Birth Of A Learning Law

Stephen Grossberg‡

Department of Cognitive and Neural Systems
and
Center for Adaptive Systems
Boston University
677 Beacon Street
Boston, MA 02215†

October, 1997

Technical Report CAS/CNS–TR–97–017
Boston, MA: Boston University

Submitted to the *INNS Newsletter*

Send requests for reprints to:
Professor Stephen Grossberg
Department of Cognitive and Neural Systems
Boston University
677 Beacon Street, Room 201
Boston, MA

‡Supported in part by the Defense Advanced Research Projects Agency and the Office of Naval Research (ONR N00014-95-1-0409, ONR N00014-95-1-0657 and ONR N00014-92-J-1309).

†Acknowledgments: The author wishes to thank Robin Amos for his valuable assistance in the preparation of the manuscript.

Today we use the basic equations of neural networks with such familiarity that we often don't think about the many assumptions that go into them. This may be due to the much greater acceptance of neural networks now than when these equations were introduced 20 to 40 years ago, when they represented a radical paradigm shift. Having had the privilege of introducing several of these equations into the literature, I thought it might be of interest to review the conceptual foundations on which they are based, partly because the full implications of these hypotheses have still not been fully exploited.

One such equation is the learning law for an adaptive weight or long-term memory trace *z(t)* (also sometimes denoted by *w(t)* or *m(t)):*

$$dz/dt = f(x)[-Az + g(y)],$$

where x is the activity of a presynaptic (or postsynaptic) cell, y is the activity of a postsynaptic (or presynaptic) cell, and z is the adaptive weight at the synapse of an intervening pathway, or axon. This apparently simple equation was, for example, used to introduce Self-Organizing Maps (Grossberg, 1976a, 1976b, 1978; Kohonen, 1982, 1984), Adaptive Resonance Theory (Carpenter and Grossberg, 1987; Grossberg, 1976c, 1980), and Counter Propagation networks (Grossberg, 1976b; Hecht-Nielsen, 1987), among other models, and variants of it have been used to model neurophysiological data about the hippocampus and visual cortex (e.g., Artola and Singer, 1993; Levy, Brassel, and Moore, 1983; Levy and Desmond, 1985; Singer, 1983), among other structures.

Although the equation is mathematically simple, it represented a radical break with a number of traditions in psychology, neuroscience, and artificial intelligence when I first derived it as a student in 1958. The size of this break helps to explain why the equation, and variants thereof, took a decade to get published in the scientific literature (e.g., Grossberg, 1969a, 1969b, 1972c, 1974) — but that it another story — and why it took more than a decade more to start being used frequently in the neural network literature, where it goes by such varied names as the outstar learning law, the instar learning law, the gated steepest descent law, Grossberg learning, Kohonen learning, and mixed Hebbian/anti-Hebbian learning. I will use the mathematically most descriptive term, *gated steepest descent*, to discuss it below.

At least four major hypotheses, or research themes, are embodied by the gated steepest descent law:

l. Real-Time Processing. When gated steepest descent was first introduced, the very fact that it was represented by a differential equation was considered highly controversial. This was so, in part, because most researchers of mind, brain, and artificial intelligence were not familiar with thinking about how learning evolves in an individual learning subject (or system) moment-by-moment in real time. Resistance from many experimentalists was great because they were not comfortable thinking in terms of models at all, let alone models that used mathematics in a substantive way. Such resistance still exists today, but with the advent of the connectionist cognitive science and computational neuroscience movements, it has abated in the intervening years. I have elsewhere written about some of the historical factors that, I believe, contributed to the ferocity of this resistance by experimentalists for almost a century (e.g., Grossberg, 1980, Section 1; Grossberg, 1982b, Introduction; Grossberg 1988, Sections 1–5).

Resistance was particularly great from practitioners of Artificial Intelligence, whose emphasis on symbolic processing by a serial computer was carried to an almost religious fervor in those days. Many neural network researchers are now aware how the opposition of the AI researchers Marvin Minsky and Seymour Papert to Perceptrons inhibited the development and acceptance of all neural network ideas for two decades. These AI researchers did not understand that neural networks could be used to learn *how* symbols form using a continuous description of the learning process. By now, AI has also begun to incorporate neural networks into its research program, and various new algorithms synthesize concepts from neural networks, expert production systems, and even fuzzy logic. One such algorithm is Fuzzy ARTMAP (Carpenter *et al*., 1992). It is thus no longer possible to credibly argue that neural networks can just do A and expert production systems can just do B, where A and B have disjoint capabilities.

