Appeared in Y.C. Lee (ed.) *Evolution, Learning and Cognition*, 391–403, 1988 World Scientific

# SELECTED BIBLIOGRAPHY ON CONNECTIONISM

Oliver G. Selfridge, Richard S. Sutton, and Charles W. Anderson*

GTE Labs, Waltham, MA.

## Introduction

The topic of this annotated bibliography is connectionism, a field of computer science that has enjoyed a vast resurgence in the last ten years. Properly speaking, connectionism should be regarded as part of Artificial Intelligence, or AI, and up to some years ago it was usually so treated. The earlier entries below will make that clear. Another term for connectionism is *Neural Networks*, and it is being widely used, especially among the new efforts that are arising, including start-up companies.

Connectionism has two chief interests: one is the efficiency or novelty of certain kinds of computation; the other is models of real brains or real neural networks—the kinds made of flesh. The former is our concern here. It has been claimed that connectionism can exhibit a new kind of computing, which is "non-von-Neumann," and can thereby provide new capabilities that cannot be matched in other ways. We want to point readers towards publications that can give them background and insights about the real issues and the state of the field and its prospects. We do not cover the recent work on implementation technologies. Our audience here is anybody who wants or needs more than buzzword knowledge about the field, including researchers, students, and managers in computer science and technology.

The entries have been selected according to their relevance to learning machines that we now recognize as connectionist. Entries are ordered by date of publication.

---

* The authors gratefully acknowledge the helpful comments of A. Barto, M. Steenstrup, J. Franklin, and H. Klopf.

McCulloch, W.S. and Pitts, W.H., "A logical calculus of ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, 1943, *5*, 115–133; reprinted in McCulloch's *Embodiments of Mind*, Cambridge, MA: M.I.T. Press, 1965.

> This early paper lays out the ideas behind connectionism with austere and literate precision; though in places it is not easy reading. It shows that a simple model neuron, working in discrete time and emitting a purely binary signal, can be assembled in numbers to form a Turing machine; that is, that it can compute anything that is computable at all. An awesome piece of work, considering that the junior author, who was responsible for all the mathematics and many of the ideas, was barely twenty years old.

Pitts, W.H. and McCulloch, W.S., "How we know universals: The perception of auditory and visual forms," *Bulletin of Mathematical Biophysics*, 1947, *9*, 127–147.

> A companion paper to the previous one. It shows that neural networks—that is, connectionist mechanisms—can compute features or *concept membership* in the current AI sense.

Hebb, D.O., *The Organization of Behavior: A Neurophysiological Theory*, New York: Wiley, 1949.

> This connectionist classic deals broadly with the problem of relating psychology to neurophysiology. The most lasting specific contribution has been Hebb's neurophysiological learning rule—that a synapse becomes strengthened whenever the pre- and post-synaptic neurons are simultaneously active. Hebb argued that neurons following this rule would group themselves together to form *cell assemblies*, which would then be capable of further learning and more complex behavior. One difficulty in the ideas is that the cell assemblies seemed not to behave very differently from the neurons they were assembled from. In later years, Hebb seemed to abandon his old ideas more willingly than some of his readers.

Farley, B.G. and Clark, W.A., "Simulation of self-organizing systems by digital computer," *I.R.E. Transactions on Information Theory*, 1954, vol 4, 76–84.

> The earliest publication we know of presenting simulation results with connectionist systems. The learning problem was primarily one of pattern classification, but there was also discussion of what

we would now recognize as reinforcement learning. It was presented at what is regarded as the opening guns of AI, the Western Joint Computer Conference session in 1954. See also Clark and Farley's "Generalization of pattern recognition in a self-organizing system" (*I.R.E. Transactions on Inf. Theory*, 1955, *5*, 86–91), which includes a summary of the results of their earlier paper.

Rosenblatt, F., "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological Review*, 1958, *65*, 386–408. See also Rosenblatt's *Principles of Neurodynamics*, New York: Spartan, 1962.

