The Pattern Extraction Hierarchy Architecture: a Connectionist Alternative to the von Neumann Architecture, in Biological and Artificial Computation: from Neuroscience to Technology, eds. Mira., J., Morenzo-Diaz, R., and Cabestanz, J., (1997), pp 634-43, Berlin: Springer.

# The Pattern Extraction Architecture: a Connectionist Alternative to the Von Neumann Architecture

**L. Andrew Coward**

**7518 Cottonwood Court, Garland, Texas 75044**

**andrew_coward@nt.com**

**Abstract**

A detailed connectionist architecture is described which is capable of relating psychological behavior to the functioning of neurons and neurochemicals. The need to be able to build, repair and modify current electronic systems with billions of hardware components has been met through a seldom appreciated aspect of the von Neumann architecture: the hardware architecture is compatible with a simple functional architecture which can support precise translation between functional descriptions at many levels of detail down to the individual hardware components, through the use of a common functional element, the instruction. Existing neural network models have been developed to simulate different aspects of cognition, but do not offer a behavioral architecture analogous with the von Neumann architecture. The brain has experienced intense evolutionary selection pressures analogous with the requirements to build, repair and modify electronic systems. These pressures have resulted in a neural architecture which is compatible with a simple functional architecture based on the use of the common functional element, the pattern extraction. The pattern extraction functional architecture is the basis for intellectual understanding of brain functioning. Physiological structures can be understood as an efficient partitioning of function within the constraints imposed by the properties of neurons. Behavioral phenomena such as declarative memory can be understood in neuron terms, including the reasons for distribution of memory traces. A range of physiological and psychological evidence is discussed. Electronic simulation demonstrates that key psychological functions can be emulated by the architecture.

## Introduction

A range of neural networks have been proposed as models for different aspects of biological brain operation. These models employ the perceptron to model the neuron. Information is coded in the connection weights of inputs to the perceptrons, and learning proceeds by gradual adjustment of the weights.

There are several objections to these models. Firstly they require biologically implausible control signals or prior knowledge of the learning task for successful learning. Even Kohonen's topological map scheme for unsupervised learning (Kohonen 1988) has been criticized on the basis that the patterns which are presented to the model must be preselected (Pfeiffer 1996). Secondly, gradual adjustment of weights has difficulty in accounting for permanent, instantaneously created declarative memory.

The most fundamental objection is the gap between high level behavioral descriptions and the problems addressed by neural networks, which tend to be the modeling of sensory or motor functions etc. There has been little work addressing the constraints imposed on the functional partitioning of behavior by the use of perceptron type models. Such work would lead to functional separations which could be compared with physiological structure. Kohonen (1995) has written "Some researchers [have] the goal ... to develop autonomous robots; accordingly, the main functions to be implemented ... are sensory functions, motor functions, decision making, verbal behavior." This list of functions may or may not be a plausible partitioning of the operation of an "autonomous robot" given the use of perceptron models of neurons.

## Constraints on Possible System Architectures

Electronic systems are operational today which contain billions of individual hardware elements. The largest telecommunications switches, able to handle 100 thousand telephone subscribers, contain close to 5 billion transistors.

In such systems, the relationship between a high level function (for example, creating a conference connection between three telephones) and the operation of individual transistors can be very complex. However, there are three factors which require a relatively simple hierarchy of relationship between descriptions in terms of function and descriptions in terms of component states. These factors are the need to build many copies of a system, the need to repair component failures, and the need to add features.

To illustrate the requirement, imagine a system which had been created by repeated cycles of random connection of randomly selected transistors, followed by test, until a system was found which performed as desired. The only way to build a copy of such a system would be to duplicate it component by component, connection by connection, simplification like "do the following x times" would be rare. Furthermore, slight differences between the original component and a copy might have large, unpredictable effects on system function. Failures, experienced as a loss of a system function, could not be easily related to the individual component which failed. Adding a new feature (for example, making it possible to have conference connections between more than three telephones) would not be achievable with a simple set of component changes.