Resistance was also great from other behavioral and neural modelers. In psychology, the reigning learning model at the time was the Stimulus Sampling Theory of William Estes (e.g., Atkinson and Estes, 1963). This theory used finite Markov chains to describe learning as a process by which abstract stimulus features are transformed from an unlearned state to a learned state by a "stimulus sampling operation". Here the emphasis was on statistical data from groups of subjects learning in discrete time, and not on learning by an individual subject in real time. This movement ultimately hit a brick wall because of this limited perspective. Many respected neural modelers also used discrete time models, particularly those whose models grew out of linear algebra, as in the early models of Jim Anderson and Teuvo Kohonen.

In my own work, a real-time network framework was identified, after an intense intellectual struggle, to explain data about human and animal learning. One early analysis focused upon how associations are learned among events that occur sequentially in time, a process that is of equal importance in biology and technology. Even in the 1950's, there were plentiful data about how such learning occurs in humans and animals. I was particularly drawn to the bowed serial position curve of human verbal learning and animal discrimination learning. This bow says that, when a list of events is learned under appropriate conditions, associations near the beginning and end of the list are easiest to learn, and those near the middle are hardest to learn, much as we can often more easily remember how sequences of events begin and end, whereas the details in the middle may be muddled. Incorrect learned associations near the beginning of the list tend to be learned in the forward direction in time, whereas incorrect associations near the end of the list tend to be learned in the backward direction in time (!). In addition, the distribution of these learning errors depends upon the rate with which the events are presented.

Analyzing how "time" could run both forwards and backwards, and how learned errors could be distributed across many list events in both the forwards and backwards directions forced me into a continuous time description of learning dynamics within a neural network (e.g., Grossberg, 1969c; Grossberg and Pepe, 1971), which represented a radical break with classical ideas about the internal representation of space and time; see Grossberg (1974, Section II; 1982a) for reviews. At about the same time, I realized that same laws could also explain a lot of data about animal learning, both classical (or Pavlovian) learning and instrumental (or Skinnerian) learning (e.g., Grossberg, 1971, 1972a, 1972b, 1975). Both types of learning, moreover, used variants of gated steepest descent.

The fact that both cognitive data from humans and reinforcement data from animals could be explained by the same set of laws motivated me to derive them from general principles; e.g., Grossberg (1972b; 1974, Section II). I called this model the Additive Model because it adds up nonlinear contributions to the neuron activity. The Additive Model was used for many applications since the 1960's, including the introduction of Self-Organizing Maps (see below). Because new contributors to the field often entered it without ever reading its formative literature, they sometimes misname this model the "continuous Hopfield model", after Hopfield's first use of it in 1984 (Hopfield, 1984), despite the fact that the Liapunov function for this model had also earlier been discovered and generalized (e.g., Cohen and Grossberg, 1983) based on global Liapunov methods that I introduced in the 1970's; see Grossberg (1988, Section 9) for a review. I therefore think that the term Additive Model best reflects the history leading to this model.

Many neural modelers even today have not squarely faced key issues about real-time learning. In particular, real-time learning laws cannot achieve their maximum power unless they are embedded within architectures which enable learning to remain stable through time, free from catastrophic forgetting, particularly when the amount of data becomes large and can change its statistical properties through time. The popular and useful Back Propagation and Self-Organizing Map models do not have this property. In my own work, this realization came forcefully to me in the early 1970's through the following events. In the 1960's and early 1970's, I showed that one needed to combine associative and competitive processes to get good real-time learning results; e.g., Grossberg (1972c). Even this was highly nonobvious in those days. von der Malsburg then adapted the Additive Model that I used in Grossberg (1972c) to introduce the first Self-Organizing Map (von der Malsburg, 1973). This was, of course, a major contribution. As I have reviewed elsewhere (e.g., Grossberg, 1987), Malsburg's model was, however, neither real-time nor local. Based on my earlier theorems about associative learning and on-center off-surround feedback networks, I was able to show in Grossberg (1976a, 1976b, 1978a) how to define real-time and local Self-Organizing Maps, and mathematically proved various of their properties (e.g., Baysian tracking of feature density and self-normalization by the adaptive weights) that are still used today. I also generalized the scope of the model from Malsburg's famous example of self-organized orientation maps in visual cortex to maps that could categorize arbitrary data structures as part of a process of "universal recoding".

