Spatio-temporal influences at the neural level of object recognition

Guy Wallis
Max-Planck-Institut für biologische Kybernetik, Spemannstraße 38, 72076 Tübingen, Germany

Received 1 August 1997, in final form 21 January 1998

Abstract. In late 1988, Miyashita published work reporting recordings of single cells in the inferotemporal cortex of the macaque monkey (Miyashita 1988 Nature 335 817–20). He described the responses of neurons to a sequence of random fractal pattern images, and how many of the neurons tested were seen to respond strongly to a subset of the images on the basis of sequence presentation order, i.e. appearance in time, rather than their spatial similarity. In this work, I describe a local, Hebb-like learning rule which in conjunction with a simple feedforward neural architecture is capable of replicating the type of temporal-order association apparent in the cells from which he made recordings. The paper also advances reasons for requiring such learning by describing its possible role in establishing transformation invariant representations of objects.

1. Introduction

The primate temporal lobe has long been implicated in the recognition of objects (Ungerleider and Mishkin 1982, Goodale and Milner 1992). Single cell recording in the inferior part of the temporal lobe (IT cortex), has revealed neurons responsive to images of faces (Desimone 1991, Rolls 1992) and other, more abstract image features (Tanaka et al 1991, Miyashita and Chang 1988). Despite the apparent selectivity of these cells, they are not the Gnostic or Cardinal cells of earlier theories (Konorski 1967, Barlow 1972) since, in general, they respond not to one, but rather to a subset of the presented stimuli. As a result, it has been suggested that object encoding is achieved via small ensembles of firing cells which both efficiently and robustly code for individual objects by invariantly encoding sets of salient features (Young and Yamane 1992, Abbott et al 1996).

Of particular relevance to this work is the robustly invariant selectivity of these cells to a preferred stimulus object as it undergoes large rigid transformations such as changes in size, rotation in depth and/or translations. Quite how a single neuron learns such a transform invariant representation of a set of objects remains unclear. Any solution basing selectivity solely upon the similarity in physical appearance of the objects at each viewpoint cannot hope to capture all of the invariances which have been described in the literature. It is this major question which I intend to address in this paper.

One particular set of single cell recordings which tells us a great deal about learning in temporal lobe neurons was reported by Miyashita (1988), who was able to show that the selectivity of neurons in IT cortex are affected by temporal correlations in the appearance of stimuli as well as their physical appearance. His observations form the basis for the neural learning theory described and implemented here, and I shall be presenting a paradigm in
which neurons are able to associate images appearing in sequences in the manner which he describes.

The main thrust of this paper, however, is to describe why Miyashita’s findings may result from the workings of a learning rule involved in training neurons to set up transformation invariant representations of objects in IT cortex. The manner in which the learning rule achieves this is described in detail in the discussion section and in earlier, related theoretical work (Földiák 1991, Wallis and Rolls 1997). By replicating Miyashita’s neurophysiological data, my aim is to lend credence to this earlier work, and in so doing to convey how a rule which draws its power from temporal as well as spatial correlations, not only explains Miyashita’s results, but could underlie the learning of transformation invariant recognition.

2. Short- and long-term memory mechanisms in the temporal lobe

2.1. Introduction

In this section I will be describing in more detail some of the findings upon which the simulations here are based. I will say more about Miyashita’s work and the type of testing paradigm which he used. I will also be discussing a recurrent neural network model described by Griniasty et al (1993) and simulated by Brunel (1996), which was able to replicate the neural response properties which Miyashita and his colleagues report.

2.2. Short-term memory and the delayed match-to-sample paradigm

Although it is Miyashita’s discussion of long-term modifications to IT neurons which is of most interest to the work described in this paper, it is important to first consider an earlier paper which he published with Chang (Miyashita and Chang 1988) concerning short-term memory effects. The two began their experiments by training macaque monkeys to observe randomly generated colour fractal patterns. The experimental paradigm which they used took the form of a delayed match-to-sample (DMS) task, in which the monkey briefly saw a stimulus, then saw nothing for a fixed delay period of 16 s, and then finally, either a matching stimulus, or a distractor stimulus (see figure 2). The animals' task was then to indicate whether the subsequent pattern was the same as the first or not.

