Modelling the acquisition of goal-directed behaviors by populations of neurons

Emmanuel Guigon [*], Yves Burnod

INSERM CREARE, Neurosciences & Modélisation, Institut des Neurosciences, Université Pierre et Marie Curie, 9, quai Saint-Bernard, 75005 Paris, France

Abstract

Recent neurophysiological studies have revealed the patterns of neuronal activity during the acquisition of goal-directed behaviors, both in single cells, and in large populations of neurons. We propose a model which helps three sets of experimental results in the monkey to be understood: (1) activity of single cells vary greatly and only population activities are causally related to behavior. The model shows how a population of stochastic neurons, whose behaviors vary widely, can learn a skilled conditioned movement with only local activity-dependent synaptic changes. (2) typical changes in neuronal activity occur when the rules governing the behavior are changed, i.e. when the relationship between cues and actions to reach a goal changes over time. There are two types of neuronal patterns during changes in reward contingency: a monotonic increasing pattern and a non-monotonic pattern which follows the change in the way the reward is obtained. Units in the model display these two types of change, which correspond to synaptic modifications related to the encoding of the behavioral significance of sensory and motor events. (3) These two patterns of neuronal activity define two populations whose anatomical distributions in the frontal lobe overlap with a gradient organized in the rostro-caudal direction. The model consists of two artificial neural networks, defined by the same set of equations, but which differ in the values of two parameters (P and Q). P defines the adaptive properties of processing units and Q describes the coding of information. The model suggests that a balance in the relative strengths of these parameters distributed along a rostro-caudal gradient can explain the distribution of neuronal types in the frontal lobe of the monkey.

Keywords: Neuronal populations; Goal-directed behavior; Learning; Connectionist modelling; Computer simulation

1. Introduction

A major adaptive property of brain circuits is their great capacity to learn new behaviors. This capacity is generally distributed over large popu-

lations of neurons, but it remains unclear how these populations are programmed to produce the motor signals needed to execute a given behavior. Several recent studies have begun to characterize the changes in cerebral activity with on-line acquisition or modification of goal-directed behaviors, both at the level of neurons with single unit recordings in monkeys (Burnod et al., 1982; Thorpe et al., 1983; Niki et al., 1990; Mitz et al.,

* Corresponding author. Tel.: (+33-1) 44 27 37 47; Fax: (+33-1) 44 07 15 85.E-mail: guigon@ccr.jussieu.fr.

1991), and at the level of neuronal populations, with cortical fields in monkeys (Sasaki and Gemba, 1982) and brain potentials in humans (Papakostopoulos, 1978; Chiarenza et al., 1983).

Changes in neuronal activity during the acquisition of goal-directed behaviors should reflect both behavioral changes and plasticity phenomena at the synapses, neurons and circuits. This intuitive formulation of the problem of the relationship between plasticity and behavior leads to the question of the rules used by the brain to build internal representation of goal-directed behaviors. This question is approached in this paper by neural network modelling (Rumelhart and McClelland, 1986; Hertz et al., 1991). Neural network models help to formulate hypotheses on the adaptive properties at the neuronal level by relating local processing and global performance. In these models, the basic processing unit is a neuron-like element which receives a set of inputs from other units and adjusts its output in response to inputs. Units are grouped into subsets (such as layers, maps) and connected according to specific rules (random connections, full or partial connections between subsets, etc.) to form a network. A network usually contains input and output units, which define a processing pathway and provide a way to place constraints on the network. The ability of such a network is given by the set of modifiable connection weights between units, which shape unit activation. During a training session where inputs and/or outputs of the network are constrained to particular values, connection weights are modified according to a learning rule. Neural network models provide two types of results. First, it is possible to analyze the global behavior of the network (performance, generalization, resistance to input breakdown and synaptic weighting). The results can be compared with the outcomes of psychophysical and learning experiments. The second concern the properties of neurons in the network (e.g., their selectivity) that give information on the internal representations used and on the building of these representations by learning (Zipser and Andersen, 1988; Burnod et al., 1992; Guigon et al., 1995).

We propose a neural network model, which defines a relationship between the activity of a single neuron and the behavior of a population of neurons during the acquisition of two goal-directed tasks. The model is consistent with available experimental results on neuronal activities in monkeys, and helps to explain three observations which have been frequently made in study of goal-directed behaviors.

The first observation is that the command of movement is not produced by a single neuron that control the whole behavior, but is distributed over a large population of broadly tuned neurons (Georgopoulos et al., 1986). We propose a network (A) which learns the distributed control of the direction of planar arm movements. After learning, the network produces movements in the direction defined by a goal, and population activity in the network can be viewed as an internal representation of the intended movement.

The second observation is that typical changes in neuronal activity occur when the rules governing the behavior are changed, i.e. when the relationship between cues and actions to reach a goal changes over time. There are two populations of neurons in the frontal lobe during theses changes (Niki et al., 1990). The first population appears to encode the impending behavioral response, while the second is more responsible for the association between guiding sensory cues and behavioral responses (Niki et al., 1990; di Pellegrino and Wise, 1991). We propose a network (B) that can adapt to changes in reward contingencies during new learning and reversal learning experiments and displays patterns of neuronal activity similar to those observed experimentally.

