Connectionist Models of Orientation Identification

ALEXANDRE POUGET & SIMON J. THORPE

We have used connectionist simulations in an attempt to understand how orientation tuned units similar to those found in the visual cortex can be used to perform psychophysical tasks involving absolute identification of stimulus orientation. In one task, the observer (or the network) was trained to identify which of two possible orientations had been presented, whereas in a second task there were 10 possible orientations that had to be identified. By determining asymptotic performance levels with stimuli separated to different extents it is possible to generate a psychophysical function relating identification performance to stimulus separation. Comparisons between the performance functions of neural networks with those found for human subjects performing equivalent tasks led us to the following conclusions. Firstly, we found that the 'psychometric functions' generated for the networks could accurately mimic the performance of the human observers. Secondly, the most important orientation selective units in such tasks are not the most active ones (as is often assumed). Rather, the most important units were those selective for orientations offset 15° to 20° to either side of the test stimuli. Such data reinforce recent psychophysical and neurophysiological data suggesting that orientation coding in the visual cortex should be thought of in terms of distributed coding. Finally, if the same set of input units was used in the two-orientation and the 10-orientation situation, it became apparent that in order to explain the difference in performance in the two cases it was necessary to use either a network without hidden units or one with a very small number of such units. If more hidden units were available, performance in the 10-orientation case was found to be too good to fit the human data. Such results cast doubt on the hypothesis that hidden units need to be trained in order to account for simple perceptual learning in humans.

KEYWORDS: Orientation, visual cortex, coding, back-propagation, neural networks, psychophysics, neurophysiology.

1. Introduction

One of the central problems in perception involves trying to understand how the perceptual capacities of observers can be related to the neurophysiological response properties of neurons in sensory systems. For example, it has been known since the early 1960s that the visual cortex contains neurons which selectively respond to contours with particular orientations (Hubel & Wiesel, 1962; Henry *et al.*, 1974; Howard, 1982; Orban, 1984; DeValois & DeValois, 1988). However, the way in which the activity of such neurons is used to code the orientation of a stimulus

Alexandre Pouget, Computational Neurobiology Laboratory, The Salk Institute, San Diego, USA; Simon J. Thorpe, Institut des Neorosciences, Université Pierre et Marie Curie, Paris, France.

remains controversial. One popular view, based on Selfridge's 'Pandemonium' model (Selfridge, 1959) is to think of these orientation-selective neurons as 'feature detectors' and to suppose that perceived orientation is determined by the ones which are most active. A clear example of this 'peak-activity' coding position is given in a recent review article on computational maps in the brain in which Eric Knudsen and his coworkers (Knudsen *et al.*, 1987, p. 59) proposed the following formulation:

> Whether the desired information is the orientation or direction of motion of a visual stimulus . . . , the answer is always represented as the location of a peak of activity within a population of neurons.

However, recent psychophysical and neurophysiological data cast doubt on such a formulation, at least in the case of orientation perception. It would appear that it is perhaps not the most active neurons which determine perceived orientation, but rather the relative levels of firing in neurons tuned to orientations offset to either side of the orientation to be judged.

Regan & Beverley (1985) have studied the effect of adaptation on orientation discrimination. They showed that after adaptation to a vertical grating (1) discrimination of orientations around the vertical is facilitated and (2) discrimination of orientations between 10° and 15° to either side of the vertical is impaired. These results suggest that for orientation discrimination, the most important neurons are not those whose peaks are centered on the test orientations, but those that are tuned for orientations 10°–15° on either side of the test orientations. Moreover, the facilitation observed after adapting the neurons tuned to test orientations show that this neuronal population is actually impairing the discrimination performance of the subject in normal conditions. A number of physiological studies have led to similar conclusions. They have looked at the ability of cortical neurons to signal small changes in stimulus parameters such as orientation, spatial frequency, contrast and vernier offset (Tolhurst *et al.*, 1982; Parker & Hawken, 1985; Swindale & Cynader, 1986; Bradley *et al.*, 1987; Skottum *et al.*, 1987; Vogels & Orban, 1990). The essential feature of such studies is that they take into account not just the selectivity of cortical neurons, but also the statistical reliability of their responses. By determining the minimum stimulus change that produced a reliable change in firing at various points on the cells tuning function it was possible to show that neuronal sensitivity was higher where the slope of the tuning curve is steeper. Such results strongly suggest that the most useful neurons for signalling a change in the stimulus are not the most active ones, but the ones centered on offset values.

This brief review of some of the recent psychophysical and neurophysiological data relevant to orientation coding illustrates how the whole question of coding by neurons is currently in a state of flux. However, it is often the case that psychophysical and neurophysiological data are treated largely independently, and it is generally difficult to gain an overall grasp of how these different types of data fit together. In our opinion, connectionist models offer a powerful way to tackle such questions as the relationship between neurophysiological data and psychophysical performance. The basic strategy consists of simulating a neural net in which the input layer is composed of units with properties that roughly correspond to the response properties of neurones in the visual system as determined using single unit recording techniques. The output layer contains units that are trained to classify stimuli. Intermediate layers containing hidden units may or may not be used. The network is then trained on a task which is equivalent to the sort of task given to human observers in a psychophysical experiment, and thus the performance of the network and the human can be directly compared.

Several things can follow from the use of such a comparative approach. Firstly, if the performance of the network can be made to match that of the human observer we are in a position to propose a reasonably realistic model of how the task is performed in humans. Secondly, by analyzing the way in which the connections in the network have been set up as a result of training, it may be possible to gain insights into the strategies used by the network. Finally, by comparing the relative performance of networks and human subjects over a range of related tasks it may be possible to further restrict the range of models that are consistent with the data.

In this paper we would like to illustrate how we have applied this approach in the case of orientation identification (see also Thorpe & Pouget, 1989). Orientation is in many ways an ideal test case for such an approach. A great deal has been learnt about the coding of orientation by visual cortical neurons and so the response properties of the units in the input layer can be specified with a fair deal of confidence (see Orban, 1984). Note that this is generally not the case in most connectionist models of psychological processes such as language in which the properties of input layer units are often largely determined on the basis of educated guesswork. In addition, there is an extensive psychophysical literature on orientation perception (for reviews, see Howard, 1982; DeValois & DeValois, 1988; Graham, 1989).

