

## Econets: Neural networks that learn in an environment

Domenico Parisi, Federico Cecconi and Stefano Nolfi

Institute of Psychology, National Research Council, 15 viale Marx, 00137 Rome, Italy

Received 1 September 1989, in final form 30 March 1990

**Abstract.** Ecological networks are networks that learn in an environment. It is the environment, and not the researcher, that determines the conditions in which learning takes place such as which input patterns are seen, what the teaching input is, etc. Furthermore, input patterns at time  $N + 1$  are often a function of the output of the network at time  $N$ . Two hypotheses are explored with reference to ecological networks. One is that predicting the sensory consequences (input) for an organism of the organism's actions (output) on the environment is one of the basic tasks of this type of network—basic for constructing an environmental map or world model. The other is that learning to predict the sensory consequences of the organism's actions favourably predisposes the organism to learn to attain goals with those actions. Some data from simulations that support these two hypotheses are reported.

### 1. Ecological networks

Artificial neural network models aspire at biological plausibility. Hence, they are open to criticism when one can demonstrate that a particular feature of such models is biologically implausible. One type of biological implausibility concerns whether the 'brain-like' style of computing (Rumelhart 1989) that characterises neural networks is an appropriate model of how real brains compute mental functions. For example, doubts as to the neurophysiological plausibility of current neural network models have been raised with regard to the back-propagation of error antidromically on the same connections that propagate activation and inhibition (Crick 1988, but see Stork 1989) or the assumption of a simple algebraic sum of activations and inhibitions to calculate the net input to a unit (see the review by Segal (1988) of the book by McClelland and Rumelhart (1988)).

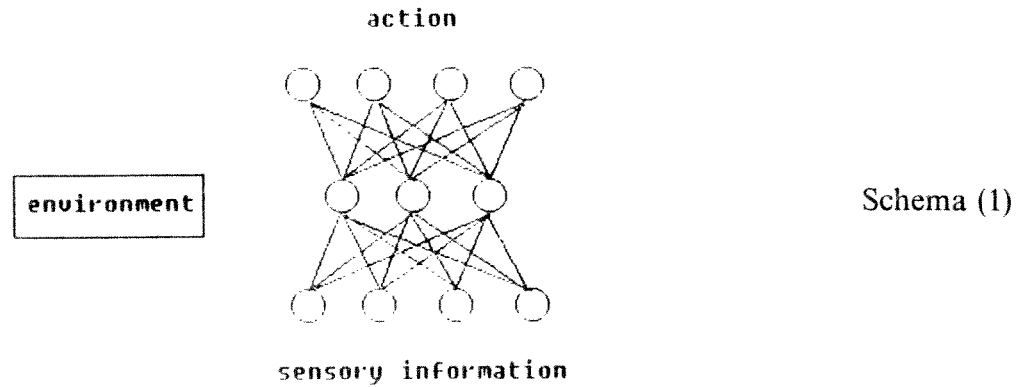
Another kind of biological implausibility of artificial neural network models concerns their behavioural appropriateness. Current neural models tend to be behaviourally implausible if one considers the conditions in which learning takes place. Many experiments and theorising on neural nets seem to assume that learning takes place in a vacuum. It is the experimenter who arbitrarily decides which input patterns will be seen by a network, in which order, with which frequency, what the teaching input is for computing error, which patterns are learning patterns and which are used to test the generalisation power of the network, etc. Moreover, the output of the network does not have any influence on which input pattern the network sees in the next cycle. This approach may be indicated for artificially manipulating experimental conditions and for conducting parametric studies but it runs the risk of missing critical aspects of the learning and behaviour of organisms.

In the biological world, by contrast, organisms live and learn in an environment. The environment has a structure that determines, in interaction with the behaviour of the organism, the conditions in which the organism's learning takes place. Most learning can be interpreted as the progressive construction, in an organism's nervous system, of an internal model of the environment in which the organism lives. This model is used by the organism to pursue its goals in the environment in an efficient way. If there is no environment, learning cannot but be arbitrary and will not exhibit many of the features of biological learning.

Ecological networks, or econets, are networks that learn in an environment. The experimenter defines the structure of the environment and from that point on it is the environment itself, and the behaviour of the organism in it, that determine the conditions in which learning takes place, i.e. which input patterns (sensory information) arrive at the network, in which order, with which frequency, what the teaching input is, etc. Furthermore, ecological networks pursue goals in the environment in which they live, which implies that they act on the environment producing changes in the environment itself or in the relationships between the organism and environment. Hence, the kind of sensory information impinging on the network as input patterns is normally a function of both the structure of the environment and the output of the network in the preceding cycle(s).

A fundamental difference between non-ecological and ecological networks is that non-ecological networks discover regularities in the input as such whereas ecological networks discover regularities also in the effects that the network's actions have on the environment as these effects are revealed by subsequent input. As Jordan (1989, p 62) has observed, what is emphasised in most current research is 'the role of environment as a source of stimuli, rather than as a recipient of actions'. But one should add that what is especially neglected is the role of the environment in mediating between the organism's actions and the subsequent stimulation to the organism. Even in research that emphasises the role of the environment as recipient of actions, such as Williams (1988), this only means that the environment has the role of correcting the network's output (actions) using some supervised learning method. As a matter of fact, the types of tasks that have been traditionally studied with neural systems tend not to be 'ecological'. Consider pattern classification; a network must learn to classify together different input patterns by providing the same response to all patterns belonging to the same class and a different response to the patterns belonging to another class. One typical way of doing this is through supervised learning in which the network is explicitly taught the response appropriate to each class of patterns. However, in real environments it is not clear where this explicit teaching input may come from. Organisms are not usually told what is the response appropriate for a particular class of stimuli but they have to find out by themselves. An ecological approach to pattern recognition would try to generate pattern recognition indirectly by having the network discover which input patterns require the same or similar actions (outputs) on the part of the network if the organism must attain some goal with those actions, and which patterns require a different type of actions. (For an ecological network approach to pattern recognition see Floreano *et al* 1989.)

The basic architecture of an ecological network includes some input units for encoding sensory information from the environment to the organism and some output units for encoding the organism's actions on the environment. The network's actions cause changes in the environment or in the relationships between the organism and the environment. These changes, in their turn, determine in part or completely the sensory information which arrives at the network in the successive cycle. The form of such a network is the following:



The environment is simulated by an algorithm which incorporates the environment's structure and, on the basis of this structure and the network's output (the organism's actions), computes the next input to the network (sensory information to the organism).

(Note that in schema (1) the role of sensory information with respect to action is not necessarily that of cause to effect as in a stimulus-response situation, but it can be that of helping the organism to select the appropriate action which has another, perhaps internal, cause.)

Two simple situations can help illustrate (1). In both examples the organism lives in a bi-dimensional world and it has a facing direction. (For research using similar situations, see Patarnello and Carnevali (1988) and Booker (1988).) In the first example (see figure 1) the organism can make one of four possible actions: it can move forward a fixed limited amount in the facing direction, it can turn right or left 90°, or it can do nothing. These four actions are coded in the output units. The environment contains a single static object (say, a piece of food) and at each cycle the network receives sensory information from this object specifying the object's position relative to the organism. This information, which is coded in the input units, specifies (a) the distance of the object from the organism, and (b) the angle of the object with reference to the facing direction of the organism.

Since the organism turns and moves, the sensory information from the food object changes. The input patterns which are presented to the network at each cycle are not established arbitrarily and *a priori* by the experimenter but are determined by the output of the network in the preceding cycle—and by the structure of the environment, e.g. where the food is located within it.

