circ$  (Poggio 1984). This tripartite division is an idealization, and many disparity-tuned neurons are often difficult to fit neatly into any classification scheme. LeVay and Voigt (1988), in their study of disparity tuning in cat visual cortex, emphasize the large number of cells with intermediate properties.

### Psychophysical Data

Humans can discriminate small differences of depth near the plane of fixation with an accuracy that is typically around 5 seconds. This is smaller by a factor of about 50 than the width of the narrowest cortical disparity tuning curves and is a factor of 6 smaller than the width of a photoreceptor. Stereoacuity falls off rapidly away from the plane of fixation. The disparity discrimination curve in Figure 2B plots the smallest discriminable change in disparity as a function of stimulus disparity. Disparity increment threshold curves have been measured using a variety of stimuli with similar results, including line patterns (Westheimer 1979), random-dot stereograms (Schumer and Julesz 1984), and difference of Gaussian stimuli (Badcock and Schor 1985).

Another aspect of depth perception is interpolation. In random-dot stereograms, the surface of the square floating in depth appears to be solid, even though the dots may be quite sparse and smooth surfaces are perceived even for complex shapes. This suggests that when a stereogram dot is matched, it influences the perceived depth of neighboring blank locations. Psychophysical experiments have been performed to measure the interactions occurring between two nearby locations (Westheimer 1986a,b; Westheimer and Levy 1987). As shown in Figure 4A, the disparities of two lateral lines (labeled *a*) were set to a series of values by



**Figure 2.** Comparison of a model disparity discrimination curve with a human psychophysical curve. (A) The smallest discriminable change in disparity is plotted as a function of a pedestal disparity for a model based on the tuning mechanisms shown in panel B and the psychophysical curve measured by Badcock and Schor (1985). (B) The smallest population (17 units) judged sufficient to give an adequate representation of the data. Tuning curve width increased with peak location, so that the steepest portions of the near and far curves all fall near zero disparity. Since discriminability depends on tuning curve slope, this organization produced highest discriminability at zero. This population gives a rough indication of the minimum size needed to encode disparity.

the experimenter. The disparities of two nearby inner lines (labeled *b*) were kept at zero. The basic observation was that the presence of depth at *a* warped the perceived depth at *b* to nonzero values. The amount of warping was quantified by having the subject adjust the disparity of the middle line to produce the same apparent depth as the lines at *b*. For small separations, moving the lines at *a* in depth dragged the perceived depth of *b* in the same direction. As the separation increased, this attractive interaction decreased and then reversed so that the two lines appeared to repel each other. This suggests that there are excitatory and inhibitory interactions between pools of neurons representing depth at neighboring locations. In the following section, depth discrimination and depth interpolation are considered in a reference frame relative to fixation. Issues related to depth constancy and absolute depth estimation are discussed in a later section.

### MODELING DEPTH DISCRIMINATION

Our model does not attempt to describe how the disparity tuning observed in neurons is synthesized from monocular inputs. Nor do we model the matching process between images to the two eyes, or what aspects of the images may act as tokens during matching. This is the correspondence problem, which we assume the circuitry in the brain has already solved, since neurons have been found that can correctly compute the disparity for lines and random-dot stereograms. We start with model neurons that have the same types of disparity tuning curves found in cortical neurons and ask whether they can account for the psychophysical performance of the visual system.

#### Threshold for Discrimination

Figure 1D shows a subset of two idealized disparity tuning curves from a much larger population. For disparity  $d_1$ , each unit responds at a given level. When the disparity changes to  $d_2$ , the response of one unit increases and the response of the other unit decreases. A discriminable change in disparity has occurred if the net change in response summed over all units in the population is significant relative to the noise in the units. Signal detection theory can be used to determine the probability that this change was not produced by chance (Green and Swets 1966). The threshold for discriminability is defined as the value of disparity difference that produces a probability of 0.75 correct discrimination. We applied this method to the idealized tuning curves shown in Figure 2C.

We first attempted, unsuccessfully, to reproduce the human discrimination threshold curve in Figure 2B using just three tuning curves, one from each class. The resulting discrimination curve calculated from signal detection theory had prominent spikes because there was insufficient overlap between mechanisms. More importantly, the best discrimination threshold obtainable with three mechanisms and using the noise observed in neurons was 70 seconds, well below the hyperacuity range. This suggests that there must be more than three units engaged in encoding disparity at a particular location in the visual field.

The next step was to add additional tuning curves to the population. The following rule was found to yield good results: Make the bandwidth of each tuning curve proportional to the disparity of the tuning curve peak. This always placed the steepest portion of the near and far tuning curves near zero disparity, producing fine discrimination at the point of fixation. Thorpe and Pouget (1989) reached a similar conclusion about the importance of the slope of the tuning curve for identifying orientation. The smoothness of the discrimination curve improved as more tuning curves were added; satisfactory results were achieved with a minimum of 17 mechanisms, as shown in Figure 2. It is interesting to note that the fine stereoacuity at zero disparity is produced not by the narrow tuned mechanisms, but by the

near and far mechanisms, all of which have their steep portions at zero. Although tuned mechanisms also had steep slopes, they were not concentrated at any one disparity value.

No special significance should be placed on the exact number of mechanisms we used; 17 was just a rough estimate of the minimum size of the population encoding disparity. In addition, no claim is made that the tuning curves presented here are unique. However, more tuning curves can be used. The curve generated by 200 units retained the same shape as that produced by 17 units, but it was shifted down to about 1.0 seconds because of the increased probability summation in the larger population (Lehky and Sejnowski 1990). Any larger population would push stereoacuity to unrealistically low levels. These bounds refer only to the final output that can be assayed by perceptual reports; in particular, this estimate does not include additional binocular units that might be needed for solving the correspondence problem.

#### Predictions of Interval and Rate Codes

Can models based on interval encoding (Fig. 1A) or rate encoding (Fig. 1B) also account for these data? When some approximation is made to the narrow disparity tuning curves used by Marr and Poggio (1976), as well as many other investigators, the resulting disparity discrimination curve does not resemble the data. The problems are (1) insufficient overlap between mechanisms, leading to a "spiky" appearance of the curve at the fine level, and (2) uniform widths in their tuning, leading to the essential flatness of the curve at a gross level. These are problems independent of the exact shape of the tuning curves. The only way to overcome both of these difficulties is to broaden the tuning curves to overlap more, in effect turning the interval code into a population code. Rate encoding, on the other hand, could account for the psychophysical data very well. The disparity response curve in Figure 1B has a steep slope near zero disparity, leading to fine discriminability, and flattens out for larger disparity values (both positive and negative), where discriminability is poor. With the appropriate flattening function, a V-shaped discrimination curve can be generated. Rate encoding offers the most parsimonious accounting for the psychophysical disparity discrimination data considered in isolation. Unfortunately, there is no evidence for neurons having such monotonic disparity responses, so this form of encoding must be rejected.