The third observation is that there is not a clear delineation between these populations. Instead, they seem to overlap with a gradient organized in the rostro-caudal direction. Networks A and B show how the two properties can be derived from the same basic mechanism by changing the value of two parameters. The proposal of the model is that these parameters vary along a rostro-caudal gradient in the frontal lobe, allowing the cortical networks to learn both to represent the intended movements (network A, more caudal) and to change their relations with cues in order to adapt to changes in reward contingencies (network B, more rostral).

2. Methods

The model is formally presented. A model neuron is first defined by its excitability and plasticity properties (Eqs. 1 and 2). Two network models (A and B) are then described by (i) a subset of interconnected model neurons, (ii) a set of input and output pathways carrying information from and to the external world, (iii) a task which represents the desired function of the network, (iv) and a training protocol which indicates the interactions with the external world which are necessary to learn the task.

Neural network modelling

Neuronal processing can be approximated by nonlinear interactions between inputs which signal for different types of information. This approximation is derived from the integrative properties of dendritic trees (Shepherd and Koch, 1989; Mel, 1993) and from experimental results on the combination of information in different cortical areas (Andersen et al., 1985; Caminiti et al., 1991; Burnod et al., 1992; Trotter et al., 1992). Another characteristic of neuronal processing is the stochastic behavior of neurons, which reflects the equivalence between the firing rate of a neuron and its probability of firing (Sejnowski, 1986).

A model neuron has n binary input pathways x_i and an output pathway y. Its output is given by

$$y(t) = F[a(t)]Q(t) \qquad (1)$$

where

$$a(t) = \sum_i \alpha_i(t) x_i(t) + \sum_{i>j} \beta_{ij}(t) x_i(t) x_j(t)$$

and where Q(t) corresponds to the influence of the population of neurons on the neuron. The α_i and β_{ij} are modifiable "learning" coefficients in the interval [0,1]. The variable t is a quantized time, which describes the course of processing. F is the stochastic output function defined by

$$F(u) = 1 \text{ with prob } g(u) \text{ and } 0 \text{ with prob } 1\text{-}g(u)$$

where

$$g(u) = 1/[1 + exp(-\gamma(u - \phi))]$$

where γ and ϕ are parameters. Eq. 1 states that the probability of a neuron discharging is an increasing function of the quantity of presynaptic activity.

Neurons should have a way to measure the consequence of their activations in relation to the expected goal in order to learn to reach that goal. We propose a local rule for modifying of learning coefficients that has this property, independently of the nature of the goal (reward, sensory effect, etc). The rule describes the change in the learning coefficients α_i and β_{ij} following ongoing variations in the activity of input and output pathways between time t and t + 1

$$\Delta\alpha_i(t,t+1) = y(t)\{\lambda x_i(t+1) - \mu[x_i(t+1)]\}$$

$$\Delta\beta_{ij}(t,t+1) = x_j(t)y(t)\{\lambda x_i(t+1)$$

$$-\mu[1 - x_i(t+1)]\} \qquad (2)$$

where λ and μ are parameters. Eq. 2 states that a postsynaptic activity y(t) followed by a presynaptic activity $x_i(t+1)$ is a critical pattern in order to modify learning coefficients. The presynaptic activity reinforces previous activity by eliciting an increase in the coefficient by λ. On the other hand, the coefficient decreases by μ when no presynaptic activity follows the postsynaptic activity. The relative values of λ and μ (described by the quantity $P = \mu/\lambda$) define the influences of reinforced and non-reinforced activities.

Eq. 2 can be described intuitively in the following way. If the output activity of the unit is always followed, through an external feedback loop, by the same reafferent input (pathway i), the input-output relation in the unit can represent a one-to-one sensory-motor correspondence which always predicts the sensory effect of the output activity, independently of any other conditions. In this case, the learning coefficient α_i increases, and, after learning, the input x_i can trigger a strong output (full anticipation of the effect). When the reafferent activity occurs only sometimes, it means that another condition is required, which may be represented by another input (pathway j). If the conjunction of the same two inputs has an higher probability of being related to a strong output, the learning coeffi-

cient β_{ij} increases and the activity in one pathway acts as a gating signal on the other input. This rule has been used to learn the coordinate transformation for visually-guided arm reaching movements (Burnod et al., 1992). Since the computation of the appropriate motor command requires information on both the target position and the arm position, conditional relationships between the two types of information were learned using reafferent visual information corresponding to the sensory effect of a motor command.

Two networks (A and B) are described in the following. Eqs. 1 and 2 define the properties of the processing units in both models, but for two different sets of value of the parameters P and Q.