Our second example concerns an organism which does not move or turn but has a

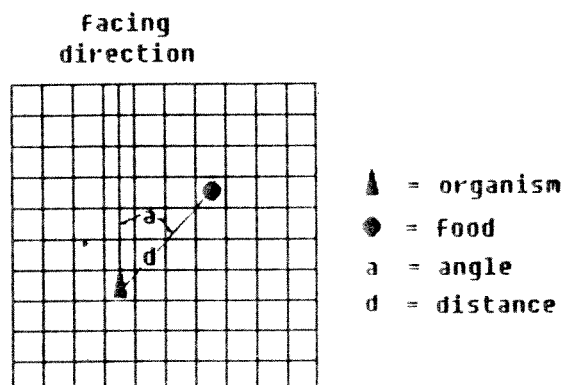
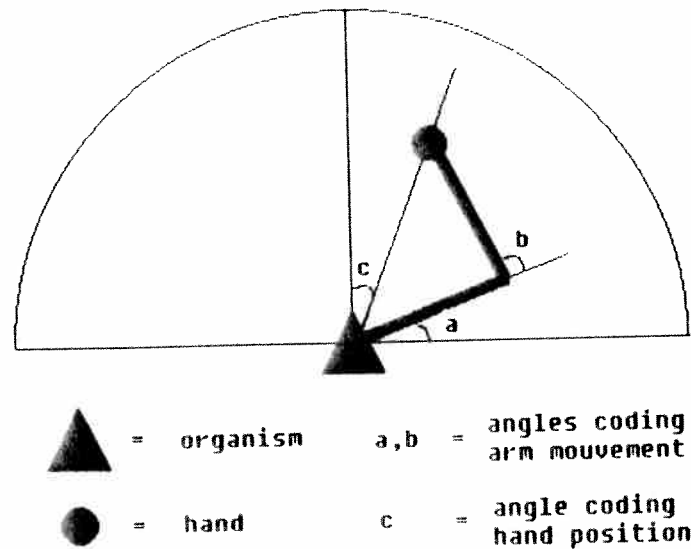


Figure 1. Example 1. The organism may move forwards, left or right, or may remain stationary.



**Figure 2.** Example 2. The organism cannot move but has a two-segment arm that can be moved.

two-segment arm that can be moved (see figure 2). The movements of the arm are specified by the two angles that the first and the second segment of the arm make with some fixed reference line. These two angles, which together specify a new position of the arm given a preceding position, are coded in the network's output units. The sensory information (input patterns) concerns the position of the endpoint of the farther segment of the arm, let us say, of the arm's hand. As in the preceding case, the network is informed at each cycle about the distance of the hand from the organism and about the hand's angle with reference to the organism's facing direction.

In this case too the input to the network at cycle  $N$  depends on the network's output in the preceding cycle  $N - 1$ . Since the organism moves its arm, this movement changes the arm's endpoint (the hand). The difference with respect to the first example is that in the present case the object (the hand) moves and the organism does not, whereas the opposite was true in the preceding example. However, since it is the organism that causes the movement of the hand by moving its arm, in both cases the organism's actions have sensory consequences.

It is interesting to consider how ecological networks relate to time. Time appears to be much more important for ecological networks than for ordinary, non-ecological ones. In ordinary networks without recurrent connections the relationship of an activation cycle to the following cycle is an extrinsic one. Each cycle is independent of the preceding or following one. In these networks, not only is it the experimenter who arbitrarily decides what the successive input to the network will be and therefore indirectly determines what the network will do in the next cycle, but the sequence of outputs of the network (that is, the outputs of successive activation cycles) is not a temporal sequence having a meaning as a sequence but merely a plurality of outputs.

On the other hand, one might say that the temporal dimension is intrinsic to the functioning of ecological networks. From the input point of view it is clear why ecological networks are temporal networks since each member of a sequence of inputs is determined either by the intrinsic temporal structure of the environment or by the output of the network (action on the environment) in the preceding cycle, or by both. But even more important for understanding the intrinsically temporal nature of ecological networks is their output side. Ecological networks have at least a portion of their output units that must necessarily be interpreted as 'actions of the organism on the environment'

and therefore have a causal effect on the environment, producing changes in it or in its relationship with the organism. Now, a sequence of actions of an ecological network has an intrinsic temporal meaning in that the behaviour of the network is likely to be evaluated not in terms of single actions but in terms of entire sequences of actions. We will call a sequence of actions of an ecological network a 'trajectory'. Typically, the final action of a trajectory is the one that attains some goal, e.g. that which determines a quantitative change in the parameter in terms of which the fitness or adaptiveness of a network is being evaluated. For example, if the fitness is 'number of food elements eaten', an action which directly causes the eating of a food element can be the final action of a trajectory which has included 'approaching' actions in addition to the final 'eating' action. (The final action of a trajectory used to be called the 'consummatory act' in animal psychology.)

The environment in which an ecological network lives can contain natural objects, other organisms (e.g. conspecifics), instruments and artefacts. Furthermore, the notion of environment that is needed for ecological networks must be sufficiently general and abstract to include the organism's body as part of the organism's environment. According to this abstract notion, the environment of an organism is whatever systematically mediates the relationship between the organism's actions and the sensory consequences (input patterns) of these actions for the organism itself. An object is part of an organism's environment if an action of the organism changes the sensory information coming from that object to the organism. It is clear, from this definition, that the organism's body is part of the organism's environment. If the organism moves one part of its body this may result in proprioceptive or visual information from that body part to the organism. Another kind of effect is tactile information when the body's moving part touches an external object. This effect is doubled if the touched object is another part of the organism's own body. Then, tactile sensory consequences from the organism's actions result from both the touching (and moving) part and from the touched part.

Another reason for including the body of an organism as part of the organism's environment is that as the external environment may change and the organism must adapt to these changes, so the organism's body can change, as in the course of growth, and the organism must adapt to these changes in its body. For example, if a network has learned to guide an arm to reach objects, a subsequent increase in the length of the arm will require some new adaptation on the part of the network.

It is important to keep in mind that an organism's body is part of the organism's environment since it might be that one of the first tasks for a developing organism is to construct an internal model of that very special part of the environment which is its own body. The second of the two organisms that we have described above, the one which moves its arm, receives sensory input from an object which is a part of the organism's own body, i.e. from its hand.

## **2. Learning to predict the sensory consequences of one's own actions**

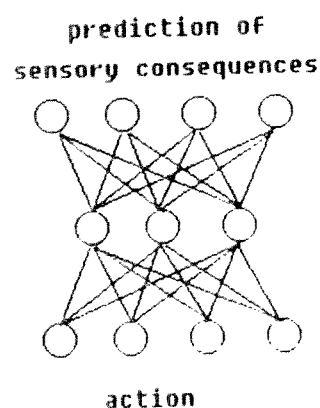
One fundamental task for ecological networks is to learn to predict the sensory consequences of the organism's own actions. The action output of an ecological network causes changes in the environment (e.g. by displacing or modifying objects) or in the relationship between the organism and the environment (e.g. when the organism turns or displaces itself). In both cases the actions of the organism have sensory consequences, in the sense that the sensory input from the environment in the next cycle is partially or completely determined by the particular action executed by the organism in the

preceding cycle. A prediction task, in the sense intended here, is learning to predict what these sensory consequences are, that is, to specify the nature of these consequences before a planned action has been executed. This kind of prediction task must be neatly distinguished from the task of predicting the environmental event that will follow a current event (see e.g. Elman 1988, Sutton and Pinette 1985). This second type of prediction task, which may be involved in classical conditioning (see Sutton and Barto 1981), in event expectation, and in the learning of sequences of events, can be very important in itself and can be intertwined in interesting ways with predicting the consequences of one's own actions. However, what we are concerned with here is this latter type of prediction only.

To construct a predicting network we will have to slightly complicate the basic design of an ecological network as described above. This basic design involved a set of input units coding sensory information from the environment and a set of output units coding actions of the organism. The environment is simulated by an algorithm that, on the basis of the environment's structure and the network's output at cycle  $N$ , determines the network's input at cycle  $N + 1$ .