Testing proceeded from trial to trial with a consistently ordered choice of test stimuli being maintained throughout the experiment. Examples of the types of patterns used appear in figure 1 and an overview of the testing regime appears in figure 2. After several months of exposure to these stimuli, the monkeys were performing the task at 85% accuracy. Cells were then located in the inferior part of the temporal lobe (IT), and tested for their responsiveness to the 97 test images, as well as to sets of totally new fractal images generated by the same image generation algorithm.

Miyashita and Chang’s initial interest was in the dynamics of neural responses during the delay period. They described extended neural activity throughout this period, which they interpreted as supporting short-term memory of the original image, for comparison with the second image.

One interpretation of Miyashita and Chang’s results is that they reveal the working of an attractor neural network, in which attractor states of patterns close in temporal presentation order overlap (Griniasty et al 1993). Griniasty et al make a convincing case for the role of recurrent connectivity in the maintenance of neural activity during the delay period, such as appears during DMS and paired associate working memory tasks. They have made various
Figure 1. Fifteen example fractal images, generated using the algorithm described by Miyashita et al. (1991).

Figure 2. Overview of the testing paradigm used by Miyashita (1988) showing (a) the presentation timing and (b) the repeating sequence of 97 fractal image test stimuli.

proposals about the implementation of the learning in the network, but in a later simulation of the model Brunel (1996), temporal lobe neurons were interconnected via a mixture of Hebbian and anti-Hebbian synapses.

Since it is also my intention to replicate Miyashita’s results, I shall be discussing Brunel’s model in some detail. However, before doing so, it is instructive to consider Miyashita and Chang’s DMS task itself, and reasons for supposing that the behaviour of neurons in this task are very different to behaviour under normal viewing.

When simply viewing stimuli, neural activity only proceeds for a few hundred milliseconds after the removal of the activating stimulus (Rolls and Tovee 1994). The probable explanation for this is that the maintenance of activity seen in Miyashita and
Chang’s work, is not due to recurrent connections in the temporal lobe, but rather between the temporal lobe and prefrontal cortex (Desimone et al 1995, Fuster et al 1985). The fact that there is much less recurrent feedback in IT neurons under normal viewing conditions, calls into question the existence of a learning, recurrent network of the sort Brunel has described. I will raise this and other points in more detail when describing an alternative, much simpler architecture in subsection 3.3.

2.3. Long-term memory in the temporal lobe

After describing maintained neural activity during the delay period, Miyashita went on to search IT for long-term representations of the fractal patterns which he had used in the DMS task (Miyashita 1988).

Miyashita discovered that many neurons within IT cortex had developed selectivity for small sets of the fractal images. Although the experimental paradigm did not explicitly require the overall test sequence to be remembered, Miyashita also noted that these neurons consistently responded well to single images which neighboured one another in the test sequence. For example, one neuron might respond preferentially to images 5, 6 and 7, whereas another neuron would respond to images 37, 38 and 39. The fact that the images were generated randomly, meant that there was no particular reason (on the grounds of spatial similarity) as to why these images should have become associated together by a single neuron. Instead, the results indicate the importance of temporal order in controlling the learning of neural selectivity. As well as providing an explanation for the maintained activity of IT neurons during the delay period, Brunel’s model is also able to capture this remarkable long-term ‘side-effect’ of the short-term memory task. Temporal correlations present in the image presentation order were encoded by allowing neurons active during the delay period to remain active for a short period after the presentation of the next stimulus, just long enough for afferent synapses to modify.

Brunel’s explanation for the formation of the long-term representation shares clear parallels with the model which I shall be describing in this paper, although important differences do exist in the architecture and form of Hebbian learning as will become apparent in the next section. The most important difference, however, is in my explanation for why I think the long-term representations of the stimuli changed. I will explain more about this theory, in terms of a mechanism for solving invariant object recognition, in the discussion section.

3. Learning from spatio-temporal associations

3.1. Introduction

Having discussed the background to the work, this section proposes a biologically justifiable learning mechanism and neural architecture, for implementing the temporal aspect of the learning which Miyashita describes.

3.2. Learning rule

The simple premise of the simulations described here, is that connections between neurons depend not only upon current input and current neural activity, as Hebb proposed (Hebb 1949), but also upon previous neural activity as well.

