# A1.1 The historical background

## *J G Taylor*

### Abstract

The brief history of neural network research presented in this section indicates that, although the initial revolution in neural networks lost its early momentum, the second revolution may well avoid the fate of the first. The subject now has strengths that were absent from its earliest version: these are discussed, and especially the fact that the biological origin of the subject is now giving it greater stability. The new avenues opened up by biologically motivated research and by studies in other areas such as statistical mechanics, statistics, functional analysis and machine learning are described, and future directions discussed. The strengths and weaknesses of the subject are compared with those of alternative and competing approaches to information processing.

### A1.1.1 Introduction

The discipline of neural networks is presently living through the second of a pair of revolutions, the first having started in 1943 with the publication of a startling result by the American scientists Warren McCulloch and Walter Pitts. They considered the case of a network made up of binary decision units (BDNs) and showed that such a network could perform any logical function on its inputs. This was taken to mean that one could 'mechanize' thought, and it helped to support the development of the digital computer and its use as a paradigm for human thought. The result was made even more intriguing due to the fact that the BDN is a beautifully simple model of the sort of nerve cell used in the human brain to support thinking. This led to the suggestion that here was a good model of human thought.

Before the logical paradigm won the day, another American, Frank Rosenblatt, and several of his colleagues showed how it was possible to train a network of BDNs, called a *perceptron* (appropriate for a device which could apparently perceive), so as to be able to recognize a set of patterns chosen beforehand (Rosenblatt 1962). C1.1.1

This training used what are called the connection weights. Each of these weights is a number by which one must multiply the activity on a particular input in order to obtain the effect of that input on the BDN. The total activity on the BDN is the sum of such terms over all the inputs. The connection weights are the most important objects in a neural network, and their modification (so-called *training*) is presently under close study. The last word has clearly not yet been said on what is the most effective training algorithm, and there are many proposals for new learning algorithms each year. B3

The essence of the training rules was very simple: one would present the network with examples and change those connection weights which led to an improvement of the results, so as to be closer to the desired values. This rule worked miracles, at least on a set of rather 'toy' example patterns. This caused a wave of euphoria to sweep through the research community, and Rosenblatt spoke to packed houses when he went to campuses to describe his results.

One of the factors in his success was that he appeared to be building a model duplicating, to some extent, the activity of the human brain. The early result of McCulloch and Pitts indicated that a network of BDNs could solve any logical task; now Rosenblatt had demonstrated that such a network could also be trained to classify any pattern set. Moreover, the network of BDNs used by Rosenblatt, which possessed a more detailed description of the state of the system in terms of the connection weights between the model neurons than did the *McCulloch–Pitts network*, seemed to be a more convincing model of the brain. B1.2

### A1.1.2 Living neurons

To justify such a strong claim it is necessary to expand the argument a little. *Living neurons* are, in fact, composed of a cell body and numerous outgrowths. One of these, which may branch into several collaterals, is called the axon. It acts as the output line for the neuron. The other outgrowths are called the dendrites; they are often covered with little 'spines', where the ends of the axons of other cells attach themselves. The interior of the nerve cell is kept at a negative electric potential (usually about $-60$ mV) by means of active pumps in the cell wall which pump sodium ions outside and keep slightly fewer potassium ions inside. This electrical balance is especially delicately assessed at the exit point of the axon. If the cell electrical potential becomes too positive, usually by about $+10$ to $+15$ mV, then there will be a sudden reversal of the potential to about $+60$ mV, and an almost as sudden return to the usual negative resting value, all in about 2 to 3 ms. A1.2

This sequence of potential changes is called an action potential, which moves steadily down the axon and its branches (at about 1 to 10 m $s^{-1}$). It is this action potential that is the signal sent from one nerve cell to its neighbors. The generation of the signal by the neuron is achieved by the summation of the signals coming to the cell body from the dendrites, which themselves have been affected by action potentials coming to them from nearby cells. The strengths of the action potentials moving along the axons are all the same. It is by means of rescaling the effects of each action potential as it arrives at a synapse or junction from one cell to the next (by means of multiplication of the incoming activity of a nerve impulse by the appropriate connection weight mentioned earlier) that a differential effect is achieved for each cell on its neighbors.