Rosenblatt and the perceptron are the names that today we most associate with the early surge and then ebbing of interest in connectionism. Probably Rosenblatt and his group at Cornell Aeronautical Laboratory were responsible for more hyperbole per actual man-month of work than any other group in history—though some today may pose competition. At the core of the perceptron work is the *convergence theorem*, which states that a certain kind of perceptron will eventually learn any predicate it is capable of representing. As others noted later, the primary limitations of this result are 1) the word *eventually*, 2) that many predicates cannot be represented, and 3) that a perceptron must be explicitly told the correct behavior in order to learn.

Selfridge, O.G., "Pandemonium: A paradigm for learning," *The Mechanisation of Thought Processes*, London: H.M. Stationery Office, 2 vols., 1959; reprinted in *Pattern Recognition; Theory, Experiment, Computer Simulations, and Dynamic Models of Form Perception and Discovery*, Uhr, L., ed., New York: Wiley, 1966.

A statement of the importance of features in recognition, and suggesting that the hierarchy of features is a dominant and natural structure; the interplay between layers has a connectionist flavor and function. This paper and Samuel's paper (below) were the first to discuss the idea of generating new features from combinations and mutations of old features that have already proven useful.

Samuel, A.L., "Some studies in machine learning using the game of checkers," *IBM Journal on Research and Development*, 1959, *3*, 210–229; reprinted in *Computers and Thought*, Feigenbaum, E.A. and Feldman, J., eds., New York: McGraw-Hill, 1963.

Probably the most famous learning paper in AI. Although Samuel saw his work as an alternative to the "Neural-Net Approach" that

was popular at the time, it would fit well into 1980's connectionism, and is the basis for several modern learning procedures.

Widrow B. and Hoff, M.E., "Adaptive switching circuits," *1960 WESCON Convention Record Part IV*, 1960, 96–104. See also *Adaptive Signal Processing*, by Widrow, B. & Stearns, S.D., Englewood Cliffs, NJ: Prentice-Hall, 1985.

This paper introduced the ADALINE, one of the most effective and best understood of connectionist units, and one of the very few that have already served in useful applications. The ADALINE continues to be widely used and to provide a theoretical base for new learning procedures (for example, *back-propagation*). *Adaptive Signal Processing* is an excellent textbook presentation of the ADALINE results obtained over the years by Widrow *et al.* at M.I.T. and then at Stanford.

Minsky, M.L. and Selfridge, O.G., "Learning in random nets," *Information Theory, Fourth London Symposium*, London: Butterworths, 1961.

This was the first real critique of Rosenblatt's perceptrons, and pointed out that the perceptron as he had defined it, far from being able to make general abstractions, could not even generalize towards the notion of binary parity; it also analyzed other claims to convergence and suggested the roles that connectionist mechanisms might play in larger systems.

Minsky, M.L., "Steps toward artificial intelligence," *Proceedings of the Institute of Radio Engineers*, 1961, *49*, 8–30; reprinted in *Computers and Thought*, Feigenbaum, E.A. & Feldman, J., eds., New York: McGraw-Hill, 1963.

This excellent early paper in AI includes a large section on what is now termed connectionism. It should be remembered that Minsky wrote a connectionist doctorate thesis in 1954 at Princeton University ("Theory of neural-analog reinforcement systems and its application to the brain-model problem," available from University Microfilms, Ann Arbor, MI).

Nilsson, N.J., *Learning Machines*, New York: McGraw-Hill, 1965.

This early book is a well-written exposition on linear separability in hyper-spaces; it is still one of the best teaching tools for

understanding the basic processes of simple pattern recognition programs.

Minsky, M.L. and Papert, S., *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA: MIT Press, 1969.

This properly famous book analyzed one-layer perceptrons and proved that they are inherently incapable of making some global generalizations on the basis of locally learnt examples: in particular, connectivity of a binary picture. It is a thoughtful, thorough, and well written book. However, the limitations discussed are all of perceptrons as computational mechanisms, not as learning mechanisms; that is, the limitations are on what they can compute, not on what they can learn in a practical amount of time. As Minsky and Papert note, the latter is often the more pressing concern. For example, a perceptron has no computational difficulties in learning to recognize shapes independent of their size, position, and orientation, but it can do so only after experience with each shape in all possible sizes, positions, and orientations. Although this book is often said to have killed the early perceptron work, it had already been nearly abandoned by the time the book appeared.