There is, however, a serious problem in trying to produce connectionist models of psychophysical tasks. This stems from the fact that there is a major methodological difference between the performance measures that are typically used in human psychophysics and those used to assess the capacities of artificial pattern classification systems such as neural nets. In the case of a machine vision system or a neural network, one will typically train the system to classify a set of stimuli, and performance is defined as the percentage of correctly identified stimuli following training. Surprisingly, identification accuracy is only relatively rarely used as a performance measure in human psychophysics. Instead, subjects are typically presented with two stimuli, either simultaneously or one after the other, and required to make a judgment about whether the two stimuli were the same or not. Such measurements of detection or discrimination are concerned with the ability of the observer to signal a change in the stimulus—not his ability to identify or classify them.

It might be added that there is a considerable amount of work in the experimental psychology literature involving identification paradigms, but they nearly always involve complex stimuli such as words, or line drawings, for which no relevant neurophysiological data are available. When such neurophysiological data is available, as is the case for such stimulus parameters as orientation, color, luminance, movement and stereo, the vast majority of psychophysical data comes from paradigms involving detection and discrimination tasks.

The approach that we have adopted in the present study, is to try and make the psychophysical task used to test human subjects more like the sort of pattern classification task used with artificial systems. For this we have developed a variety of experimental protocols in which subjects are trained to make absolute judgments of a variety of stimulus parameters, including orientation, luminance and color. In this paper we will discuss the application of connectionist models to two tasks involving absolute identification or orientation.

2. Orientation Identification by Humans

The orientation identification paradigms that we have developed for use in humans were designed to parallel those used to assess the performance of pattern classifiers.

The subjects are trained to classify stimuli differing in orientation. In one series of experiments there were just two possibilities—test gratings were presented that were either oriented to the left or to the right of vertical, and the subject was trained to indicate which by moving a joystick to the left or right. In a second series of experiments, 10 different orientations had to be identified by typing a number between 0 and 9. In both cases feedback was provided during training and the learning phase was continued until no further improvement in performance is seen. At this point, the feedback was removed and the performance of the subject measured during 400 trials.

Such protocols should be fairly natural to anyone familiar with the pattern classification literature. However, it should be pointed out that when such procedures are used with human subjects, additional precautions have to be taken to avoid 'cheating'. Humans naturally try to solve such problems by making trial to trial comparisons. In other words, it may be possible to perform such a task without actually identifying each stimulus in turn, but rather by comparing each stimulus with the previously presented one. The use of such a trial-to-trial comparison strategy is less likely in the task with 10 different orientations, since there are too many stimuli to keep track of the changes. However, when there are only two stimuli the problem is more serious—the subject may simply compare one stimulus with the next and ask whether it had rotated to the left or to the right? To limit this possibility, a series of distractor gratings with random orientations were presented between each test stimulus. Since the subject was not told whether a particular stimulus was a test or a distractor until after the stimulus had disappeared, each stimulus in turn had to be analyzed, thus overloading the subject's short-term memory and ensuring a true identification task.

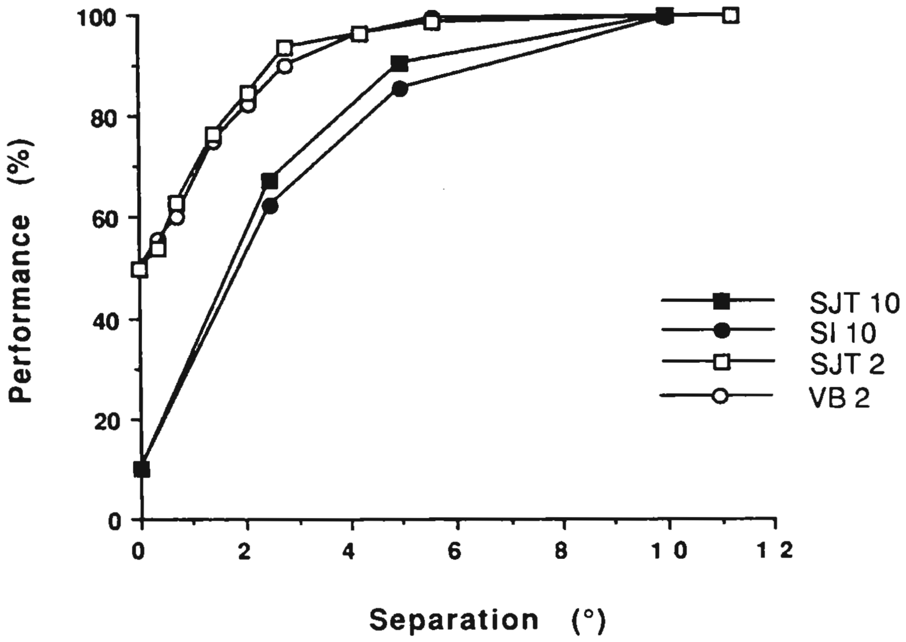

Figure 1. Psychometric functions for human subjects in two absolute identification tasks. Open symbols show identification rate (% correct) as a function of the separation between the test stimuli in a task with two possible orientations. Solid symbols show results in a task in which there were 10 different orientations to be identified.

In both the two-orientation and the 10-orientation tasks the stimuli were presented to the subjects in complete darkness in order to eliminate the possibility that the test stimuli could be compared with an external reference. Training was continued with feedback until no further progress was possible and then performance was determined during a sequence of 400 trials with no feedback. The whole training and testing procedure was repeated for a variety of different values for the separation between the stimuli in order to determine a psychometric function relating identification to orientation difference. In the case of 10 orientations, the training, as well as the testing phase, turned out to be extremely time consuming. Therefore, only a few data points were obtained; clearly not enough to draw any definite conclusion, but enough to get the general shape of the psychometric curve.

Figure 1 summarizes the results of these studies. In the task involving two orientations (open symbols), it can be seen that with the smallest separations (less than 1°) performance is close to chance which corresponds to a value of 50% correct with two possibilities. As separation is increased, performance improves rapidly before reaching virtually 100% for values of around 6°. In the case of the task with 10 possible orientations (solid symbols), chance performance is 10% correct, but again there is a rapid improvement with increasing separation before reaching virtually perfect performance at around 10°.

The question we then asked was whether we could use neural network modeling to understand (i) the form of these psychometric functions and (ii) the relative performance levels in the two tasks.