A predicting network in a sense inverts the respective roles of sensory information and of actions. The input of a predicting network becomes the specification of an action that the organism is planning to do and the network's output is a prediction of what the sensory consequences of this action will be, i.e. what the next sensory input will be when the planned action will have been executed. Such a prediction task can be taught using back-propagation without violating the principles of ecological networks. In many uses of back-propagation it is assumed, rather implausibly, that a 'tutor' is continuously available to provide a teaching input to the learning organism. In a predicting network it is the environment itself which acts as such a tutor. For each input (planned action of the organism) the algorithm which simulates the environment computes the actual sensory consequences for the organism and these are used as teaching input to be compared with the predicted sensory consequences (output of the network).

The schema of this basic predicting network is the following:

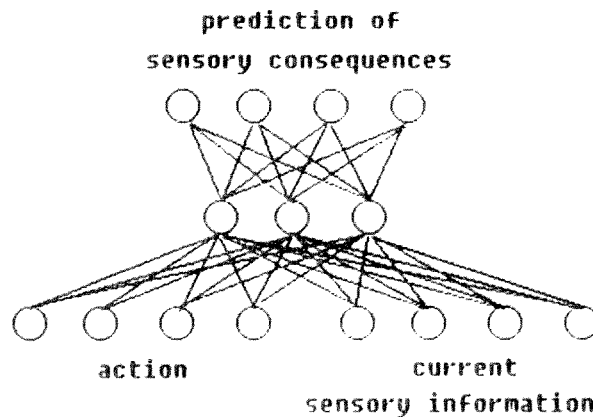


Schema (2)

Schema (2) refers to a situation in which a given action has sensory consequences that are independent of the current sensory information. One example of this situation might be an infant in its first year of life learning to predict the kind of sounds which are produced by its various phono-articulatory actions. What sound results from a phono-articulatory action is largely independent of the sensory situation prevailing when the action is executed. However, in most other cases the sensory consequences of an action depend both on the specific action and the current sensory data. This is true, for example,

for the two situations that we have described above. If a network must predict the new distance and the new angle of the object as the organism turns or moves in space, it is necessary for the network to know both what the specific action planned by the organism is and the current distance and angle of the object. In the same way, if we want a network to learn to predict the new position of the arm's endpoint (hand) as the organism moves its arm, we will have to supply the network with information both on the planned arm movement and on the current position of the hand.

The schema for these more complex but more realistic predicting networks is the following:

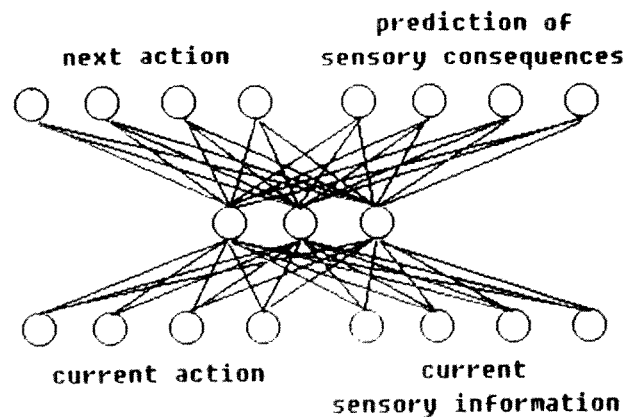


Schema (3)

It is important to ask where the two inputs to this network may come from. The teaching of a prediction task can be realised in two different ways, one more ecological than the other. The less ecological procedure assumes that it is the experimenter who tells the network each time which action to do. The experimenter prepares an *a priori* file of (e.g. randomly selected) actions and asks the network to learn to predict the sensory consequences of each of these actions (see e.g. Jordon 1989). Similarly, but independently, the current sensory information accompanying each input action can be arbitrarily specified by the experimenter. In other words, in this non-ecological approach it is the experimenter who decides each time the action whose sensory consequences must be predicted by the network and the sensory situation in which such an action is executed.

The alternative, more ecological, approach excludes the experimenter and assumes that it is the network itself that decides which action to take next. Similarly, it is the environment and the network's previous action which specify the sensory circumstances in which the action is executed. As we saw, it is a defining characteristic of ecological networks that they generate actions as output. What we want is that the same network that generates actions as output has a second kind of output, i.e. a prediction of the sensory consequences of the actions that the network itself generates. The action which is decided at cycle  $N$  (output) becomes the currently planned action (input) at cycle  $N + 1$ . As for the other input to network (3), i.e. the current sensory information, it is the environment itself which is the source for this information. In fact, the same angle and distance of the object (or hand) which is computed by the algorithm as the actual position of the object (or hand) after the organism's action is executed, and which is used as teaching input for correcting the prediction error, becomes the next current sensory information.

Hence, the more complete network that we propose to study has the following architecture:



Schema (4)

We assume that a network like (4) is the 'nervous system' of simple organisms like those that we have described above. In the next two subsections we report some results of simulations on teaching this type of network to predict the sensory changes resulting from the organisms' actions.

### 2.1. Predicting how the sensory information from an object changes as the organism displaces itself in space

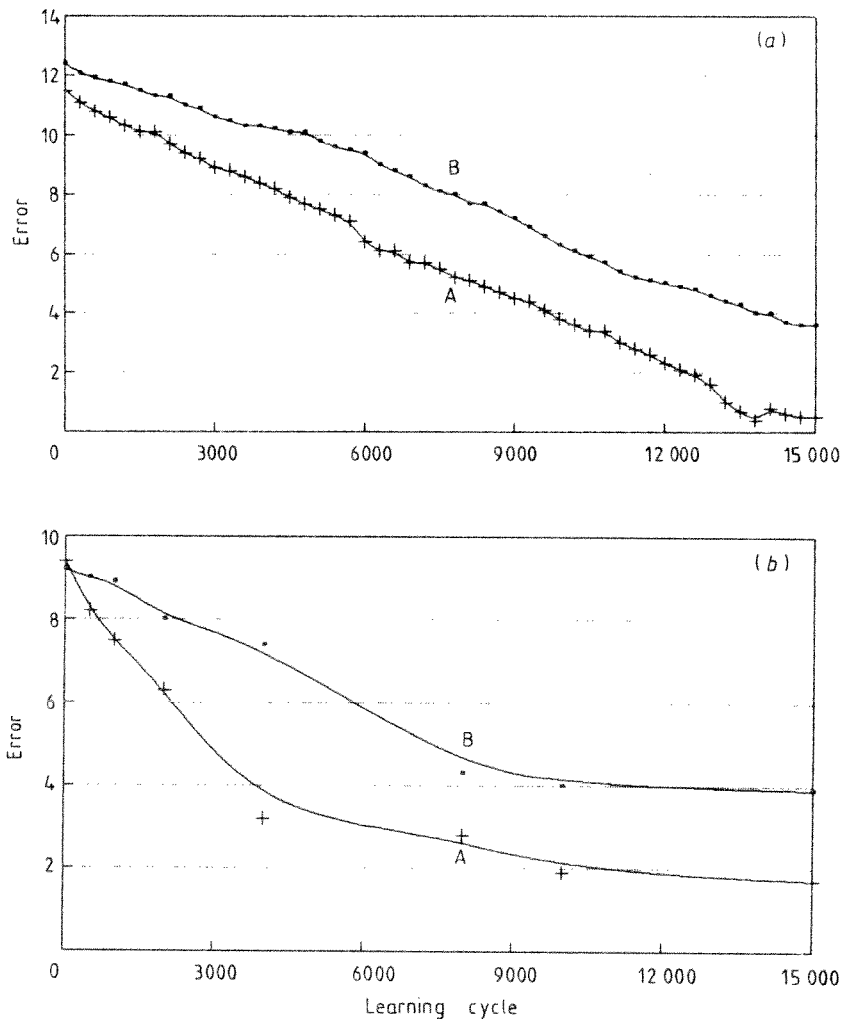
The organism lives in a bidimensional environment which is a grid of  $10 \times 10$  cells. The environment contains the organism and a single object, both occupying one cell. The organism has a facing direction and a repertoire of only four possible actions: it can move one cell in its facing direction, it can turn  $90^\circ$  left or right, or it can stay still. The sensory information from the object is (a) the Euclidean distance of the object from the organism, and (b) the angle of the object with the organism's facing direction.