The above description of the actions of the living nerve cells in the brain is highly simplified, but gives a correct overall picture. It is seen that each nerve cell is acting like a BDN, with the decision to respond being that of assessing whether or not the total activity from its neighbors arriving at its axon outgrowth is above the threshold mentioned earlier. This activity is the sum of the incoming action potentials scaled by an appropriate factor, which may be identified with the connection weight of the BDN. The identification of the BDN with the living nerve cell is thus complete. A network of BDNs is, indeed, a simple model of the brain.

### A1.1.3 Difficulties to be faced

This, then, was the first neural network revolution. Its attraction to many (although not all) was reduced when Marvin Minsky and Seymour Papert showed in 1969 that perceptrons are very limited. They have an Achilles heel: they cannot solve some very simple pattern classification tasks, such as separating the binary patterns (0, 0), (1, 1) from the patterns (1, 0), (0, 1), known as the parity problem, or XOR. To solve this problem it is necessary to have neurons whose outputs are not available to the outside world. These so-called 'hidden neurons' cannot be trained by causing their outputs to become closer to the desired values given by the training set. Thus, in the XOR case, the input–output training set is (0, 0), 0; (1, 1), 0; (0, 1), 1; (1, 0), 1. The desired outputs of 0 or 1 (in the various cases) for the output neurons are not provided for any hidden neuron. Yet in the case of any linearly inseparable problem, such as XOR, there must be hidden neurons present in the network architecture in order to help turn the problem into a linearly separable one for the outputs.

In addition, there was a further important difficulty which was emphasized by Minsky and Papert, who gave a very thorough mathematical analysis of the time it takes to train such networks, and how this increases with the number of input neurons. It was shown by Minsky and Papert (1969) that training times increase very rapidly for certain problems as the number of input lines increases.

These (and other) difficulties were seized upon by opponents of the burgeoning subject. In particular, this was true of those working in the field of artificial intelligence (AI) who at that time did not want to concern themselves with the underlying 'wetware' of the brain, but only with the functional aspects—regarded by them solely as logical processing. Due to the limitations of funding, competition between the AI and neural network communities could have only one victor.

### A1.1.4 Reawakening

Neural networks then went into a relative quietude, with only a few, but very clever, devotees still working on it. Then came new vigor from various sources. One was from the increasing power of computers, allowing simulations of otherwise intractable problems. At the same time, the difficulty of training hidden neurons was solved by the *backpropagation algorithm*, originally introduced by Paul Werbos (1974), and independently discovered by Parker (1985) and LeCun (1985); it was highly publicized by the PDP Group with Rumelhart and McClelland (1986). Backpropagation allowed the error to be transported back from the output lines to earlier layers in the network so as to give a very precise modification of the weights on the hidden units. It was possible to simulate ever-larger problems using this training scheme, and so begin to train neural networks on industrially interesting problems. C1.2.3

Another source of stimulus was the seminal paper of John Hopfield (1982) and related work of Grossberg and collaborators (Cohen and Grossberg 1983) in analyzing the dynamics of networks by introducing powerful methods based on Lyapunov functions to describe this development. In all, this work showed how a network of BDNs, coupled to each other and asynchronously updated, can be seen to develop in time as if the system were running down an energy hill to find a minimum. Hopfield (1982) showed, in particular, how it is possible to sculpt the energy landscape so that there are a desired set of minima. Such a network leads to a content-addressable memory, since a partially correct starting activity will develop into the complete version quite quickly.