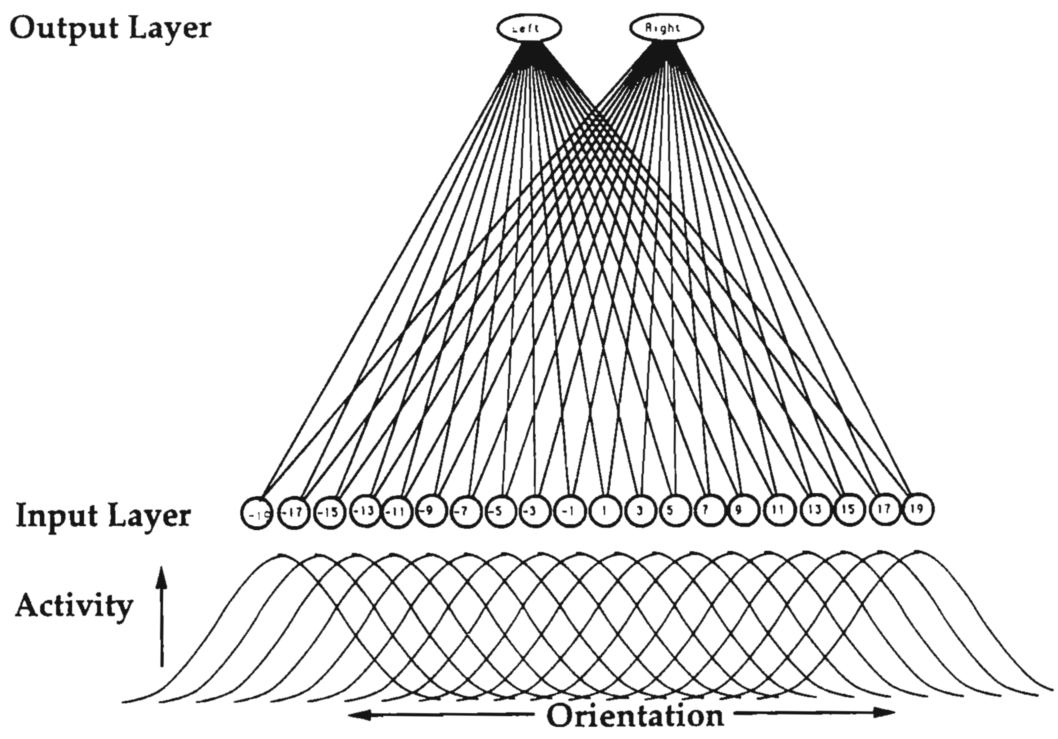

Figure 2. Basic structure of the neural net used to simulate performance in the two-orientation task. The input layer is composed of a series of orientation tuned units, each with a Gaussian filter shape. For simplicity, we show only a subset of the 60 input units that were used in the first simulation. The output layer contains units that were trained to respond to the different test stimuli. In the task with 10 orientations, the output layer was composed of 10 units. In some simulations, a layer of hidden units was also included.

3. Orientation Identification by Neural Networks

The basic structure of the neural network used in this study is illustrated in Figure 2. It consists of a network with two layers (i.e. no hidden units). The input layer consists of a variable number of orientation-tuned channels, with properties similar to psychophysical channels in the human visual system. The output layer contains either two or 10 units depending on the number of orientations to be identified. In some simulations, an additional processing layer was added with a variable number of hidden units. All the networks used were completely connected, feed-forward nets trained with the standard back-propagation algorithm (Rumelhart *et al.*, 1986). We used our own simulator and all simulations were run on a PC AT.

During training, either two or 10 training orientations were presented to the network and each output unit was trained to respond when a particular test stimulus was present. During testing, the activation levels of the output units were compared and the one with the most activity determined the response of the network. Clearly, at the start of training, and with random connectivity, the network will perform at chance (i.e. 50% of the patterns will be correctly classified in the case of the two orientation task, and 10% in the case of the 10 orientation case). However, after training, performance can reach 100%, depending on the particular situation.

The performance of the network is critically dependent on the properties of the units in the input layer. There are a number of parameters which influence the performance of the network which in general were chosen to be as biologically realistic as possible.

3.1. Shape of the Selectivity Function for the Input Channels

The activity in each of the input units in response to a stimulus with a particular orientation is determined by its selectivity function. On the basis of neurophysiological and psychophysical data (for references, see Henry *et al.*, 1974; DeValois & DeValois 1988) we opted for a Gaussian selectivity function, with a standard deviation σ of 10°, which corresponds to a bandwidth at half-height of roughly 23°. This value corresponds to the most selective neurons which, we believe, are most likely to be involved in fine discrimination. Thus the mean activity, A_i^{mean}, in channel i, centered on orientation θ_0, in response to a stimulus with orientation θ is given by the formula

$$A_i^{\text{mean}} = \exp\left[-(\theta-\theta_0)^2/(2\sigma^2)\right]$$

where σ is the standard deviation of the tuning curve. Thus mean activity is equal to 1.0 when the units preferred orientation is presented and decreases progressively with orientations further away from the preferred value.

3.2. Noise in the Input Units

The activity of real neurons is always influenced by noise. Neurophysiological data indicates that the variance of this noise is roughly proportional to activity (Tolhurst *et al.*, 1982). Thus in our simulations, we added gaussian noise to the activity levels of the input units, by using the equations

$$A_i = A_i^{\text{mean}} + N$$

$$N = \sigma_{\text{noise}} Z$$

$$\sigma_{\text{noise}} = (K A_i^{\text{mean}})^{1/2}$$

Z being determined according to a normal distribution. Here A_i is the activity of channel i for a given trial, N is the noise, Z a number drawn from a normal distribution around 0 and K a constant. The standard deviation of this noise is given by taking the square root of the units mean response multiplied by a constant K. Tolhurst *et al.* (1982) found that in anesthetized cats, the variance was in general about three times the activity level. More recent studies have reported somewhat smaller values in the awake monkey (Vogels *et al.*, 1989). However, in our computer simulations we have assumed that noise in the input units has been reduced by averaging together the outputs of a number of orientation-tuned units—if the outputs of N neurons with independent noise are averaged together, the variance is reduced by a factor of N. In fact, the noise level was typically set with a K value of about 0.1. This corresponds to the effect of averaging together the activity of 20–30 neurons for each input unit.

