RUMELHART,D.E., McCLELLAND,J.L. and the PDP RESEARCH GROUP (Eds.)
*Parallel Distributed Processing - Explorations in the Micro-
structure of Cognition*, vol. I : *Foundations*, vol. II : *Psy-
chological and Biological Models*.   Cambridge, Mass. : MIT
Press, 1986.

Pre-History of PDP

The work reported on in the 2 volumes under review originates with
the McCulloch and Pitts'(1943) simulation of a highly simplified
neural network, whose elements are connected to each other by
excitatory or inhibitory fibres. Here, the units have Boolean
(1-0) values, and the connections have weights that take real
numbers as values lying within a fixed region (min.,max.). Mc-
Culloch and Pitts were able to show that all Boolean functions
could be simulated by well-selected connections: the units could
be divided in input units receiving stimuli from outside, output
units delivering stimuli to the outside world, and intermediary
ones in 1,2 or n layers. Although far removed from a realistic
model of the brain, about whose functioning our knowledge was poor
at that moment, this work was provocative enough to make people
think about learning processes by using such neuron-like networks.
Hebb, a neurologist, proposed the best known learning rule in 1949
which one could call the *Hebb rule*. In its simplest quantitative
version, it goes like this: "When A and B are simultaneously exci-
ted, increase the strength of the connection between them" (p.36,
Vol.1). Though biologically plausible, this rule is too simple;
nevertheless, its essence persists with important modifications
(to be discussed below)in many learning paradigms (p.152, Vol.1).
Hebb's proposal remained a speculation until Frank Rosenblatt
pioneered, in his *Principles of Neurodynamics* (1962), two methods
for examining the power of semi-neural networks provided with a
learning procedure: formal mathematical analysis, and digital
computer simulation. The formal mathematical analysis resulted in
the 'perceptron learning theorem' (I):

> 'Given an elementary α-perceptron, a stimulus world W, and
> *any classification C(W)for which a solution exists*; let all
> stimuli in W occur in any sequence, provided that each sti-
> mulus must reoccur in finite time; then beginning from an
> arbitrary initial state, an error correction procedure will
> always yield a solution to C(W) in finite time,..' (p. 154,
> Vol.1, *my emphasis*).

Roughly, the solution will be a division of the class of impact patterns into subclasses, whose members would be associated with the same output values when they belong to the same subclasses, and with different output values when they belong to different subclasses.

The classical problems of the theory of knowledge reappear in a new version by the fact that all learning rules must form a spectrum: *at one end*, there is an error-correcting teacher which tells how to change the values of the weights in accordance with the response of the model; *at the other*, completely unsupervised learning occurs. The necessary existence of rationalism (teacher) and empiricism (unsupervised modification) makes intermediary systems the only interesting ones. F.Rosenblatt introduced a rule for these perceptrons, which he called the C-type, that increments weights on lines active in the 1-set (patterns providing a 1 output) and decrements weights on lines active in the 0 set.

This did not work. Sooner or later all the input patterns were classified in one set, either the 1-set or the 0-set, and remained so. A less rigid decision rule was needed. Rosenblatt introduced mechanisms to limit weight growth in such a way that the set that was to be reinforced positively at active lines compensated the other by giving up some weight from all its lines. (A special case of this C' would be: lower the magnitude of all weights by a fixed fraction of their current values before incrementing the magnitude of some weight-lines.) C' was able to capture the similarity of *some* patterns and this was a major result.

However, in a 1959 conference, Rosenblatt became involved in a major controversy with Minsky and McCarthy. These AI specialists were only interested in simulating *what* the brain did, and did not think it necessary that 'Intelligent Machines' worked in a way *similar* to that of the brain . Rosenblatt, enthusiastic about what a brain-like neural network could do, overestimated the powers of his brainchild. In 1969, *Perceptrons* (Minsky and Papert) showed with mathematical necessity that a *one layered* perceptron (one input, one inside layer, one output layer) was very severely limited (the classes 'for which a solution existed' were too small).It could not even distinguish patterns with an even number of 1's from patterns with an uneven number of 1's (the problem is actually difficult because *very* similar patterns have to be distinguished); and could not separate connected from non-connected forms.