After deriving these positive results, however, I showed, by example, how easily the learned map categories could be destabilized in a dense and changing input environment. Adaptive Resonance Theory was introduced to self-stabilize the learning of these maps (e.g., Grossberg, 1976c, 1978a, 1980). A lot of work has since been done (e.g., Carpenter and Grossberg, 1991) to build ever-stronger ART models, but many researchers are still working with models whose learning becomes unstable when it takes place in an information-rich world whose statistical rules can change through time. I believe that the time is ripe to incorporate self-stabilizing mechanisms more broadly into our learning architectures, in order to more fully realize the promise of real-time processing.

2. Nonlinear Processing. The gated steepest descent law is also nonlinear. Many people were, at first, opposed to using nonlinear processing. One notable example concerned the Brain-State-in-a-Box (BSB) model of Anderson *et al.* (1977). Jim Anderson had always tried to keep his models as linear as possible, starting with his linear associator models. I very much shared this perspective, because I believed that one should always derive the simplest, indeed the minimal, model that is consistent with the task at hand. In this spirit, my derivations of the Additive Model emphasized the need to keep the model as linear as possible; e.g., Grossberg (1974, Section IIE).

On the other hand, many models introduced insufficient nonlinearity to achieve crucial computational properties. For example, the BSB model amplified noise as part of its effort to carry out a winner-take-all operation. I had earlier proved how to design a winner-take-all network that suppressed noise (Grossberg, 1973), and critiqued the BSB model accordingly (Grossberg, 1978b). My network combined the shunting nonlinearity of neuronal membrane equations with on-center off-surround feedback interactions. Using this network, I also proved how to get another property that BSB could not achieve; namely, the property of partial contrast (what Kohonen called "bubbles" a decade later) by using a sigmoid signal function. This property was used to get the ordering property of Self-Organizing Maps. Such experiences made me realize that introducing the right nonlinearities could be crucial for achieving important computational properties.

Over the past 30 years, mathematical theorems have demonstrated how carefully chosen nonlinear feedback processes can generate basic properties like noise suppression, automatically gain-controlled normalization, winner-take-all or partial choice, stable learning, unbiased associative pattern learning, fast synchronous processing, and the like. Despite successes of this kind, many researchers in the field still do not use enough nonlinear feedback in their work. That is one reason, I believe, why many researchers still use learning models that suffer catastrophic forgetting. Without an appropriate type of nonlinear top-down learned feedback, one cannot escape this instability in response to a nonstationary input environment. Many behavioral and brain data are now implicating a particular type of ART top-down feedback circuit to achieve this goal; see Chey *et al.* (1997), Gove *et al.* (1995), Grossberg (1995, 1997), and Grossberg and Merrill (1996) for some recent uses of this circuit. This convergence suggests that a simple type of top-down on-center off-surround feedback circuit may provide a lot of additional computational power.

3. Parallel Processing and Stimulus Sampling. The gated steepest descent law embodies a type of "stimulus sampling" operation that enables cells to selectively process information only when they are sufficiently active. When I introduced this concept into the neural network literature, I had been motivated in part by the work of Estes and his colleagues on Stimulus Sampling Theory. In addition, analyses of many learning situations made it clear that a sampling operation was essential in order to achieve effective *parallel processing* of information in real time. Without such an operation, too many cells could unselectively learn about events with which they were not temporally correlated. Many cognitive science and AI practitioners thought, instead, in terms of serial processing, and this slowed down the acceptance of a sampling operation that could achieve task-dependent selectivity in a parallel processing environment.