Mendel, J.M. and Fu, K.S., eds., *Adaptive, Learning and Pattern Recognition Systems*, New York: Academic Press, 1970.

Though twenty years old, this thorough study of adaptive techniques in pattern recognition problems is a standard. Starting with simple linear separability, it examines gradient techniques, adaptive optimization, and reinforcement learning with good mathematical support.

Klopf, A.H., "Brain function and adaptive systems–A heterostatic theory," *Air Force Cambridge Research Laboratories Research Report*, AFCRL-72-0164, Bedford, MA., 1972. An updated version is available as Klopf's *The Hedonistic Neuron: A Theory of Memory, Learning, and Intelligence*, Washington DC: Hemisphere/Harper & Row, 1982.

Klopf's primary contribution was to recognize that something was missing from the then-current stock of connectionist learning methods. That something was the ability to learn in environments in which you were told *how well* you were doing, but not exactly *what* you should be doing (or should have done); that is, the ability to do

reinforcement learning rather than supervised or error-correction learning.

Arbib, M.A., *The Metaphorical Brain*, New York: Wiley, 1972.

This book laid out the imperatives for modern connectionism clearly and convincingly, and helped set the stage for the current renewed interest in the area.

Sommerhoff, G., *The Logic of the Living Brain*, London: Wiley, 1974.

A little known but excellent analysis of basic connectionist concepts and assumptions. For example, there are sections on what it means for a connectionist system to be goal-directed and to have an internal model of the world, sections on expectation, attention, and the apparent stability of the visual scene, and sections on the difference between error signals and goal signals.

Uttley, A.M., *Information Transmission in the Nervous System*, London: Academic Press, 1979.

A compact presentation of Uttley's pioneering work in connectionism. He is best known for his early work on *conditional probability machines* (1956) and for the *informon*, a connectionist learning unit using the negative of the Hebb rule, and which he related to animal learning theories.

Fukushima, K., "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, 1980, *36*, 193–202.

An often-cited problem in using connectionist networks for pattern classification is their inability to generalize to shifted or rotated versions of trained patterns. A brute-force approach to this problem is to provide sets of units that extract identical features from different parts of an input field. Fukushima's *Neocognitron* (a development of his earlier *Cognitron*) is constructed of multiple layers of such unit sets and demonstrates limited shift-invariance. The generality of this approach may be limited by the large number of units required.

Sutton, R.S. and Barto, A.G., "Toward a modern theory of adaptive networks: Expectation and prediction," *Psychological Review*, 1981, *88*, 135–170.

A study of connectionist learning elements as models of *Pavlovian conditioning*, the simplest and best understood kind of associative learning in nature. This paper points out that while the Hebb rule is a very poor model of animal behavior, the ADALINE rule is equivalent to a popular and successful psychological theory, the *Rescorla-Wagner model*. The paper also proposes a new connectionist learning element that improves over both in some ways.

Albus, J.S., *Brains, Behavior, and Robotics*, Peterborough, NJ: McGraw-Hill/BYTE, 1981.

This book lays out a connectionist approach to robotics, hierarchical control, and cerebellar modelling, which was pursued by the author throughout the 1970's.

Hinton, G.E. and Anderson, J.A., eds., *Parallel Models of Associative Memory*, Hillsdale, NJ: Lawrence Erlbaum, 1981.

A collection of articles from an informal conference held at the University of California at San Diego in 1979. This conference can be said to mark the onset of renewed interest in connectionism within cognitive psychology. Included are articles by Hinton, Rumelhart, Kohonen, Anderson, Sejnowski, Feldman, Willshaw, Geman, Fahlman, and Ratcliff.

Hopfield, J.J., "Neural networks and physical systems with emergent collective computation abilities," in *Proceedings of the National Academy of Sciences, USA*, 1982, *79*, 2554–2558. See also Hopfield, J.J. and Tank, D.W., "Computing with neural circuits: A model," *Science*, 8/8/86, *233*, 625–633.