The idea that temporal delays might play a role in cortical learning is not new, and
appears in models of classical conditioning (Klopf 1972, Klopf 1988, Sutton and Barto 1981), route learning and navigation (Blum and Abbott 1996, Schölkopf and Mallot 1995) and stereo matching (Stone 1996). Indeed, the model of Griniasty et al described in this paper is clearly related to an earlier model proposed by Herz et al (1989) which is capable of learning and reproducing temporal sequences, such as the sounds which constitute a word or tune.

However, unlike many of the models mentioned above, in the theory described here the temporal order of stimulus presentation is only important in setting up the representation and is itself in no way stored or reproducible. It may be deduced from the selectivity of a neuron to individual images, but no one image will elicit the seeing of another. It is also important to realize that, unlike some of the models above, the aim of training is to produce neurons which respond fully to a single isolated image, rather than requiring a sequence of images to activate them.

The actual learning rule used here was originally proposed by Sutton and Barto (1981), although its relevance to object recognition learning (Földiák 1991) and specifically in setting up representations in the primate temporal lobe, is more recent (Földiák 1992, Wallis and Rolls 1997). The learning rule works by replacing the current neural activation term found in standard Hebbian learning paradigms, with a running, time averaged measure called the ‘trace’ value. The use of this memory trace has lead to it being called the trace rule.

Various biological bases for this temporal trace have been advanced:

- The persistent firing of neurons for as long as 100–400 ms observed after presentations of stimuli for 16 ms (Rolls and Tovee 1994) could provide a time window within which to associate subsequent images. It is suggested that this would, in natural circumstances, be time enough for new views of the effective stimulus object to be seen and learnt. This supposes that firing at the soma should not only propagate along the axon but also be capable of affecting learning in the dendritic tree. Evidence to support this claim has in fact recently been reported in rat neocortical layer V pyramidal neurons (Markram et al 1995).

- The binding period of glutamate in the NMDA channels, which may last for 100 ms or more, may implement a trace rule by producing a narrow time window over which the average activity at each presynaptic site affects learning (Rolls 1992, Rhodes 1992, Földiák 1992).

- Chemicals such as nitric oxide may be released during high neural activity and gradually decay in concentration over a short time window during which learning could be enhanced (Földiák 1992, Montague et al 1991).

- Researchers have reported a short-term synaptic enhancement which increases firing as a result of repeated input within a short time window of approximately 100 ms, see Fisher et al (1997) for a review. This phenomenon has been shown to facilitate sensitivity to temporal sequences (Buonomano and Merzenich 1995), which could in turn, lead to long-term synaptic enhancement of the sort simulated with the trace rule.

The precise mechanisms involved may alter the form of the trace rule which should be used. A trace implemented by extended cell firing should be reflected in representing the trace as the cell’s firing rate, rather than as some form of internal signal maintained within the cell. In the simulations described in this paper, the trace of neural activity is assumed to be stored postsynaptically and internally, with its value being defined by current and previous firing of the cell, but without the value itself affecting the overall firing of the neuron. Other implementations are certainly feasible, indeed Földiák (1992) describes an alternative presynaptic trace rule which models individual NMDA channels.
The formulation of the trace learning rule used can be summarized as follows:

\[ \Delta w_{ij}(t) = \alpha y_i(t) \cdot x_j \quad \text{and} \quad \sum_j w_{ij}^2 = 1 \quad \text{for each } i \text{th neuron} \]

\[ \overline{y}_i(t) = (1 - \eta)y_i(t) + \eta \overline{y}_i(t-1) \quad \text{and} \quad y_i = \Phi \left( \sum_j x_j w_{ij} \right) \]

where \( x_j \) is the \( j \)th input to the neuron; \( y_i \) is the output of the \( i \)th neuron; \( w_{ij} \) is the \( j \)th weight on the \( i \)th neuron; \( \eta \) governs the relative influence of previous cellular activity and the new input; and \( \overline{y}_i(t) \) represents the value of the trace of the \( i \)th cell at time \( t \). The value \( \alpha \) controls how quickly the weights adapt, and was set at 0.001 in all of the simulations described here.