At least two implications of this type of parallel processing are still of current interest: synchronization and the microanatomy of associative learning. The synchrony issue is discussed in Section 4. The microanatomy issue concerns the need to dissociate *read-out* of previously learned information from a synapse, and *read-in* of new information to the synapse. The nonlinear sampling function $f(x)$ of the learning law is necessary, but not sufficient, to achieve this dissociation. Such a dissociation is needed to prevent learning from being destabilized by a flood of irrelevant signals and noise in a complex parallel processing environment. In particular, I claim that a network should be designed to read-out information whenever the presynaptic signals are strong enough, but to read-in information only after it has undergone a context-sensitive cooperative-competitive decision process. The case of Self-Organizing Maps provides a classical example of this. Other examples include reinforcement learning. In various of these examples, I predicted that retrograde spikes from cell bodies to their dendrites could be used as a teaching signal (e.g., Grossberg, 1975, Figure 24; Grossberg, 1982b, Figure 16; Grossberg and Merrill, 1992; Grossberg and Schmajuk, 1987). The basic idea is that local read-outs of learned associations from synapses onto cell dendrites get accumulated at the cell bodies, which then interact via cooperative-competitive feedback processing across the whole network of cells. The winning cells can then deliver the most effective retrograde spikes up their dendritic trees to trigger further learning at their dendritic synapses. Recent data have tended to support this prediction (Christie, Magee, and Johnston, 1996), and how this process is accomplished by the brain is a topic of great current interest in experimental neuroscience. I believe that using the *read-in-out dissociation property* can also help to stabilize learning in a parallel processing environment.

The dissociation property has been implicitly used since the introduction of gated steepest descent, without making explicit use of dendritic spikes. This has typically been achieved by realizing the learning process algorithmically, rather than in real time. For example, the instar version of the law (where x represents postsynaptic activity and y presynaptic activity) enables cells which survive a competitive process to sample only the input patterns that enabled them to win the competition. The outstar version of the law (where x represents presynaptic activity and y postsynaptic activity) then enables an active cell to learn and recall a distributed pattern of sampled information. The instar law was used to define a Self-Organizing Map in Grossberg (1976a, 1976b, 1978a) and Kohonen (1982, 1984). The outstar law has been used even longer for associative learning, including learning of temporally occurring sequences of events (e.g., Grossberg, 1969c) and learning of spatially distributed information (e.g., Grossberg, 1968, 1969b). Both laws were brought together to learn "counter propagation" maps in Grossberg (1976b) to show how "universal recoding" (viz., mapping from an arbitrary m-dimensional input to an arbitrary n-dimensional output via a learned category node) could be achieved. Hecht-Nielsen (1987) later gave the counter propagation model its name, generalized it to a version with feedback, and used it to work out some interesting applications.

The difference between the dates 1976 and 1987 is of historical interest, because in the interim, Rumelhart, Hinton, and Williams published their article on back propagation, which had earlier been discovered at various times by Shun-Ichi Amari, Paul Werbos, and David Parker. Many advocates of backpropagation claimed that earlier neural networks could not learn such maps. As with so many other claims about the field at this time, this one was also incorrect. In fact, the instar and outstar laws were also both used to introduce ART (e.g., Grossberg, 1976c and later, Carpenter and Grossberg, 1987). Here the instar law learns the bottom-up adaptive filter weights and the outstar law learns the top-down expectation weights. In my 1976 "universal recoding" papers (Grossberg, 1976b, 1976c), the basic idea was to use instars and outstars in an ART module to learn a self-stabilizing category map in response to m-dimensional input vectors, and then to use the stable category nodes to sample and learn n-dimensional output vectors using outstars.

When I first introduced gated steepest descent in the 1960's (e.g., Grossberg, 1969b; 1974, Section VI), I described it as "perfect memory until recall" learning, since it allows memory to remain unchanged until a sampling gate opens, whereupon learning and forgetting can resume. This idea helped to explain how memories can last so long without a loss of future plasticity. It was my first example of how to deal with the *stability-plasticity dilemma*.. This issue raised the question of how to design an architecture in which the learning gate does not open at inappropriate times, which could cause catastrophic forgetting. This road led to ART as well as to the read-in-out dissociation property. It provides a good example, I think, of how a simple learning equation can force one to think about more global architectural issues, that are still being worked out today, 30 years after the law was introduced.

