A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons

David Zipser* and Richard A. Andersen†

* Institute for Cognitive Science, University of California, San Diego, La Jolla, California 92039, USA
† Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Neurons in area 7a of the posterior parietal cortex of monkeys respond to both the retinal location of a visual stimulus and the position of the eyes and by combining these signals represent the spatial location of external objects. A neural network model, programmed using back-propagation learning, can decode this spatial information from area 7a neurons and accounts for their observed response properties.

This article addresses the question of how the brain carries out computations such as coordinate transformations which translate sensory inputs to motor outputs. Visual inputs are collected in the coordinate frame of the retina on which the visual environment is imaged, but motor movements such as reaching are made to locations in external space. Changes in eye position will alter the retinal locations of targets while their spatial locations remain constant. As a result, visual inputs must be transformed from retinal coordinates to coordinates that specify the location of visual objects with respect to the body to perform accurately directed movements.

Lesions to the posterior parietal cortex in monkeys and humans produce profound spatial deficits in both motor behaviour and perception. Humans with lesions to this area can still see but they appear to be unable to integrate the position of their bodies with respect to visual inputs. The lesion data further suggest that it is the inferior parietal lobe, which comprises the posterior half of the posterior parietal cortex,
Fig. 1  a, Experimental protocol for determining spatial gain fields, with the projection screen viewed from behind the monkey's head. These experiments were carried out several years before the start of the modelling described here, but some of the data are being presented for the first time. To determine the effect of eye position, the monkey with head fixed, fixates on a point, f, at one of 9 symmetrically placed locations on the projection screen. The stimulus, S, is always presented at the same retinal location, chosen as the maximum-response zone of the retinal-receptive field. The stimulus consists of 1- or 6-degree diameter spots flashed for 500 ms. Each measurement is repeated 8 times. b, Peri-stimulus histograms of a typical gain field determination. The nine histograms are located in the same relative positions as the fixations that produced them. The vertical line indicates the time of visual stimulus onset. c, A graphic method for illustrating these data in which the diameter of the darkened inner circle, representing the visually evoked gain fields is calculated by subtracting the background activity recorded 500 ms before the stimulus onset from the total activity during the stimulus. The outer circle diameter, representing the total response gain fields, corresponds to the total activity during the stimulus. The annulus diameter corresponds to the background activity that is due to an eye-position signal alone, recorded during the 500 ms before the stimulus presentation.

which is involved in spatial processes.

Anatomical and physiological experiments in macaque monkeys indicate that the inferior parietal lobe contains at least four separate cortical fields. Area 7a contains visual and eye-position neurons, area 7b contains somatosensory and reach-related cells, area MST contains visual motion and smooth pursuit eye movement activity (refs 12–14; R. H. H. Wurtz and W. T. Newsome, personal communication); and area LIP contains visual and saccade-related activity. It has been proposed that the area most likely to perform spatial transformations is area 7a. Most of the cells in this region were found to receive a convergence of both eye-position and retinal signals. The interaction between eye-position and visual responses was non-linear, and in most cases the visual response could be modelled as a gain that was a function of eye position multiplied by the response profile of the retinal receptive field. Thus the visual receptive field remained retinotopic, but the magnitude of the response was modulated by eye position. This modulation can be shown to produce a tuning for the location of targets in head-centred coordinates that is eye-position-dependent; the cells will fire most for a particular location in craniofacial space, but only when the eyes are at the appropriate positions in the orbits. No cells were found that coded target location over all eye positions (eye-position-independent coding), indicating that this information can only be contained in the pattern of activity of a population of neurons. Here we describe a neural network model that shows how eye-position-independent location can be extracted from a population of area 7a neurons. The model also reproduces the non-linear interactions of eye position and retinal position information seen in actual area 7a neurons, and demonstrates response properties, such as large receptive fields, which are strikingly similar to those observed in single-unit recording studies.