Network A

This model illustrates the behavior of a population of neurons in which all the neurons contribute to a continuously graded output (direction of movement). The starting point is the experimental results which show that neurons in several cortical areas have directional tuning properties

and that this directional tuning is distributed in a regular way within the population and within the space of the movement (Georgopoulos et al., 1982, 1986; Caminiti et al., 1991). The population consists of n ($n = 16$) neurons (N_i, $1 \leq i \leq n$), each tuned for a preferred direction of movement $\mathbf{p}_i$ (bold characters are used for vectors) in the 2D horizontal plane, in such a way that the vectors $\mathbf{p}_i$ are uniformly distributed in this plane. This means that, at each time t, the neuron N_i contributes to the direction of the movement in proportion to its activity $y_i(t)$ along its preferred direction $\mathbf{p}_i$. The resulting movement is characterized by the direction $\mathbf{d}(t)$ according to

$$\mathbf{d}(t) = (1/n)\Sigma_i y_i(t)\mathbf{p}_i \tag{3}$$

Eq. 3 is a population code, as proposed by Georgopoulos et al. (1986). The values of the parameters are $\lambda = 0.5$, $\mu = 0.15$ (P = 0.3) and

$$Q(t) = cos(\mathbf{p}_i, \mathbf{d}(t))$$

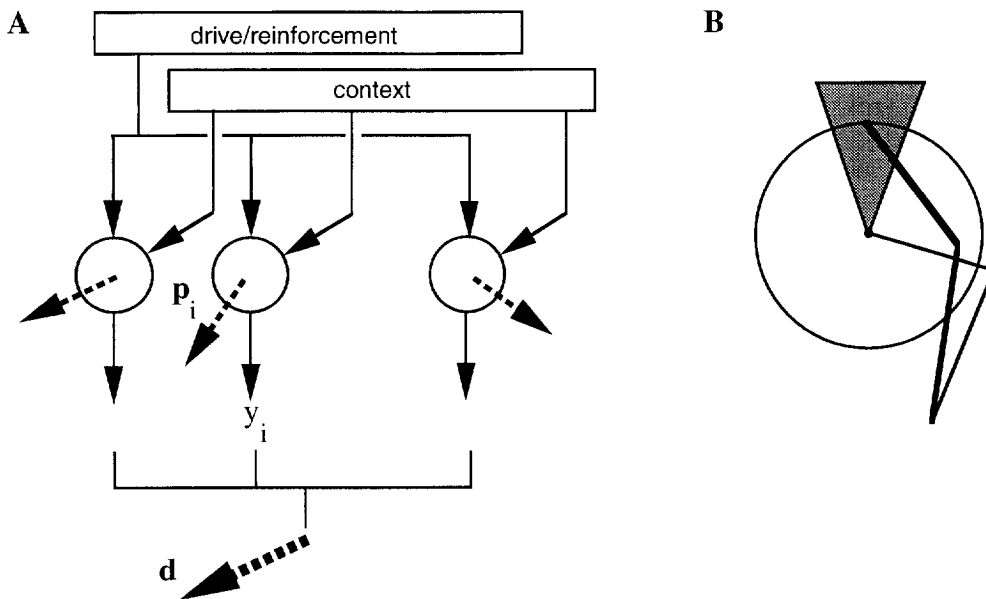which describes lateral interactions between neurons.

Fig. 1. Diagram of network A and the task used with this network. A: network architecture. Each neuron has two input pathways (context and drive/reinforcement) and one output pathway. Thin dashed arrow at the center of a neuron represents the preferred direction of the neuron (see text). Thick dashed arrow is the population vector. B: the goal of the task is to move the arm (represented by two connected lines) between an initial central position (thin line) to a peripheral position (thick line) in the shaded angular sector.

The direction of movement at each time is defined by the activity in two afferent pathways which indicate the actual sensory cues (context) and the goal to reach (drive/reinforcement). The network architecture is shown in Fig. 1A. Each neuron receives (1) a context signal. This signal is different at each time and for each unit. It corresponds to the ability of the network to generate all possible behaviors (all directions of movement); (2) a drive/reinforcement signal. This signal is the same for all neurons: it indicates both an internal motivational state related to the expectation of a reward and the effective reward.

A simple task, a movement of the arm in a specific direction in order to get a reward, is used to train the network. The population of neurons controls the direction of arm movements in the horizontal plane (Fig. 1B). The desired behavior of the population is to move the arm from an initial central position to the periphery in such a way that the final hand position is in a given angular sector ([60°, 115°] in Fig. 1B). Starting from a random behavior, the network is trained with repeated pairings of the drive/reinforcement input and some randomly selected context inputs. Whenever a combination elicits a movement in the correct direction, a reinforcement is given as a reafferent input in the drive/reinforcement pathway. The network adaptation results from activity-dependent modifications of learning coefficients according to Eq. 2.

Network B

The efficacy of a goal-directed action may depend upon previous processing, for example to take into account a sensory cue. Furthermore, the relationships between actions and goals are not fixed, but change with changes in the external world. These alterations are explored using experimental paradigms such as new learning and reversal learning, where the behavioral significance of an action is modified in the course of a conditioning program. Network B illustrates the behavior of a population of neurons in which each neuron (or subpopulation of neurons) can define the association between a cue and an action, where cues and actions are related by

arbitrary and changing rules. Considering the same network architecture as for network A, we assume that the n neurons (N_i, $1 \leqslant i \leqslant n$) contribute by their output to the independent concurrent actions (A_i, $1 \leqslant i \leqslant n$). Thus only one neuron N_i is active at a given time, and the context input to this neuron acts as a cue for action A_i.