### MODELING DEPTH INTERPOLATION

In this section, the discrimination model is extended to include interactions between nearby patches of the visual field, which requires the units to influence each other through a network. The opponent spatial organization of depth attraction and repulsion in Westheimer's psychophysical data (Westheimer 1986a;

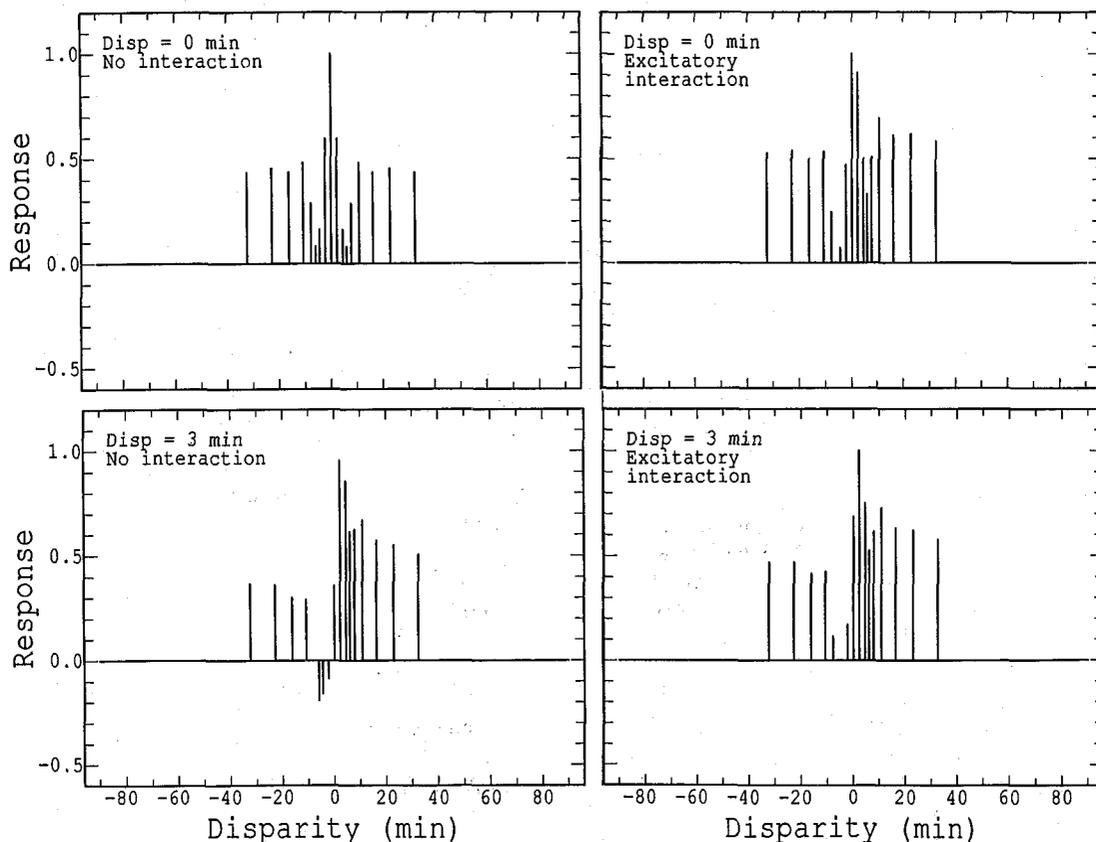
see introduction) immediately suggests an old idea in neuroscience: short-range excitation and long-range inhibition between neurons (Ratliff 1965). Assume that the entire population of 17 disparity-tuned units used previously (Fig. 2C) is replicated at each spatial location. A unit at one location interacts with units at neighboring locations to form a network. Assume further that a unit interacts only with other units (at different locations) tuned to the same disparity. If units tuned to the same disparity are spatially close, there is mutual excitation; however, if they are farther apart, there is inhibition.

Each encoding population represents disparity for some patch of visual field. For present purposes, the size of these patches may be considered the area subserved by a single cortical column, but this is an empirical question, and in any case, the spatial scale does not affect the formal structure of the model. Similarly, the scale of the lateral interactions is also an empirical question. With these lateral interactions included, the model neurons will adjust their activity levels through a relaxation process. In this manner, the lateral spatial

interactions transform an initial pattern of activity at each position into a new pattern. The responses of a neural population can be shown in an "activity diagram," such as those in Figure 3, which shows the response of each unit by the height of a line relative to spontaneous activity. The line is at a position along the horizontal axis corresponding to the peak of the tuning curve for that unit.

### Interpretation of the Population Code

What disparities do the new patterns of activity represent? One possibility is to assign an interpretation based on some weighted average within the population (Georgopoulos et al. 1986). However, this method assigns a unique depth to each point and would run into trouble with transparent surfaces. Another approach, which we adopt, is to consider the pattern of activity in a population as forming a "representational spectrum" irreducible to anything simpler. An interpretation of the pattern is defined by template matching: First, for every possible disparity, a canonical activity pattern is



**Figure 3.** Activity diagrams, showing the patterns of activity when the population of 17 units (Fig. 2C) was presented with different disparities. The height of each line indicates a unit's response. The position of a line along the disparity axis indicates the value of the tuning curve peak for that unit. Each disparity produced a unique pattern of activity, which can be thought of as a representational spectrum. The two left-hand panels show activity patterns when there were no lateral interactions, such as when a single disparity stimulus is presented in isolation. (*Top left*) Response to a disparity of 0.00 min. (*Bottom left*) Response to a disparity of 3.00 min. The two right-hand panels show new activity patterns arising when two disparity stimuli were presented simultaneously at nearby positions, with excitatory interactions between positions. (*Top right*) New pattern in response to 0.00 min, which should be compared to the top left panel. (*Bottom right*) New pattern in response to 3.00 min, which should be compared to bottom left panel.

established, which is simply the pattern produced by a single disparity in the absence of any perturbing lateral influences. Second, the arbitrary activity pattern is matched to the canonical pattern with the minimum root-mean-square (RMS) error between the arbitrary patterns and the canonical patterns.

The mutual excitation between two positions causes an "attractive" effect in the apparent disparities at both positions, as shown in Figure 4B. On the other hand, if

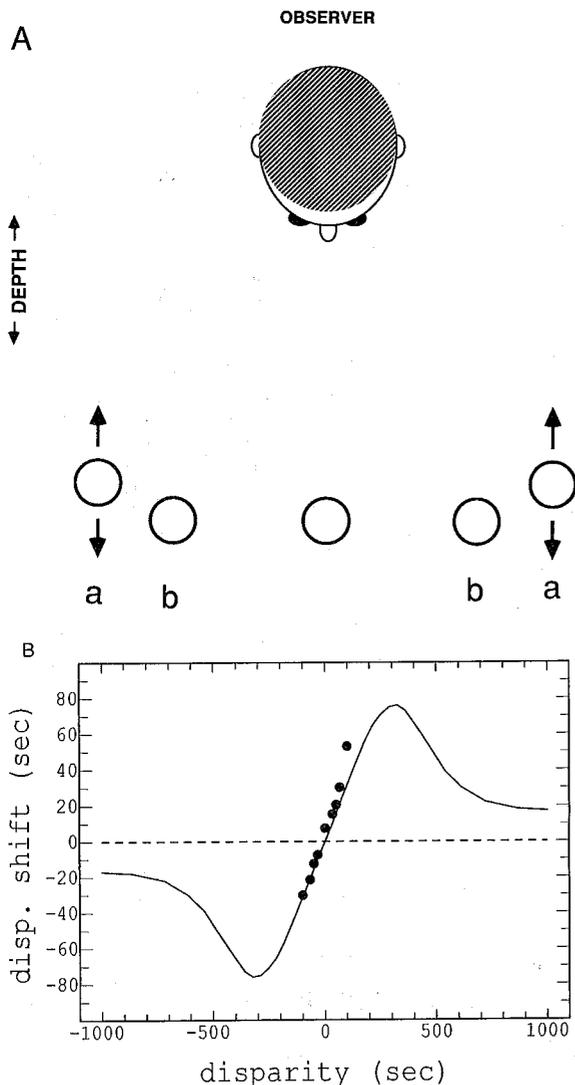
the interactions are inhibitory, then there is a "repulsive" effect. The effects described by Westheimer (1986a) are duplicated in the model by using excitatory connections to produce the short-range depth attraction or by using inhibitory connections to produce the longer-range depth repulsion. The model predicts further that if data were collected for larger disparities, the curve would fold over (Fig. 4B).

