

- unconscious from conscious cognition. *Journal of Experimental Psychology: General* **124**: 22–42
- Greenwald A G, Spangenberg E R, Pratkanis A R, Eskanazi J 1991 Double-blind tests of subliminal self-help audiotapes. *Psychological Science* **2**: 119–22
- Holender D 1986 Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences* **9**: 1–66
- Jacoby L L, Whitehouse K 1989 An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General* **118**: 126–35
- Kunst-Wilson W R, Zajonc R B 1980 Affective discrimination of stimuli that cannot be recognized. *Science* **207**: 557–58
- LeDoux J E 1995 Emotion: Clues from the brain. *Annual Review of Psychology* **46**: 209–35
- MacLeod C 1998 Implicit perception: Perceptual processing without awareness. In: Kirsner K, Spelman C, Maybery M, O'Brien-Malone A, Anderson M, MacLeod C (eds.) *Implicit and Explicit Mental Processes*. Erlbaum, Mahwah, NJ, pp. 57–78
- Marcel A J 1980 Conscious and preconscious recognition of polysemous words: Locating the selective effects of prior verbal context. In: Nickerson R S (ed.) *Attention and Performance*. Erlbaum, Hillsdale, NJ, Vol. VII, pp. 435–57
- Marcel A J 1983 Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology* **15**: 197–237
- Merikle P M, Daneman M 1996 Memory for unconsciously perceived events: Evidence from anesthetized patients. *Consciousness and Cognition* **5**: 525–41
- Merikle P M, Daneman M 1998 Psychological investigations of unconscious perception. *Journal of Consciousness Studies* **5**: 5–18
- Merikle P M, Joordens S 1997 Measuring unconscious influences. In: Cohen J D, Schooler J W (eds.) *Scientific Approaches to Consciousness*. Erlbaum, Mahwah, NJ, pp. 109–123
- Morris J S, Öhman A, Dolan R J 1998 Conscious and unconscious emotional learning in the human amygdala. *Nature* **393**: 467–70
- Peirce C S, Jastrow J 1884 On small differences of sensation. *Memoirs of the National Academy of Sciences* **3**: 73–83
- Poetzl O [1917] 1960 The relationships between experimentally induced dream images and indirect vision. *Monograph 7. Psychological Issues* **2**: 46–106
- Pöppel E, Held R, Frost D 1973 Residual visual function after brain wounds involving the central visual pathways in man. *Nature* **243**: 295–6
- Reber A S 1993 *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*. Oxford University Press, New York
- Reingold E M, Merikle P M 1988 Using direct and indirect measures to study perception without awareness. *Perception and Psychophysics* **44**: 563–75
- Sidis B 1898 *The Psychology of Suggestion*. Appleton, New York
- Vokey J R, Read J D 1985 Subliminal messages. Between the devil and the media. *American Psychologist* **40**: 1231–9
- Weiskrantz L 1995 Blindsight—not an island unto itself. *Current Directions in Psychological Science* **4**: 146–51
- Weiskrantz L, Warrington E K, Sanders M D, Marshall J 1974 Visual capacity in the hemianoptic field following a restricted occipital ablation. *Brain* **97**: 709–28

R. Reber and W. Perrig

Perceptrons

Perceptrons was the generic name given by the psychologist Frank Rosenblatt to a family of theoretical and experimental artificial neural net models which he proposed in the period 1957–1962. Rosenblatt's work created much excitement, controversy, and interest in neural net models for pattern classification in that period and led to important models abstracted from his work in later years. Currently the names (single-layer) Perceptron and Multilayer Perceptron are used to refer to specific artificial neural network structures based on Rosenblatt's perceptrons. This article references the intellectual context preceding Rosenblatt's work and summarizes the basic operations of a simple version of Rosenblatt's perceptrons. It also comments briefly on the developments in this topic since Rosenblatt.

Rosenblatt's first (partial) report in January 1957 was titled *The Perceptron: A Perceiving and Recognizing Automaton*. A subsequent report, in January 1958, titled *The Perceptron: A Theory of Statistical Separability in Cognitive Systems* was adapted for publication (Rosenblatt 1958). A compendium of research on perceptrons by Rosenblatt and his group is *Principles of Neurodynamics* (Rosenblatt 1962).