The introduction of an energy function quickly alerted the physics community, ever eager to sharpen their teeth on a new problem. This led to the spin glass approach, with the global ideas on phase transitions and temperature entering the field of neural networks for the first time. A spin glass derivation was also given by Amit (1989) of the capacity limit of $0.14N$ as the limit to the number of patterns which can usefully be stored in a network of $N$ neurons (and which was originally found experimentally by Hopfield (1982)). Gardner then introduced the general notion of the 'space' of neural networks (Gardner 1988), an idea that has been explored more fully by the recent developments of differential geometry by the work of Amari (1991). It is clear that the statistical mechanical approach is still flourishing, and is leading to many new insights. For example, it has become clear how the presence of temperature allows the avoidance of spurious states brought about by the form of the connection weights; these false states are made unstable if the network is 'hot' enough, and only the correct states are recalled in that case. It has also become clear as to what was the source of the limit on the storage capacity of these networks, and how this might be increased by choosing suitable connectivity to obtain the full capacity $N$ (Coombes and Taylor 1993).

Another very important historical development was the creation of the *Boltzmann machine* (Hinton and Sejnowski 1983), which may be regarded as the extension of the *Hopfield network* to include hidden neurons. The name was assigned since the probability distribution of the states of the network is identical to the Boltzmann distribution. The Boltzmann machine learning algorithm, based on the Kullback–Liebler metric as a distance function on the probability distributions of the states, allowed this probability distribution to move more closely to an external one to be learned. However, the learning algorithm is slow, and this has prevented many useful applications. C1.4 B1.3

A further network which proved very attractive to those entering the field was the *self-organizing map*. This had been developed by several workers (Willshaw and von der Malsburg 1976, Grossberg 1976) and reached a very effective form for applications in terms of the self-organizing feature map (SOFM) of Kohonen (1982). This allowed the weights of a single-layer network to adapt to an ensemble of inputs so as to learn the distribution of those inputs in an ordered fashion. Numerous developments have occurred in this approach more recently (Ritter *et al* 1991). C2.1.1

The other question, of the scaling of training times as the size of the input space increases, which was raised by Minsky and Papert, is still unsolved. Papert, in a recent paper (Minsky and Papert 1989), wrote '…the entire structure of recent connectionist theories might be built on quicksand: it is all based on toy-sized problems with no theoretical analysis to show that performance will be maintained when the models are scaled up to realistic size. The connectionist authors fail to read our work as a warning that networks, like brute force, scale very badly'. This is a warning not to be taken lightly. It is being met by various methods and devices: accelerator cards, ever faster and smaller *hardware devices*, and a deeper understanding of the theory behind neural computation. It is to be noted in this respect that accelerator cards may offer time saving and tractable training sessions on large databases but still may not help the convergence to significant solutions. It may be that the second neural network 'revolution' is only just beginning, but it is very clear that the scaling problem is in the forefront of researchers' minds. E1

## A1.1.5 Forms of networks and their training

In order to understand in more detail the way that greater strength is being brought to the subject of neural networks, it is important to point out the two extremes that now exist inside the discipline itself. At one end is the work of those mainly concerned with solving industrial problems. These include engineers, computer scientists, and people in the industrial sector. To them, neural computing is only one of a spectrum of adaptive information processing techniques. At the other extreme are those interested in understanding living systems, such as biologists, psychologists, and philosophers, together with mathematicians and physicists who are interested in the whole range of the subject as throwing up valuable and interesting new problems.

The styles of approach of the two extremes are somewhat different. The subject of artificial neural computing is based on networks, some of which have been mentioned earlier, which use the rather simple BDNs defined above. There are two extremes of the architectures of the networks: *feedforward networks* B2.3 (input streams steadily through the network from a set of input neurons to a set of output ones) and *recurrent networks* (where there is constant feedback from the neurons of the network to each other, as in B2.3 the Hopfield network mentioned earlier). This is mirrored in the differences between the topologies such networks possess; one is the line, and the other the circle, which cannot be topologically deformed into each other. As is to be expected, there are two extreme styles of computation in these networks. In the feedforward case the input moves through the network to become the output; in the recurrent network the activities in the network develop over time until it settles into some asymptotic value which is used as the output of the network. The network thus relaxes into this asymptotic state.