The model can be easily expanded to three locations. When disparity stimuli are present at only the two lateral locations, the interactions between the populations will indirectly stimulate an intermediate population. This is similar to what happens in a random-dot stereogram, where a solid surface is perceived even though a large fraction of the surface is blank. What pattern of activity was induced in the intermediate position? The answer depended on the disparity gradient (Burt and Julesz 1980): For small disparity differences, the intermediate point was best matched by a disparity that is the average of the two lateral disparities. A very different result occurred when there was a large disparity difference between lateral positions. There was no longer a unique minimum error for the disparity match at the intermediate position; rather, there were two minima. Patterns having two minima can represent two disparities simultaneously at a single position. This nonuniqueness was only seen when there were mutually excitatory interactions between neurons.

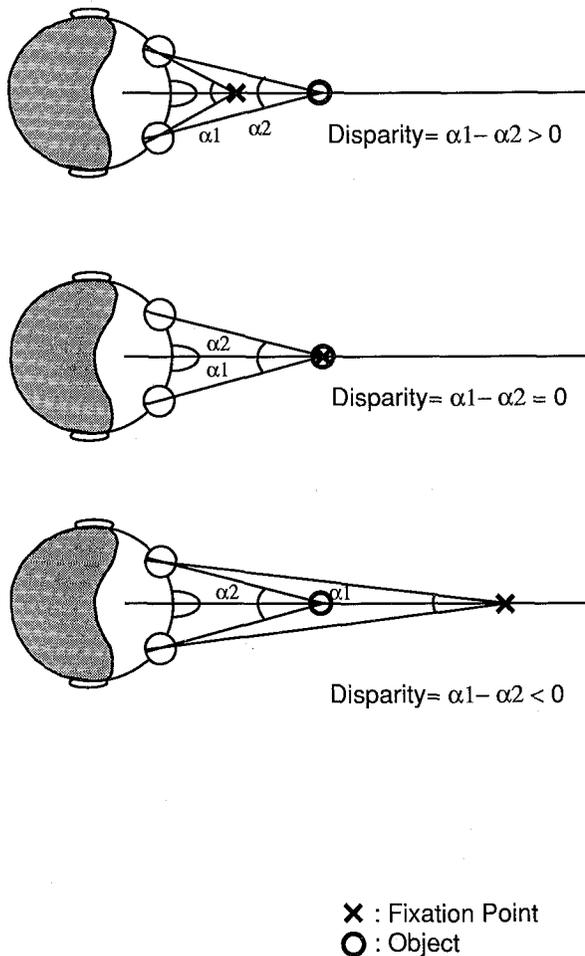
There are two circumstances in which multiple disparities could occur simultaneously at a single point. The first is at a depth discontinuity in the surface of an object, and the second is when there are transparent surfaces. What these situations have in common is that nearby points in the visual field may have small spatial separations, yet belong to surfaces at radically different depths. A transition from depth averaging to transparency has been observed psychophysically by Parker and Yang (1989). They measured the depth percept that resulted when random-dot stereograms were constructed by intermixing dots from two depth planes. For small disparity differences, the percept was that of a single surface at the average of the two depths, but for large differences, the percept was that of two simultaneous depths and the appearance of transparency. The disparity difference at the transition from averaging to transparency increased rapidly as a function of the mean disparity of the two sets of dots in both the experiments and the model.

#### MODELING DEPTH CONSTANCY

An object fixed in the environment does not seem to move around, despite jumps in its retinal location as we continually shift our gaze. This phenomenon, called space constancy, suggests that our internal representation of the world compensates for eye movements. This compensation must also exist for depth as well as location in the visual field, since we also shift the fixation point of our two eyes in depth. Disparity is a relative depth cue as illustrated in Figure 5. When a subject



**Figure 4.** Comparison between human psychophysical measurements of depth interpolation and predictions of a network model. (A) Schematic of the setup used by Westheimer (1986a) to measure lateral interactions between nearby stimuli at different depths. The circles represent lines seen by the observer. When the outer lines *a* were moved in depth, the apparent depth of the nearby lines *b* were dragged along with them, although the physical disparities of *b* remained unchanged. For small spatial separations between *a* and *b*, *b* was dragged in the same direction as *a* (attraction), and for larger separations, *b* was repulsed. The central line was used to monitor the apparent depth at *b*. (B) Stereo interpolation data (●) from Westheimer (1986a). The apparent shift in disparity at *b* is shown as a function of a real shift in the disparity at *a*. The line shows predictions of the model.



**Figure 5.** Schematic drawing of the interaction between disparity and vergence angle. The disparity ( $\alpha_2 - \alpha_1$ ) of an object depends on the eye vergence  $\alpha_1$ . A neuron that is tuned to the depth of an object (O) should respond to (1) far disparities, (2) disparities around zero, and (3) near disparities, depending on the point of fixation of the eyes (×).

fixates in front of an object, it has a negative disparity, whereas the disparity of the same object is positive when the subject fixates behind it. Although the position of an object along the depth axis is not explicitly represented on the retina, depth can nevertheless be recovered from a variety of other cues, such as eye vergence (Foley and Held 1972; Foley and Richards 1972; Gogel 1961, 1962; von Hofsten 1977), accommodation (Ittelson and Ames 1950), and vertical disparity (Bishop 1989). None of these cues alone can account for depth constancy, which suggests that they normally work in combination, together with monocular cues for depth. However, there are circumstances when each of these cues individually is used to estimate depth, but the extent to which each cue is used under normal circumstances is still debated.

A neural network model for space constancy in the plane of fixation has been proposed by Zipser and Andersen (1988). Physiological data suggest that area 7a in the parietal cortex of the monkey is involved in this transformation (Andersen and Mountcastle 1983).

The goal of our model is to understand the mechanism of space constancy for depth based on the interaction between disparity and vergence cues. Such interactions can be studied by recording from neurons in alert monkeys and determining the disparity tuning curves for different vergences. In all previous studies, the vergence angle was kept constant. If a neuron were coding depth rather than disparity per se, one should expect a modulation of its disparity tuning with vergence. We make specific predictions for what should be observed.