Rosenblatt's research was done in the context of: the neobehaviorism of Hull (1952), the general principles of neuromodeling of behavior proposed by Hebb (1949), the Threshold Logic Unit (TLU) neuron model of McCulloch and Pitts (1943), proposals for brain modeling based on neurophysiology (Eccles 1953) and symbolic logic, switching theory and digital computers (Shannon and McCarthy 1956), the stimJEUlus sampling models of Estes and Burke (1953), the linear operator probabilistic learning models of Bush and Mosteller (1955), and statistical classification procedures (e.g., Rao 1955).

The basic building block of a perceptron is an element that accepts a number of inputs x_i , $i = 1 \dots N$, and computes a weighted sum of these inputs where, for each input, its fixed weights β can be only +1 or -1. The sum is then compared with a threshold θ , and an output y is produced that is either 0 or 1, depending on whether or not the sum exceeds the threshold. Thus

$$y = \begin{cases} 1 & \text{if } \left(\sum_{i=1}^N \beta_i x_i \right) \geq \theta \\ 0 & \text{if } \left(\sum_{i=1}^N \beta_i x_i \right) < \theta \end{cases}$$

A perceptron is a signal transmission network consisting of sensory units (S units), association units (A units), and output or response units (R units). The 'retina' of the perceptron is an array of sensory elements (photocells). An S-unit produces a binary