Network training can be classified into three sorts: *supervised*, *reinforcement* and *unsupervised*. The B3.1, C3 most popular of the first of these, backpropagation, has been mentioned earlier as the way to train neural networks to solve hard problems like parity, which needs hidden nodes (with no output that might be specified directly by the supervisor or teacher). It uses a set of training data which is assumed to be given, so that the (usually) feedforward network has a set of given inputs and outputs. When a given input is applied to the untrained network, the output is not expected to be the desired one, so that an error is obtained. That is used to assign changes, usually small ones, to the connection weights to all the neurons (including the hidden ones) in the network. This process of change is repeated many times until an acceptably low error level is obtained.

The second training method uses a reward given to the network by the environment on its response to a given input. This reward may also be used to determine modifications to the weights to achieve a maximum reward from the environment. Thus, this form of learning is ‘with a critic’, to be compared to supervised learning, which is ‘with a teacher’. Finally, there is unsupervised learning, which is closer to the style of learning in biological systems (although reinforcement learning also has strong biological roots). In this method correlations between signals are learned by increasing the connection weight between two neurons which are both active together.

At the other end of the subject of neural computation is investigation of nervous systems of the many species of animals, in an attempt to understand them. Since even a single living neuron is very complex, this approach does not aim for application in the marketplace, although simplified versions of mechanisms gleaned from this area of study are turning out to be of great value in commercial applications. This is true, for example, for models of the eye or ear, and also in the area of control, where reinforcement training (related to conditioned learning) has led to some very effective industrial control systems (White and Sofge 1992). The biological neural networks which are of interest are also extremely complex as nonlinear dynamical systems or mappings, although there is steady progress in their unraveling.

The most important lesson to be learned from these studies, besides the detailed network styles being used, is that the brain has developed a very powerful modular scheme for handling the scaling problem mentioned earlier. Exactly how this works is presently under extensive scrutiny, in particular, through the use of noninvasive techniques (EEG, MEG, PET, MRI). The causal chains of activations of various brain regions is being discovered as a subject performs a particular information processing task; the results are allowing more global models of the brain to be constructed.

### A1.1.6 Strengths of neural networks

In the face of the difficulties neural networks are still facing, of slow training, incompletely understood complexity and the highly nonlinear neural network system involved, as mentioned earlier, there are several features which will ensure the continued strength of the subject as a viable discipline.

Firstly, increases in computing power that were almost undreamed of several years ago, with gigabytes of memory and giga-interconnection updates per second. That may still be some way from the speed and power of the human brain. But if only specialized devices are to be developed, the total complexity of the human brain need not be a deterrent from attaining a lesser goal.

Secondly, there are developments in the theoretical understanding of neural networks that are impressive. Convergence of training schedules and their speed-up is presently under active investigation. The subject of dynamical systems theory is being brought to bear on these questions, and impressive results are being obtained. The use of concepts like attractor, stability, circle maps and so on are allowing a strong framework to be built for neural networks; in particular, the manner in which the dynamics of learning appears to display the general features of a sequence of phase transitions, as new features of the complexity of the training set are able to be discovered by the network, and new specialized feature detectors in the hidden layers emerge in the training process.