4. Spatial Pattern Learning and Synchronization. Gated steepest descent fully made sense to me only after I had proved some theorems showing that "the unit of long-term memory is a spatial pattern"; in other words, that the functional unit of learning, both in the instar and outstar modes, is a distributed pattern of activation across a network, and that network design should be aimed at controlling the transformation of these distributed patterns through time. This mathematical result grew out of the intuition that the brain is designed to achieve *behavioral* success, and that the functional processing units that determine behavioral success are often distributed activation patterns. It is hard now to express how radical this idea seemed in the late 1950's when I started using it to model human and animal learning data. It took until the 1960's for me to rigorously prove that the functional unit is a distributed activation pattern, both during both temporal learning (e.g., Grossberg, 1969c) and spatial learning (e.g., Grossberg, 1968, 1969b). This fact is now taken so much for granted that many people seem not to realize that it was considered a major discovery not so very long ago. Perhaps partly for that reason, even today various of its implications have not been fully developed.

One implication of the fact that distributed spatial patterns are the units of learning is that *synchronous processing* is needed to define these patterns. This issue was raised already in the 1960's by my theorems on associative pattern learning and was discussed in terms of "order-preserving limit cycles" in the first ART paper (Grossberg, 1976c). How synchronous processing is achieved in the brain is still a topic of great current interest both experimentally (e.g., Eckhorn *et al.*, 1988; Gray and Singer, 1989) and in models (e.g., Baldi and Meir, 1990; Eckhorn *et al*., 1989; Grossberg and Somers, 1991; Grunewald and Grossberg, 1997; Konig and Schillen, 1991; Terman and Wang, 1995).

Another implication of spatial processing is that the same learning law should incorporate both Hebbian and non-Hebbian properties, as has been increasingly supported by neural data during the past fifteen years. Hebbian learning only allows adaptive weights to increase; this leads to either weight saturation or explosion. Anti-Hebbian learning only allows adaptive weights to decrease; this eventually leads to a collapse of processing. The gated steepest descent law allowed adaptive weights to either increase or decrease in order to track the activation pattern while the sampling source was active. Although this law was available when von der Malsburg was introducing his version of the Self-Organizing Map, he used instead a purely Hebbian learning law. It was this choice that led him to give up real-time and local processing in order to learn the map. Restoring this law was one of the steps that enabled me to define a real-time local Self-Organizing Map.

The selection of a learning law thus needs to be done in parallel with the design of the information processing architecture in which it lives. This proposal runs counter to beliefs that are still held in cognitive science and AI, wherein it is often thought that an information processing architecture can be designed first, and that learning can be tacked on later. Many examples now show that learning laws which might be suitable in one type of architecture might be the wrong laws for a different type of architecture. For example, the learning laws for sensory and cognitive processes, as in the ART model (e.g., Carpenter and Grossberg, 1987), are often computationally *complementary* to the learning laws for spatial and motor processes, as in the VAM model (Gaudiano and Grossberg, 1991). This conclusion is also contrary to the belief that a single type of learning architecture, such as Back Propagation or Self-Organizing Maps, can be adapted to solve all of our problems. Thinking deeply about the conceptual foundations that support each of our equations — in particular, knowing and understanding the intellectual struggles that led to their discovery — can help us to better define their appropriate uses and limits, as well as to discover the additional concepts that will be needed to make further progress in our work. I hope that the above review has clarified how issues raised by the steepest descent learning law are still being developed, more than 30 years after its discovery.

References

Anderson, J. A., Silverstein, J. W., Ritz, S. R., and Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, **84**, 413–451. [Steve: I'm not sure if this is the correct paper, see pg 6, par 2]

Artola, A. and Singer, W. (1993). Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. *Trends in Neurosciences*, **16**, 480–487.

Atkinson, R. C. and Estes, W. K. (1963). Stimulus sampling theory. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.) **Handbook of mathematical psychology**, Vol. II., Chapters 9–14. New York, NY, Wiley, 121–265.

Baldi, P. and Meir, R. (1990). Computing with arrays of coupled oscillators: An application to preattentive texture discrimination. *Neural Computation*, **2**, 458–471.

Carpenter, G. A. and Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, **37**, 54–115.