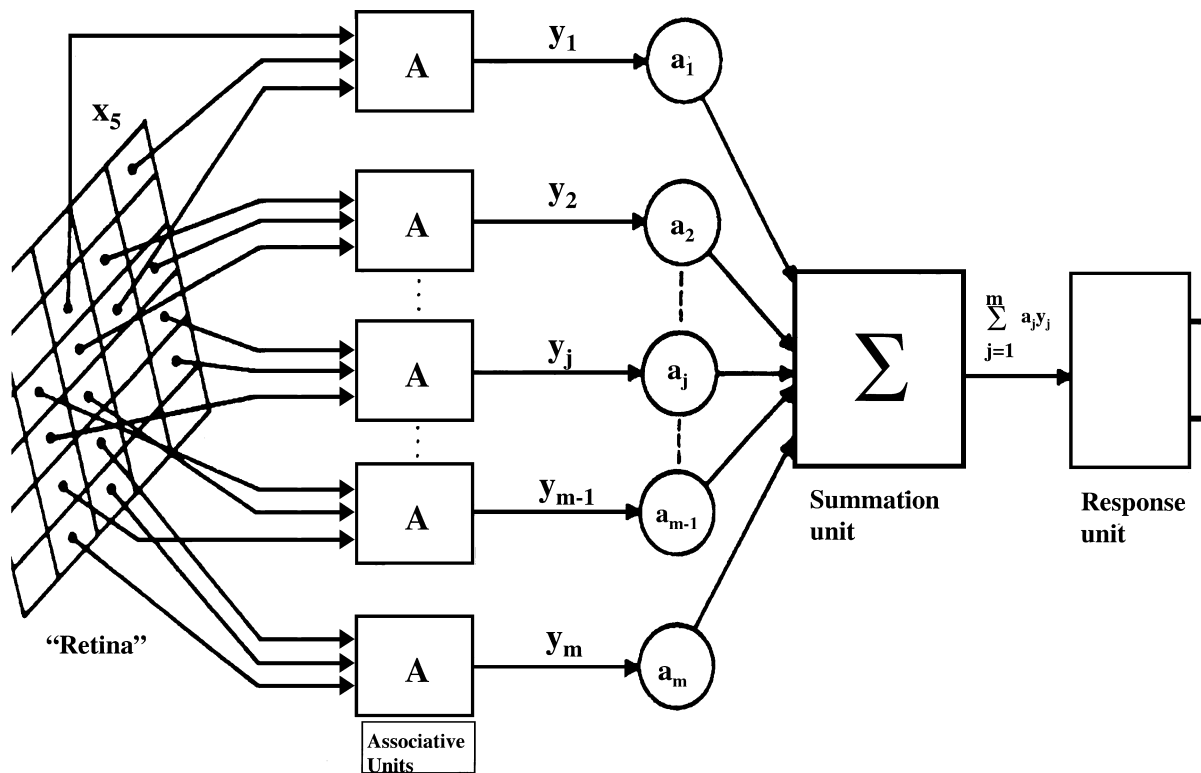


Figure 1
The mark 1 perceptron

output depending on whether or not it is excited. A randomly selected set of retinal cells is connected to the next level of the network, the A units. As originally proposed there were extensive connections among the A units, the R units, and feedback between the R units and the A units. A simpler version omitting the lateral and feed-back connections is shown in Fig. 1 This simplification is the perceptron considered here.

Each A unit behaves like the basic building block discussed above, where the $+1$, -1 weights for the inputs to each A unit are assigned randomly. The threshold θ for all A units is the same. The binary output of the k^{th} A unit ($k = 1, \dots, m$) is multiplied by a weight a_k , and a sum of all m weighted outputs is formed in a summation unit that is the same as the basic building blocks with all weights equal to $+1$. Each weight a_k is allowed to be positive, zero, or negative, and may change independently of other weights. The output of this block is again binary, depending on a threshold, t , that is normally set at 0. The binary values of the output are used to distinguish two classes of patterns that may be presented to the retina of a perceptron. The design of this perceptron to distinguish between two given sets of patterns involves adjusting the weights a_k , $k = 1, \dots, m$, and setting the threshold θ .

Rosenblatt (1962) proposed a number of variations of the following procedure for 'training' perceptrons. The set of 'training' patterns, that is, patterns of known classification, are presented sequentially to the retina, with the complete set being repeated as often as needed. The output of the perceptron is monitored to determine whether a pattern is correctly classified. If not, the weights are adjusted according to the following 'error correction' procedure: If the n -th pattern was misclassified, the new value $a_k(n+1)$ for the k -th weight is set to

$$a_k(n+1) = a_k(n) + y_k(n) * \delta(n)$$

where $\delta(n)$ is 1 if the n -th pattern is from class 1 and $\delta(n)$ is -1 if the n -th pattern is from class 2. No adjustment to the weight is made if a pattern is classified correctly.

Rosenblatt conjectured that, when the pattern classes are 'linearly separable' the error correction 'learning' procedure will converge to a set of weights defining the separating hyperplane that correctly classifies all the patterns. The shortest proof of this perceptron convergence theorem was given by A. J. Novikoff. Subsequent contributions related the simple perceptron to statistical linear discriminant functions

and related the error-correction learning algorithm to gradient-descent procedures and to stochastic approximation methods that originally were developed for finding the zeros and extremes of unknown regression functions (e.g., Kanal 1962).

The simple perceptron described is a series-coupled perceptron with feed-forward connections only from S units to A units and A units to the single R unit. The weights a_k , the only adaptive elements in this network, are evaluated directly in terms of the output error.

Minsky and Papert (1969) further reduced the simple perceptron to a structure with sampled connections from the 'retina' directly to the adjustable weights. This reduced structure is referred to as a *single-layer perceptron*. There is no layer of 'hidden' elements, that is, additional adjustable elements for which the adjustment is only indirectly related to the output error. A feed-forward perceptron with one or more layers of hidden elements is termed a *multilayer perceptron*. Rosenblatt investigated *cross-coupled perceptrons* in which connections join units of the same type, and also investigated *multilayer back-coupled perceptrons*, which have feedback paths from units located near the output. For series-coupled perceptrons with multiple R units, Rosenblatt proposed a 'back-propagating error correction' procedure that used error from the R units to propagate correction back to the sensory end. But neither he nor others at that time were able to demonstrate a convergent procedure for training multilayer perceptrons.