Thirdly, there are several different disciplines which are seen to have a great deal of overlap with neural networks. Thus the branch of statistics associated with regression analysis is now recognized as having been extended in an adaptive manner by the use of neural network representations of time series (Breiman 1994). Computer-intensive techniques, such as bootstrapping, are proving of great value in neural networks for tackling problems with small data sets. *Pattern recognition*, for example, also has important overlaps with the discipline in the areas of classification and *data compression*. Neural networks can extend these areas to give them an adaptability that is proving to be very important, such as in learning the most important features of a scene by means of adaptive principle component analysis (PCA) (Oja 1982). Statistical mechanics (especially spin glasses) has already been noted above as leading to important new insights into the problems of storage and response of neural networks. Machine learning is also of importance for the subject, and under the 'probably approximately correct' (PAC) approach has allowed the study of the complexity of neural networks needed to solve a given problem. B1.5, B6 F1.5

Fourthly, the field of function approximation has led to the important 'universal approximation theorem' (Hecht-Nielsen 1987, Hornik *et al* 1989). This theorem states that any suitably smooth function can be approximated arbitrarily closely by a neural network with only one hidden layer. The number of nodes required for such an approximation would be expected to increase without bound as the approximation was made increasingly better. The result is of the utmost importance to those who wish to apply neural networks to a particular problem; it states that a suitable network can always be found. This is also true for trajectories of patterns (Funahashi and Nakamura 1993).

There is a similar, but more extended result, for the learning of conditional probability distributions (Allen and Taylor 1994), where now the universal network has to have at least two layers to be able to have a smooth limit when the stochastic series being modeled becomes noise-free. Again, this is very important in the modeling by neural networks of *financial series* which have considerable stochasticity. G6.3

Fifthly, and already discussed briefly above, is the emerging subject of computational neuroscience. This attempts to create simple models of the neural systems which are important in controlling the response patterns of animals of a given species. This has a vast breadth, encompassing as it does the million or so species of living animals, culminating with man. It is a subject with vast implications for mankind, especially from the medical benefits that better understanding of brain processes would bring, both to those in the field of mental health and in the more general area of understanding of healthy living systems.

The field of computational neuroscience has led to useful devices by the route of 'reverse engineering'. In this, algorithms are developed for information processing based on simple models of the neural processing occurring in the living system. Thus it is not only the single neuron which is proving of value in reverse engineering, as it has already for the development of artificial neural networks (and where also it continues with the incorporation of increasingly complex neurons to achieve more powerful artificial neural networks). It is increasingly occurring in the reverse engineering of the overall architecture of artificial networks from that of living neural networks. This approach has also proved of value at the hardware level, as well as generating new styles of artificial neural computation. Thus, in the first category, is the work of Carver Mead and his colleagues at the California Institute of Technology in the United States (Mead 1989). They have built both a silicon retina and a silicon ear, using VLSI designs based on the known functions of these devices in living systems and their approximate wiring diagrams.

The retina has lateral inhibitory connections between the first (horizontal) layer of cells and the input cells, which leads to a very elegant method of reducing redundancy (say, in patches of constant illumination) of visual inputs. It is also possible to extend this modeling to later layers in the retina, and also to proceed further into the early layers of the visual cortex. The latter appears to use a decomposition of the input into some overcomplete set of functions, such as might arise from differences of Gaussians or similar functions with localized values. This leads into the field of wavelet transforms, another theoretical area proving to be of great value in developing new paradigms for neural networks (Szu and Hopper 1995).

The manner in which more global brain processing can be understood has been developed over the last few years by Teuvo Kohonen in the SOFM mentioned earlier (Kohonen 1982). In more detail, this algorithm is based on the idea of competition between nearby neurons, ending up in one neuron winning and the others being turned off by lateral inhibition from that winner. This winner is then trained by increasing the connection weights to it so that it gives a larger output. This means rotating the weights on the winning neuron so that they are more closely aligned to the input. The same is done for the neurons in a small region round the winner. If this is done repeatedly for a set of training inputs the network ends up representing the inputs in a topographic fashion over its surface (assuming the network is laid out in a two-dimensional fashion). If the inputs have features which are more than two dimensional then the resulting map may have folds in it; such discontinuities are seen, for example, in the map of rotation sensitivity for cells in the visual cortex.