The von Neumann architecture is the basis for almost all commercial systems. It has a capability to support the construction of a hierarchy of functional descriptions at many different levels of detail. These descriptions are precisely translatable between levels, down to the level of the individual hardware components used. They are capable of being partitioned in different ways within a level to allow different system functions to depend in different ways on a common set of more detailed building blocks. In addition, these descriptions can be mapped into the physical partitioning of the hardware functionality and into the logical partitioning of the data used by the system.

There are in fact three interlocking architectures in a complex system. The hardware architecture partitions the system into physically separate functions. The data architecture partitions representations of the external world into elements which can be considered separately. The functional architecture is the hierarchy of functional descriptions which makes it possible to identify the individual components associated with system functions, and vice versa. Failure to maintain a simple functional architecture results in systems which are very difficult to build, test, repair, or modify.

The requirement for precise translation and flexible partitioning is severe, and a major factor in the ability of von Neumann based systems to meet the requirement is the use of the instruction concept. Functional descriptions are built at every level from instances of this concept. This common base allows translatability and permits flexible partitioning. A limitation is that the implicit time sequence of descriptions leads to considerable difficulty in designing parallel processing systems.

Analogous pressures to those which lead to simple functional architectures in electronic systems apply to the brain. There are requirements to construct many copies from DNA 'blueprints', to recover from damage, and to add features by a process which only allows simple mutations. These requirements have given an immense evolutionary advantage to brains with a simple functional architecture.

What is the neural architecture of which this functional architecture is a key component? The first step in answering this question is to recognize that a condition for the existence of life is the existence of repetition in the environment. A behavior found to be successful in a given combination of environmental conditions is repeated when the combination of conditions is perceived to repeat. A combination of environmental conditions which repeats is a pattern, and the fundamental role of repetition has resulted in a neural architecture in which pattern extraction plays a similar role to that of instruction in von Neumann systems. The simple functional architecture based upon pattern extraction, although it has emerged as a result of natural pressures, can be the basis for intellectual understanding of the brain, because it provides a simple way to relate psychological behavior to functional descriptions in terms of neurons and neurochemicals. The neural architecture was originally described in Coward 1990. The work reported here is a refinement of that architecture.

## Neuron Models

Standard neural networks use neuron models of the perceptron type. Such models have a range of states (not firing, firing at a range of rates) which can be communicated to the rest of the network via connections to other neurons. These connections stimulate or inhibit the firing of their targets to different degrees (or weights). The next firing state of a neuron is determined by its current state and by the ratio of the weighted sum of its active inputs to a fixed, internally determined threshold.

The information content of such networks is coded in the input weights, and learning proceeds by slight modification of weights. There is a layering of neurons, with a layer receiving external inputs separated from a layer generating external outputs by one or more 'hidden' layers. A layer receives the bulk of its inputs from the immediately preceding layer.

An additional neuron model introduced in Coward 1990 has inputs of differing weights from large numbers of other neurons. Neurons are in a passive (or 'virgin') state until a single, unique learning event. This learning event is triggered by a high level of firing of regular neurons in the neighborhood of the virgin neuron but no output firing from the neighborhood. The combination of these conditions lowers thresholds in all local virgin neurons until enough virgin neurons fire to generate an output from the neighborhood. Learning occurs by a large, instantaneous, permanent weakening of the inactive input connections to any virgin neurons which fired, and the setting of the threshold of the neuron at a level which will cause it to fire in the future if a similar combination of active inputs occurs. This mechanism provides a permanent record of a pattern extracted at a particular instant from a perceived object.

New patterns become part of the set which can trigger learning in the same neighborhood in the future. Category learning by similarity starting from a single example is therefore possible as described in more detail later. At that point the neighborhoods will be referred to as modules. While the relative weights of the imprinted inputs remains constant, it is possible for some level of addition of relevant inputs to occur later as described in Coward 1996. This addition plays a role in the management of learning.