Minsky and Papert (1969) proved various theorems about single-layer perceptrons, some of which indicated their limited pattern-classification and function approximating capabilities. For example, they proved that the single-layer perceptron could not implement the Exclusive OR logical function and several other such predicates. Later, many writing on Artificial Neural Networks (ANN) blamed this book for greatly dampening interest and leading to a demise of funding in the USA for research on ANN's. The section on 'Alternate Realities' in Kanal (1992) details why the blame is misplaced. As noted there, by 1962 (see, e.g., Kanal 1962) many researchers had moved on from perceptron type learning machines to statistical and syntactic procedures for pattern recognition.

Minsky and Papert's results did not apply to multilayer perceptrons. Research on ANN's, biologically motivated automata, and adaptive systems continued in the 1970s in Europe, Japan, the Soviet Union, and the USA, but without the frenzied excitement of previous years, which also came back starting in the early 1980s. In a 1974 Harvard University dissertation Paul Werbos presented a general convergent procedure for adaptively adjusting the weights of a differentiable nonlinear system so as to learn a functional relationship between the inputs and outputs of the system. The procedure calculates the derivatives of some function of the outputs, with respect to all inputs and weights or parameters of the

system, working backwards from outputs to inputs. However, this work, published later in a book by Werbos (1994), went essentially unnoticed, until a few years after Rumelhart et al. (1986), independently popularized a special case of the general method. This algorithm, known as error backpropagation or just backpropagation, adaptively adjusts the weights to perform correct pattern classification using gradient descent and training samples. It propagates derivatives from the output layer through each intermediate layer of the multilayer perceptron network. The resurgence of work on multilayer perceptrons and their applications in the decades of the 1980s and 1990s is directly attributable to this convergent backpropagation algorithm.

It has been shown that multilayer feedforward networks with a sufficient number of intermediate or 'hidden' units between the input and output units have a 'universal approximation' property: they can approximate 'virtually any function of interest to any desired degree of accuracy' (Hornik et al. 1989). Several modifications of the basic perceptron learning procedure that make perceptron learning well behaved with inseparable training data, even when the training data are noisy and not error free, have been proposed in recent years and various ANN learning procedures have been shown to related to known statistical techniques. The reference list includes some recent books with tutorial material covering Perceptrons and related ANN's.

See also: Artificial Neural Networks: Neurocomputation; Connectionist Models of Concept Learning; Connectionist Models of Development; Connectionist Models of Language Processing; Neural Networks and Related Statistical Latent Variable Models; Neural Networks: Biological Models and Applications; Perception: History of the Concept

Bibliography

- Anderson J A 1995 *An Introduction to Neural Networks*. MIT Press, Cambridge, MA
- Bush R R, Mosteller F 1955 *Stochastic Models for Learning*. Wiley, New York
- Chauvin Y, Rumelhart D E (eds.) 1995 *Backpropagation: Theory, Architectures, and Applications*. Erlbaum, Mahwah, NJ
- Eccles J C 1953 *The Neurophysiological Basis of Mind*. Clarendon Press, Oxford, UK
- Estes W K, Burke C J 1953 A theory of stimulus variability in learning. *Psychological Review* 6: 276-86
- Haykin S 1999 *Neural Networks. A Comprehensive Foundation*, 2nd edn. Prentice Hall, Upper Saddle River, NJ
- Hebb D O 1949 *The Organization of Behaviour*. Wiley, New York
- Hornik K, Stinchcombe M, White H 1989 Multilayer feedforward networks are universal approximators. *Neural Networks* 2: 359-66