3.3. Separation between the Response Peak of Input Units

In the simulations described here, we generally used a fixed value for the separation between neighboring input units of 2°. Thus to cover a range of orientations from −59° to +59° (0° corresponding to the vertical) we used an input layer containing 60 units. Given the 23° bandwidth of the neurons in question, this results in considerable overlap between adjacent units. There is no claim that 2° is the actual separation between orientation channels in humans. Presumably, there are channels tuned to any orientation but using smaller difference did not lead to any change in our results.

4. Identification with Two Orientations

Can the connectionist model of orientation identification account for the shape of the psychometric function found in humans? To test this we trained the network with a range of orientation separations. Before the training started, the connection strengths between the input units and the output units were set at small random values. Then the two training stimuli were each presented 2000 times. Each time a test stimulus was presented, an activity level was calculated for each of the 60 input units and target values of 1 and 0 were assigned to the two output units, depending on which test stimulus had been presented. The back-propagation algorithm was then used to modify the connection strengths between the input and output units. After each orientation had been presented 2000 times, the network was tested with 800 random test stimuli in order to determine the percentage of correct responses.

Separation values of 0.25°, 0.5°, 1.0°, 1.3°, 1.6°, 2.0°, 2.3°, 2.6°, 3°, 4° and 8° were used to generate the psychometric function illustrated in Figure 3. Clearly, the shape of the psychometric function generated by the network corresponds very well to the functions of the two human subjects (VB and SJT). This therefore confirms that it is possible to generate reasonable connectionist models of human psychophysical data.

The next question involves trying to understand the strategy developed by the network. Let us consider the weights developed by the network when the separation between the two test stimuli was set at 2° (one stimulus at −1°, the other at +1°). This value is close to the human psychophysical threshold of 1.4° and therefore is a good test of how the network 'solved' the problem under difficult conditions.

The results are shown in Figure 4, which represents the connection strengths of the 60 input units separately for the two output units. One is immediately struck by

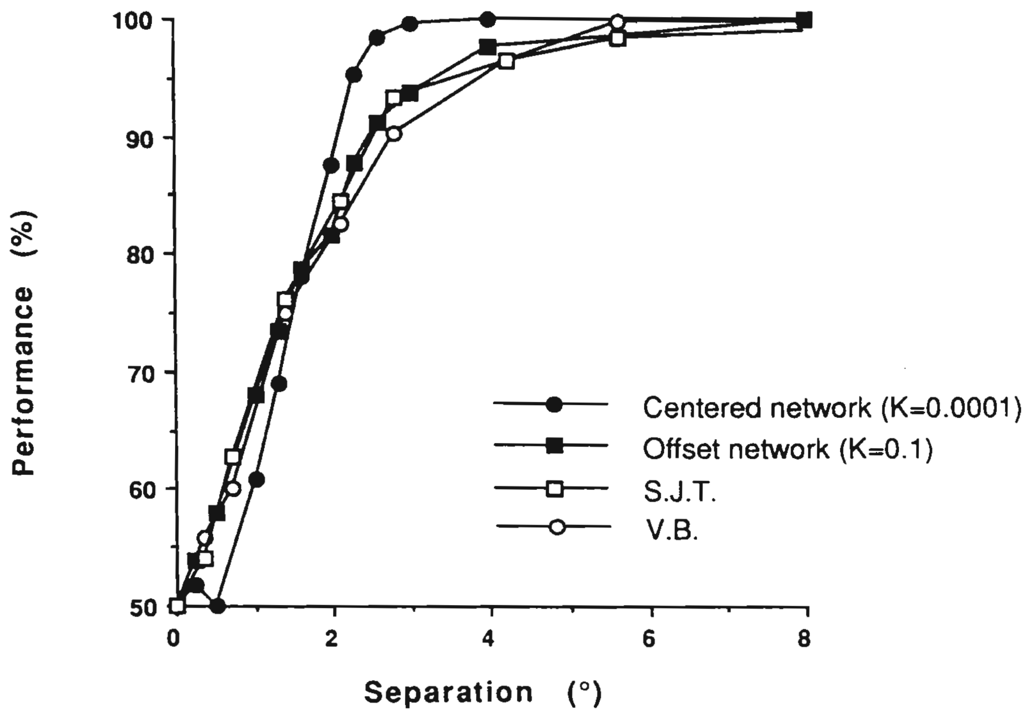

Figure 3. Comparison of the psychometric functions for the network (solid symbols) and two human observers (open symbols). When the network uses the offset channels (solid squares), the fit with human data is very good for a noise level of $K=0.1$. In contrast, a strategy based on only the most active input units (solid circles) requires a much lower noise level of 0.0001 to reach similar threshold and has a sigmoid psychometric function which fails to explain the data.

the symmetry between the weight functions for the two units. If we consider output unit A, we can see that all the units to the left of 0° have positive weights whereas the units to the right have negative weight values. Thus the unit can be thought of as comparing the relative activity of units preferring orientations to the left of vertical with those preferring orientations to the right. A similar description applies to the right hand unit except that the values of the connection strengths have opposite signs.

One feature of the weight distribution is particularly striking. The units with the highest connection strengths are not those which were most activated by the test stimuli. Given that the test stimuli had orientations of $+1$ or $-1°$ relative to vertical, it is clear that the most active units would have been those centered on $-1°$ and $+1°$. However, the weights of these two units are virtually zero. The units which have the most influence on the output units are those which respond best to orientations 15°–20° to either side of the vertical.

Why is this so? The main reason is the Gaussian shape of the sensitivity functions for the input units. Such functions are virtually flat for orientations close to the optimal value—as a result, in the presence of noise, the activity of such neurons carries no information about a subtle change in orientation like that used in this simulation. It is where the slope is the largest that the most significant changes occurs between two close orientations. A Gaussian function with a bandwidth of 25° has