Learning is successful with random inputs to virgin neurons, provided the number of inputs is large enough (Coward 1996). Fewer resources are used if there is a statistical bias in favor of inputs which frequently fire regular neurons at the same level in the same module. Coward (1990) therefore proposed that the function of REM sleep is to impose this statistical bias. A rerun of past neuron firing, with a bias towards the recent past as the best available indicator of the immediate future, allows virgin neurons to accept inputs from axons which frequently fire regular neurons at their level in the module. The learning process is thus a variant of the Hebb mechanism, but taking place in two functionally separate stages.

One additional feature of these neurons is that global parameters can lower the threshold of all regular neurons within a particular set of modules. This lowering increases the strength of outputs from modules within the set and, as discussed later, plays a role in global modulation of behavior. Some perceptron type neurons exhibit the same mechanism for similar reasons.

## Pattern Extraction Hierarchy Architectural Description

*Data Architecture*

A data architecture defines the partitioning and flow of information about the external world within a system. In general the data architecture for the pattern extraction hierarchy architecture is similar to the data architecture implied in standard neural networks and is illustrated in figure 1. The primary data connectivity is from one neuron layer to the next, but lateral connections play major functional roles (see below). A neuron can be regarded as being programmed with a pattern which is the combination of the set of patterns

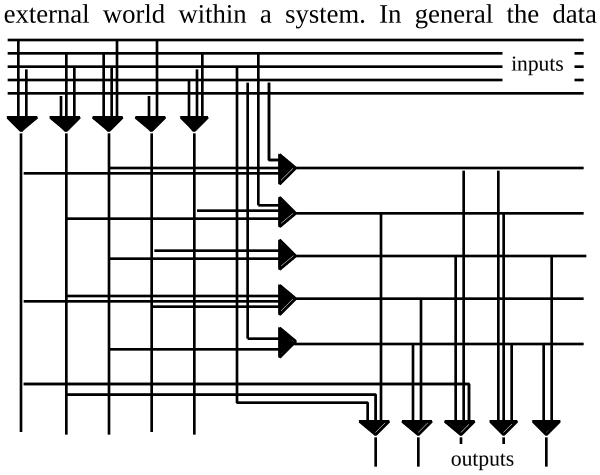


**Figure 1. Data Architecture**

extracted at the previous level by its inputs. The input weights indicate the relative importance of each input pattern. The firing of a neuron indicates the presence of a high enough proportion of its currently programmed pattern to require attention under current conditions, with the rate of firing indicating the level of attention required. The proportion is the current neuron threshold.

Patterns close to sensory input can be functionally interpreted as simple sensory characteristics of objects or internal body conditions, while deeper in the hierarchy the patterns are complex combinations of characteristics, or relationships between objects. Deeper still the patterns acquire the functional role of behavioral recommendations (' *this* is present, therefore do *that* '). Patterns evolve by weight adjustment or by imprinting; the mechanism depends on the functional domain in which the neuron is located.

*Functional Architecture*

The functional architecture is shown in figure 2. Sensory preprocessing extracts patterns from raw input. These sensory patterns are stable with respect to external objects (an example would be object color independent of illumination) and are generally established by the imprinting mechanism in the developmental phase. The massive cell death in that period is associated with the elimination of imprinted patterns which are determined not to repeat over long periods and are therefore not object specific.

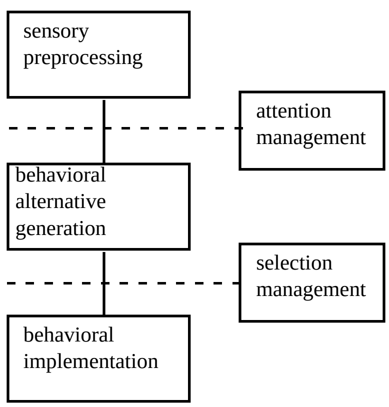


**Figure 2. Pattern Extraction High Level Functional Architecture**

Attention management allows the set of sensory pattern associated with a single object to proceed to behavioral alternative generation. This function generates a set of alternative responses to the object. Selection management selects a consistent set of behaviors to proceed to behavioral implementation, where a portfolio of muscle movements are managed to produce the selected behavior.