For the sake of simplicity, we conserved only a subset of neurons that will be involved in a specific change in reward contingency. We consider a network of $n = 3$ neurons, N_1, N_2, N_3, corresponding to actions A_1, A_2, A_3. The network is first trained to execute the task described in Fig. 2A. The cue C_1 elicits the rewarded action A_1. The cue C_2 elicits the non-rewarded action A_2. The cue C_3 is never delivered (Fig. 2A). From this state, we produce a change in the conditioning program. The desired behavior of the network is to execute the new task described in Fig. 2B, which is derived from the task of Fig. 2A by a change in reward contingency. Learning the new task requires both reversal learning (the cue C_1 is now associated with a non-rewarded action) and new learning (the cue C_3 now elicits the rewarded action A_3).

In this case, the values of the parameters are $\lambda = 0.05$, $\mu = 0.04$ ($P = 0.8$), and $Q(t) = 1$, since the outputs of the network are independent. The network is trained with repeated pairing of the drive/reinforcement input and context inputs according to the current rule (Figs. 2A and 2B).

Computer simulations

The computer simulations of the neural network models A and B were run. All learning coefficients were initially at 0.5. The training consisted of 30 blocks of 10 trials, where a trial corresponds to (1) the activation of the network, (2) the measure of network output (direction of the resulting movement in network A and action in network B), and (3) the delivery of a reward for a correct output.

The results of each trial were noted as correct or not, and the activity and the variation in the learning coefficients in each neuron were recorded. The performance rate was computed using the percentage of correct responses in suc-
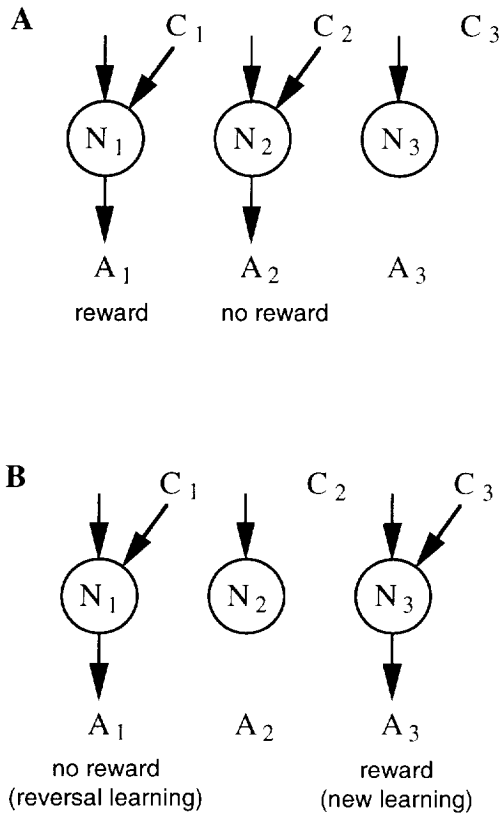
Fig. 2. Schematic description of the two tasks used to illustrate acquisition of a change in reward contingencies. The neurons are depicted as in Fig. 1A, with a context and a drive/reinforcement pathway. A: the network first learns the association between three cues (C_1, C_2, C_3), three actions (A_1, A_2, A_3) and a reward, according to a specific rule. B; a new rule is defined which involves both new learning and reversal learning. The goal of the network is to learn the new rule starting from the initial behavior described above.

cessive blocks of 10 trials (trials with no movement were not taken into account). Individual trial activity and variation in the learning coefficients were average by block of 10 trials for each neuron. Histograms were constructed which indicate the change in the performance of the network, the changes in the activity and learning coefficients in a unit (Figs. 3 and 4).

3. Results

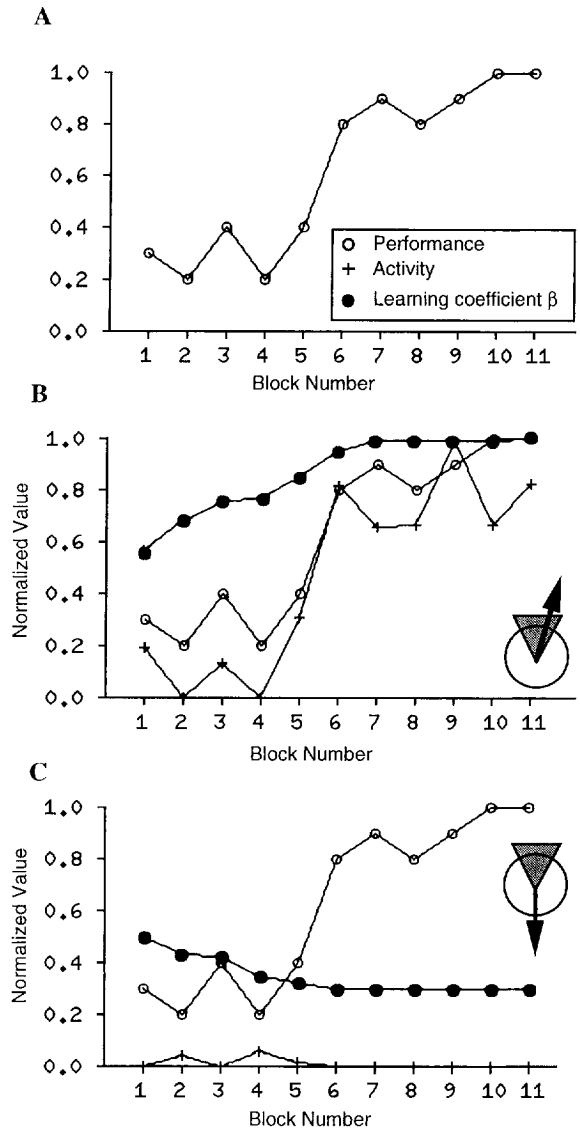This section provides a quantitative evaluation of networks A and B. The results concern the