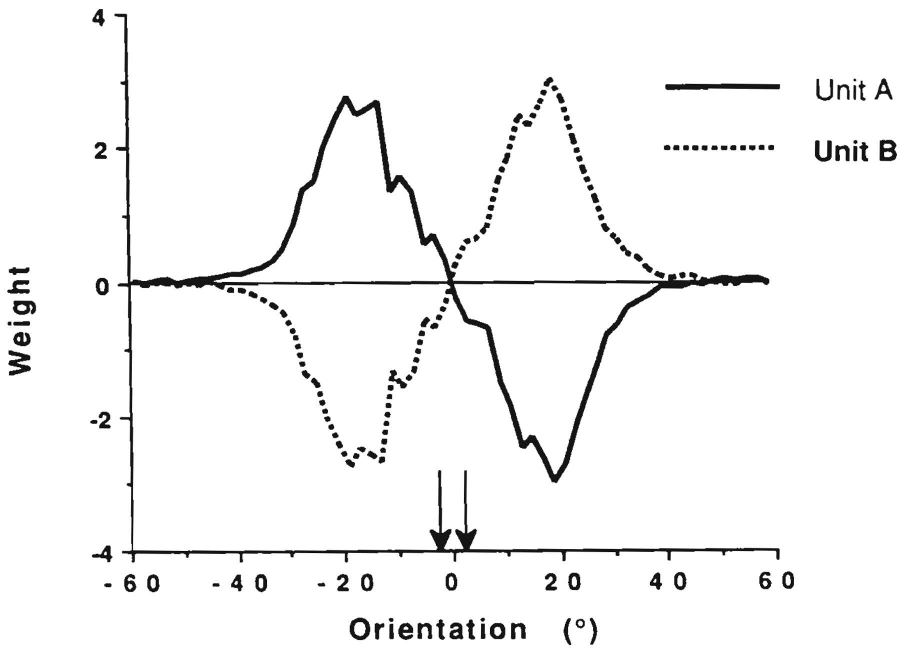

Figure 4. Connection strengths between the 60 input layer units and the two output units A and B after training with two test orientations separated by 2°. The weight sent by a particular input was plotted as a function of the preferred orientation of this unit. When the two orientations presented (arrows) are separated by 2°, the input units that are used, as indicated by the weight values, are not the one whose activity are maximum for these orientations but the one whose peaks are offset on the right and the left by 15°–20°.

very large slope between 10° and 15° on either side of the peak of the response. Because the change in activity between the two orientations in one channel determines the rate of decrease or increase of the weights, channels offset by 10°–15° on either side of the test orientations develop the largest weights. Notice that this result is completely dependent on the presence of noise. Without noise, any change in activity would be significant and the network could pick up any channel in order to perform the task. This is indeed what we observed when we ran simulations with zero noise (not shown).

We have also looked at the weight pattern when the orientation difference is larger. The results with a 30° orientation separation are shown in Figure 5 where it can be seen that the centered channels are increasingly used as the difference between the orientation increases. Therefore, the network used the offset channel only when required to distinguish between two close orientations. When the two orientations are far apart, the weights are maximum for the centered channels because they exhibit the largest change in activity between the two orientations.

In any case, the results of the simulation study tie in nicely with the psychophysical experiments of Regan & Beverley (1985). As was mentioned in an earlier section, they found that adapting the neurons most sensitive to vertical orientations produced no impairment in the subject's ability to detect very small changes in orientations around the vertical. By contrast, such adaptation did impair judgments 10°–15° to either side—a value which fits well with the position of the most important units in the present simulations. However, our model does not

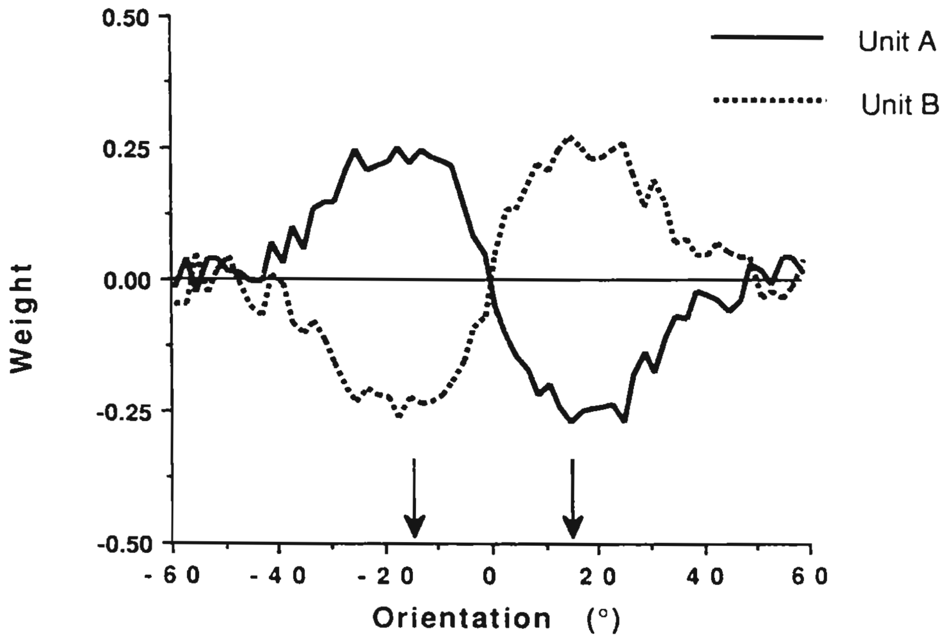

Figure 5. Connection strengths between the 60 input layer units and the two output units A and B after training with two test orientations separated by 30°. The largest weights correspond to the units whose response peaks are centered on the orientation to be discriminated (arrows).

account for the facilitation effect observed by Regan & Beverley when the orientation adapted is similar to the orientations to be discriminated. As we will see in the discussion, this may result from differences between discrimination and identification tasks.

4.1. Comparison between Two Different Strategies

To make sure that the shape of the psychometric curve is characteristic of the strategy used by the network, we determined the same curve for a network that was forced to use the centered channel. To do so, we used a network with only two input units whose tuning curves were always centred on one of the two possible orientations presented. The position of the peak of the tuning curves was changed when the difference between the two orientations was modified. For instance when $+1°$ and $-1°$ were presented the two input units were respectively centered on $+1°$ and $-1°$. When $+2°$ and $-2°$, the properties of the input units were modified so that their peaks were respectively at $+2°$ and $-2°$. In such a configuration, the network had no other choice than using centered channels. The bandwidth of the tuning curves was kept the same, but the noise parameter had to be reduced to 0.0001 in order to obtain the same threshold as in the previous stimulation. Figure 5 indicates the new curve and the comparison with the first network. To obtain the curve, we trained and tested 11 different networks with the same set of separations as the one used for the first network. However, the learning turned out to be much slower for this new network. Therefore, we trained it on 100 000 examples instead of 2000 before testing the performance.

As can be seen from Figure 3, the psychometric curve corresponding to the second network exhibits a sigmoid shape characteristic of signal detection theory.