As discussed below, the imprinting model applies in behavioral alternative generation, and in sensory preprocessing particularly during early development. A more perceptron like model applies in behavioral implementation and in the management functions.

The power of the architecture will now be illustrated by descending through several layers of descriptive detail to the neuron operational levels in two areas: behavioral alternative generation and selection management.

The next level of detail for behavioral alternative generation is shown in figure 3. The function contains a number of regions, each of which generates behavioral recommendations of a different type. Thus in response to perceiving a dog, alternative recommendations might be 'kick the dog' from the aggressive region, 'avoid the dog' from a fearful region, 'pat the dog' from a cooperative region, 'say "dog" ' from a speech generating region, and 'focus attention on the dog' from

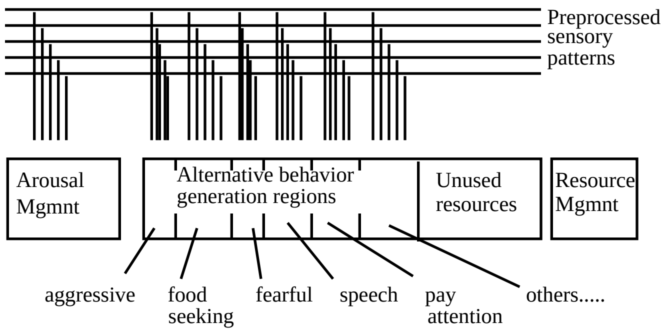


**Figure 3. Functional Model for Behavior Generation**

an attention management region. These alternatives can be visualized as parallel cascades of neuron firing all driven by the same set of sensory patterns. Arousal management uses a subset of the preprocessed sensory patterns, including patterns extracted from internal physiological states, to control the probability of a strong cascade emerging from different regions. Thus anger lowers the threshold of all neurons in the aggressive region and increases the strength of any cascade. Hunger lowers neuron thresholds in a food seeking region. One feedback route is that an aggressive recommendation including a recommendation to become more angry. Generation of behavioral recommendations consumes neuron resources. Resource management assigns more resources in response to the extraction of a resource-depleted pattern . This management function takes the form of a resource map with an active boundary, and resources are assigned from the unused resources corresponding with the current position of the active boundary (Coward 1990).

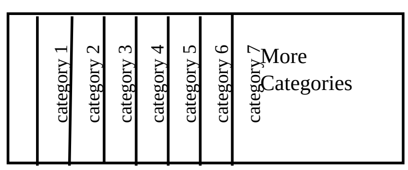


**Figure 4. Functional Model: Behavior Generation Region**

The next level of detail is generation of behavioral alternatives within a region as shown in figure 4. A category module extracts the presence of an object similar to an existing category, or can be created in response to an object which has no such similarity. Detailed category output can be unique to an individual object in a particular situation since patterns extracted from the individual object and situation will be included in the neurons imprinted to generate the output. This uniqueness allows control of the particular type of (for example) aggressive action to be taken. Some categories may also input to other categories: a leg may be an object/action recommendation in its own right or a component pattern to a dog category.