Carpenter, G. A. and Grossberg, S. (Eds.), **Pattern recognition by self-organizing neural networks**. Cambridge, MA, M.I.T. Press.

Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., and Rosen, D. B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, **3(5)**, 698–713.

Chey, J., Grossberg, S., and Mingolla, E. (1997). Neural dynamics of motion grouping: From aperture ambiguity to object speed and direction. *Journal of the Optical Society of America*, **14(10)**, 2570–2594

Cohen, M. and Grossberg, S. (1983). Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC–13(5)**, 815–826.

Christie, B. R., Magee, J. C., and Johnston, D. (1996). The role of dendritic action potentials and Ca^{++} influx in the induction of homosynaptic long-term depression in hippocampal CA1 pyramidal neurons. *Learning and Memory*, **3**, 160–169.

Eckhorn, R., Reitboeck, H. J., Arndt, M., and Dicke, P. (1989). A neural network for feature linking via synchronous activity. In R. M. Cotterill (Ed.) **Models of Brain Function**. 255–272, New York, NY, Cambridge University Press.

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., and Reitboeck, H. J. (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex?, *Biological Cybernetics*, **60**, 121–130.

Gaudiano, P. and Grossberg, S. (1991). Vector associative maps: Unsupervised real-time error-based learning and control of movement trajectories. *Neural Networks*, **4**, 147–183.

Gove, A., Grossberg, S. and Mingolla, A. (1995).Brightness perception, illusory contours, and corticogeniculate feedback. *Visual Neuroscience*, **12**, 1027–1052.

Gray, C. M., and Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of CAT visual cortex. *Proceedings of the National Academy of Sciences*, **86**, 1698–1702.

Grossberg, S. (1968). Some nonlinear networks capable of learning a spatial pattern of arbitrary complexity. *Proceedings of the National Academy of Sciences*, **59 (2)**, 368–372.

Grossberg, S. (1969a). Embedding fields: A theory of learning with physiological implications. *Journal of Mathematical Psychology*, **6**, 209–239.

Grossberg, S. (1969b). On learning and energy-entropy dependence in recurrent and nonrecurrent signed networks. *Journal of Statistical Physics*, **1**, 319–350.

Grossberg, S. (1969c). On the serial learning of lists. *Mathematical Biosciences*, **4**, 201–253.

Grossberg, S. (1971). On the dynamics of operant conditioning. *Journal of Theoretical Biology*, **33**, 225–255.

Grossberg, S. (1972a). A neural theory of punishment and avoidance (I) *Mathematical Biosciences*, **15**, 39–67.

Grossberg, S. (1972b). A neural theory of punishment and avoidance (II) *Mathematical Biosciences* , **15**, 253–285. Reprinted in Grossberg, S. (1982b). **Studies of mind and brain.**

Grossberg, S. (1972c). Neural expectation: Cerebellar and retinal analogs of cells of fired by learnable or unlearned pattern classes. *Kybernetik*, **10**, 49–57.

Grossberg, S. (1972d). Pattern learning by functional-differential neural networks with arbitrary path weights. In K. Schmitt (Ed.), **Delay and functional-differential equations and their applications**. New York, NY, Academic Press. Reprinted in Grossberg, S. (1982b). **Studies of mind and brain.**

Grossberg, S. (1973) Contour enhancement, short-term memory and constancies in reverberating neural networks. *Studies in Applied Mathematics*, **52**, 217–257.

Grossberg, S. (1974). Classical and instrumental learning by neural networks. In R. Rosen and F. Snell (Eds.) **Progress in theoretical biology**. **Vol. 3.** 51–141. New York, NY, Academic Press, Reprinted in Grossberg, S. (1982b). **Studies of mind and brain.**

Grossberg, S. (1975). A neural model of attention, reinforcement and discrimination learning. *International Review of Neurobiology*, **18**, 263–327. Reprinted in Grossberg, S. (1982b). **Studies of mind and brain.**

Grossberg, S. (1976a). On the development of feature detectors in the visual cortex with applications to learning and reaction-diffusion systems. *Biological Cybernetics*, **21**, 145–159.

Grossberg, S. (1976b). Adaptive pattern classification and universal recoding, Part I: Parallel development and coding of neural feature detectors. *Biological Cybernetics*, **23**, 121–134.