The recurrent form of the trace rule has the effect of producing a decaying, weighted average of previous cell activity. The parameter \( \eta \) serves to control the precise weightings and the time period over which averaging occurs. The pairwise averaging proposed by Griniasty et al is one possible weighting function, but the trace rule is preferred here simply because its natural decaying form lends itself to straightforward implementation locally within a neuron, permitting a wide range of potential mechanisms for its implementation— as described above.

The final output of a neuron depends upon an internal transfer function and the activity of neighbouring cells, the overall effect of which is represented by the function \( \Phi \) in the preceding equation. The internal transfer function (i.e. neural activation function) used in these simulations, is sigmoidal in form. If we take \( N_y \) as the number of neurons within a pool of competing neurons, \( N_x \) as the number of inputs per neuron, and \( z_{\text{max}} \) as the maximum activation of any neuron within the pool, then the activation function is implemented as follows:

\[ z_i = \sum_{j=1}^{N_x} x_j w_{ij} \quad \text{and} \quad \overline{z}_y = \frac{1}{N_y} \sum_{n=1}^{N_y} z_n \]

\[ \gamma_i = \frac{\exp \left( m \frac{z_i - z_{\text{av}}}{z_{\text{max}} - z_{\text{av}}} + c \right)}{\exp \left( m \frac{z_i - z_{\text{av}}}{z_{\text{max}} - z_{\text{av}}} + c \right) + 1} \quad \text{with } c = -1.5 \text{ and } m = 3.0. \]

The parameters \( m \) and \( c \) control the degree of saturation and the gradient of the activation function. Competition within the network is implemented using the softmax algorithm (Bridle 1990), which provides a scalable generalization of winner-take-all inhibition:

\[ y_i = \frac{e^{\gamma_i s}}{\sum_n e^{\gamma_n s}} \quad \text{with } s = 10. \]

The parameter \( s \) controls the strength of the competition. In the limit \( s \to \infty \) competition approximates ever more closely to winner-take-all.

As is evident from the above equations, the neurons simulated here use simple, average firing rates to communicate information. No attempt is made to replicate the complex temporal pattern of spikes emanating from real neurons, and as such the network neurons are an extremely simplified model of cortical neurons. In defence of this and similar models, recent studies of spike trains have shown that in IT neurons, over 80% of the information about a visual stimulus contained in a spike train can be described by the first principal component of the spike train signal, which itself is very highly correlated to firing rate in the vast majority of neurons (Tovee et al 1993).
3.3. Network architecture

The recurrent network format used by Brunel (1996) was avoided here for several reasons:

- The only evidence for lengthy, maintained activity is in a working memory task in which no other stimuli are shown during the delay period. Under normal viewing conditions, this activity is rapidly attenuated (Baylis and Rolls 1987, Miller and Desimone 1994).
- Single cell recording evidence from the ventral stream V1–V2–V4–IT suggests that the speed with which neural activity progresses through the processing stages leaves no scope for time consuming recurrent processing (Rolls and Tovee 1994, Thorpe and Imbert 1989).
- Recording evidence also fails to reveal any gradual adaptation in neural activity such as one would expect in a recurrent attractor network architecture (Thorpe and Imbert 1989).

Instead, a two-layer feedforward network was constructed (see figure 3). The first layer acts as a local feature extraction layer and contains three channels, each consisting of a $32 \times 32$ grid of neurons. The three channels are used to segregate each image into its three constituent colour pixel values (red, green, blue). The $32 \times 32$ layer-1 neurons are arranged in a lattice of 64 ($4 \times 4$) competitive pools. The softmax algorithm then implements soft competition between the 16 pool members. Each pool receives input from a corresponding $4 \times 4$ patch of the $32 \times 32$ pixel input image.

Competition in the network is restricted to within the competitive pools, and unlike the Brunel model, is fixed rather than plastic. As described above, competition is implemented using the softmax algorithm. Softmax is convenient, in that many researchers are familiar with it, but it has the disadvantage of not being readily implemented via simple cortical circuitry. Fortunately, it turns out that the precise details of the competition are unimportant to the functioning of the model. Indeed, the model described here has also successfully been implemented using a much simpler form of ‘winner-take-most’ inhibition, similar to that used by Rumelhart and Zipser (1985), which is more amenable to implementation with real, inhibitory interneurons.