The next level of detail is the internal structure of a category module as shown in figure 5. The memory trace of, for example, a new dog, consists of patterns extracted from previously perceived dogs plus a small set extracted from the new dog. This trace exists within all the behavioral generation regions which generated a recommendation, each such region therefore contains an independent module which can be regarded as a 'dog recognizing' module. The new patterns in each module add to that module's definition of 'dog'. An unknown animal would not create a cascade in any existing module strong enough to produce an output or trigger imprinting. This condition triggers assignment of a new, randomly connected module in which imprinting occurs until an output results. This animal thus becomes the initial member of a new category. Experimental connections are made to implementation, which will be evolved by pleasure and pain as discussed later. The new category will be evolved by perception of any similar animals in the future. In practice provision must be made for merging duplicate categories within a region when, for example, the two animals of a particular type first seen are extremes within the type (Coward 1996). The same reference describes electronic simulation demonstrating that imprinting within the proposed architecture is an effective means of sorting and recording perceived objects in categories on the basis of similarity, with no guidance or feedback of any kind, and that dream sleep significantly improves the efficiency of the process.

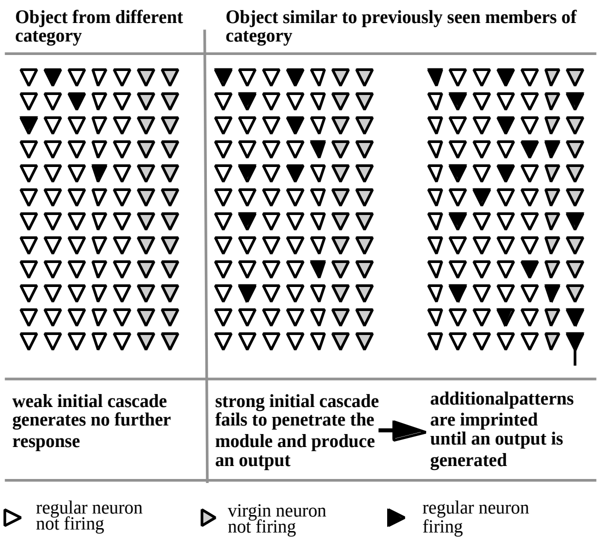


**Figure 5. Learning within a Category Module**

Turning now to the more detailed descriptions of the selection management function as shown in figure 6, the alternative recommendations are directed to the selection management function as well as towards behavioral implementation, but are blocked in the latter direction. Each alternative enters a pipe within selection management which is specific to the type of behavior. Competition between the active pipes in general results in an output from one (or no) pipe. An output opens the gate to allow the recommendation which provided the successful input to proceed towards behavioral implementation. At the next level of detail shown in figure 7, there is cross inhibition between all pipes. A recommendation in one pipe inhibits all other pipes. There are two control functions. One modulates total activity: if the cross inhibition process is resulting in more than one recommendation

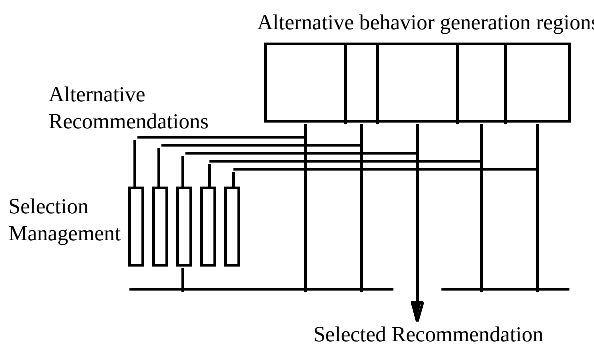


**Figure 6. Selection Management High Level Functional Architecture**

getting close to output, all recommendations are reduced in strength until a single clear recommendation emerges, or none. Faults in this mechanism would lead to rapid, erratic shifts in behavior as conflicts are resolved at the muscle control level. The other control function is pleasure/pain. If patterns of success are extracted from the result of a behavior, pleasure is triggered which weakens all recently active inhibitive connections in the successful pipe. Pain strengthens such connections. The effect is to modulate the probability of future acceptance of similar behavioral recommendations.