The neural modelling technique we use differs significantly from most previous approaches, which first found an algorithm to accomplish the computation, and then specified neural models to implement the algorithm. Our approach is based on the use of a neural network training procedure, called 'back propagation', which can programme artificial neural networks to compute arbitrary functions (refs 19–23; S. R. Lehky and T. J. Sejnowski, personal communication). Unlike computer programming, programming by training uses only examples of input and output. This means that it is not necessary to know in advance the algorithm that the network will use; the learning process will discover an appropriate algorithm. Back propagation networks have internal or hidden units that are free to take on the response properties to best accomplish the computational task being learned. It is the properties of these hidden units that we find resemble those of cortical neurons. The back-propagation procedure accomplishes learning by adjusting the strengths of the synapses within the network.

Experimental results from area 7a

The experimental data that must be accounted for by any model of area 7a were collected previously in an extensive series of studies with awake, unanaesthetized monkeys. Here we describe new analyses of these data that facilitate a comparison of area 7a cells with the units generated by training the network model.

Three of the major classes of area 7a neurons are of interest here: the eye-position cells responding to eye-position only (15% of all cells sampled from area 7a), the visual cells responding to visual stimulus only (21%), and the spatially tuned cells responding to both eye position and visual stimulus (57%). Neurons in the first two classes presumably represent the eye position and retinal location information available to area 7a as input. The interaction of eye position and visual information found in the third class of cells produces a representation for the head-centred location of visual targets that is eye-position-dependent.

The experimental protocol involved recording neuronal activity extra-cellularly from awake, unanaesthetized monkeys trained in various visuospatial tasks (Fig. 1). The eye-position sensitivity was tested by having the animal fixate on a small point at different eye positions, with the head fixed in otherwise total darkness. The eye-position cells typically showed a linear increase in activity for a range of horizontal or vertical eye positions, although some cells showed more complex eye-position coding. An ensemble of 30 eye-position unit responses is shown in Fig. 4c. The receptive fields of the visual cells were tested by flashing a spot stimulus at different locations in the visual field while the animal fixated on a target at a single eye position. Surfaces were fit to these data points using a gaussian interpolation. These cells typically had large receptive fields equally distributed across the visual field for the population of neurons with a single peak of activity. The shape of the receptive fields approximated a symmetrical gaussian with a 1/e width of 15 degrees (see legend to Fig. 4).

The spatially tuned neurons were the largest group and showed a convergence of eye position and retinal position information. The receptive fields are very large (often over 80 degrees in diameter) and have one or more peaks that form a smoothly changing, hilly landscape. A set of 12 retinal receptive fields from spatially tuned cells, arranged according to peak eccentricity and complexity, is shown in Fig. 2.

As mentioned above, the evoked visual response of spatially tuned neurons varies as a function of eye position. This effect was examined by collecting data under the condition in which the visual stimulus always appears at the peak location in the retinal receptive field, but with the animal fixating at nine
different eye positions\textsuperscript{10}. These plots are referred to as spatial gain fields. Figure 1 demonstrates the experimental protocol for mapping spatial gain fields.

The majority of the spatial gain fields are roughly planar. This planar behaviour is evident in Fig. 1c, where the darkened inner circles are proportional to the magnitude of the visual evoked response, the outer circle diameter to the total activity during the flashed stimulus, and the annulus diameter the background activity due to the eye position alone. The data for the visually evoked response in Fig. 1c can be fitted by a plane tilted up for eye positions to the left and also tilted up for downward eye positions. Analysis using linear regressions in the two dimensions of eye position indicated that the gain fields for the visually evoked activity (represented by the dark inner circles) were planar, or had a large planar component, for 55% of the neurons. Interestingly, 80% of the total response gain fields (represented by the outer circles) were planar or largely planar.

A useful way to further characterize the nonlinear combinations of eye and retinal information is to compare the contributions of each to the total response of the cells. This can be done simply by comparing the dark inner circles, which represent the visual contribution to the response, with the white annuli, which represent the eye position contribution to the response. This comparison shows three basic types of gain fields. For 28% the background and evoked activities change in a parallel fashion (Fig. 3b, e, f). In most of the cells (43\%), the evoked activity changes with eye position while the background activity, if any, remained constant (Fig. 3a, c, d); three-quarters of these cells had very low or undetectable background activity. The remaining 28\% of the neurons showed the interesting property that the background and evoked activities changed in different directions, so that the activity of either alone was grossly non-planar, but the overall activity was planar (Fig. 3g, h, i).