John Hopfield has almost single-handedly created an enormous amount of activity and interest in connectionist systems among physicists and the wider lay public. In this paper, he introduced the idea of *computational energy*, a new way of understanding the computation performed by networks with feedback and effectively symmetric connections. This idea was used subsequently, for example, in the development of the Boltzmann Machine (see Ackley *et al.* below). In the Hopfield and Tank article, energy analyses are used to design networks and weight settings to solve particular problems; for instance, samples of the traveling salesman problem. This work should be taken as an excellent illustration of the energy-function design methodology, not as a demonstration of the competitiveness of such networks on combinatorial optimization problems.

Feldman, J.A. and Ballard, D.H., "Connectionist models and their properties," *Cognitive Science*, 1982, *6*, 205–254.

A scholarly argument for connectionist models as opposed to information processing models in cognitive science, and a presentation of the general "University of Rochester" connectionist model. The Rochester model emphasizes the computational advantages of massive parallelism even when *not* coupled with learning and distributed representations. Because it features a large variety of specialized types of units, the Rochester model also challenges the dogmatic assumption of many connectionists that all units should be simple and identical.

Anderson, J.A., "Cognitive and psychological computation with neural models," *IEEE Transactions on Systems, Man, and Cybernetics*, 1983, *SMC-13*, 799–814.

An excellent summary of Anderson's connectionist modelling approach to cognitive psychology, including a presentation of his "brain-state-in-a-box" model, one of the earliest (1979) and simplest extensions of associative network ideas to include feedback and nonlinearity.

Barto, A.G., Sutton R.S., and Anderson, C.W., "Neuronlike elements that can solve difficult learning control problems," *IEEE Transactions on Systems, Man, and Cybernetics*, 1983, *SMC-13*, 834–846.

This paper contains the best demonstration of the abilities of the reinforcement-learning units developed by Barto *et al.* at the University of Massachusetts. A network consisting of two units is shown to be able to solve a broomstick balancing problem that Perceptrons or ADALINEs cannot solve, and to do so much more efficiently than a non-connectionist system previously developed for this task.

Kohonen, T., *Self-Organization and Associative Memory*, Berlin: Springer-Verlag, 1984.

An excellent review of the ongoing work of this pioneer in associative memory and connectionism.

Ackley, D.H., Hinton, G.E., and Sejnowski, T.J., "A learning algorithm for Boltzmann machines," *Cognitive Science*, 1985, *9*, 147–169.

Introduced the first effective supervised-learning algorithm applicable to networks with interior or "hidden" units.

Lee, Y.C., Doolen, G., Chen, H.H., Sun G.Z., Maxwell, T., Lee, H.Y., and Giles, C.L., "Machine learning using a higher order correlation network," in *Evolution, Games and Learning*, *Proceedings of the Fifth Annual International Conference of the Center for Nonlinear Studies*, 276–306, Amsterdam: North-Holland, 1985.

Describes an approach to extending the linear computation of most connectionist units to include higher-order nonlinear terms. The extension allows the solution of nonlinear problems without resorting to multi-layer networks, and can result in spectacularly effective generalization if the selection of higher-order terms is done on a task-specific basis. However, the complexity of the individual units grows exponentially with their order and thus must be limited; multiple layers are still required to solve problems with high-order non-linearities.

Rumelhart, D.E., McClelland, J.L., and The PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*, Cambridge, MA: Bradford, 1986.

McClelland, J.L., Rumelhart, D.E., and The PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2: Psychological and Biological Models*, Cambridge, MA: Bradford, 1986.

These two volumes form a splendid reference set for the field and provide a multi-disciplinary review of many of the underlying ideas. Among the many excellent articles, two must be specially mentioned: "Learning internal representations by error propagation," by Rumelhart, D.E., Hinton, G.E., & Williams, R.W., which introduces *back-propagation*, the most efficient known learning procedure for multi-layer networks; and "On learning the past tenses of English words," by Rumelhart, D.E. & McClelland, J.L., which presents a controversial connectionist model that challenges rule-oriented conceptions of language learning.

These important books share with many of the others discussed here a lack of exploration of the limits of their connectionist mechanisms. Keith Holyoak points out some of the limits of these

two books in an appreciative and informative review in *Science* (5/22/87, *236*, 992–996), such as their inability to learn sequences of action in which early components receive no direct feedback.

Minsky, M.L., *Society of Mind*, New York: Simon and Schuster, 1986.