The network which simulates the organism's nervous system is a feed-forward network with a single layer of hidden units. There are four input units, four output units, and seven hidden units. Two input units code the currently planned action. The coding is binary: 11 is 'move forward', 10 is 'turn  $90^\circ$  left', 01 is 'turn  $90^\circ$  right', 00 is 'stay still'. The other two input units code the current distance and angle of the object, respectively. The coding in this case is continuous, with angle and distance normalised to values between 0 and 1. Of the four output units, two code the next action and two code a prediction of the new angle and distance resulting from executing the currently planned action. The coding is identical to that of the corresponding input units.

There is no teaching for the action output units. The ability to predict how the object's position relative to the organism changes with the organism's displacements is taught using the backpropagation algorithm (Rumelhart *et al* 1986). The activation level of the prediction output units as computed by the network is compared with the correct activation level which is computed by the environment-simulating algorithm, and the resulting error is used to modify the weights of the connections from the hidden units to the prediction output units and from the input units to the hidden units. Notice, however, that the weights from the hidden units to the action output units are never modified but they remain always the same weights that were initially assigned in a random way to the network.

We ran three different networks with the same architecture but a different initial assignment of weights. Each network was taught 15 000 cycles of back-propagation, i.e.





**Figure 3.** Prediction error curves for (A) self-generated movements and (B) externally imposed movements for (a) body rotations and displacements, and (b) arm movements.

1000 actions  $\times$  3 epochs  $\times$  5 worlds, each epoch being defined by a new starting position of the organism and each world by a new position of the food element. The learning rate was 0.2 and momentum was 0.9.

To compare prediction learning based on actions generated by the experimenter with the same learning based on actions generated by the network itself, we ran two different kinds of simulations. In one simulation the output of the action units was completely ignored. We prepared a file of randomly selected actions and the network learned to predict the sensory consequences of these actions (externally generated actions). In the other simulation the output of the action units in cycle  $N$  became the currently planned action (input) in cycle  $N + 1$  (internally generated actions). In other words, in this second simulation the network learned to predict the sensory changes resulting from actions that it itself had generated. In both simulations, the actual sensory consequences as computed by the environmental algorithm were the next sensory input.

The results are displayed in figure 3(a) which shows the prediction error averaged for the three networks in each of the two conditions. The networks learn to predict in both simulations but the prediction error decreases more quickly when the network is learning to predict on the basis of internally generated actions than with externally generated actions.

## 2.2. *Predicting how the sensory information from the organism's own hand changes as the organism moves its arm*

This organism does not move but it has a two-segment arm that can be moved. As the arm moves, the endpoint of the farther segment, i.e. the arm's hand, also moves and the organism must learn to predict where the hand will end up when a currently planned movement of the arm is executed.

The environment of this organism is a bidimensional semicircular world with the organism sitting at the centre of the circle. The organism has a fixed 'visual field' of  $180^\circ$  and since the radius of the circle is equal to the length of the organism's arm when the arm is completely stretched, the hand has access to all points within the semicircular world.

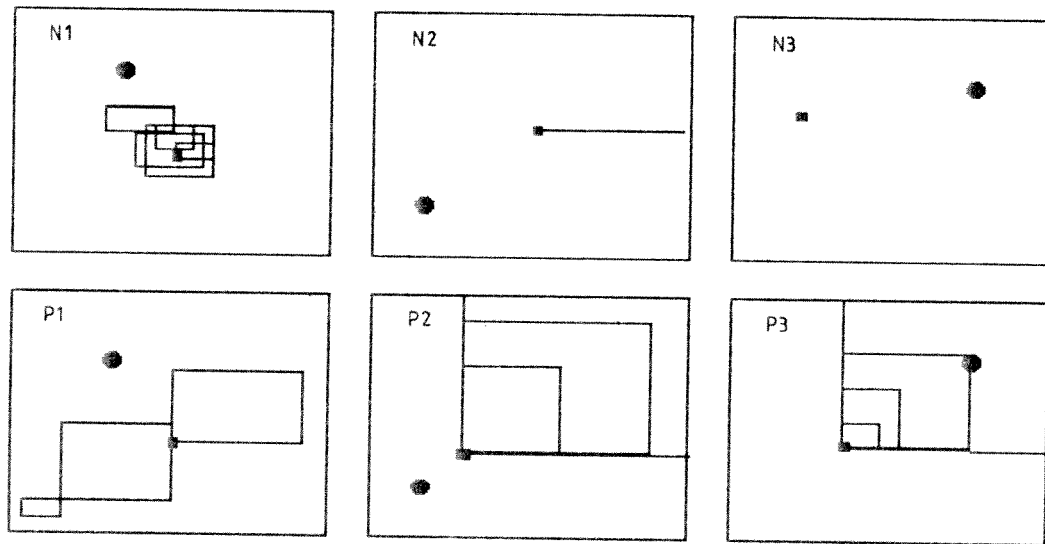
The architecture of the network is identical to that of the preceding simulations. The movements of the arm are defined by the two angles that must be added to or subtracted from the current angles of each of the arm's two segments to generate a new arm configuration. Two input units code the two angles that define the currently planned arm configuration. The arm movements are somewhat restricted in that each angle cannot exceed  $30^\circ$ . Within this restriction each angle is normalised to an activation value between 0 and 1. The position of the hand is expressed as its Euclidean distance from the organism's body and its angle with respect to the organism's fixed facing direction (North). Two other input units code the current angle and distance of the hand in the same manner as the object's angle and distance in the preceding simulations. Two output units code the next movement of the arm and two additional output units the predicted position of the hand.

There were 15 000 cycles of back-propagation subdivided in five epochs, each epoch being defined by a new starting position of the hand. As in the preceding simulations, the learning rate was 0.2 and momentum 0.9.

Two sets of simulations were run, one set with internally generated actions and the other with externally generated actions. The prediction error for the two sets of simulations (average of three networks in each condition) as a function of learning cycle is shown in figure 3(b). As in the preceding case, predicting the sensory consequences of one's actions is easier if such learning is based on actions that the network itself has generated.

As this point one should ask why prediction learning based on internally generated actions is constantly better than the same learning based on externally generated actions. Consider the architecture of our network (4). The connection weights which are responsible for generating a prediction are (a) the weights from the input to the hidden units, and (b) the weights from the hidden units to the prediction output units. On the other hand, the weights that generate the next action are (a) the weights from the input to the hidden units, and (b) the weights from the hidden units to the action output units. Hence, the two type of tasks, prediction and action, share a common set of connection weights, i.e. the (a) weights from the input to the hidden units, although they are also based on distinct weights, the (b) weights. Now it appears that if predictions are learned on the basis of actions generated by (at least some of) the same connection weights that also generate predictions, they become easier to learn.