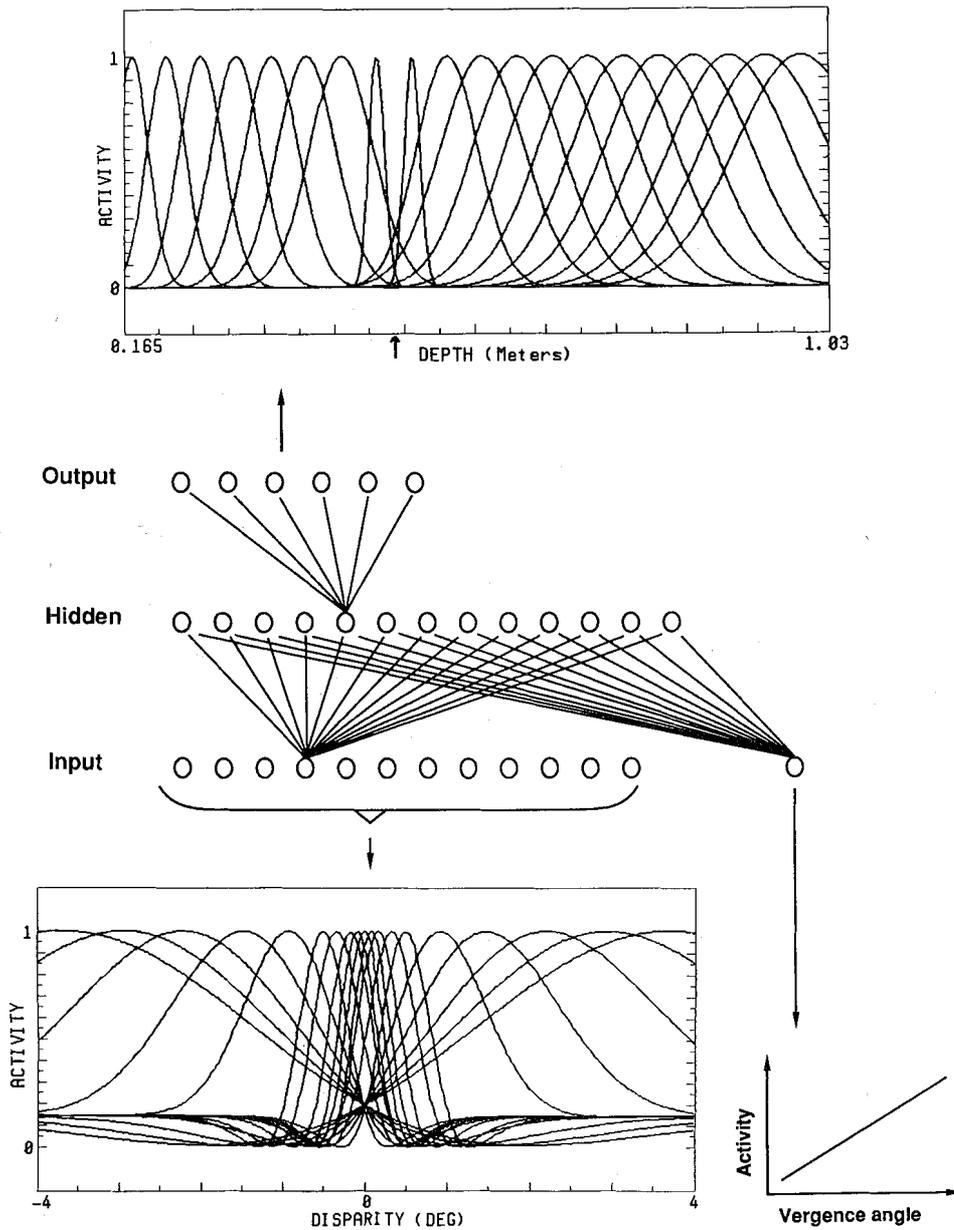
### Network Architecture

A three-layer, feedforward network, completely interconnected between layers (Fig. 6), was trained with the backpropagation algorithm (Rumelhart et al. 1986) to recover absolute depth on the output layer from combinations of 5 eye positions and 21 disparity values. Vergence values were chosen so that the fixation point varied between 20 cm and 50 cm from the subject. Disparity values were limited to the interval  $-40^\circ$  to  $+4^\circ$ . One input unit coded eye vergence with its activity level directly proportional to the vergence angle. This unit was similar to the vergence neurons reported in the parietal cortex (Sakata et al. 1980; Joseph and Giroud 1986) and in the oculomotor nuclei (Mays 1984). The remaining 19 input units encoded disparity with the same representation used in the above model of depth discrimination. There were 25 hidden units in the networks reported here, each of which received inputs from all the input units and projected to all the output units. The 40 output units were trained to compute absolute depth according to a Gaussian curve centered on a position specific to each unit (Fig. 6). Peaks were spread along the whole range of depth tested. Other output representations that were studied gave similar results.

Depth estimation is more accurate at short range. This was taken into account by varying the tuning curve bandwidth with the position of the peak, so that units coding for short distances were more finely tuned than those coding long distances (Fig. 6). We also made the tuning curve bandwidth vary with eye position because stereoacuity is highest near the point of fixation (Westheimer 1979). Thus, an output unit with a peak at 50 cm was trained to produce a narrower Gaussian curve when the fixation point was at 50 cm than when it was elsewhere. This second type of modulation of the output units was effective only for output units in a narrow range around the point of fixation. Although the bandwidths of the output units varied with position and point of fixation, their centers were always fixed.