Grossberg, S. (1976c). Adaptive pattern classification and universal recoding, Part II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, **23**, 187–202.

Grossberg, S. (1978a). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen and F. Snell (Eds.) **Progress in theoretical biology**. **Vol. 5.** 233–374. New York, NY, Academic Press. Reprinted in Grossberg, S. (1982b). **Studies of mind and brain.**

Grossberg, S. (1978b). Do all neural networks really look alike? A comment on Anderson, Silerstein, Ritz, and Jones. *Psychological Review*, **85**, 592–596.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, **87**, 1–51.

Grossberg, S. (1982a). A psychophysiological theory of reinforcement, drive, motivation, and attention. *Journal of Theoretical Neurobiology*, **1**, 286–369. Reprinted in S. Grossberg, (1987) **The adaptive brain, I**. Amsterdam, Elsevier/North-Holland.

Grossberg, S. (1982b). **Studies of mind and brain.** Boston, Kluwer/Reidel.

Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science*, **11**, 23–63.

Grossberg, S. (1988). Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks*, **1**, 17–61.

Grossberg, S. (1995). The attentive brain. *American Scientist*, **83**, 438–449.

Grossberg, S. (1997). Pitch-based streaming in auditory perception. In N. Griffith and P. Todd (Eds.), **Musical Networks: Parallel Distributed Perception and Performance**. Cambridge, MA, M.I.T. Press

Grossberg, S. and Merrill, J. W. L. (1992) A neural network model of adaptively timed reinforcement learning and hippocampal dynamics. *Cognitive Brain Research*, **1**, 3–38.

Grossberg, S. and Merrill, J. W. L. (1996). The hippocampus and cerebellum in adaptively timed learning, recognition, and movement. *Journal of Cognitive Neuroscience*, **8**, 257–277.

Grossberg, S. and Pepe, J. (1971). Spiking threshold and overarousal effects in serial learning. *Journal of Statistical Physics*, **3**, 95–125.

Grossberg, S. and Schmajuk, N. A. (1987). Neural dynamics of Pavlovian conditioning: Conditioned reinforcement, inhibition, and opponent processing. *Psychobiology*, **15**, 195–240.

Grossberg, S. and Somers, D. (1991). Synchronized oscillations during cooperative feature linking in a cortical model of visual perception. *Neural Networks*, **4**, 453–466.

Grunewald, A. and Grossberg, S. (1997). Cortical synchronization and perceptual framing. *Journal of Cognitive Neuroscience*, **9**, 117–132.

Hecht-Nielsen, R., (1987). Counter propagation networks. *Applied Optics*, **26**, 4979–4984.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences*, **81**, 3058–3092.

Kohonen, T. (1982). A simple paradigm for the self-organized formation of structural feature maps. In S. Amari and M. A. Arbib (Eds.) **Competition and cooperation in neural networks**, New York, NY, Springer-Verlag.

Kohonen, T. (1984). **Self-organization and associative memory**. New York, NY, Springer-Verglag

König, P. and Schillen, T. B. (1991). Stimulus-dependent assembly formation of oscillatory responses: I. Synchronization. *Neural Computation*, **3**, 155–166.

Levy, W. B., Brassel, S. E., and Moore, S. D. (1983). Partial quantification of the associative synaptic learning rule of the dentate gyrus. *Neuroscience*, **8**, 799–808.

Levy, W. B. and Desmond, N. L. (1985). The rules of elemental synaptic plasticity. In W. B. Levy, J. Anderson, and S. Lehmkuhle (Eds.) (1985). **Synaptic modification, neuron selectivity and nervous system organization**. Hillsdale, NJ, Erlbaum, 105–121.

Singer, W. (1983). Neuronal activity as a shaping factor in the self-organization of neuron assemblies. In E. Basar, H. Flohr, H. Haken, and A. J. Mandell (Eds.) (1983). **Synergetics of the brain**. New York, NY, Springer-Verlag, 89–101.

Terman, D. and Wang, D. (1995). Global competition and local cooperation in a network of neural oscillators. *Physica D*, **81**, 148–176.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetick*, **14**, 85–100.