The modification of connections between layer-1 neurons and the input, proceeds as described in the previous subsection, with the parameter $\eta$ set to zero, which has the effect of rendering a memoryless trace rule—exactly equivalent to standard, weight bounded Hebb learning. Above layer-1 there is a second layer consisting of a single pool of 16 neurons. Competition within this pool is once again implemented using the softmax algorithm. Each neuron in layer-2 is fully connected to all $32 \times 32$ neurons in all three channels of layer-1. Connections between the layer-1 and layer-2 neurons are once again modified using the trace learning rule, but with $\eta = 0.7$.

4. Simulating fractal image learning

4.1. Introduction

Having established a learning rule and network architecture, this section describes a series of experiments carried out to discover whether the results described by Miyashita can indeed be replicated. In particular, the aim is to see whether neurons can associate fractal images on

† The model hence makes no attempt to simulate either true retinal colour sensitivity, or colour opponency.
Figure 3. The two-layer network used in the simulations, containing three separate input layers, one for each colour channel. A square in the first layer represents a single inhibitory group of $4 \times 4$ neurons. In the second layer the squares represent each of the 16 output neurons.

Figure 4. The response of two neurons to all 97 test images and 97 novel fractal images. The contrast in the degree of clustering between novel and test images and the amplitude of responses, demonstrates effective learning.

the basis of their presentation in time as opposed to simple spatial similarity. The section is divided into two experiments: the first establishes the ability of the learning rule to replicate the results, and the second the effectiveness of the trace rule over standard Hebbian learning.
4.2. Experiment I: methods

The type of stimuli used during the experiment appear in figure 1. They were generated using the algorithm described by Miyashita et al. (1991) in order to reproduce his experimental stimuli as closely as possible. Simulations involved repeated exposure of the network to the DMS task. In each training cycle a stimulus was presented to the network just long enough to establish a firing pattern in the layer-2 neurons: this was the equivalent of the 200 ms period used by Miyashita to present a test stimulus. After this, the input image was removed, but the activity in the layer-2 neurons maintained.

The reason for maintaining activity in the neurons during the delay period was to replicate the maintained activity described in temporal lobe neurons during a DMS task (Desimone et al. 1995, Fuster et al. 1985, Baylis and Rolls 1987, Miller and Desimone 1994), and mentioned in subsection 2.2. It is important to note that activity was maintained at its initial level during the delay period, and not adapted or processed in any way, as would be the case in a recurrent attractor network architecture of the type described by Brunel. This removes the need for a learning, recurrent connection and supposes that the effect of the frontal lobe is to maintain the current activity in IT rather than modify it. The consistent value of the neural output \( y_i \) in turn causes the trace value \( \tau_i \) to saturate to a stable value. However, this does not affect the weights \( w_{ij} \), because the inputs \( x_j \) are all zero (as a result of there being no image present during the delay period).

After the delay period, a second stimulus was presented which was either the same as the first stimulus, or chosen at random from the 96 other images, both with probability 0.5, in accordance with the DMS testing format described by Miyashita and Chang. Activity was similarly maintained onto the next training cycle in which the next image in the set sequence was trained. After running all 97 images in the DMS task 800 times, the delay period response of each neuron was recorded for each of the 97 images, yielding a final response value for each neuron to each stimulus. To complete the testing, a novel set of 97 images were generated using the same fractal algorithm and likewise tested on the network.

4.3. Results

Since the aim of these simulations has been to replicate Miyashita’s findings, I have chosen to analyse the results of the training in the same manner as he does in his paper. Miyashita’s goal was to show that image sequence order had affected the choice of which stimuli should be associated by individual neurons. The first piece of evidence which he provided was plots of the individual cell’s responses to the 97 individual training images as well as 97 novel images generated with the same fractal generation algorithm. I have plotted the same style of response chart in figure 4, showing the responses of two layer-2 cells to the 97 trained stimuli and the 97 novel stimuli. The bunching of strong responses along the image number axis, for trained stimuli, clearly demonstrates the preference of these neurons for groups of stimuli which neighbour one another in terms of sequence ordering, i.e. which appeared closely in time. The more sporadic form of the responses to the novel images also confirms that this is an effect of learning. These results are in general agreement with the form of responses appearing in Miyashita’s paper.