Fig. 3. Quantitative description of the behavior of network A. A: changes in performance level (○). Each horizontal division represents a block of 10 successive trials (trials are not taken into account if no movement occurs). The corresponding ordinate value is the proportion of correct trials in the block. B: normalized neuronal activity (crosses) and normalized learning coefficient β (●) for a neuron, whose preferred direction is indicated by the thick arrow at the right of the graph. C: same as (B) for a neuron with a different preferred direction. The parameters are: $\lambda = 0.5$, $\mu = 0.15$, $\gamma = 8.0$, $\phi = 0.5$, $w_0 = 0.5$.

global behavior of the network (as measured by the number of correct trials) and the behavior of units in the network (measured by neuron activation).

Network A

The performance of the network obtained by computer simulations is shown in Fig. 3A. The monotonic increase in the performance rate (open circles) reflected the increase in the number of movements in the intended angular sector (see Fig. 1B). Two observations were made.

First, changes in neuronal activity were different for different units, as shown for two units in Figs. 3B and 3C. A unit whose preferred direction was in the intended angular sector (Fig. 3B) displayed a learning-dependent increase in activity (crosses) correlated with the increase in the learning coefficient β (solid circles) and with the change in performance (open circles). There was no change in the activity of a unit whose preferred direction was not directed toward the correct sector (Fig. 3C). The model thus showed that the excitability of neurons which have been involved in the production of the intended response was modified in such a way as to increase their involvement in future responses. In the same way, the activities of neurons involved in the production of incorrect responses were depressed. As a result, the correct response was more surely and more frequently evoked.

Second, increasing activity in network was related to the variation in learning coefficients and occurred in parallel with the acquisition of the task. However, although the learning coefficients varied smoothly, activity and performance curves had an S-shape, with a strong nonlinear increase in the middle of the training period (Fig. 3B). The development of performance also seemed to precede the development of activity (Fig. 3B).

Network B

The results are shown in Fig. 4 and concern the variations in the performance rate, in the neuronal activities and in the learning coefficients during the acquisition of the tasks defined in Fig. 2. The vertical dashed line on each graph corresponds to the time of change in reward
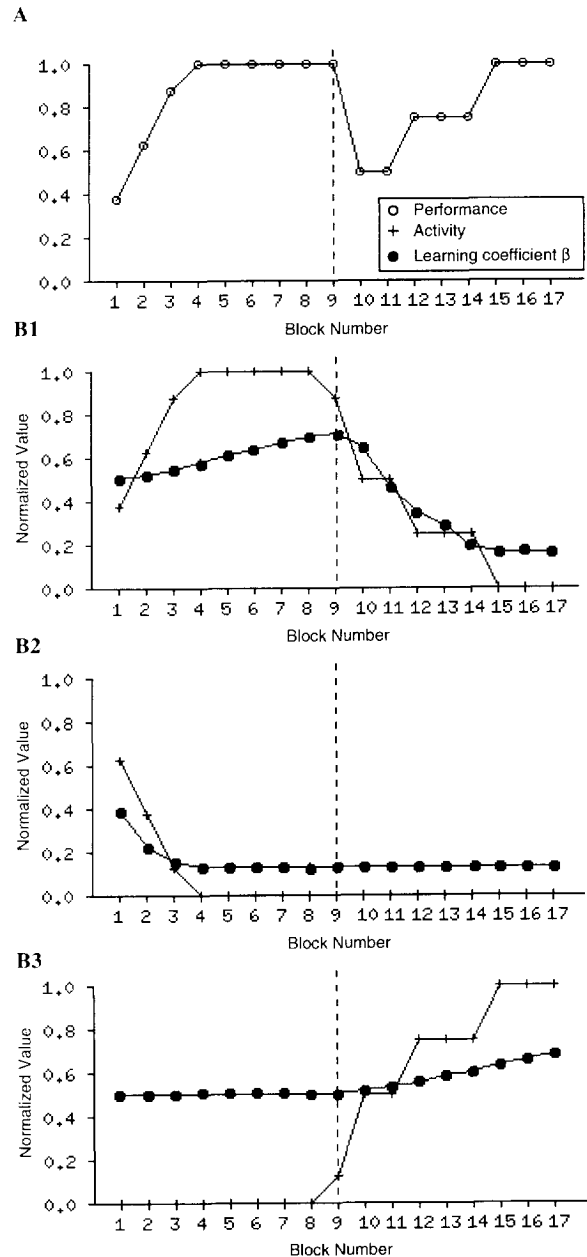

Fig. 4. Quantitative description of the behavior of network B. Same conventions as in Fig. 3. A: changes in performance level. B: normalized neuronal activity and normalized learning coefficient for the neuron N_1 of Fig. 2. C: same as (B) for the neuron N_2. C: same as (B) for the neuron N_3. The parameters are: $\lambda = 0.05$, $\mu = 0.04$, $\lambda = 8.0$, $\phi = 0.35$, $w_0 = 0.5$.