- Hull C L 1952 *A Behaviour System*. Yale University Press, New Haven, CT
- Kanal L 1962 Evaluation of a class of pattern-recognition networks. In: Bernard E E, Kare M R (eds.) *Biological Prototypes and Synthetic Systems*. Plenum Press, New York (reprinted in Sethi I, Jain A 1995 *Neural Networks and Pattern Recognition*. Elsevier, Amsterdam)
- Kanal L N 1992 On pattern, categories, and alternate realities, 1992 KS Fu award talk at IAPR, The Hague. *Pattern Recognition Letters* **14**: 241–55
- McCulloch W S, Pitts W A 1943 A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* **5**: 115–33
- Minsky M, Papert S 1969 *Perceptrons*. MIT Press, Cambridge, MA
- Rao C R 1955 *Advanced Statistical Methods in Biometric Research*. Wiley, New York
- Rosenblatt F 1958 The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* **65**: 386–408
- Rosenblatt F 1962 *Principles of Neurodynamics*. Spartan, New York
- Rumelhart D E, Hinton G E, Williams R J 1986 Learning internal representations by error propagation. In: Rumelhart D E, McClelland J L et al. (eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press, Cambridge, MA, Vol. 1, pp. 318–62
- Shannon C E, McCarthy J M (eds.) 1956 *Automata Studies*. Princeton University Press, Princeton, NJ
- Werbos P 1994 *The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting*. Wiley, New York

L. N. Kanal

Perceptual Constancy: Direct versus Constructivist Theories

Fortunately, our perception of the environment surrounding us is usually quite veridical, accurate, and stable. While laypersons take this ability of ours for granted, students of perception have pointed out that it is actually quite a feat, since the stimulation reaching our senses changes continuously. If we take the visual perception of size as an example, we find that the size of the retinal image of, say, another person changes with that person's distance from us. In spite of these changes we perceive that person as being of constant size and not shrinking or expanding with changes in distance. This 'feat' of perception has aroused a great deal of interest and theories have been proposed to account for it. Prominent among these are two contrasting views of how it is achieved. One view, the 'constructivist,' maintains that the perceptual system is 'intelligent.' In order to attain veridical perception it utilizes high-level mental processes to correct for the changes in stimulation reaching our senses. In contrast, the 'direct' approach claims that there is no need to posit such mental processes to explain veridical

perception, maintaining that the sensory information reaching our senses suffices to explain perceptual veridicality and that what is needed is a much deeper analysis of that sensory information.

1. Stability and Constancy in Perception

The ability of our perceptual system to overcome the effects of changes in stimulation and maintain a veridical and stable percept is called *perceptual constancy* (Epstein 1977, Walsh and Kulikowski 1998). In discussing the constancies it is common to make a distinction between the *distal* stimulus and *proximal* stimulus. The distal stimulus is the actual physical stimulus, the physically objective dimensions of the viewed object. The proximal stimulus, in the case of vision, is the very image that falls on the retina. It changes with changes in position or lighting of the physical stimulus. Constancy can be defined as the ability to correctly perceive the distal stimulus; that is, the stable properties of objects and scenes, in spite of changes in the proximal stimulus.

One example of constancy is *shape constancy*, which refers to the ability of the visual system to ascertain the true shape of an object even when it is not viewed face on but slanted with respect to the observer. In an example of shape constancy the distal stimulus might be a door half open, while the proximal stimulus will be the trapezoid shape it projects on the retinas. Shape constancy in this case would be the perception of the distal rectangular shape of the door in spite of the fact that the proximal shape is far from rectangular. Another example of constancy is *lightness constancy*. Lightness refers to the perceived reflectance of an object; high reflectance is perceived as white and very low reflectance as black, with intermediate reflectances as various shades of gray. We are capable of perceiving the distal lightness of a surface in spite of the fact that the proximal amount of light reaching our eyes changes with changes in the amount of light illuminating that surface. For example, if we hold a piece of chalk in one hand and a piece of charcoal in the other, the chalk will be perceived as white and the charcoal as black. This will be true if we observe the two in a dimly lit room or in bright sunshine, in spite of the fact that the amount of light reaching the eyes from the charcoal in the sunshine might be greater than that from the chalk in the dim room lighting.

The perceptual constancy that has received the greatest amount of research attention is *size constancy* (Ross and Plug 1998). If the distance is not very great we usually perceive the distal size of objects in spite of the changes in their proximal size as their distance from us changes. The two theoretical approaches mentioned above, the constructivist and the direct, deal with size constancy differently. According to the constructivists, size constancy is achieved through some process whereby the perceptual system perceives the object's distance and then takes the distance

11221