**The neural network model**

We used a three-layer network (illustrated in Fig. 4) that was trained to map visual targets to head-centred coordinates, given any arbitrary pair of eye and retinal positions. The first, or input, layer has two sections, an array of units on which the visual stimulus is represented, and a set of units representing eye position. The second layer consists of the hidden units, which map the input to the output. Each hidden unit receives input

---

**Fig. 2** The receptive fields of spatially tuned neurons from area 7a, arranged in rows with the eccentricity of the field maxima increasing to the right, and in columns with the complexity of the fields increasing downwards. Receptive fields were sampled at 17 radially spaced points, with one sample taken at the centre of the field, and four samples taken on each of four circles of radius 10, 20, 30 and 40 degrees. All the fields in row a have single peaks. Those in row b have a single large peak but some complexities in the field. The fields in row c are the most complex with multiple peaks. The data have been normalized so that the highest peak in each field is the same height.

**Fig. 3** The spatial gain fields of 9 neurons (a–i) from area 7a in the format of Fig. 1c.
Fig. 4. a, Back-propagation network used to model area 7a. The visual input consists of 64 units with gaussian receptive fields with 1/e widths of 15 degrees. The centre of each receptive field occupies a position in an 8 by 8 array with 10 degree spacings. The shading represents the level of activity for a single-spot stimulus, with darker shading representing higher rates of activity. The units have been arrayed topographically for illustrative purposes only; this pattern is not an aspect of the model as each hidden unit receives input from every one of the 64 retinal input units. The eye position input consists of 4 sets of 8 units each with two sets coding horizontal position (one for negative slope and one for positive slope) and two sets coding vertical position. Shading represents the level of activity. The interconnects have been ordered for illustrative purposes only and do not represent information available to the hidden layer. Each eye position cell projects to every unit in the hidden layer. Two output representations were used; the gaussian output format is shown on the right and the monotonic format on the left. The gaussian format units have gaussian shaded receptive fields plotted in head-centred coordinates. They have 1/e widths of 15 degrees and are centred on an 8 by 8 array in head coordinate space with 10 degree spacings. The monotonic format units have firing rates that are a linear function of position of the stimulus in head-centred coordinates. There are four sets of 8 units with two sets of opposite slope for vertical position and two sets for horizontal position in head-centred coordinates. Again, shading represents the degree of activity and the topographic ordering is for illustrative purposes only. The small boxes containing 'W' indicate the location of the synapses whose weights are trained by back propagation. Each hidden unit projects to every cell in the output layer. The output activity of the hidden and output units is calculated by the logistic function: output = 1/(1 + e^-net), where net = (weighted sum of inputs) + bias. The arrow for the connections represents the direction of activity propagation; error was propagated back in the opposite direction.

The back-propagation procedure guarantees that the synaptic weight changes will always move the network towards lower error by implementing a gradient descent in error in the multi-dimensional synaptic weight space.

b, Area 7a visual neuron receptive field with a single peak near the fovea. Receptive fields were plotted using the same method as in Fig. 3. Visual cells that had no eye-position-related activity or modulation of their responses by eye position were used to model the retinal input to the network. c, A composite of 30 area 7a-eye-position units, whose firing rates are plotted as a function of horizontal or vertical eye deviation. The slopes and intercepts are experimental values for eye-position neurons.

from every input unit and projects to every unit in the third output layer. The output of each unit of the hidden and output layers is computed as an S-shaped (logistic) function of the synaptic strength weighted sum of its inputs, plus a bias term. The training paradigm uses back propagation learning, which consists of choosing an input and desired output, applying the input to the first layer of the network and propagating the activity it generates through the network to the output units. The actual output is then subtracted from the desired output to generate an error. This error is used to adjust the weights of synapses on the output layer units and hidden layer units in a manner prescribed by the back-propagation procedure. Training begins with all weights randomized, resulting in large errors, and the training cycle is repeated until the error is reduced to desired levels.