contingencies. The performance rate is shown in Fig. 4A. It increased during the acquisition of the first task, started to decrease when the rule of the task changes and then increased once again. Fig. 4 B1, B2, and B3 correspond to neurons N_1, N_2, N_3, respectively (Fig. 2). We observed that each neuron had a specific behavior. After a steady increase during the first task, repetitive non-rewarded executions of action A_1 lead to a decreasing activity in neuron N_1 (Fig. 4B1). In this way, the non-rewarded behavior was extinguished, while a new rewarded behavior (action A_3) related to cue C_3 gradually developed (Fig. 4B3). Action A_2 was gradually lost since it was not rewarded (Fig. 4B2). Note the complementary contributions of neurons N_1 (Fig. 4B1) and N_3 (Fig. 4B3) to change in the behavior of the network (Fig. 4A).

4. Discussion

Learning a distributed representation of an intended movement

The main feature of network A is the distributed coding of a single parameter (direction of movement) in a population of neurons. Learning a specific value of this parameter requires the simultaneous tuning of many neurons in the population. The learning rule (defined by Eq. 2) allows the neurons to improve their contribution to the correct intended response, either in a positive way when their preferred direction is close to the intended direction, or in a negative way. We have introduced the quantity P, which defines the balance between depression and potentiation of learning coefficients. This quantity should be small in this model, since neurons whose preferred direction is close to the intended direction can be nonetheless active (and thus become depressed) for incorrect movements.

The model suggests that changes in activity are likely to be described by changes in conditional probability encoded in learning coefficients β of the network. This interpretation provides a simple way to relate distributed information processing in a neuronal network and the execution of a global action, and is consistent with experimental

data. Burnod et al. (1982) have described the short-term changes in neuronal activity in motor and parietal areas during the operant acquisition of self-initiated flexion movements. Monkeys were trained to perform a flexion movement from a starting position to a final predetermined angular sector without visual control. They observed that the relationships between the neuronal activity of movement-related cells and the movement changed with the performance rate: as the performance improved, the neuronal activity became more closely time-locked with the onset of the movement. This modification reflects an increase in the conditional probability of discharge of neurons in relation to the beginning of the movement.

Another important aspect of the model is that the behaviors of the network at early and late phases of training are quite similar, in the sense that the network is able to sometimes produce the correct response at the beginning of the training. The main difference between early and late training period is thus not really a matter of ability to succeed in a given trial, but rather of ability to reach a good performance criterion. This interpretation is consistent with paradoxical observations that show that the development of performance seems to precede the development of activity. A similar observation has been made in experimental studies, when recordings are made during training periods (Sasaki and Gemba, 1982; Watanabe, 1990; Mitz et al., 1991). Sasaki and Gemba (1982) recorded cortical field potentials in monkeys during the learning of a skilled conditioned movement. They observed that activity related to the production of the response (early and late precentral potentials) appeared only when the monkey starts to perform correctly (Sasaki and Gemba, 1982). Some activity became differentiated earlier than the performance, but it was true only for sensory processes.

Learning changes in reward contingencies

Network A illustrates the learning of skilled arm movement. The results of network B show the effects of changing the way a reward is obtained, assuming that some behaviors (for in-

stance, skilled arm movements) are already known by the network. This change involves both new learning (since a cue which was never seen before was introduced) and reversal learning (since a cue previously associated with a rewarded action was associated with a non-rewarded action in the new task). The main characteristics of neuronal activity in the network are (1) a monotonic increase in activity associated with new learning (Fig. 4B3), and (2) a non-monotonic change in activity when the current behavior is suddenly not rewarded (Fig. 4B1).

Niki et al. (1990) described the changes in neuronal activity in the frontal cortex during new learning (discriminations of new visual stimuli) and reversal learning. Monkeys were first trained on a GO/NO-GO discrimination task with several pairs of stimuli. Long-lasting alterations of activity were found when new stimuli were used: activity increased for one stimulus, but did not change for the other stimulus. Niki et al. (1990) observed that such learning-dependent changes occurred in neurons whose activity was related to the forthcoming movement.

Niki et al. (1990) observed two characteristic patterns of changes during reversal learning: (1) Neurons which showed opposite discharge patterns for correct and incorrect trials (Type 1) keep the same level of activity when the behavioral responses are reversed; (2) Neurons which showed the same discharge patterns for correct and incorrect trials (Type 2) reversed their activity with the reversal. Type 1 reflects a dependence on the impending response whereas Type 2 is related to stimulus-response association. Similar results were obtained by Watanabe (1990). Indeed, Watanabe (1990) observed that changing the "associative" significance of a visual cue (i.e., does it predict a reward?) elicits learning-dependent decrease in neuronal activity in the units examined in prefrontal and premotor cortices.

These non-monotonic changes may contribute to both the development of new behaviors and the transition between two behaviors following changes in reward contingencies. Decreasing activity related to unrewarded behavioral actions is responsible for the suppression of these actions. Subsequent increases in neuronal activity reflect

the construction of a new rewarded behavior (eg., the integration of new environmental cues).

The quantity P must be high in network B. Neurons which have been involved in a non-rewarded action must be strongly depressed in order to rapidly suppress this action and to select new rewarded actions.