An interesting consequence of this architecture is that the actions generated by a network will tend to change in the course of prediction learning. It is true that there is no teaching on the action output units and therefore the connection weights from the hidden units and these output units do not change. However, as we have indicated, an action is generated by these networks based both on these fixed weights and on the weights of the

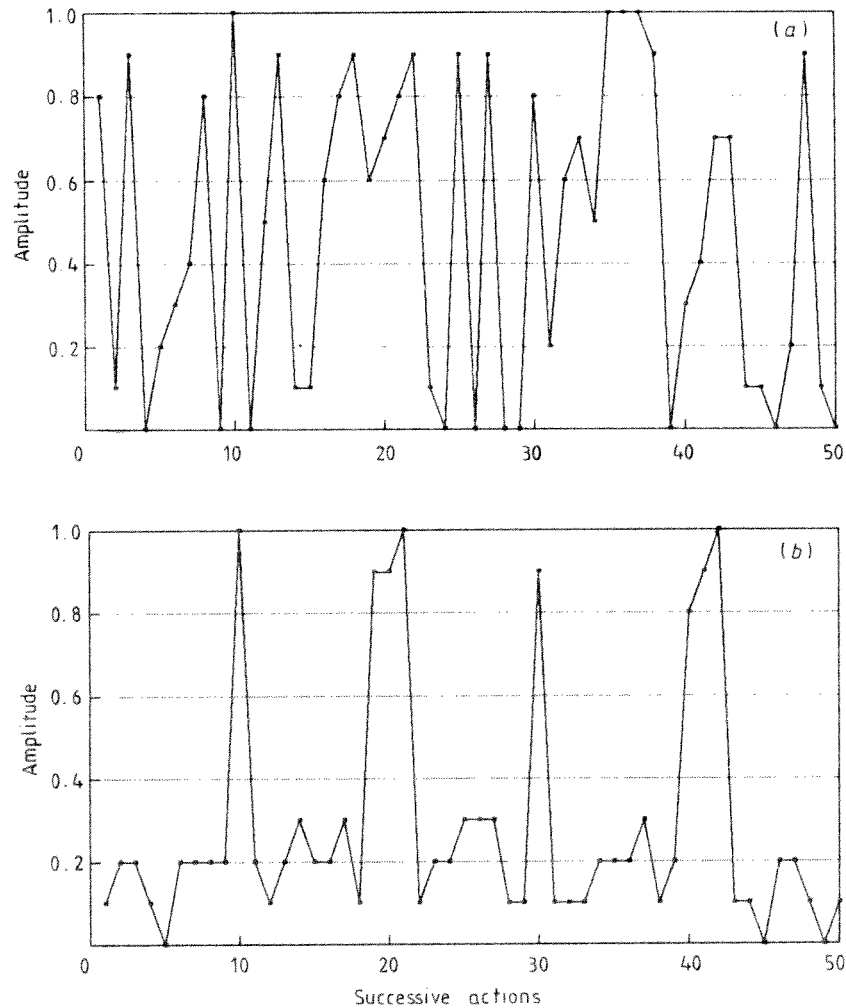


**Figure 4.** Trajectories of 50 successive body rotations or displacements for three naive (N) and three predicting (P) networks. The starting position is indicated by the rectangle and the food location by the circle. (Note that N2 is generated by an organism that always selects 'go forward' and N3 by an organism that always selects 'turn 90° right'.)

connections from the input units and the hidden units—and these weights do change as a consequence of prediction learning. As a matter of fact, we have observed some changes in the trajectories generated by our networks during prediction learning. These changes emerge if one compares a sequence of actions as these actions are generated by a network prior to prediction learning with a sequence of actions which are generated by the same network after it has learned to predict. Notice that these changes are a spontaneous by-product of prediction learning since no one is teaching the network to act in any particular way. Figure 4 shows a trajectory of 50 actions generated by each of the three networks of the first experiment prior to prediction learning and a trajectory of 50 actions generated by the same networks after prediction learning with internally generated actions. The first set of trajectories are based on the random weights initially assigned to the networks and the second set of trajectories are a result of the weights as some of them (those from the input to the hidden units) have been modified due to prediction learning. While prior to prediction learning the organisms' movements are quite stereotypic (e.g. 'turn right' or 'go forward' all the time), they become more differentiated after prediction learning, with the organisms exploring a larger portion of the surrounding space.

Changes also emerge in the kinds of arm movements that are generated by the networks in the second experiment before and after prediction learning. To compare the motor behaviour of these organisms before and after prediction learning we have examined the amplitude of their arm movements as measured by the two angles that define each arm movement. Figure 5 shows the amplitude of the two angles summed together (with identical sign) and normalised to values between 0 to 1 for 50 successive arm movements generated by one of the three networks before and after prediction learning. Prior to prediction learning the arm movements appear to be more sweeping and jerky so that the hand ends up in widely separated regions of the space. After learning the arm movements are less sweeping and more finely adjusted with periodical explorations of more distant portions of the space. The less sweeping arm movements are also typical of the other networks after they have learned to predict.

Given these changes in the actions that are generated as a network learns to predict



**Figure 5.** Trajectories of 50 successive arm movements for (a) one naive and (b) one predicting network. The trajectories are represented by the normalised sum of the amplitudes of the two angles defining each movement.

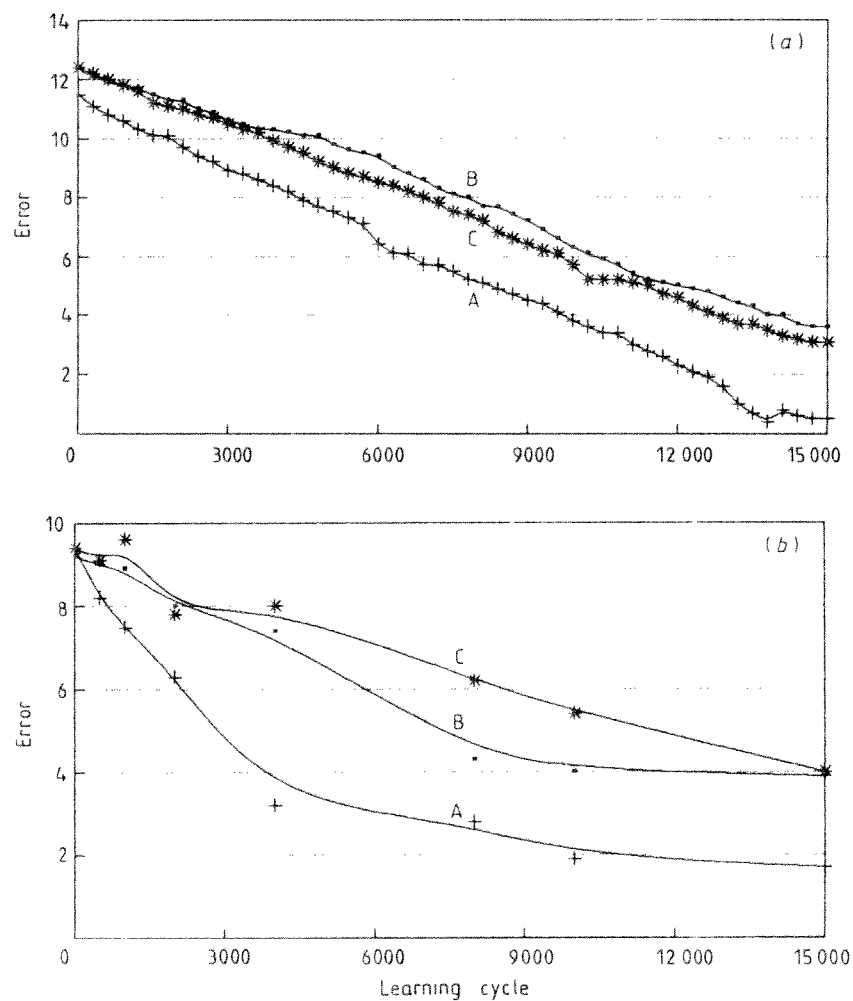
their sensory effects, one might advance the hypothesis that prediction learning is easier for internally than for externally generated actions because internally generated actions are progressively changed in a manner that tends to facilitate prediction learning. One might even think that if a network has to learn to predict the effects of some actions and it is free to decide which actions these are, it would progressively move toward very simplified trajectories. For example, a network that governs the displacements of an organism in the first experiment should tend to choose always the 'stay still' action because this would make the prediction task maximally easy: there would be no changes in the current sensory information from one cycle to the other. While it is interesting that the trajectories that actually emerge do not have this simplified character, the question can be explicitly posed: is the advantage in learning with internally generated actions due to the changes that are observed in these actions or to the fact that internally generated actions are generated by the same network that must learn to predict their sensory consequences?