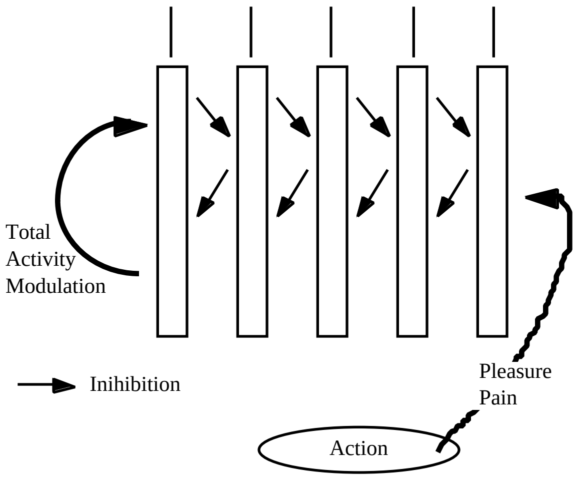


**Figure 7. Selection Management Detailed Functional Architecture**

At a detailed level, perceptron type neurons are appropriate for the function, except that their thresholds must be increased or decreased in accordance with the activity modulation function, and the learning mechanism does not need a detailed feedback of an expected result. Connections from layer to layer within a pipe are stimulative.

A simple simulation of this function has demonstrated that because output from a category in the behavioral generation function is generally unique to the individual object, a simple pain mechanism acting on provisional connections to behavioral implementation eliminates categorization errors (Coward 1996). Such categorization errors could be of the type identifying a “dog-like cat” incorrectly as a dog.

*Physiological Architecture*

The mapping of functional architecture into brain physiology is shown in figure 8. The thalamus attention gateway uses the primal sketches described by Marr (1980) to compete for access of all the sensory input within the domain defined by the primal sketch to the cortex generating recommendations for bahavior (Coward 1997). The mechanisms described by Taylor et al (1993) form a plausible detailed

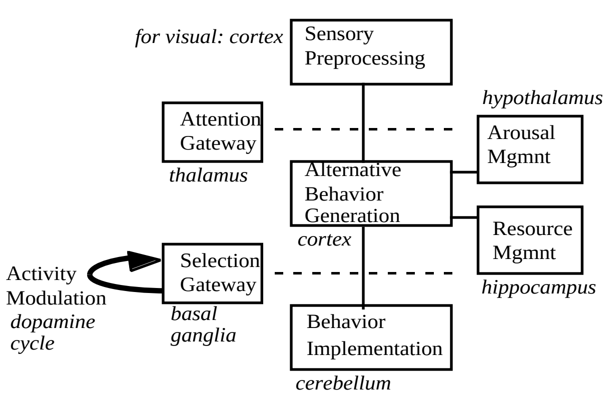


**Figure 8. Physiological Architecture**

description of this function. The role of the hippocampus as a map of cortex resources assigning them to cortex modules as required results in the development of an implicit time sequence of memory. Damage to the hippocampus or to structures linking it to the cortex can therefore result in time related deficits such as the loss of a time slice of memory and the inability to create new memories observed in Korsakov's syndrome (Coward 1990). Defective operation of the dopamine feedback loop in the basal ganglia results in the rapid, erratic behavioral shifts observed in Tourette's syndrome (Coward 1990).

## Physiological and Psychological Evidence

Some of the most persuasive evidence is the way in which the pattern extraction hierarchy architecture makes it possible to create a systematic, integrated description of an extensive range of psychological and physiological evidence (Coward 1990). A wide range of detailed pattern management mechanisms can be expected at the neuron level, but there is a clear requirement for an imprinting mechanism triggered by firing of other neurons in the neighborhood for neurons in the cortex and the hippocampus to account for declarative memory.

Permanent, instantaneous weakening of the weights of all inactive inputs could be followed by elimination of the synapses. There is some evidence that learning by initial excess and subsequent reduction of neuron connectivity occurs in the human cortex (Huttenlocher et alii. 1982). Synaptic density has been observed to be greater in rats raised in a stimulating environment (Turner et al. 1988) and there is evidence of a higher proportion of synapses in cortex regions undergoing active learning (Greenough et alii 1988). There is also evidence that synaptic plasticity depends on the activity of nearby neurons (Fazeli 1992).