One can search for other tricks that nature may use, and attempt to incorporate them into suitable machines. Thus there are presently attempts to build a ‘vision machine’ by means of the sensitive response of sets of coupled oscillators to their inputs. Yet again this also leads to some very important mathematical problems in understanding the response patterns of many physical systems.

It also leads to the more general question of whether or not it is possible to use the finer details of the temporal structure of neural activity. An extreme case of this is the use of information by coincidence of a number of nerve impulses impinging on a given cell. Suggestions of this sort have been around for a decade or more, but it is only recently that the improvement in computing power has allowed increasing numbers of simulations to test this idea.

As is well known, chaos and fractals are a key aspect of any physical phenomena. Will they prove to be of importance in improving neural networks? Some, especially Walter Freeman (1995) from Berkeley in connection with olfaction, suggest that such is the case, and that strange attractors may be used to give a very effective method of searching through, or giving access to, a large region of the state space of a neural network. That possibility has not yet been achieved in detail; however, see Quoy *et al* (1995) for an interesting attempt to achieve a useful speed-up by ‘living on the edge of chaos’ for a neural network. But the question is an important one and again indicates the breadth of possibilities now coming under the banner of neural networks.

### A1.1.7 Hybrids and the future

From what has been sketched above about the past and some of the avenues being explored in the present for neural networks, it is clear that the subject now has such breadth and depth that it is unlikely to run out of steam as it did earlier. Indeed, it is becoming increasingly clear that artificial neural networks (ANNs) can be seen to be one of a number of similar tools in the tool-kit of anyone tackling problems in information processing. Along with *genetic algorithms*, *fuzzy logic*, belief networks, and other areas D2, D1 (such as parallel computing), ANNs are to be used either on their own or in hybrid systems wherever and however is most appropriate. The past divisions, noted above as having existed between different branches of information processing, seem to have been removed by these developments. Moreover, new techniques are being developed to allow the parallel use of these various technologies, or even better, in a manner that allows them to help each other. Thus genetic algorithms are being used to help improve the architecture of a neural network, where the fitness function used to select better descendants at each stage of the generation process is the error on the training set (in the case of a supervised learning problem). Similarly, it has proved of value to obtain help from fuzzy logic to allow for rough initial settings of the weights in a network.

There are some general rules for determining when a neural network is most appropriate for a particular task, compared with one of the other methods mentioned earlier. If the data are noisy, if there are no rules for the decisions or response that are required, or if the training and response must be rapid (something missing from genetic algorithms, for example), then ANNs may be the best bet. It is also necessary to comment finally on the present situation in the relation between ANNs and AI mentioned earlier. As noted above for other adaptive techniques, the move is now to combine an ANN solution for part of a problem with results obtained from a knowledge-based expert system (KBES). That has been done successfully in *speech recognition*, where the Kohonen network mentioned earlier is good for individual phoneme F1.7.2, G1.4 recognition, but not so good for words (due to difficulty in incorporating context into the ANN). A KBES approach, with about 20 000 expert rules, then allows the total system to be far more effective. Similar greater efficiency can also be obtained using hybrid systems with *time-delayed neural networks* (which B4.10.2, C1.2.8 involve inputs that are delayed or lagged relatively to each other, so as to cover a spread of input times).

It is clear that a more realistic and effective approach is arising in the relationship between the different branches of information processing. Undoubtedly this use of the best of all possible worlds will increase. But at the same time the neural network approach, in the context of obtaining a better understanding of the human brain, will also give ever increasing powers to the ANN approach. In the end one can only see that as being the most effective (provided there is the computing power) method for many of the deeper problems facing the information industry. Nor is there any serious alternative to the further development of neural network models of ourselves to understand the higher levels of human cognition, including human consciousness.