This 1969 result *"killed"* the Hebb-Rosenblatt tradition. It showed that a general learning theorem (similar to I) could *not* be proved for n-layered networks (n≥2), and suggested that for multilayered perceptrons (with or without feedback) no training rules

that could be close to the natural ones were in evidence.
  The pre-history of PDP ends in 1969.

Reasons for a Revival

And yet, the books under review begin where Rosenblatt left off,
and *show* - even though there is no general theorem analogous to I
at the moment - that many interesting tasks may be learned by
multilayered perceptrons. Obviously, the authors are aware that it
is only a progress report; but believe that the progress made so
far is both important and impressive. This reviewer agrees. Before
presenting some of the proposals together with their limitations,
and before inquiring into the philosophical importance of this
research, we need to understand why this radical reversal (so far
removed from the rule-directed, sequential, production process
automata current in AI) came about.

1. From psychology, both pattern recognition and language lear-
ning theorists remarked that in order to recognize a face or a
word, the simultaneous action of many feature-detectors (some
sensitive to angles, others to letters, others to syllables; some
sensitive to contours, some to slope, some to continuity, some to
area etc.) were needed. The sequential paradigm of "classical AI",
where one task must end before another could begin, introduced
too many complications. Moreover, the analyses of typing, memori-
zing and balancing all showed that even if frames, scripts or
schemata were non-serial concepts, their implementation needed the
simultaneous activation of widely diverse units in widely diverse
contexts.
2. From neuroscience came several important and general messa-
ges: neurons are slow (p.130, Vol.1: "the basic hardware of the
brain is 10^6 slower than that of serial computers"), and yet many
perceptual memory and reasoning tasks must be done in no more than
100 steps (a few hundred milliseconds). Besides, most neurons
probably compute very simple functions. So, the brain must work by
means of *massive parallelism* (made possible by 10^{10}–10^{11} neurons).
This large number also sets an upper limit for the number of ope-
rations that can be performed in an allotted time. However, the
brain is massively connected: an average cortical neuron may make
from 1000 to 100.000 synapses on the dendrites of other neurons.
This connectivity suggests statistical behavior *and* a small number
of layers (no neuron with more than 4 synapses away from another).
The units do not store knowledge: it is stored in the connection
strengths. And these connection strengths are not messages, but
simple, signed numbers with limited precision. This picture of the

brain also suggests that destruction of neurons only leads to a continuous decrease in the quality of functioning, while it would provoke disastrous discontinuities in serial computers. Luria (p. 135, Vol.1) stresses that neurosurgical data suggest distributed, not central control.

3. Even if the computer is no longer the inspiration, but the brain is, for the view that intelligent behavior is parallel distributed processing, one of its main tools is, precisely, *computer simulations* of the PDP hypotheses: developments in computer hardware itself make it more and more probable that huge parallel computers will be available to *test* the theories of the PDP. in fact, in the present work that moves so clearly away from the computer inspired branch of programming science, there is a chapter explaining how programs are to be written for parallel network simulating systems (ch. 13).

4. On the conceptual level, the fact that multi-layered linear or non-linear, feedback or feedforward networks can be *shown* to escape the restrictions of the Minsky-Papert result is certainly one of the main stimulants of the PDP revival – roughly 20 years after its supposed disappearance. (One may even say that it is *precisely* the Minsky-Papert precision and elegance, which *encouraged* the search for its limitations over the years.) Moreover, most psychologists and, I may add, most philosophers believe that a synthesis of empiricism and rationalism, of inneiism and experience is needed. The PDP prototype of the brain leads exactly to the same conclusion: the activation of the units, the strength of the connections between them, and the learning rule must be fixed, to a certain degree, from the beginning (rationalism, innatism). But different models may do this to a different extent and with different rigidity. *It thus becomes possible to experimentally determine the optimal mix of constraint and variability for specific tasks!* Philosophy becomes science.