At a structural level, the observation of a predominance of local connectivity resulting in a column like appearance in the cortex in some areas is as expected for category modules. The observation that long range connectivity in the cortex is stimulative but a proportion of local connectivity is inhibitive (Douglas and Martin 1991) is also as expected. One role of the local inhibitive connectivity is inhibition of imprinting if a module output is being produced. Observation of "multiple, parallel, segregated circuits ... [which] receive inputs from several separate areas, traverse specific portions of the basal ganglia and thalamus and project back upon one of the cortical areas providing input" (Parent and Hazrati 1995) is the structure predicted by the architecture.

At a phenomenological level, the architecture provides a straightforward explanation for the inability of local damage to remove declarative memory, and its impact on motor abilities and personality. The memory trace of a particular encounter with, say, a dog, is distributed over all the regions in which a category module produced a behavioral recommendation and distributed within each module. Local cortex damage would generally be within one region of one behavioral type, leading to personality shift and/or motor deficits, but to eliminate the total memory trace of an event, damage would have to occur to similar modules in multiple regions.

The proposed role of REM sleep in memory would predict (Coward 1990) that REM sleep deprivation coupled with major activity in a particular domain should result in behavior atypical of that domain. This prediction has not been tested experimentally, but anecdotal evidence that intractable labor or political disputes are sometimes settled under conditions of protracted negotiation with sleep deprivation is in agreement with the prediction.

## Conclusions

Evolutionary pressure has resulted in a neural architecture with a relatively simple functional architecture based upon the concept of pattern extraction. This functional architecture can be the basis for intellectual understanding of the brain, relating psychological phenomena to neural physiology. The same architecture can also be the basis for construction of an electronic system which would exhibit equivalent phenomenology to the human brain.

## References


Coward, L. A. (1990), Pattern Thinking, New York: Praeger.

Coward, L.A.(1996), 'Understanding of Consciousness through Application of Techniques for Design of Extremely Complex Electronic Systems' Towards a Science of Consciousness , Tucson, Arizona.

Coward, L.A.(1997), 'An Integrated Theory of the Phenomenal Content of Consciousness', to be published.

Douglas, R.J. and Martin, K.A. (1991), 'A Functional Microcircuit for Cat Visual Cortex', Journal of Physiology, 440, 735-769.

Fazeli, M.S. (1992), Synaptic Plasticity: on the trail of the retrograde messenger, Trends in Neuroscience 15, 4, 115-7.

Greenough, W.T., Hwang, H.M. and Gorman C. (1988), 'Evidence for Active Synapse Formation, or Altered Postsynaptic Metabolism, in Visual Cortex of Rats raised in Complex Environments', Proceedings of the National Academy of Sciences USA .

Huttenlocher, P.R., de Courtin, C., Gary, L.J. and van der Loos, H (1982), 'Synaptogenesis in the Human Visual Cortex - evidence for synapse elimination during ordinary development', Neurosciences Letters 33, 247-52.

Kohonen, T. (1988), Self Organization and Associative Memory, Berlin: Springer-Verlag.

Kohonen, T (1995), Self-Organizing Maps, Berlin: Springer-Verlag.

Marr, D. (1982), Vision , New York: W.H. Freeman.

Pfeifer (1996), Symbols, Patterns and Behavior: Beyond the Information Processing Metaphor, in Encyclopedia of Microcomputers, 17, 253-75, New York: Marcel Dekker.

Taylor, J.G. and Alavi, F.N., (1993), 'Mathematical Analysis of a Competitive Network for Attention' in Mathematical Approaches to Neural Networks ed. J.G. Taylor (Elsevier).

Turner, A.M. and Greenough, W.T. (1985), 'Differential Rearing Effects on Rat Visual Cortex Synapses I: Synaptic and Neuronal Density and Synapses per Neuron', Brain Research 329, 195-203.