### Properties of the Hidden Units

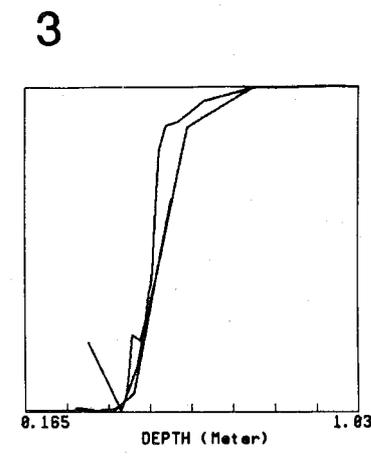
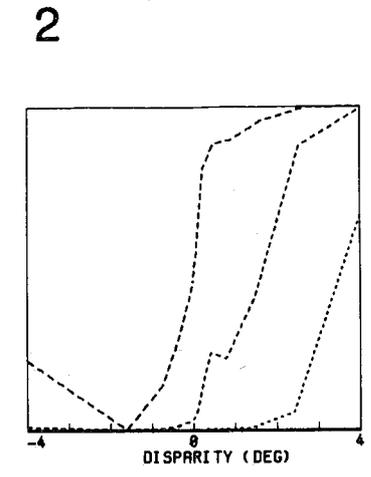
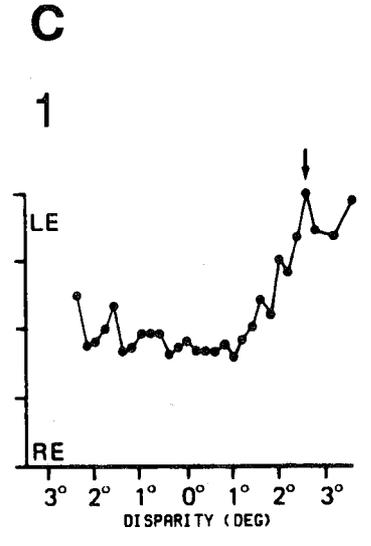
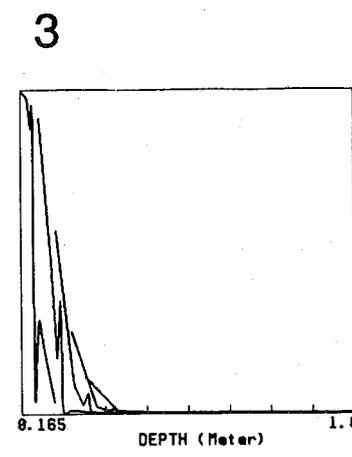
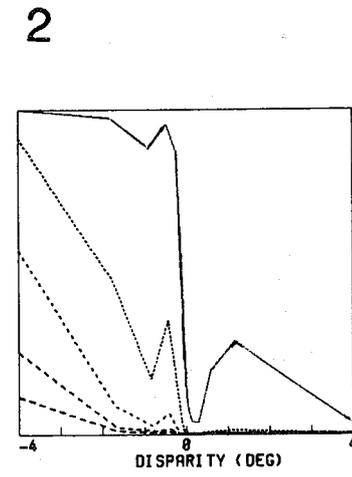
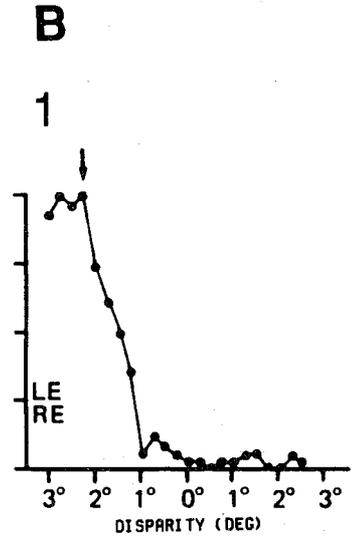
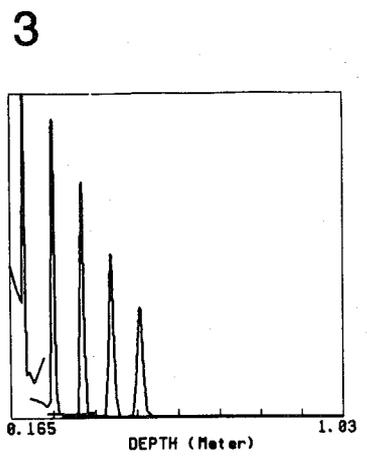
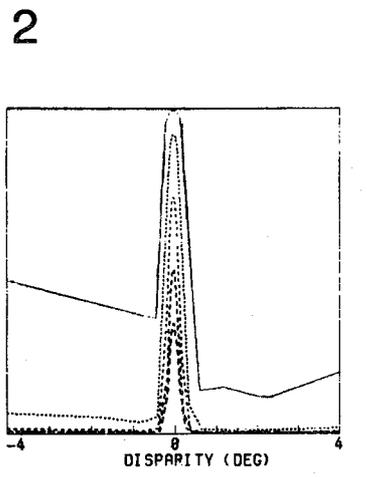
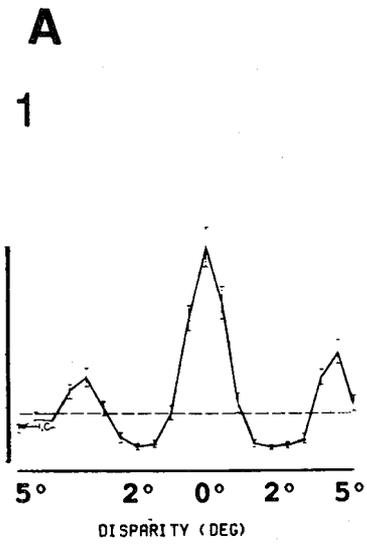
We studied the hidden units of a mature network that was trained to perform accurately the transformation from relative to absolute depth. The disparity tuning curves were very similar to those of neurons observed

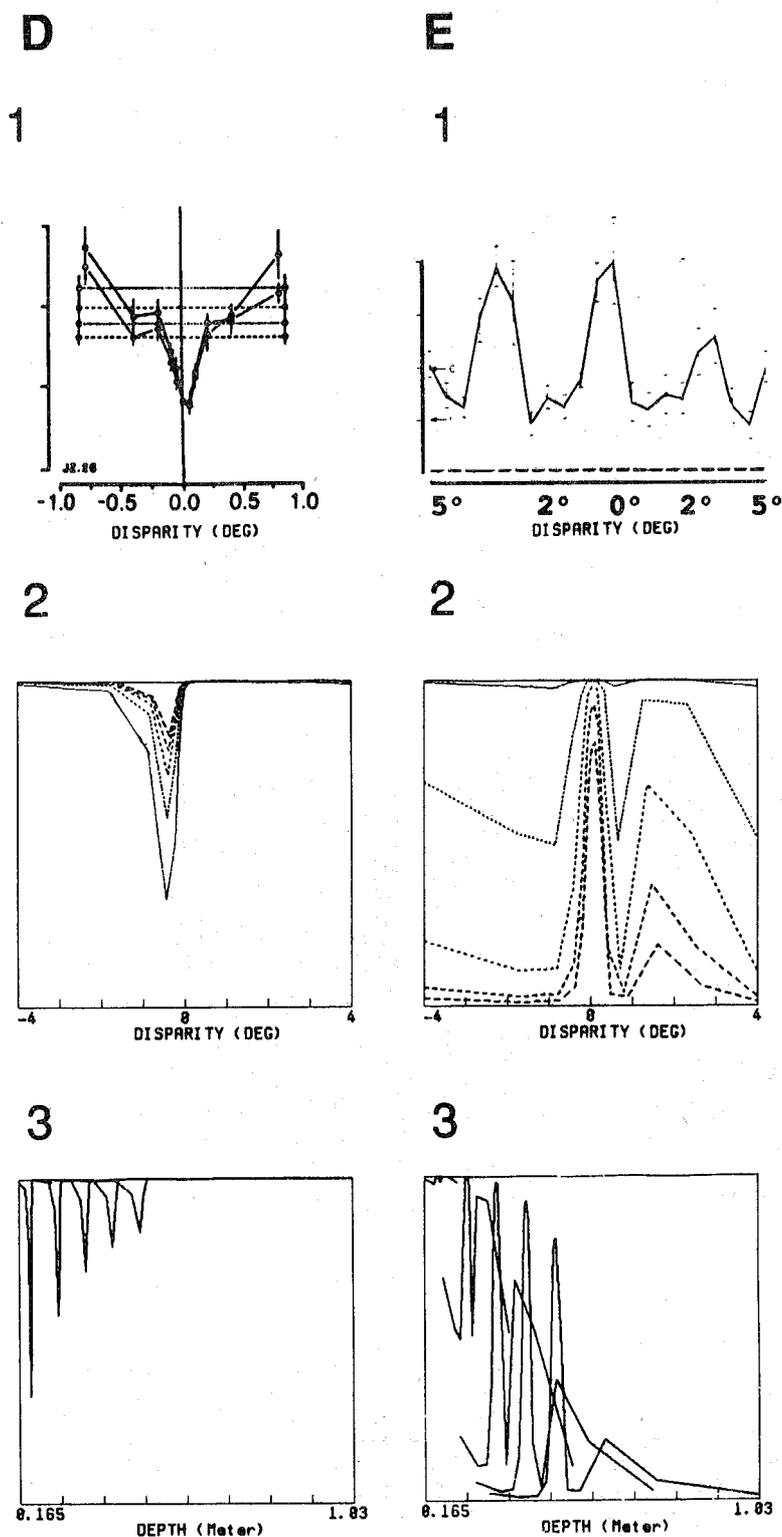


**Figure 6.** Network architecture for estimating depth from disparity and vergence. The input layer has one unit linearly coding eye vergence (*bottom right*) and a set of units coding disparity using a distributed representation (*bottom left*). The transformation between the input and the output is mediated by a set of hidden units. The output layer is trained to encode distance in a distributed manner (*top*). After training, the tuning curves for each output unit was a gaussian function of depth centered on a value specific to each unit (only every other depth tuning curve is shown here). The bandwidth of the curves increased with depth, except around the fixation point where the curves were narrower. These two types of bandwidth modulation produced depth estimates with relative accuracy similar to that found in humans.