To answer this question we have compared how networks learn to predict based on their own internally generated actions and how they learn on the basis of actions that are generated by another network already trained to predict. If the critical factor in facilitating learning is the changes in motor behaviour which are a by-product of

prediction learning, then prediction learning should be equally easy in the two circumstances—or even better for the network trained with actions generated by an already trained network since in this case the actions incorporate the changes right from the beginning. On the other hand, if what makes internally generated actions a better basis for prediction learning is that fact that they are internally generated, then learning on the basis of actions generated by an already trained network should be no better than learning based on randomly generated actions.

We have trained a set of network to predict the sensory consequences of their own actions. After training each network generates a sequence of actions which is then used as the input file for a new set of networks that must learn to predict. The results are shown in figure 6(a) for the networks of the first experiment and in figure 6(b) for the networks of the second experiment. The error curve for the networks trained using actions generated by already trained networks is as bad as that of the networks trained with randomly generated actions.

We conclude that learning to predict the sensory effects of actions is facilitated by the fact that the same network (i.e. the same connection weights) that must learn to predict is also generating the actions whose sensory effects it must predict. Prediction learning causes changes in the actions that are generated by a network but these changes do not by themselves facilitate prediction learning. If a network is trained to predict the sensory



**Figure 6.** Prediction error curves for (A) self-generated movements, (B) random movements, and (C) movements generated by another network, for (a) body rotations or displacements and (b) arm movements.

effects of actions which incorporate these changes (in that they have been generated by another network already trained to predict) but have not been self-generated, it will not learn any better than a network trained on the basis of externally chosen random actions.

### **3. Learning to predict the sensory consequences of one's own actions makes it easier to learn to attain goals with those actions**

A network which learns only to predict the sensory consequences of its own actions and stops there would be the nervous system of a purely speculative or contemplative organism. What is the practical advantage of such learning? The ability to predict the sensory consequences of one's own actions implies possession of an internal model of the environment in which those actions must be executed, a model which is necessary to attain some goals in that environment. One might think that such an internal model is at least in part constructed through the kind of prediction learning that we have discussed in the preceding section. If this is true, then learning to predict the sensory consequences of one's own actions should make it easier to attain goals with those actions. In learning to predict, an organism learns what changes in the environment or in its relationship with the environment will follow various actions in its repertoire. To attain a goal the organism must select an action that has a desired consequence. Hence, the two tasks are clearly related. However, learning to predict has the advantage that the organism can utilise the teaching input automatically provided by the environment and, furthermore, there is not much cost in making a prediction error. On the other hand, as we will discuss below, learning to attain goals with one's own actions is less likely to be based on external supervision and, furthermore, 'errors' in this case can be much more costly than prediction errors. We conclude that at least for more advanced organisms, i.e. organisms with a more complex, learned, flexible behaviour, prediction learning will be an important occupation, especially during development, and that the functional significance of prediction learning is to help the organism learn to attain goals with its actions. (For related research that connects prediction learning with goal attainment learning (see Jordon (1989) and Kuperstein (1988).

One problem that must preliminarily be solved when talking of goal attainment is how to make a network learn to reach goals with its actions. Consider a network that controls the organism which turns and moves in a bidimensional environment. The environment contains an object, say a food element, and the network learns to predict the position of the food relative to the organism, expressed in terms of distance and angle from the organism. Assume now that we want the organism to learn to approach the food element, i.e. to select the next action in such a way that this action is more likely to bring the organism closer to the food element and that a sequence of actions is more likely to bring the organism to where the food is located. (Since a food element disappears when the organism steps on it, we will say that the organism 'eats' the food.) Remember that during prediction learning the network was free to choose the next action at its will and there was no teaching of any kind as to which action to select. However, as we saw, as a spontaneous by-product of prediction learning the network was induced to modify somewhat its trajectories.

But now we want to change the action output of the network in such a way that the organism exhibits efficient food-approaching trajectories. To use a supervised teaching method such as back-propagation for this type of learning would not be very plausible for ecological networks. Each action generated by the network would have to be compared with the 'correct' action in those circumstances, i.e. with the action most likely

to bring the organism nearer to food. However, it would not be clear at all where this teaching input would come from. Furthermore, in many circumstances we, as experimenters, may not be able to identify the best solution to a problem and it would be much better to have a network discover its own solutions and perhaps come up with some solutions that we would not have thought of. (For a discussion of the problems of teaching complex trajectories to reach a final goal, see Tesauro and Sejnowski (1989).)

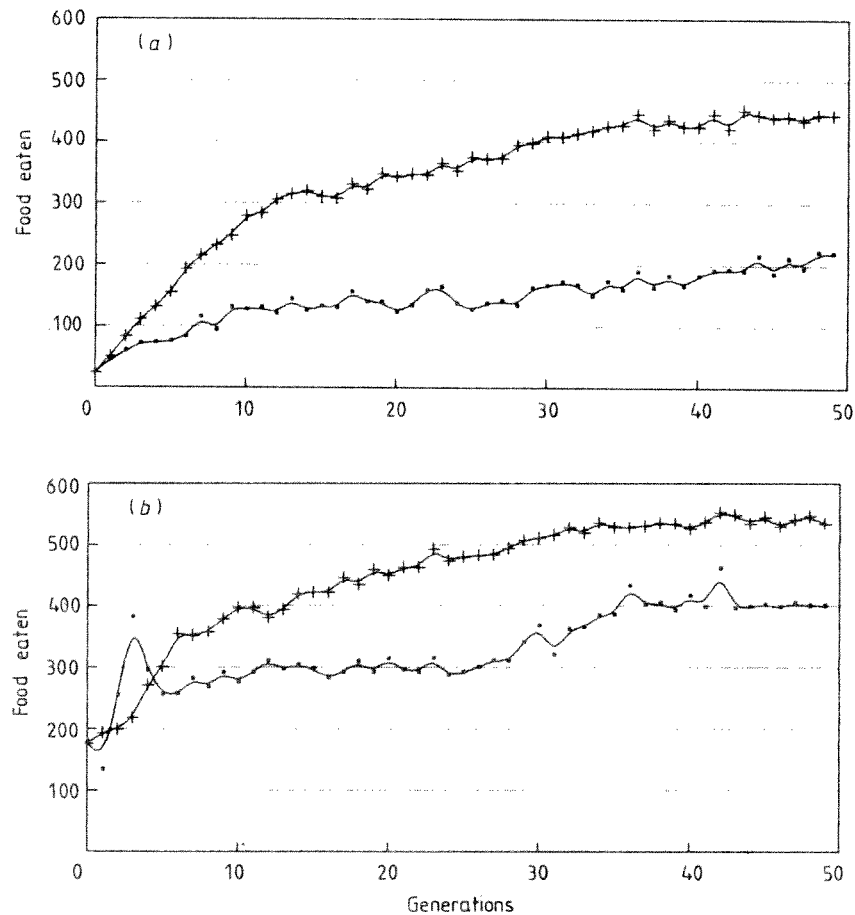
Other weaker supervised teaching methods such as reinforcement learning (Williams 1988) would partially but not entirely solve the problem. In reinforcement learning the teaching input is not a description of the correct output, as in back-propagation, but only a quantitative evaluation of the output generated by the network. This might be appropriate for the final 'consummatory' act of eating food but not for the selection of actions in approaching food.

For the above reasons we have discarded supervised teaching methods for goal attainment learning and we have made recourse to one form of evolutionary learning which is based on random mutation and selective reproduction. In other works, to teach networks to generate food-approaching trajectories we have used a method which does not involve any direct weight change in individual networks but rather is based on weight mutation and selection in whole generations of networks (Nolfi *et al* 1989).