This recent popular work is at once exciting and disappointing. Minsky is articulate, witty and visionary, and he possesses a superb ability to evoke penetrating insights. Much of the discussion implicitly endorses the questions and drives that have motivated connectionists. The book as a whole seems to us to be uneven and patchy; for example, he underestimates the differences among people in how they perceive the world and behave in it. But he wisely reiterates and re-emphasizes the complexity and richness of human thought.

Sejnowski, T.E. and Rosenberg, C.R., "Parallel networks that learn to pronounce English text," *Complex Systems*, 1987, *1*, 145–168.

This paper describes *NETtalk*, a multi-layer back-propagation network that learns to convert text to its phonemic representations, using a human expert for a teacher, and with some residual error. Coupled with a commercial phoneme-to-speech box, *NETtalk*'s learning behavior makes an impressive demonstration that has captured the imagination of the public—*NETtalk* has been widely discussed in the popular press including the TODAY show. In *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*, 1987, Rosenberg uses standard clustering methods to analyze the features learnt by *NETtalk*, showing one way to gain insight into the functioning of a trained network.

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

This paper is the best presentation of Grossberg's *Adaptive Resonance Theory* (ART), which Grossberg himself introduced a decade earlier. An ART network is a mechanism for performing unsupervised clustering of input patterns. Such clustering is widely recognized as a useful technique for reducing the dimensionality of the input to a system; but additional mechanisms are required in order to relate changes in the system to its goals, a point that is

not apparent from the descriptions of ART in the literature. That is, the clustering is responsive merely to the metric induced by the particular representations, and not at all to the designer's or user's purposes. Also missing in the literature are comparisons with other implementations of clustering methods. Grossberg *et al.* at Boston Unversity have analyzed this and other connectionist networks as systems of differential equations, which may facilitate their direct realization in parallel hardware.

Lippmann, R. P., "An introduction to computing with neural nets," *IEEE ASSP Magazine*, April 1987, 4–22.

This is a gentle introduction to a few of the popular methods for using networks as pattern classifiers. They are related to standard pattern classification techniques.

Elman, J.L. and Zipser, D., "Learning the hidden structure of speech," Technical Report 8701, Insitute for Cognitive Science, University of California at San Diego, La Jolla, CA, 1987.

This paper is a good example of the current connectionist attacks on real problems; it illustrates both what can be done and the limitations of the techniques. It is well written and clear, and it does not make claims beyond what it shows. The task is discrimination of spoken consonant/vowel pairs. The preprocessing seems to have been of over-riding importance, including sampling, A/D conversions, sophisticated normalizations, and, in the most convincing experiments, Fourier transforms. The (few) hidden units are found to "encode the input patterns as feature types." Some feature types turn out to be easily comprehensible, but others are harder to interpret. It appears that they are, however, only simple combinations of the presence or absence of particular input values. The experimental procedure uses but a single male speaker.

Hinton, G. E., "Connectionist learning procedures," Technical Report CMU-CS-87-115, Carnegie-Mellon University, Pittsburgh, PA, 1987; to appear in *Artificial Intelligence*, 1988.

This technical report reviews learning methods for multilayer networks and some early work with single-layer networks. Research issues are briefly discussed. Examples of supervised, unsupervised, and reinforcement learning methods are described. This paper is the most concise treatment covering these three forms of learning.

Sutton, R.S., "Learning to predict by the methods of temporal differences," Technical Report TR87-509.1, GTE Laboratories, Waltham, MA, 1987; to appear in *Machine Learning*, 1988.

> From the Abstract: "This article introduces and provides the first formal results in the theory of *temporal-difference methods*, a class of statistical learning procedures specialized for prediction—that is, for using past experience with an incompletely known system to predict its future behavior ... It is argued that most problems to which supervised learning is currently applied are really prediction problems of the sort to which [these] methods can be applied to advantage."

Anderson, J.A. and Rosenfeld, E., eds., *Neurocomputing: Foundations of Research*, Cambridge, MA: MIT Press, to appear in 1988.

> A collection of 43 connectionist reprints, including many of those listed here, and "with a general introduction, introductions to each paper, and reference materials."