The retinal position and eye position inputs to the network are modelled using characteristics of the cells in the posterior parietal cortex that respond to visual stimuli only and eye position only. The visual input consisted of 64 gaussian-shaped receptive fields, with 1/e widths of 15 degrees and with each peak separated by 10 degrees in an 8 by 8 array. The eye position input consisted of four sets of 8 units with single sets for positive and negative slopes for horizontal and vertical eye position.

We used two representations of location in head-centred coordinates at the output layer. One (output 2 in Fig. 4a) was a gaussian format in which each unit had a gaussian receptive field similar to the representation of the retinal input, but coding location in head-centred rather than retinal coordinates. The other (output 1 in Fig. 4a) was a monotonic format, in which the activity of each neuron is a linear function of the location of the stimulus in head-centred coordinates. For the gaussian format, a 64-unit array similar to the retinal input array was used and for the monotonic format, a 32-unit array similar to the eye position input array was used. The gaussian and monotonic formats were chosen because they represent the most common types of coding formats found for brain cells. Also the monotonic format has the interesting feature that it has the same representation as the eye-position code at the input. Thus, if the animal foveates the visual stimulus, the resulting eye-position signal could be used as the teacher to indicate the correct location of the stimulus in head-centred coordinates.

The model network was trained using randomly selected pairs of input eye positions and retinal positions. The teacher signal (desired output) used to train the output units was the true spatial location in head-centred coordinates implied by the inputs, and was represented in either the monotonic or gaussian format. The network trained quickly: after ~1,000 trials, accuracies equivalent to the distance between retinal unit centres were reached. When training was continued, error continued to decrease, but at a lower rate.

© 1988 Nature Publishing Group
Fig. 5  Hidden unit retinal receptive fields generated by the back propagation model. These plots were generated by holding the eye-position input to the network constant and simulating visual stimulation at the same 17 retinal positions used in the experiments on area 7a. The hidden unit activities were normalized and plotted in the same way as the experimental data shown in Fig. 3. The data shown here are from a series of 4 training sessions using networks with 25 hidden units and the monotonic format output. Similar results were obtained for the gaussian format output. All the fields, except for C-10, C-20 and C-30, are from networks that have received 1,000 learning trials. The remaining three are from untrained networks, resulting only from the random synaptic weights assigned at the start of a training run. Very complex fields are only rarely found in trained networks. No hidden unit with a single peak at 10 degrees appeared in this data set and such units are very rare in trained networks. No spatially tuned neurons with central receptive fields were found in area 7a, and no such fields appeared in the trained model. But central receptive fields are found among the visual neurons in 7a, and this kind of unit was among those used as input to the model network.

Agreement of model with experiment
To evaluate the model, we first compare the experimental and model retinal receptive fields. The model receptive fields were categorized according to their complexity, and the eccentricity of their activity maxima (Fig. 5), as had been done previously for the experimental receptive fields (Fig. 2). Comparison of the top lines in Figs 2 and 5 shows that the trained models generate single-peak receptive fields resembling those observed experimentally at all eccentricities except 10 degrees. The fully trained model also produces moderately complex fields like those found in line 2, but rarely produces receptive fields as complex as those in the bottom line of Fig. 2. This kind of highly complex field is not distinguishable from the untrained model receptive fields shown in the bottom of Fig. 5. The comparison process contains an element of subjectivity, but it demonstrates that the trained model generates retinal receptive fields remarkably similar to the experimentally observed fields. The gain fields generated by the model are shown in Fig. 6. All the total-response gain fields, whether generated by the monotonic or the gaussian output format, were planar in shape. This result compares with 80% of the experimental fields in this planar class. When the model gain fields are examined in more detail, taking into account the non-linearities of the visual response fields, there are significant differences between the eye position and retinal output formats. For example, when trained with the monotonic output format, 67% of the visual response gain fields were planar, but when trained with the gaussian output format only 13% fall in this class. These figures compare with 55% in this class for the experimental data. The irregular visual response gain fields generated by the gaussian output format are more radically irregular than those generated by the monotonic output format. Thus it appears that to account for the details of the visual response gain fields, it may be necessary to use both types of output representation. It should be pointed out, however, that whereas the visual receptive fields and the total gain fields