An initial generation of  $N$  networks, each with its initial random matrix of connection weights, is allowed to move freely in a certain number of food-containing environments. Each network 'lives' alone in an environment containing multiple food elements. However, sensory information only comes in any particular time from the nearest food element. Given the different random assignment of connection weights, some networks will happen to have weights that given some sensory input from a food element will tend to generate actions that bring the organism nearer to the food while other networks will not exhibit this type of approaching behaviour. At the end of their life, which has the same length for all networks, the networks are rank ordered in terms of a 'fitness' measure, i.e. how many food elements they have eaten. Only the best are allowed to reproduce by generating copies of their weight matrix (offspring). In addition, the weight matrix of each offspring receives a limited amount of random mutation resulting in some offspring likely to be worse and some better than their parents. However, selective reproduction ensures that the mutations that result in better food-approaching trajectories are more likely to be retained than other mutations.

The method has many similarities to the genetic algorithms of Holland (1975). There are two main differences. The first is that the material on which selection operates is a population of binary digit vectors in Holland's algorithms and a population of connection weight matrices in our case. One can specify a network's weight matrix using a binary digit vector but then the problem of mapping binary numbers onto continuous-value weights must be solved. The other difference is that Holland uses crossover, inversion, and mutation as genetic operators whereas we only use mutation. The reason we restrict ourselves to mutation is that crossover and inversion appear to be more appropriate for vectors while mutation is perfectly appropriate for weight matrices.

We have used our evolutionary method both for teaching organisms that move in space to approach food and for teaching organisms with a moving arm to move their arm so that its endpoint (hand) reaches an object. Since we were interested in the role of prediction ability in goal-attainment learning, we have run two different simulations in each case. In one simulation there is no teaching of the prediction ability but only evolutionary emergence of the ability to attain goals (eating food and reaching objects, respectively). In the second simulation, in addition to subjecting populations of networks to random mutation and selective reproduction to evolve food-approaching and object-



**Figure 7.** Increase of food eaten by 50 successive generations with (upper curves) and without (lower curves) individual prediction learning for (a) the average and (b) the best individual.

reaching behaviour, we also teach individual networks during their 'lifetime' to predict the sensory consequences of their actions. Although inheritance is Darwinian and not Lamarckian, i.e. the weight matrix which is copied at reproduction is the parent's original matrix at birth, not the modified matrix that results from prediction learning at the end of life, we expect that prediction learning will accelerate the evolutionary emergence of the goal-attainment capacity.

There were 100 networks per generation and the best 20 networks each had five offspring (copies of their weight matrix) at the end of their lifetime so that a new generation of 100 networks came into existence. Five randomly selected connection weights were mutated in each offspring matrix by adding a quantity randomly selected between +1.0 and -1.0 to these weights. Each individual network had a total lifetime of 5000 actions in five worlds. For the organisms displacing themselves a world was a  $10 \times 10$  cell environment containing ten randomly distributed pieces of food. Sensory information at any particular time specified angle and distance from the nearest food. For the organisms with a moving arm a world was the semicircular environment that has already been described, with the addition of 20 randomly distributed objects. The network for the latter organisms included two additional input units for coding angle and distance of the object that was nearest to the 'hand' at any particular time.

The results are presented in figures 7 and 8. Figure 7 shows the evolutionary increase in the number of food elements eaten by the moving organism with and without prediction learning during life, for (a) the average and (b) the best individual across 50 generations. Figures 8(a) and 8(b) give the same results for the arm-moving organism

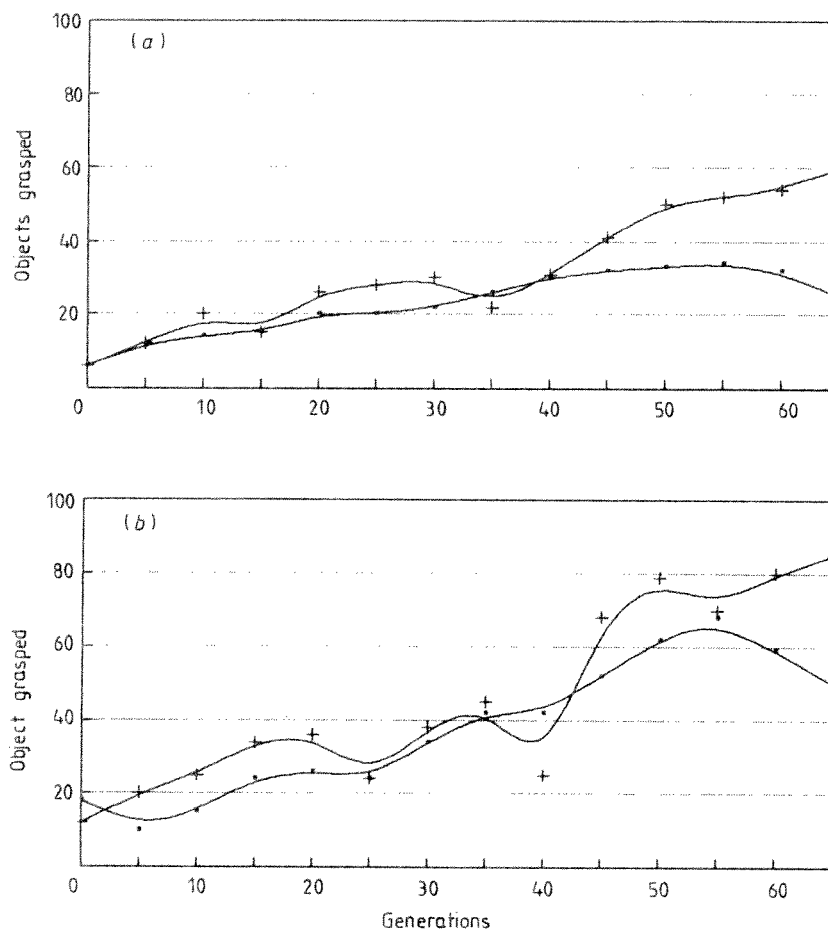


with regard to the number of objects grasped across 65 generations. As these figures show, the evolutionary increase in goal-attainment capacity is greater in both cases for the organisms that learn during their life to predict the sensory consequences of their actions.

#### 4. Discussion

Neural networks can be taught to predict the sensory consequences of their planned actions given those actions and the current sensory situation as input. Prediction learning is easier if the actions whose sensory consequences must be predicted are generated by the same network that is learning to predict instead of being randomly generated from outside. Although a network which is learning to predict progressively changes the sequence of actions it generates, it is not these changes in themselves that explain why prediction learning is easier for self-generated actions. Actions that are generated by a network which has already learned to predict, and that therefore incorporate these changes, do not make prediction learning easier when they are used by another network for its own learning. What appears to be critical is that the same connection weights that are used for generating actions are also used for predicting what the sensory consequences of those actions are.

The use of a supervised teaching method such as back-propagation to teach this



**Figure 8.** Increase of number of objects grasped by 65 successive generations with (upper curves) and without (lower curves) prediction learning for (a) the average and (b) the best individual.

prediction ability is quite plausible behaviourally since it is the environment itself that provides the network with the necessary teaching input for comparing the predicted and the actual sensory consequences of its actions. In fact, we believe that learning to predict the sensory consequences of their own actions is an important part of what young organisms do in their everyday life. This must be especially true for organisms whose behaviour has a large learned component. Learning to predict the sensory changes resulting from one's own actions is one way—perhaps the most basic way—to construct an internal model of the environment in which an organism must pursue and possibly attain its goals. If this internal model is largely genetically determined and the organism's behaviour is restricted and unmodifiable, then there is not much reason for learning to predict the consequences of one's own actions. But if the model of the environment must be learned and the organism's behaviour must adapt to this progressively learned model, then prediction learning should emerge as a fundamental occupation of the young organism.