Adaptive properties and information coding

The adaptive properties in networks A and B depend on two parameters: P which defines the learning rate of neurons (λ and μ in Eq. 2), and Q which defines the coding of information in the neural network.

Parameters λ and μ are related to the contribution of rewarded and non-reward behaviors. The quantity $P = \mu/\lambda$ is an index of the sensitivity to external constraints defining the correctness of the behavior: if P is low, the learning coefficients show a large positive change for rewarded trials and a small negative change for non-rewarded trials (Eq. 2); if P is high, the learning coefficients change by almost the same positive or negative amount for rewarded and non-rewarded trials.

The quantity Q defines the coding of information in the network. Some brain structures are organized in a map, which reflects changes in the value of a parameter coding for a sensory or a motor feature across one or more dimensions of the biological substratum (Knudsen et al., 1987). On the other hand, some structures, like the prefrontal cortex, are not organized into maps along specific sensory or motor features. In this case, neurons appear to become specialized by learning processes (Fuster, 1988). This characteristic can be defined by the quantity Q, which reflects the degree of organization of a structure. We will say that Q is high in the former case and low in the latter.

The networks A and B illustrate two configurations of quantities P and Q. Network A corresponds to low P and high Q. In this case, neurons are broadly tuned around a preferred value and learning allows selection of a range of values which are effective for producing a given behavior. There is no absolute relationship between single neuron activity and behavior, which is due

to population coding. A small influence of non-rewarded trials is thus required to ensure smooth changes in learning coefficients and the selection of continuous values (continuous range of movement direction). It corresponds to a low value of P. The example in Fig. 1 and the corresponding results in Fig. 3 illustrate this mechanism (P = 0.3). This configuration may occur in parts of motor and premotor areas. Indeed, neuronal activities in these areas are mainly related to the impending behavioral response (movement) and do not change with change in reward contingency (Niki et al., 1990). For these structures, the high value of Q corresponds to the coding of direction of movement (Georgopoulos et al., 1982).

Network B corresponds to high P and low Q. Learning may correspond to the selection of a subset of neurons which participates in the intended behavior through their specific connections with sensory and motor structures. This selection would be efficient if the index P is high enough, allowing the extinction of non-rewarded behaviors and the production of new behaviors. In the example in Fig. 4, the index P is 0.8. Current views of the prefrontal cortex agree with this configuration (see Fuster, 1988). The influence of reinforcement contingencies is a fundamental aspect of prefrontal functions. It is characterized by the inability of prefrontal animals to adjust their behavior in response to changes in reward contingencies (for instance, during reversal discrimination; Fuster, 1988). Milner and Petrides (1984) have shown with the Wisconsin Card Sorting Test that humans with prefrontal lesions failed to change their behavior when the rule was changed. The anatomical organization of the prefrontal cortex is consistent with a low value of Q. The prefrontal cortex is generally described as a region of cross-modal integration, receiving auditory, somatic and visual information (Fuster, 1988). Although there are local anatomical specializations, prefrontal neurons appear to be active only in the execution of purposive behaviors.

The model thus suggests that two properties (defined by the quantities P and Q) allow motor/premotor and prefrontal circuits to make differential functional contributions to learning

processes. This regionalization is predicted by the patterns of connectivity of these cortical structures (Barbas and Pandya, 1987, 1989), by the differential effects of lesions (Petrides, 1986) and by the gradient distributions of neuronal properties (Niki et al., 1990; di Pellegrino and Wise, 1991). The main purpose of the model is to explain these observations by intrinsic properties (defined by P and Q) of populations of neurons in these areas.

Learning rule

The adaptive properties in these models are defined by Eq. 2. This learning rule is not a Hebbian rule as generally used in neural modelling (Brown et al., 1990). Hebbian modifications are related the conjunction of pre- and postsynaptic activities. In the present rule, the preferred pattern of activation is defined by a postsynaptic activity preceding a presynaptic activity. The rule can be viewed as an error-correcting rule, which allows the synaptic weights to be adjusted in proportion to the difference between a desired and computed value (Widrow and Hoff, 1960). However, Eqs. 1 and 2 do not give a specific role to the error signal. Any signal following the activation of a neuron can be used as an error. This is a fundamental difference from other error correction rules, which rely on the unconditional influence of a desired value as a reinforcing factor. This observation stresses that conditioning can occur on any reafferent signal matching the intended effect of an action. It has been used to learn the visuomotor transformation for positioning the hand on a target (Burnod et al., 1992): the desired hand trajectory was operantly learnt as the sensory effect (visual trajectory) of a motor command. This property is interesting with regard to the ability of humans to learn using the expected effect of their action rather than an absolute reward.

References

Andersen, R.A., Essick, G.K. and Siegel, R.M. (1985) Encoding of spatial location by posterior parietal neurons. Science, 230: 456–458.

Barbas, H. and Pandya, D.N. (1987) Architecture and frontal cortical connections of the premotor cortex (area 6) in the rhesus monkey. J. Comp. Neurol., 256: 211–228.

Barbas, H. and Pandya, D.N. (1989) Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. J. Comp. Neurol., 286: 353–375.

Brown, T.H., Kairiss, E.W. and Keenan, C.L. (1990) Hebbian synapses: biophysical mechanisms and algorithms. Annu. Rev. Neurosci., 13: 475–511.