## References

Allen D W and Taylor J G 1994 Learning time series by neural networks *Proc. Int. Conf. on Artificial Neural Networks (Sorrento, Italy, 1994)* ed M Marinaro and P Morasso (Berlin: Springer) pp 529–32

Amari S 1991 Dualistic geometry of the manifold of higher-order neurons *Neural Networks* **4** 443–51

Amit D 1989 *Models of Brain Function* (Cambridge: Cambridge University Press)

Breiman L 1994 Bagging predictors *UCLA Preprint* (unpublished)

Cohen M A and Grossberg S 1983 Absolute stability of global pattern formation and parallel memory storage by competitive neural networks *IEEE Trans. Syst. Man Cybern.* **13** 815–26

Coombes S and Taylor J G 1993 Using generalised principal component analysis to achieve associative memory in a Hopfield net *Network* **5** 75–88

Freeman W 1995 *Society of Brains* (Hillsdale, NJ: Erlbaum)

Funahashi K and Nakamura Y 1993 Approximation of dynamical systems by continuous time recurrent neural networks *Neural Networks* **6** 801–6

Gardner E 1988 The space of interactions in neural network models *J. Phys. A: Math. Gen.* **21** 257–70

Grossberg S 1976 Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors *Biol. Cybern.* **23** 121–34

Hecht-Nielsen R 1987 Kolmogorov's mapping neural network existence theorem *Proc. Int. Conf. on Neural Networks III* (New York: IEEE) pp 11–13

Hinton G and Sejnowski T 1983 Optimal perceptual inference *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (Washington)* (New York: IEEE) pp 448–53

Hopfield J 1982 Neural networks and physical systems with emergent collective computational properties *Proc. Natl Acad. Sci., USA* **81** 3088–92

Hornik K, Stinchcombe M and White H 1989 Multi-layer feedforward networks are universal approximators *Neural Networks* **2** 359–66

Kohonen T 1982 Self-organised formation of topologically correct feature maps *Biol. Cybern.* **43** 56–69

LeCun Y 1985 Une procédure d'apprentissage pour réseau á seuil asymetrique *Cognitiva* **85** (Paris: CESTA) pp 599–604

McCulloch W S and Pitts W 1943 A logical calculus of ideas immanent in nervous activity *Bull. Math. Biophys.* **5** 115–33

Mead C 1989 *Analogue VLSI and Neural Systems* (Reading, MA: Addison-Wesley)

Minsky M and Papert S 1969 *Perceptrons* (Boston, MA: MIT Press)

——1989 *Perceptrons* 2nd edn (Boston, MA: MIT Press)

Oja E 1982 A simplified neuron model as a principal component analyser *J. Math. Biol.* **15** 61–8

Parker D B 1985 Learning logic *Technical Report TR-47* Center for Computational Research in Economics and Management Science, Massachusetts Institute of Technology, Cambridge, MA

Quoy M, Doyon B and Samuelides M 1995 Dimension reduction by learning in a discrete time chaotic neural network *Proc. World Congr. on Neural Networks (1995)* (Washington: INNS) pp I-300-303

Ritter H, Martinetz T and Schulten K 1991 Neural computation and self-organising maps (Reading, MA: Addison-Wesley)

Rosenblatt F 1962 *Principles of Neurodynamics* (New York: Spartan)

Rumelhart D E and McClelland J L 1986 *Parallel Distributed Processing* (Boston, MA: MIT Press)

Szu H and Hopper T 1995 Wavelets as preprocessors for neural networks Plenary Talk *Proc. World Congr. on Neural Networks (Washington, DC, 1995)* (Washington: INNS); Kohonen T 1995 Plenary Talk *Proc. World Congr. on Neural Networks (Washington, DC, 1995)* (Washington: INNS)

Werbos P 1974 Beyond regression *PhD Thesis* Harvard University

White D A and Sofge D A (eds) 1992 *Handbook of Intelligent Control* (New York: Van Nostrand Reinhold)

Willshaw D J and von der Malsburg C 1976 How patterned neural connections can be set up by self-organisation *Proc. R. Soc.* B **194** 431–45