In fact, prediction learning can be taken as a neural network model of what is called 'exploratory behaviour' in newborn and young children (Gibson and Spelke 1983). Exploratory behaviour is usually viewed as a way of exposing oneself to a larger variety of environmental conditions (sensory input). This is consistent with a view of the progressive construction of an internal model of the environment as based on extracting information from the sensory input as such. This view is quite common in psychology and in the cognitive sciences more generally and, as we have noted, may also be implicit in much research on neural networks.

The view espoused here is more action-based. The internal model of the environment that any minimally flexible organism must construct to be able to attain goals in that environment is primarily based on extracting information not from the sensory input as such but from the systematic relationship which the organism notices between its own actions and the resulting sensory input. Our view is also more active in that the organism does not passively notice this systematic relationship but actively predicts what the sensory consequences of its actions will be and modifies itself (learns) as is necessary to attain a better prediction ability.

Another consequence of prediction learning is that the organism may change its exploratory behaviour in order to get better prediction results. In fact, one potential advantage of the present view of the construction of an internal model as prediction learning is that this view might explain some aspects of the global changes in behaviour that are observed in early life. As we have reported, a neural network that learns to predict the sensory consequences of the actions that the network itself has generated tends to modify those actions. An interesting research perspective would be to relate these changes in the motor behaviour of networks that are learning to predict to the changes from more random to more systematic movements that may be observable in infants as they learn various behaviours.

This view of the progressive construction of an internal model of the environment is obviously related to Piaget's conception of intelligence as ultimately based on action. More specifically, one might think that the type of prediction learning that we have discussed captures at least one component of the notion of a circular reaction which is used by Piaget to analyse the development of sensory-motor intelligence (Kuperstein 1988). Prediction learning appears to capture the knowledge-acquisition component of circular reactions but not their motivational component which is linked with their producing what Piaget calls 'desirable results'. Furthermore, there is nothing in our present notion of prediction learning that would allow us to distinguish between 'primary', 'secondary', and 'tertiary' circular reactions, as Piaget does.

Before ending this discussion of prediction learning we want to emphasise that we have considered only some limited aspects of predicting the sensory consequences of one's own actions. We have examined the predictions of how the position of an object relative to the organism changes as the organism moves in space and how the position of one part of an organism's own body changes with respect to the rest of the body as the organism moves that part (hand). But many other instances of this type of prediction learning await examination. We have mentioned predicting the auditory consequences of phono-articulatory actions, where the sensory effects depend only on the particular action which is executed and are largely independent of the current sensory input. More complex cases include predicting how the position of an object *A* with respect to another object *B* will change as a consequence of the organism acting on *A*, which might be critical for the development of tool use, and predicting the changes in the behaviour of other organisms that are caused by one's own actions, which might underlie much of social interaction and communication.

One justification for claiming that prediction learning is at least in part the same thing as constructing an internal model of the environment is that one can demonstrate that learning to reach some goals in an environment—which presumably presupposes at least in advanced organisms the possession of such an internal model—is facilitated if the organism knows how to predict. We have demonstrated this positive influence of prediction learning on action or goal-attainment learning in two different instances: (i) predicting the new position of an object, as the organism moves, facilitates the approach to that object, and (ii) predicting the new position of the hand, as the organism moves its arm, facilitates the movement of the arm so as to reach objects with the hand.

Of course, an internal model of the environment consists not only of the ability to predict the sensory consequences of one's own actions but also the ability to predict changes in the environment that are independent of the one's own actions. This is an important topic of research in itself that we have explicitly not discussed. However, we have some preliminary results suggesting that the positive relationship between predicting ability and goal-attainment capacity may hold in this case too. We have trained a network to rotate a single eye so as to keep the eye's line of sight on a target object that moves regularly in space. We have compared learning this task in two different situations. In one situation the network only learns to rotate the eye in order to keep it on target. In the other the network is also taught to predict where the target object is going to be in the next moment in time. Learning to keep the eye on target is quicker if the network also learns to predict the next position of the target. If predicting what will happen next is importantly related to having an internal model of the environment which is useful for attaining goals in the environment, the capacity to attain goals should be learned more easily and more quickly if the organism can predict both those changes that depend on its own actions and those that are the result of independent environmental events.

## 5. Conclusion

We have introduced the notion of an ecological network and we have attempted to show that research on ecological networks may be necessary if many important aspects of the learning and behaviour of organisms must be captured by neural network models. Basically, an ecological network is a network that lives and learns in an environment. It is the environment with its structure and dynamics that determines the conditions in which behaviour and learning take place. More specifically we have proposed two

hypotheses concerning ecological networks and we have presented some initial data from simulation experiments bearing on those hypotheses. The two hypotheses are, firstly, that a basic task of ecological networks is to learn to predict the sensory consequences of the actions that the network itself generates and, secondly, that this prediction learning helps the network in learning to attain goals with those actions. A number of very interesting directions of research are open if one complicates the extremely simple situations used in the experiments already performed.

## References

- Booker L B 1988 Classifier systems that learn internal world models *Machine Learning* **3** 161–92
- Crick F 1989 The recent excitement about neural networks *Nature* **337** 129–32
- Elman J L 1988 Finding structure in time *Report Center for Research in Language, University of California, San Diego*
- Floreano D, Verterchi D and Parisi D 1989 Emerging linguistic capacities in sensory motor networks *Report Institute of Psychology, CNR, Rome*
- Gibson E J and Spelke E S 1983 The development of perception *Handbook of Child Psychology* vol III, ed P H Mussen (New York: Wiley) pp 1–75
- Holland J J 1975 *Adaptation in natural and artificial systems* (Ann Arbor, MI: University of Michigan Press)
- Jordan M I 1989 Supervised learning and systems with excess degrees of freedom *Proc. 1988 Summer School on Connectionist Models* ed D Touretsky, G Hinton and T Sejnowski (Palo Alto, CA: Morgan Kaufmann) pp 62–75
- Kuperstein M 1988 Neural model of adaptive hand–eye coordination for single postures *Science* **239** 1308–11
- McClelland J E and Rumelhart D E 1988 *Explorations in Parallel Distributed Processing* (Cambridge, MA: MIT Press)
- Nolfi S, Elman J and Parisi D 1989 Genetic adaptation and individual learning in neural networks *Report Institute of Cognitive Science, University of California, San Diego*
- Patarnello S and Carnevali P 1988 Neural networks for behavioral experiments: training an individual to find food in an external environment *Preprint IBM Rome Research Center*
- Rumelhart D E 1984 The architecture of mind: a connectionist approach. *Foundations of Cognitive Science* ed M I Posner (Cambridge, MA: MIT Press)
- Rumelhart D E, Hinton G E and Williams R J 1986 Learning internal representations by error propagation *Parallel Distributed Processing Foundations* vol 1 ed D E Rumelhart and J L McClelland (Cambridge, MA: MIT Press)
- Segal M M 1988 Book review *Science* **241** 1107–8
- Stork D G 1989 Is backpropagation biologically implausible? *Proc. Int. Joint Conf. on Neural Networks (Washington, DC, June 1989)* (Piscataway, NJ: IEEE) pp II-241–6
- Sutton R S and Barto A G 1981 Toward a modern theory of adaptive networks: expectation and prediction *Psychol. Rev.* **88** 131–70
- Sutton R S and Pinette B 1985 The learning of world models by connectionist networks *Proc. 7th Annual Conf. of the Cognitive Science Society* pp 54–64
- Tesauro G and Sejnowski T J A parallel network that learns to play backgammon *Artif. Intellig.* **39** pp 357–90
- Williams R J 1988 Connectionist learning through gradient following *Evolution, Learning, and Cognition* ed Y C Lee (Singapore: World Scientific)