The second source of evidence provided in his paper yields a more global and quantitative comparison, namely a graph of the autocorrelation function of all responsive neurons. If the tendency of cells is to respond to images appearing in close succession in the presentation sequence (i.e. similar serial position numbers), then there should be high correlation between neural response to sequential stimuli, and this correlation should
smoothly decay with sequence based distance between stimuli. This function was calculated for all 16 output cells across all stimuli, the results of which appear in figure 5. The smoothly decaying curve seen for the learned stimuli demonstrates a strong correlation between responses to neighbouring images in the sequence and is in stark contrast to the correlation for responses to the novel stimuli†. Correlation becomes indistinguishable from zero at around five image steps away from the central stimulus, which is in close agreement with the results provided by Miyashita in his paper. For the untrained stimuli, however, there is no evidence of any order based structure in neural selectivity, demonstrating the effectiveness of the training used.

4.4. Experiment II: methods

The previous subsection was effective in demonstrating the ability of the trace rule and the simple feedforward architecture to reproduce the association of multiple fractal images on the basis of the temporal order of their presentation. Before concluding that the trace rule played a crucial role in this, it is important to repeat the study with a memoryless version of the trace rule, in the expectation that the previous temporal influence on neural learning will disappear. To this end, the exact same training procedure used in experiment I was repeated, but with the second-layer neurons now being trained with the same trace value \( \eta = 0 \) as in the first layer. The result of this change is that learning now proceeds exactly as weight bounded, standard Hebbian learning in both layers.

† Correlation values with a maximum of 0.5 were obtained using the biologically relevant alternative to softmax competition mentioned earlier, underlining the fact that the use of softmax is not critical to the functioning of the network.
Spatio-temporal influences in object recognition

Figure 6. Average autocorrelation function for the responses of all 16 cells to the trained and novel test sets using normal Hebbian learning. In this case there is no apparent relation in response of neurons to stimuli which were seen in a sequence during training, with the correlations being indistinguishable from the untrained set.

4.5. Experiment II: results

The results appear in figure 6. The total absence of presentation order correlations in this case, demonstrates the crucial role played by the trace rule in setting up the response properties of the output neurons.

5. Discussion

This paper has demonstrated that a local, Hebb-like learning rule can train neurons to associate images appearing in time, in accordance with single cell recording data described by Miyashita (1988). By reproducing Miyashita’s results, the paper consolidates earlier work concerning the role played by the trace rule in learning invariant representations of objects throughout the ventral stream of primate visual cortex (Földiák 1991, Wallis and Rolls 1997). This earlier work argued that our environment is so structured, that potentially spatially dissimilar images regularly seen in close temporal succession are likely to belong to the same object, and should therefore be associated together. This piece of information about environmental structure then takes the form of a tendency, or prior in the sense of Bayesian statistics, of the human visual system to associate images of objects with one another over short periods of time.

The essence of how the trace rule can use this regularity to establish view-invariant representations of objects, can be seen by considering the situation in which a single neuron is strongly activated by some element of a real world object. In such a case, the short-term average activity of the neuron will be high, and if a new aspect of the object is seen before
the effects of this activity die away (in the order of 0.5 s), then not only will the initially active afferent synapses modify onto the neuron, but so also will the synapses activated by the transformed version of this stimulus. In this way, the cell will learn to respond to either appearance of the original stimulus. Making such associations works in practice because it is very likely that within short time periods, different aspects of the same object will be being inspected. The cell will not, however, tend to make spurious links across stimuli that are parts of different objects because of the unlikelihood in the real world of one object consistently following another.

Dong and Atick (1995) have already successfully measured temporal as well spatial correlations in natural scenes. Their results indicate a relatively smooth transformation of natural scenes over time, in accordance with the theory discussed here. Having said that, if trace rule theory is to work, one would expect to see temporal correlations of object appearance in a scene to last even longer than the spatio-temporal ones which they report, forming a second layer of temporal smoothness beyond the level of pixels, at the level of objects. If, as seems reasonable, such temporal regularity exists, and the length of time over which any one object is seen, $\tau$, is assumed to be Jeffrey’s distributed (i.e. $P(\tau) \propto 1/\tau$ meaning that objects tend to be seen for short periods but are occasionally seen for much longer periods), then the trace rule acts as an optimal, linear, temporal filter (Wallis and Baddeley 1997).