This was expected since the network is precisely trying to distinguish between two noisy Gaussian distributions. For very small differences between the orientations, the two curves are clearly different. The one corresponding to the offset channels strategy is linearly increasing with the difference in the orientations, whereas on the other curve, the performance stays close to the chance level until a certain threshold is reached and then reaches 100% very rapidly. It is therefore clear that the two strategies can be distinguished on the basis of the psychometric curve shape.

Another difference between the two strategies lies in the difference in the value of K needed in order to achieve a similar threshold. When the centered channels are used, K had to be reduced by a factor of 1000 compared with the first simulation. Although a factor of less than 30 was expected since the number of input units was decreased from 60 to 2, a reduction factor of 1000 cannot be explained solely on this basis. To assess the effect of the number of input units in the noise level, we ran several simulations with only two channels whose peak responses were kept fixed at $\pm 15°$. In this situation, we only had to reduce the noise level to 0.015 to obtain thresholds similar to those determined in the human subjects and furthermore, the shape of the psychometric curve was the same as that obtained with 60 channels (not shown). These simulations demonstrate that the centered channel strategy is much less efficient than the other one.

5. Identification with 10 Orientations

We have seen that the connectionist model can account for the performance of human observers in the two-orientation identification task. Can it also be used to understand how identification can be performed with larger numbers of stimulus classes? Unfortunately, as we have mentioned before, we do not have extensive psychophysical data in the 10-orientation case. However, we thought that it would nevertheless be interesting to see if the network model could account for the data that we possess and if the same strategy could be used to identify 10 and two orientations. We have tested this by simulating the performance of another type of net, this time with 10 output units instead of only two. For a variety of reasons, it was thought that hidden units might be particularly important for the more complicated 10-orientation problem, and so we explored the effects of introducing various numbers of hidden units.

The results of some of these simulations are shown in Figure 6. One of the curves shows the psychometric function produced by a network with 10 hidden units and a noise level K of 0.2. Clearly, the form of the human psychophysical function is again accounted for by the neural network model and again we can examine the weight structure to see what sort of strategy was developed by the network. As can be seen from Figure 7, which shows the connection strengths of hidden-input and hidden-output weights in the form of a Hinton diagram, the strategy again fits with that developed in the two orientation case. It can be seen that the hidden units typically show a weight profile which effectively computes the relative levels of activity on either side of a particular point in the orientation domain.

Thus, the connectionist model that we have used can indeed account for the performance of human subjects in the 10-orientation identification task. However, the noise level K that we used in this case was twice that used for the two-orientation case. Interestingly, when we tried to get a similar match for the same noise level, 0.1, we found that the best results were achieved with a network with three hidden units or, more surprisingly, with a two-layer network. In other words, a two-layer network

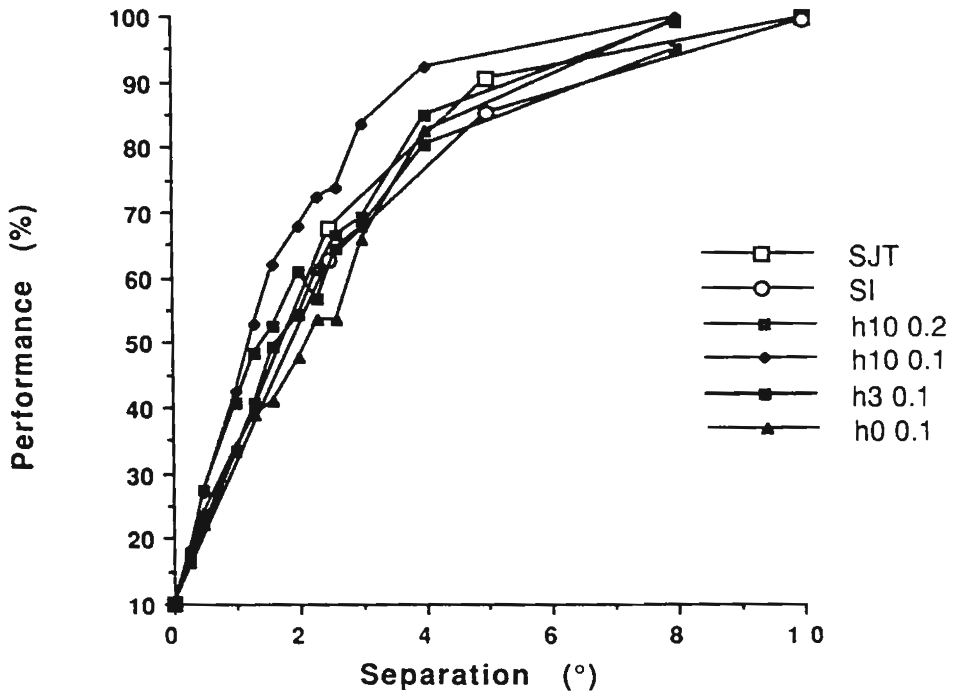

Figure 6. Comparison of the performance of various neural networks (solid symbols) with two human observers (open symbols) on a psychophysical task requiring the identification of 10 different orientations. We trained neural networks with 10, three and no hidden units using a noise level of 0.1 and another one with 10 hidden units, but this time with a noise level of 0.2. The best fit with human data was obtained with 10 hidden units and $K=0.2$ (solid squares with white dots). However, this noise value is different from the one used in previous simulations. Surprisingly, with $K=0.1$, the best matches were obtained with no hidden units (solid triangles) or only three (solid squares). With 10 (solid diamonds), the performance was too good to account for human performance.