in visual cortex. As shown in Figure 7, they could be roughly classified into three prototypical groups, near, far, and tuned (excitatory and inhibitory), with many intermediate examples (Ferster 1981; LeVay and Voigt 1988; Poggio et al. 1988). Vergence modulated the disparity tuning curves of the hidden units in various ways that fell between two general classes: In the first class, the disparity tuning of the unit remained similar for the five eye positions, but the amplitude of the response was modulated. Figure 7, A and D, shows two

examples of this class, which we call disparity gain-control neurons. These are analogous to translational gain-control neurons in area 7a, whose response is modulated by eye position but not the position of the receptive field (Andersen and Mountcastle 1983). In the second extreme class, the disparity tuning of the unit changed significantly with eye position, even though the shape of the curve remained the same, as illustrated in Figure 7C. For certain eye positions, the tuning curves of these units were completely displaced





**Figure 7.** Comparisons between typical disparity tuning curves for real neurons and for hidden units. (A) Excitatory tuned units; (B) near units; (C) far units; (D) inhibitory tuned units; (E) intermediate units with multiple peaks. (Row 1) Disparity tuning curves for five typical neurons based on recordings from primary visual cortex (B and C: Ferster 1981 [cat]; A and E: LeVay and Voigt 1988 [cat]; D: Poggio et al. 1988 [monkey]). (Row 2) Disparity tuning curves for five equivalent hidden units. Five tuning curves corresponding to the five different eye positions are shown for each unit by dashed and dotted lines. (A-2, D-2) Typical disparity gain control units; (C-2) example of a unit tuned to depth; (B-2, E-2) intermediate type of modulation. Such intermediate units are not perfectly tuned to depth, but the general envelopes of their depth tuning curves do provide useful information for depth estimation. Compare these tuning curves to those of corresponding real neurons in Row 1. (Row 3) Depth tuning curves for the corresponding hidden units. These tuning curves are predictions for what might be found in visual cortex when vergence and disparity tuning are measured in single neurons.

from zero disparity. This has been observed in the cat (see Fig. 7B,C) (Ferster 1981; LeVay and Voigt 1988) and in the monkey (Poggio et al. 1988). Tuned units tend to be in the first class, whereas near and far units tend to belong to the second class. Units exhibiting an intermediate type of modulation are shown in Figure 7, B and E. This second class of hidden units is not unexpected when one considers that the output units completely change their disparity selectivity with eye position (Fig. 5). The surprising result is that so few hidden units change their disparity selectivity with eye position.

The depth tuning curves of the hidden units are shown in Figure 7. The units in the class showing gain control (Fig. 7A) changed their depth sensitivity with vergence, which is an indirect measure of depth. For the units whose disparity selectivity changed completely with vergence (Fig. 7C), the depth tuning is almost perfect. This type of unit is therefore functionally very close to the output units. However, there are usually only 4 or 5 of these hidden units out of 25 hidden units in a single network, and the overall performance of the network does not depend critically on them.

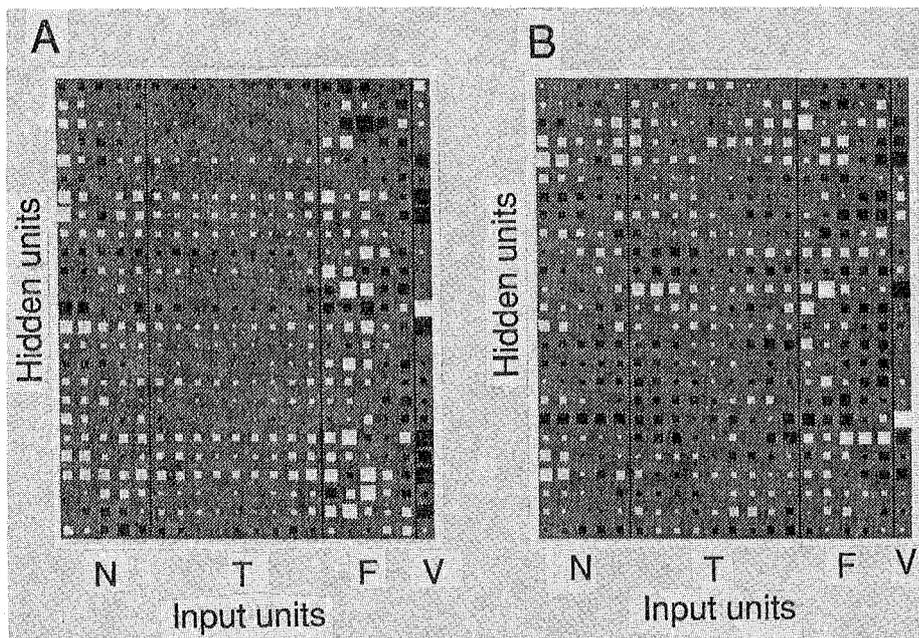
#### Effect of Higher Depth Acuity around the Fixation Point

The bandwidth of an output unit was designed to vary with the position of the tuning curve relative to the

observer and with its position relative to the fixation point. This mimics the decreasing absolute accuracy of stereopsis with increasing distance and also makes the network more accurate around the fixation point. When a network was not forced to be more accurate around the fixation point, the weights from the sharply tuned unit in the input layer to the hidden units were smaller (Fig. 8). When networks were trained without any modulation of the bandwidth of the output units, the weights were similar to those in networks with bandwidth modulation, but learning was much more difficult and the final error value was higher.

#### DISCUSSION

The central premise of our modeling was that disparity is encoded by a population of units having broad, overlapping tuning curves. In such a distributed representation, the activity of a single unit gives only a coarse indication of the stimulus parameter. This does not mean that precise information is lost, but only that the information is dispersed over a pattern of activity in the population. The concept of distributed representations arose in nineteenth century psychophysics with the idea that color is encoded by the relative activities in a population of three overlapping color channels. In our model, the parameter is "disparity" rather than "color," and more mechanisms were required to explain the data, but the idea is the same. In a similar



**Figure 8.** Hinton diagram comparing the weights from the input units to the hidden units for two different training sets. (A) Bandwidth of the output units varied only with the position of the center. The weights from the tuned units are small, indicating that these units have little influence on the hidden units. (B) Double modulation of the bandwidth of the output units with the position of the center and the position of the fixation point. Tuned input units have more influence on the hidden units as revealed by the higher values of their weights. Each row represents weights from all the input units projecting to a single hidden unit. Each column represents all the weights from a single input unit to all hidden units. The first five columns are from near units (N), the central nine weights are from tuned unit (T), the next five weights are from the far units (F), and the last column corresponds to the vergence unit (V). White weights have positive values and black weights have negative values. The area of the square is directly proportional to the value of the weight.

manner, it is possible to apply the concept to many other parameters. Some consequences of population coding were analyzed by Hinton et al. (1986) in the context of model neurons with only two levels of firing, fully on or fully off. We have extended Hinton's analysis to the case when units have continuous values and noise is the limiting factor rather than width of the tuning curve. We conclude that the population size encoding disparity for a small patch of visual field may be as small as a few tens of units or as large as a few hundred.