Burnod Y., Maton, B. and Calvet, J. (1982) Short-term changes in cell activity of area 4 and 5 during operant conditioning. Exp. Neurol., 78: 227–240.

Burnod, Y., Grandguillaume, P., Otto, I., Ferraina, S., Johnson, P.B. and Caminiti, R. (1992) Visuo-motor transformations underlying arm movements toward visual targets: a neural network model of cerebral cortical operations. J. Neurosci., 12: 1435–1453.

Caminiti, R., Johnson, P.B., Galli, C., Ferraina, S. and Burnod, Y. (1991) Making arm movements within different parts of space: the premotor and motor cortical representation of a coordinate system for reaching to visual targets. J. Neurosci., 11: 1182–1197.

Chiarenza, G.A., Papakostopoulos, D., Giordana, F. and Guareschi-Cazzullo, A. (1983) Movement-related brain macropotentials during skilled performances. A developmental study. Electroenceph. Clin. Neurophysiol., 56: 373–383.

Di Pellegrino, G. and Wise, S.P. (1991) A neurophysiological comparison of three distinct regions of the primate frontal lobe. Brain, 114: 951–978.

Fuster, J.M. (1988) The Prefrontal Cortex. Anatomy, Physiology, and Neuropsychology of the Frontal Lobe, 2nd edn., Raven Press, New York.

Georgopoulos, A.P., Kalaska, J.F., Caminiti, R. and Massey, J.T. (1982) On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. J. Neurosci., 2: 1527–1537.

Georgopoulos, A.P., Schwartz, A.B. and Kettner, R.E. (1986) Neuronal population coding of movement direction. Science, 233: 1416–1419.

Guigon, E., Dorizzi, B., Burnod, Y. and Schultz, W. (1995) Neural correlates of learning in the prefrontal cortex of the monkey: a predictive model. Cerebral Cortex, in press.

Hertz, J.A., Palmer, R.G. and Krogh, A.S. (1991) Introduction to the Theory of Neural Computation, Addison-Wesley, Reading.

Knudsen, E.I., du Lac, S. and Esterly, S.D. (1987) Computational maps in the brain. Annu. Rev. Neurosci., 10: 41–65.

Mel, B.W. (1993) Synaptic integration in an excitable dendritic tree. J. Neurophysiol., 70: 1086–1101.

Milner, B. and Petrides, M. (1984) Behavioural effects of frontal lobe lesions in man. Trends Neurosci., 7: 403–407.

Mitz, A.R., Godschalk, M. and Wise, S.P. (1991) Learning-dependent neuronal activity in the premotor cortex: activity during the acquisition of conditional motor associations. J. Neurosci., 11: 1855–1872.

Niki, H., Sugita, S. and Watanabe, M. (1990) Modification of the activity of primate frontal neurons during learning of a GO/NO-GO discrimination and its reversal. A progress report. In: E. Iwai and M. Mishkin (Eds.), Vision, Memory, and the Temporal Lobe, Elsevier, New York, pp. 305–311.

Papakostopoulos, D. (1978) The serial order of self-paced movement in term of brain macropotentials in man. J. Physiol., 280: 70–71.

Petrides, M. (1986) The effect of periarcuate lesions in the monkey on the performance of symmetrically and asymmetrically reinforced visual and auditory go, no-go tasks. J. Neurosci., 6: 2054–2063.

Rumelhart, D.E. and McClelland, J.L. (1986) PDP models and general issues in cognitive science. In: J.A. Feldman, P.J. Hayes and D.E. Rumelhart (Eds.), Parallel Distributed Processing, Explorations in the Microstructure of Cognition, Vol. 1: Foundations, MIT Press, Cambridge, pp. 110–146.

Sasaki, K. and Gemba, H. (1982) Development and change of cortical field potentials during learning processes of visually initiated hand movements in the monkey. Exp. Brain Res., 48: 429–437.

Sejnowski, T.J. (1986) Open questions about computation in cerebral cortex. In: J.A. Feldman, P.J. Hayes and D.E. Rumelhart (Eds.), Parallel Distributed Processing Explorations in the Microstructure of Cognition, Vol. 2: Psychological and Biological Models, MIT Press, Cambridge, pp. 372–389.

Shepherd, G.M. and Koch C. (1989) Introduction to synaptic circuits. In: G.M. Shepherd (Ed.), The Synaptic Organization of the Brain, Oxford University Press, New York.

Sutton, R.S. and Barto, A.G. (1981) Toward a modern theory of adaptive networks: expectation and prediction. Psychol. Rev., 88: 135–170.

Thorpe, S.J., Rolls, E.T. and Maddison, S. (1983) The orbitofrontal cortex: neuronal activity in the behaving monkey. Exp. Brain Res., 49: 93–115.

Trotter, Y., Celebrini, S., Stricanne, B., Thorpe, S. and Imbert, M. (1992) Modulation of neural stereoscopic processing in primate area V1 by the viewing distance. Science, 257: 1279–1281.

Watanabe, M. (1990) Prefrontal unit activity during associative learning in the monkey. Exp. Brain Res., 80: 296–309.

Widrow, B. and Hoff, M.E. (1960) Adaptive switching circuits. Stanford Electronics Laboratories Technical Report 1553-1, Stanford University, Stanford, CA.