Before cementing the link between object invariance learning and the temporal associations described in Miyashita’s work, there remains at least one very real problem to tackle. Making associations between stimuli over the long delay period of 16 s which he describes, would not normally be desirable, since the viewer would typically have moved his attention to a new object. This in turn might lead to the spurious linking of views of different objects. In other words, under normal viewing conditions, associations should not be made over the large delay periods used in Miyashita’s experiments. The proposed solution to this problem is held in the special task which his monkeys were set to solve. As described earlier in subsection 2.2, during the DMS task the activity of IT neurons is maintained via connections with prefontal cortex, apparently exclusively as part of solving the DMS task. In the case that the neurons in IT continue to fire throughout the delay period, the trace value of each layer-2 neuron simply saturates to its highest possible value ($3.33 \eta$ for $\eta = 0.7$) and remains at this value, as was the case in the simulations described earlier. Hence, if the memory of the activity of the neuron is explicitly maintained then such associations can indeed be made, as part of the standard operation of this simple, local learning rule.

It is worth adding that under normal viewing conditions, the lengthy reverberations reported by Miyashita (and so instrumental to Brunel’s alternative, recurrent network model) do not occur. Indeed, in another form of DMS task, in which the monkey must indicate a match in a stream of stimuli, the presence of intervening stimuli quickly abolishes any maintained activity (Baylis and Rolls 1987, Miller and Desimone 1994). This in turn precludes the feared, erroneous association of multiple views of different objects over many seconds.

As a corollary to this point, it is interesting to consider what would happen if the spatio-temporal constraints of our world were to be violated. It might be possible for humans to over-generalize between objects if views of them appear in artificial, interleaved sequences. This type of over-generalization would provide the first evidence from human subjects that object views are associated on the basis of their appearance in time, and motivates a series of psychophysical experiments described elsewhere (Wallis and Bülthoff 1997).

† The fact that monkeys can still solve the second version of the DMS task calls into question the need for maintained activity in the temporal lobe to solve the DMS task at all.
References

Blum K I and Abbott L F 1996 A model of spatial map formation in the hippocampus of the rat Neural Comput. 8 85–93
Brunel N 1996 Hebbian learning of context in recurrent neural networks Neural Comput. 8 1677–710
Buonomano D V and Merzenich M M 1995 Temporal information transformed into a spatial code by a neural network with realistic properties Science 267 1028–30
Földiák P 1991 Learning invariance from transformation sequences Neural Comput. 3 194–200
Goodale M A and Milner A D 1992 Separate visual pathways for perception and action Trends Neurosci. 15 20–5
Griniasty M, Tsodyks M V and Amit D J 1993 Conversion of temporal correlations between stimuli to spatial correlations between attractors Neural Comput. 5 1–17
Hebb D 1949 The Organisation of Behavior (New York: Wiley)
——1988 A neuronal model of classical conditioning Psychobiology 16 85–125
Miller E K and Desimone R 1994 Parallel neuronal mechanisms for short-term memory Science 254 1377–9
Miyashita Y 1988 Neuronal correlate of visual associative long-term memory in the primate temporal cortex Nature 335 817–20
Rolls E T 1992 Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical areas Phil. Trans. R. Soc. B 335 11–21
Rumelhart D E and Zipser D 1985 Feature discovery by competitive learning Cogn. Sci. 9 75–112
Schölkopf B and Mallot H A 1995 View-based cognitive mapping and path planning Adaptive Behavior 3 311–48
*Psychol. Rev.* **88** 135–70
Thorpe S J and Imbert M 1989 Biological constraints on connectionist models *Connectionism in Perspective* ed R Pfeifer, Z Schreter and F Fogelman-Soulie (Chichester: Wiley) pp 63–92
Tovee M J, Rolls E T, Treves A and Bellis R P 1993 Information encoding and the responses of single neurons in the primate temporal visual cortex *J. Neurophysiol.* **70** 640–54