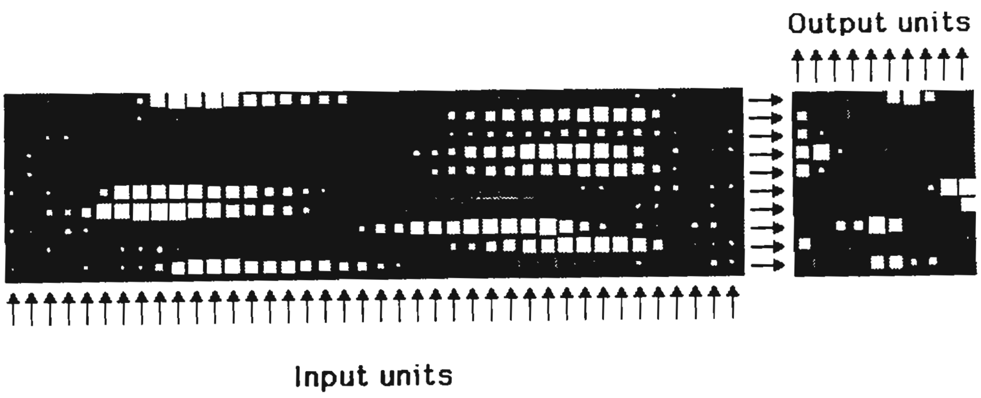

Figure 7. Hinton diagram of the connection strengths between 40 input units and the hidden unit layer in the 10 orientation task for a separation of 8° between two successive orientations. White squares correspond to positive weights and black to negative. Each line in the left rectangle shows the weight received by one single hidden unit. The weight profile resembles the one found for the output units in the two layer network used with two orientations (see Figure 4). The strategy developed by the network is therefore similar for two and 10 orientations.

with a noise level of 0.1 leads to a perfect match in the case of two orientations (Figure 3) and a reasonable one for 10 orientations (Figure 6), although the match is not as good as the one shown in Figure 3. It is difficult to draw any definite conclusion for 10 orientations, since we do not have many data points, but the results shown in Figures 6 and 7, suggest that, indeed, similar strategies may indeed be used in both the two- and 10-orientation tasks.

6. Discussion

In the simulations described in this paper we have tried to combine three different types of data with the aim of achieving a better understanding of how orientation is coded by the visual system. One source of information is neurophysiological—with the aid of single-unit recording studies it is possible to specify the response properties of neurons in the visual cortex. A second source of data comes from psychophysical studies that we have carried out in human subjects. Finally, we have attempted to use neural network modeling to see if we can account for the psychophyical performance of the human subjects on the basis of the neurophysiological data. A number of points emerge from the results of this study.

6.1. Connectionist Modeling of Psychophysical Data

The first point that needs to be stressed is that connectionist models do indeed provide a viable way of trying to bridge the conceptual gap between neurophysiological and psychophysical domains. Using simple neural nets in which the input layer contained the sort of information that is known to be available in the activation patterns of neurons in the visual cortex, it was possible to make detailed predictions concerning the level of performance that might be expected in a psychophysical task involving orientation identification. In order to use such a comparative approach it was necessary to modify the psychophysical protocol so that the task could be defined in terms of pattern classification accuracy.

6.2 Coding Schemes

The pattern of activation across cortical units can potentially be used in a wide variety of ways to code stimulus orientation. As pointed out in the introduction, orientation tuned neurons have frequently been thought of as 'feature detectors', and the assumption has often been made that the orientation of a stimulus is encoded by the location of the most active detectors. More recently, a number of authors have considered coding schemes based on population coding (Lehky & Sejnowski, 1990; Paradiso, 1988; Vogels, 1990) in which all units participate in coding. Some of these ideas can be related to the sort of distributed vector coding proposed by Georgopoulos *et al.* (1986) in the motor system. However, even these population codes often assume that the most active units make the largest contribution, and thus in a sense, they too involve determining some sort of center of gravity for the activation profile.

However, the results presented here actually suggest another possibility. When the network is trained to identify two orientations situated at $-1°$ and $+1°$, and the network has input neurons covering the range from $-59°$ to $+59°$ by 2° steps, the units which are maximally activated by the test stimuli are hardly used at all. The most important inputs are apparently those which 'prefer' orientations offset by 15°–20° to either side of the vertical. Each output unit effectively weights the relative activity of neurons to either side of the vertical. The fact that the psychometric curve for the network and human are similar suggests that human subjects may well use the same strategy. This amounts to saying that if a subject wishes to know whether a line stimulus is tilted slightly to the left or slightly to the right of vertical, he probably does it not by finding the most active 'feature detector' in the array of input neurons, but rather by measuring the relative activity in offset neurons.

It is important to realize that this sort of coding is in principle quite different from the sorts of population vector models that have been proposed recently. The most important single difference is that in the connectionist model proposed here the most active units actually have no influence on performance. One prediction of such a model is that adaptation of the most active units should not impair the accuracy of orientation judgments. As it happens, this is precisely what Regan & Beverley reported some years ago.

The method 'found' by the network for identifying the orientation of an unknown stimulus is in many ways similar to the opponent-process model of color vision proposed over 100 years ago by Hering. It is well known that the retina contains at least three kinds of cone pigments, each maximally sensitive to a different wavelength. However, it is also known that our ability to judge wavelength is not maximal at these points of maximum sensitivity, but rather in the regions of the spectrum between the cone peaks. Hering suggested that color judgments in fact make use of the relative levels of activity in the three different cone systems—not simply the identity of the most active one. In effect, this is exactly what we see in the case of orientation. The only significant difference is that whereas the wavelength selective pigments used in the case of color vision are found in the retina, orientation selectivity is something which only emerges at the level of the visual cortex. Nevertheless, the algorithm 'chosen' by the network does essentially the same thing, i.e. to calculate the relative activity in symmetrical populations of neurons.

Interestingly, Lehky & Sejnowski (1990) have recently used neural network modeling to investigate the coding of stereoscopic depth. They reported that, as in the case of orientation coding, it is not the most active units which are the most critical, since high precision appears to be obtained by comparing the relative activity in groups of non-optimally activated units. It may be this sort of coding mechanism may be very widespread within neural systems, being involved not only in the coding of color, orientation and stereoscopic depth, but many other stimulus parameters such as spatial location, spatial frequency, intensity, velocity and so on.

6.3. Hidden units

A further point concerns whether or not hidden units are required in such perceptual learning tasks. The task with two orientations is essentially simple enough that hidden units do not provide any performance improvement, but with 10 orientations, it is possible that more complicated schemes involving hidden units could provide a considerable advantage. However, our simulations suggest that if the noise level was kept constant, adding hidden units resulted in a level of performance in the 10-orientation task that was actually too good to fit the psychophysical data. This raises the possibility that learning involving the training of hidden units may not be involved in this sort of perceptual task. While it is clear that this is little more than speculation at the present time, it is perhaps worth mentioning that other studies have also indicated that human perceptual learning in tasks thought to require hidden units (such as XOR pattern classification) is in fact very poor (see, for example, Thorpe *et al.*, 1988).