### Interpreting Population Codes

There are several approaches to the problem of deciding what parameter value a pattern of activity in a population represents. One approach is what we call the "spectrum" method, used in this model of disparity and also in color vision. When assigning a color to the pattern of activity within this small population, the vector of three activities is not reduced to a single number. There is nothing simpler than the pattern itself, which forms a characteristic representational spectrum for each wavelength. In the same manner, our model represents disparity (another one-dimensional parameter) in a high (possibly several hundred)-dimensional space. A second approach is what we call the "averaging" method, used in the population code models of Gelb and Wilson (1983) and Georgopoulos et al. (1986), in which the dimensionality of a representation is reduced during the interpretation process. For example, Georgopoulos et al. (1986) used the averaging method for representing the three-dimensional direction of arm movement. The pattern of activity in a large population was interpreted by collapsing the high-dimensional representational space down to three dimensions by calculating a weighted sum of tuning curve peaks. Their interpretation of population activity is based entirely on this sum and not on any particular spectrum of activity within that population.

### Consequences of Using a Population Code for Disparity

In the Marr and Poggio model (1976), which was based on an interval code, false matches were eliminated by using inhibition to shut off all units tuned to the wrong disparities at a given location, a form of winner-take-all circuit. In contrast, the goal in a distributed code is to alter the relative firing rates to produce a new pattern of activity and not to shut off all neurons in a population except one, for a single broadly tuned unit provides little information. The choice of representation also affects the process of interpolation. Grimson (1982) and Terzopoulos (1988) used spline functions to fit through the blank regions between the surface tokens used in the stereomatching process based on an interval code. This procedure also interpolated through real depth discontinuities, shrouding sharp breaks. These models deal with the problem by adding separate

mechanisms that recognize discontinuities and attenuate the interpolation process accordingly (Koch et al. 1986). In our model, interpolation falls out automatically without any shift in parameters or any additional mechanisms.

The discrimination model was concerned in part with the representation of transparent stimuli. It is possible that analogous models can be constructed for other transparency phenomena besides those arising from depth. A specific example involves motion. Adelson and Movshon (1982) have studied the percept of two superimposed gratings drifting in different directions and found conditions under which they "cohered" to form a single drifting plaid or alternatively appeared as two transparent surfaces sliding over each other, depending on how similar the two gratings were in various respects (speed, spatial frequency, contrast, etc.). This might also be understood in terms of a distributed representation for motion formally analogous to the one used for disparity here.

### Predictions for Depth Tuning

We have trained a neural network to recover egocentric depth from disparity and vergence information and then determined the response properties of the hidden units to disparity and depth. Similar neural network models have been used to explain the response properties of known neurons in cerebral cortex (Lehky and Sejnowski 1988; Zipser and Andersen 1988). However, this is the first time that such models have been used to make predictions before the physiological results were available. These predictions will soon be tested in experiments by S.J. Thorpe et al. and J. Gnadt (pers. comm.) on alert monkeys. One of the uncertainties of the depth model was the choice of output representation. Fortunately, this choice only affected the relative number of mechanisms of each class found in the hidden layer and not their qualitative properties. We varied the sensitivity of the network to absolute depth and point of fixation by modulating the bandwidths of the output units. The flexibility of this type of distributed coding may be one of the reasons why it is widely used in the cerebral cortex to represent sensory and motor variables.

One point on which our model of stereoacuity and the model of depth estimation differed was the role of the narrowly tuned units. For depth discrimination, the broadly tuned units were primarily responsible for the highest acuity near the plane of fixation. However, in the absolute depth network, the sharply tuned units were important when the output units were modulated to produce fine discrimination around the fixation point. This indicates that different tasks may use different complements from the diverse population of disparity-selective neurons that have been observed in cerebral cortex. There is no reason to assume that the same neurons used to represent relative depth are identical with the neurons used to estimate absolute depth, which is known to be much less accurate. It may be

possible to test these predictions by selectively interfering with these tuned disparity neurons in monkeys during the performance of psychophysical tasks.

### Psychophysical Consequences

In our network model of depth estimation, we made the assumption that vergence information is accessible to the neural system that judges the depth of objects. One possible consequence of this assumption is that in the absence of other confounding cues, the apparent depth of an object should change as the vergence angle is varied by a prism, even though the actual depth of the object is fixed. Alternatively, the apparent size of an object could change. As the eyes are converged, the object should appear to shrink, since it subtends the same area of the retina, whereas evidence from vergence indicates that it is closer to the observer. Conversely, diverging the eyes should lead to an apparent increase in the size of an object. Ogle (1962) has reported observations that are consistent with these counterintuitive predictions. However, Ogle explained his results differently, and further experiments are needed to choose between various explanations. The vergence effect should also hold for depth seen in random-dot stereograms, since the same argument used for angular size should also hold for disparity. Thus, the apparent height of a single raised dot in a random-dot stereogram should appear to shrink as the eyes are verged and should appear to grow as the eyes are diverged. Experiments using random-dot stereograms of extended objects must control for other depth cues, such as vertical disparities.

In conclusion, we have studied several possible models of binocular organization in the primate visual system. Some models fit only part of the data, such as rate encoding of disparity, which can parsimoniously account for the psychophysical stereoacuity data but is inconsistent with the neurophysiology. Conversely, the psychophysics also constrains interpretation of the neurophysiology. We have further shown that the distributed representation of disparity can both smoothly interpolate between sparse data and incorporate discontinuities and transparency depending on the disparity gradient. Finally, a distributed representation of relative depth can be combined with other cues, such as vergence, to predict the absolute depth of an object. This interaction and mutual constraint between physiological and behavioral data provide a particularly rich environment for the development of neural models.

### ACKNOWLEDGMENTS

This research was supported by a grant from the Sloan Foundation to T.J.S. and G.F. Poggio and by grants from the Mathers Foundation and the Drown Foundation to T.J.S. We are grateful to Dr. G.F. Poggio for illuminating discussions about the disparity tuning and response properties of cortical neurons and to S.J. Thorpe for inspiring discussions on the model of depth constancy.