Fig. 6  Hidden unit spatial gain fields generated by the model network. Fields a-f were generated using the monotonic format output; the rest used the gaussian format output.
were virtually unaffected by changing parameters of the model, the visual response gain fields were very sensitive to parameters such as threshold value and output representation. The number of hidden units had little effect, giving similar results for simulations ranging from 9 to 36 hidden units.

The striking similarity between model and experimental data certainly supports the conjecture that the cortex and the network generated by back propagation compute in similar ways. As back propagation generates optimal solutions which produce the least error, these results also suggest that the brain chooses optimal solutions with respect to error. The similarity of the model and test results raises the question of what physiological mechanisms could subserve this equivalence. Presently the backpropagation paradigm is structured at a level higher than implementation, and obviously cannot be applied literally to the brain, because information does not travel backwards rapidly through axons. That the back-propagation method appears to discover the same algorithm that is used by the brain in no way implies that back propagation is actually used by the brain. One approach to understanding the physiological significance of these results is to generate models incorporating features found in the brain, such as Hebbian-like learning at synapses and reciprocal back-projection pathways for propagating error, and determine whether these models generate similar results. In this regard it is interesting to note that all cortico-cortical and thalamocortical connections have reciprocal feedback pathways.

An important consideration in interpreting the results is how closely does the model response actually resemble the cortical data? This is a complex issue because there is error in the experimental data, and additional errors introduced by the interpolation process used to produce the full-field views of the receptive fields. Examination of the magnitudes of these errors indicates that they could not account for the various receptive field types, or eye-position gain fields observed experimentally and in the model. The methods of comparison between model and experiment that we have used are too to some degree subjective. Perhaps a more objective comparison procedure will eventually be developed, but it is unlikely that, given the complex nature of the data to be compared, any such technique will substantially alter our conclusions concerning the degree of similarity between model and experimental data.

From the physiological perspective these experiments raise the possibility that the posterior parietal cortex learns to associate body position with visual position to localize accurately the position of objects with respect to the body. This idea about the importance of learning is reasonable considering that it would not be practical for spatial representations to be hard-wired because the body dimensions change during development. Furthermore, adaptation experiments show that distortion of space with prisms leads to rapid recalibration, suggesting that these representations are still plastic in adults. As the model, by definition, does not have a topographic organization to localize in space, it shows that the brain does not need a topographic organization to localize in space. The organization of the network is not a product of the spatial position of the cell bodies, but rather is contained in the pattern of the weights of the synaptic connections.

Finally, there is the question of where the output units of the model could exist in the brain. One possible location would be areas that receive projection from area 7a. Although eye position effects on visual responses have been described at several locations in the brain (refs 24–26; S. Funahashi, C. J. Bruce, P. S. Goldman-Rakic, and R. Lal, M. J. Friedlander, personal communication), an eye-position-independent coding has yet to be unequivocally demonstrated. But it is also possible that the final spatial output could only exist in the behaviour of the animal. For example, the muscles innervating the eye or limb are broadly tuned, and the position of the eye or limb is coded in a distributed fashion over the activity of several muscles. Thus the final spatial output may not exist in any single cell in the brain, but rather might be found only in the pointing of the eye or finger accurately to a location in space.

We thank David Rumelhart, Francis Crick and Emilio Bizzi for stimulating discussion during the development of this model and Carol Andersen for editorial assistance. D.Z. was supported by grants from the System Development Foundation, the AFOSR and the Office of Naval Research. R.A.A. was supported by the NIH, the Sloan Foundation and Whitaker Health Sciences Foundation.

Received 19 October 1987; accepted 19 January 1988.