6.4. Extensions to Other Types of Task

One of the reasons why the models used in the present study were relatively straightforward to develop is that we decided to model data obtained using an absolute identification paradigm. To what extent can the implications be extended to other tasks such as orientation discrimination? In principle, it seems likely that the strategy of using offset units can also be used in the case of discrimination tasks—indeed the psychophysical data provided by Regan & Beverley were actually obtained using a discrimination paradigm. However, there are important differences in the experimental protocol which have to be taken into account. In the identification paradigm used by Thorpe & Bonnardel, 1986 the subject sees the same two orientations, symmetric around the vertical (e.g. $-1°$ and $+1°$) which means that the separation between the two orientations and the reference orientation, here the vertical, are kept constant. The orientations are changed only when the subject performance reaches its maximum. Therefore, the subject has enough time to learn the best strategy. In a typical discrimination experiment, the orientation of the reference orientation as well as the separation are often varied frequently. In such conditions, subjects have neither the time, nor the possibility, to select the best channels. This could actually account for the facilitation effect reported by Regan & Beverly when the orientation of the adapting stimulus is similar to the two orientations to be discrimimated. Because subjects have not enough time to select the best channels, they have to use the channels that are centered as well as the offset ones. Now, the activity of the centred channels is misleading in the discrimination task, because these channels do not show any significant change in activity between the two orientations.

Learning to use the offset channels would presumably be considerably easier for human subject, as it is for the network, if the reference orientation was kept fixed. However, if multiple fixed references are used, the network may require hidden units. Devos & Orban (1990) have recently presented a simplified model of such a task on a neural network with hidden units. Unfortunately, in their paper, they do not show the pattern of the weights when learning is complete. However, the orientation tuning curves of the hidden units suggest that in their network, it is again the offset channels which are being used.

The approach used in the present study typifies that used in an increasing number of connectionist studies. As in the recent studies of Lehky & Sejnowski (1988) and Zipser & Anderson (1988) the aim has been to use connectionist modeling to gain insights into the way real nervous systems work, by explicitly examining how real neurons may be used. There can be little doubt that we are only just beginning to explore the potential of these new techniques, but as Sejnowski *et al.* (1988) made clear in their recent review article, the advent of increasingly powerful computing systems will make it feasible to simulate even highly complex brain functions.

References

Bradley, A., Skottun, B. C., Ozhawa, I., Sclar, G. & Freeman, R. D. (1987) Visual orientation and spatial frequency discrimination: a comparison of single units and behavior. *Journal of Neurophysiology*, 57, 755–772.

DeValois, R. & DeValois, K. A. (1988) *Spatial Vision*. Oxford: Oxford University Press.

Devos, M. & Orban, G. A. (1990) Modeling orientation discrimination at multiple reference orientations with a neural network. *Neural Computation*, 2, 152–161.

Georgopoulos, A., Schwartz, A. & Kettner, R. (1986) Neuronal population coding of movement direction. *Science,* **233,** 1416–1419.

Graham, N. (1989) *Visual Pattern Analyzers.* Oxford: Oxford University Press.

Henry, G. H., Dreher, B. & Bishop, P. O. (1974) Orientation specificity of cells in cat striate cortex. *Journal of Neurophysiology,* **37,** 1394–1409.

Howard, I. P., (1982) *Human Visual Orientation.* Chichester: Wiley.

Hubel, D. & Wiesel, T. (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology (London),* **160,** 106–154.

Knudsen, E., du Lac, S. & Esterly, S. (1987) Computational maps in the brain. *Annual Review of Neuroscience,* **10,** 41–65.

Lehky, S. R. & Sejnowski, T. J. (1988) Network model of shape-from-shading: Neural function arises from both receptive and projective fields. *Nature,* **333,** 452–454.

Lehky, S. R. & Sejnowski, T. J. (1990) Neural model of stereoacuity and depth interpolation based on a distributed representation of stereo disparity. *Journal of Neuroscience,* **10,** 2281–2299.

Orban, G. (1984) *Neuronal Operations in the Visual Cortex.* New York: Springer-Verlag.

Paradiso, M. A. (1988) A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biological Cybernetics,* **58,** 35–50.

Parker, A. & Hawken, M. (1985) Capabilities of monkey cortical cells in spatial-resolution tasks. *Journal of the Optical Society of America,* **2,** 1101–1114.

Regan, D. M.& Beverley, K. I. (1985) Post-adaptation orientation discrimination. *Journal of the Optical Society of America,* **2,** 147–155.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986) Learning internal representations by error propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations* (Cambridge, MA: MIT Press.

Sejnowski, T., Koch, C. & Churchland, P. (1988) Computational neuroscience. *Science,* **241,** 1299–1306.

Selfridge, O. G. (1959) Pandemonium: A paradigm for learning. *The Mechanism of Thought Processes.* London: Her Majesty's Stationary Office.

Skottun, B. C., Bradley, A., Sclar, G., Ohzawa, I. & Freeman, R. D. (1987) The effects of contrast on visual orientation and spatial frequency discrimination: A comparison of single cells and behaviour. *Journal of Neurophysiology,* **57,** 773–786.

Swindale, N. V. & Cynader, M. S. (1986) Vernier acuity of neurones in cat visual cortex. *Nature,* **319,** 591–593.

Thorpe, S. J. & Bonnardel, V. (1986) Absolute identification with two orientations. *Perception,* **15A,** 23.

Thorpe, S. J. & Pouget, A. (1989) Coding of orientation in the visual cortex: neural network modeling. In R. Pfeifer (Ed.), *Connectionism in Perspective.* Amsterdam: Elsevier, pp. 447–454.

Thorpe, S. J., O'Regan, J. K. & Pouget, A. (1989) Humans fail on XOR pattern classification problems. In Personnaz, L. & Drefus, G. (Eds) *Neural Networks: From Models to Applications.* Paris: IDSET, pp. 12–25.

Tolhurst, D. J., Movshon, J. A. & Dean, A. D. (1982) The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research,* **23,** 775–785.

Vogels, R. (1990) Population coding of stimulus orientation by striate cortical cells. *Biological Cybernetics,* **64,** 25–31.

Vogels, R. & Orban, G. A. (1990) How well do response changes of striate neurons signal differences in orientation—a study in the discriminating monkey. *Journal of Neuroscience,* **10,** 3543–3558.

Vogels, R., Spileers, W. & Orban, G. A. (1989) The response variability of striate cortical neurons in the behaving monkey. *Experimental Brain Research,* **77,** 432–436.

Zipser, D. & Anderson, R. (1988) A back propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature,* **331,** 679–684.