### REFERENCES

- Adelson, E. and J.A. Movshon. 1982. Phenomenal coherence of moving visual patterns. *Nature* **300**: 523.
- Andersen, R. and V. Mountcastle. 1983. The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *J. Neurosci.* **3**: 532.
- Badcock, D. and C. Schor. 1985. Depth-increment detection function for individual spatial channels. *J. Opt. Soc. Am.* **A2**: 1211.
- Barlow, H.B., C. Blakemore, and J.D. Pettigrew. 1967. The neural mechanisms of binocular depth discrimination. *J. Physiol.* **193**: 327.
- Bishop, P.O. 1989. Vertical disparity, egocentric distance and stereoscopic depth constancy: A new interpretation. *Proc. R. Soc. Lond. B. Biol. Sci.* **237**: 445.
- Burt, P. and B. Julesz. 1980. A disparity gradient limit for binocular fusion. *Science* **208**: 615.
- Ferster, D. 1981. A comparison of binocular depth mechanisms in areas 17 and 18 of the cat visual cortex. *J. Physiol.* **311**: 623.
- Foley, J.M. and R. Held. 1972. Visually directed pointing as a function of target distance, direction and available cues. *Percept. Psychophys.* **12**: 263.
- Foley, J.M. and W. Richards. 1972. Effects of voluntary eye movement and convergence on the binocular appreciation of depth. *Percept. Psychophys.* **11**(6): 423.
- Gelb, D. and H. Wilson. 1983. Shifts in perceived size as a function of contrast and temporal modulation. *Vision Res.* **23**: 71.
- Georgopoulos, A., A. Schwartz, and R. Kettner. 1986. Neuronal population coding of movement direction. *Science* **233**: 1416.
- Gogel, W.C. 1961. Convergence as a cue to absolute distance. *J. Psychol.* **52**: 287.
- . 1962. The effect of convergence on perceived size and distance. *J. Psychol.* **53**: 475.
- Green, D. and I. Swets. 1966. *Signal detection theory and psychophysics*. Wiley, New York.
- Grimson, W. 1982. A computational theory of visual surface interpolation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **298**: 395.
- Hinton, G.E., J.L. McClelland, and D.E. Rumelhart. 1986. Distributed representations. In *Parallel distributed processing: Explorations in the microstructure of cognition* (ed. D.E. Rumelhart and J. McClelland), vol. 1, p.77. MIT Press, Cambridge, Massachusetts.
- Hubel, D. and T. Wiesel. 1970. Cells sensitive to binocular depth in area 18 of the macaque monkey cortex. *Nature* **225**: 41.
- Ittelson, W.H. and A.J. Ames. 1950. Accommodation, convergence and the irrelatation to apparent distance. *J. Psychol.* **30**: 43.
- Joseph, J.P. and P. Giroud. 1986. Visuomotor properties of neurons of the anterior suprasylvian gyrus in the awake cat. *Exp. Brain Res.* **62**: 355.
- Julesz, B. 1960. Binocular depth perception of computer generated patterns. *Bell System Tech. J.* **39**: 1125.
- . 1971. *Foundations of cyclopean vision*. University of Chicago Press, Chicago, Illinois.
- Koch, C., J. Marroquin, and A. Yuille. 1986. Analog "neural" networks in early vision. *Proc. Natl. Acad. Sci.* **83**: 4263.
- Lehky, S.R. and T.J. Sejnowski. 1988. Network model of shape-from-shading: Neural function arises from both receptive and projective fields. *Nature* **333**: 452.
- . 1990. Neural model of stereoacuity and depth interpolation based on a distributed representation of stereo disparity. *J. Neurosci.* **10**: 2281.
- LeVay, S. and T. Voigt. 1988. Ocular dominance and disparity coding in cat visual cortex. *Visual Neurosci.* **1**: 395.
- Marr, D. and T. Poggio. 1976. Cooperative computation of stereo disparity. *Science* **194**: 283.

- . 1979. A computational theory of human stereo vision. *Proc. R. Soc. Lond. B Biol. Sci.* **204**: 301.
- Mayhew, I. and I. Frisby. 1981. Psychophysical and computational studies towards a theory of human stereopsis. *Artif. Intell.* **16**: 349.
- Mays, L.E. 1984. Neural control of vergence eye movements: Convergence and divergence neurons in the midbrain. *J. Neurophys.* **51**(5): 1091.
- Nikara, T., P.O. Bishop, and J.D. Pettigrew. 1968. Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. *Exp. Brain Res.* **6**: 353.
- Ogle, K.N. 1962. Perception of distance and size. In *The eye: Visual optics and the optical space sense* (ed. H. Davson), vol. 4, p. 265. Academic Press, New York.
- Parker, A.J. and Y. Yang. 1989. Disparity pooling in human stereo vision. *Vision Res.* **29**: 1525.
- Pettigrew, J.D., T. Nikara, and P.O. Bishop. 1968. Binocular interaction on single units in cat striate cortex: Simultaneous stimulation by single moving slit with receptive fields in correspondence. *Exp. Brain Res.* **6**: 391.
- Poggio, G. 1984. Processing of stereoscopic information in primate visual cortex. In *Dynamic aspects of neocortical function* (ed. G.M. Edelman et al.), p. 613. Wiley, New York.
- Poggio, G. and T. Poggio. 1984. The analysis of stereopsis. *Annu. Rev. Neurosci.* **7**: 379.
- Poggio, G.F., F. Gonzalez, and F. Krause. 1988. Stereoscopic mechanism in monkey visual cortex: Binocular correlation and disparity selectivity. *J. Neurosci.* **8**(12): 4531.
- Poggio, G., B.C. Motter, S. Squatrito, and Y. Trotter. 1985. Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. *Vision Res.* **25**: 397.
- Poggio, T. and B. Fischer. 1977. Binocular interactions and depth sensitivity in striate and prestriate cortex of behaving rhesus monkeys. *J. Neurophysiol.* **40**: 1392.
- Pouget, A. and T.J. Sejnowski. 1990. A neural network model for computing depth from stereopsis. *Invest. Ophthalmol. Visual Sci. (suppl.)* **31**: 96.
- Ratliff, F. 1965. *Mach bands: Quantitative studies on neural networks in the retina*. Holden-Day, San Francisco.
- Rumelhart, D.E., G.E. Hinton, and R.J. Williams. 1986. Learning internal representations by error propagation. In *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1, p. 318. MIT Press, Cambridge, Massachusetts.
- Sakata, H., H. Shibutani, and K. Kawano. 1980. Spatial properties of visual fixation neurons in posterior parietal cortex of the monkey. *J. Neurophysiol.* **43**(6): 1654.
- Schumer, R. and B. Julesz. 1984. Binocular disparity modulation sensitivity to disparities offset from the plane of fixation. *Vision Res.* **24**: 533.
- Terzopoulos, D. 1988. The computation of visible surface representations. *IEEE Trans. Pattern Anal. and Mach. Intell.* **10**: 417.
- Thorpe, S.J. and A. Pouget. 1989. Coding of orientation in the visual cortex: Neural network modeling. In *Connectionism in perspective* (ed. R. Pfeifer). Elsevier, Amsterdam.
- von Hofsten, C. 1977. Binocular convergence as a determinant of reaching behavior in infancy. *Perception* **6**: 139.
- Westheimer, G. 1979. Cooperative neural process involved in stereoscopic acuity. *Exp. Brain Res.* **36**: 585.
- . 1986a. Spatial interaction in the domain of disparity signals in human stereoscopic vision. *J. Physiol.* **370**: 619.
- . 1986b. Panum's phenomenon and the confluence of signals from two eyes in stereoscopy. *Proc. R. Soc. Lond. B Biol. Sci.* **228**: 289.
- Westheimer, G. and D. Levi. 1987. Depth attraction and repulsion of disparate foveal stimuli. *Vision Res.* **27**: 1361.
- Wheatstone, C. 1838. Contributions to the physiology of vision. 1. On some remarkable and hitherto unobserved phenomena of binocular vision. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **8**: 39.
- Zipser, D. and R. Andersen. 1988. Back propagation learning simulates response properties of a subset of posterior parietal neurons. *Nature* **331**: 679.