Another philosophical prejudice works in *favor* of the PDP model. Radical reductionism of the properties of wholes to the properties of their parts has failed in most, if not all cases. On the other side, the current software/hardware dichotomy is not satisfactory because it sounds too much like the mind/body problem. In PDP approach, we overcome both reductionism and dualism. The simultaneous cooperation of millions of units is more than the sum of its parts: "We are simply trying to *understand* the essence of cognition as a property *emerging* from the *interactions* of connected units in networks"(p. 128,Vol.1). And elsewhere: "Distributed representations give rise to some powerful and unexpected emergent properties... For example, distributed representations are good for content-addressable memory, automatic generalization,

and the selection of the rule that best fits the current situation" (p.78, Vol.1)

Most important, it seems to me, is the possibility of simulating the creative action of the brain (creating new concepts) by means of these models. A local representation system must first *decide* if this is to be done, and then locate and organize another local unit to implement the concept; this procedure is possible, but needs too much time and place. However, in a distributed system all that is required is slight modification of the connections between interacting units in such a way that a new, *stable* pattern is added to the memory. This huge problem is not solved, but significant advances have been made: *different* concepts may now be represented by the *same* units through a change in their interactions.

These two very general considerations, together with the specific arguments already mentioned, make the enterprise, even if it is in its early stages, both exciting and rewarding.

The General PDP Model and
Some of Its Divergent Realizations.

All Parallel Distributed Models exhibit some common features. We need to explain them in order to show the extent to which the work that is being carried out advances in different directions. Each network needs eight major aspects for its definition:
1.   a *set of processing units*
2.   a *state of activation* for each of them
3.   an *output function* for each unit
4.   a *pattern of connections* between the units
5.   a *propagation rule* for propagating patterns of activities through the network
6.   an *activation* rule to compute from the impinging input, together with the present activation state, a new level of activation for the unit
7.   a *learning rule* whereby patterns of connectivity are modified by experience
8.   an *environment* within which the system operates.

The present reviewer considers such a network as being analogous to the set of connected automata, called "cellular automata" by John Von Neumann. All of these 8 features are also present there; but the elements of a cellular automaton are Von Neumann automata, and this is precisely what is rejected in PDP work. Nevertheless, a comparison might be worth while. Ironically, learning in cellular automata, though studied (Burks,Ed.,1970), is as absent from the current AI, as the PDP was until these two volumes

appeared, and perhaps for the same reasons.

A distributed representation is a representation where functionally macroscopic features of the system are *not* connected to one unit, but to the interrelations between n units (n≤N, where N is the total number of units).

In the 2 volume work that we analyze, cases of both distributed and non-distributed representations are used (p. 108-109, Vol.1; Ch.14, Vol.2). The main novelty with reference to Rosenblatt Perceptrons is the presence of many units that are hidden (p.48, Vol.1) in so far as they neither receive input from nor have impact on the external world. The main tool (though not the only one) for the functioning of PDP is the vector and matrix algebra. Chapter 9 of Volume 1 gives a clear exposition of linear algebra, and shows how the responses of the composite network may be represented - in the *linear* case (a very important restriction) - by matrix multiplication: "An input vector v is first multiplied by the matrix N to produce a vector z on the set of intermediate units;... and then z is multiplied by M to produce a vector u on the uppermost set of units...u=M(Nv)" (p.395-396, Vol.1; bold face indicates vectors or matrices).

For n layers, n matrices can be multiplied. Chapters 10 and 22 (vol.1 and vol.2 respectively) use non-linear mathematics. Even in these cases, however, the main tools are vector and matrix multiplication (see note 10, p. 430, vol. 2): if the system were highly nonlinear, the present approach might have no relevance whatsoever. The authors are well aware of the fact that strong nonlinearity might be the case with respect to neural models, but their high hopes are founded on the fact that so *much* can be explained by the well-understood linear vector and matrix algebra. This reviewer -for one- is quite struck by the fact that the eigenvalues and eigenfunctions of the network play as central a role in the learning processes of the brain (ch. 22, vol. 2) as they do in quantum mechanics, and that invariance under coordinate transformations relates representations of patterns on the higher and lower levels of the network (unit states-global states) in a way that is reminiscent of the role of coordinate transformations in relativity theory (p. 411-418, vol.2). Some nonlinearity would provisionally put an end to all that. But, even though nonlinearity is certainly there, quite a lot can be accomplished avoiding it.

Earlier on, I mentioned eight aspects involved in the definition of a network. It is now time to run through them briefly:

1. the N units can be arbitrarily *ordered* and called u_1, u_2... $(=u_i)$. There is no central director and units may be active together;

2. a vector of N real numbers, a(t) determines *the degree of activity of* u_i at moment t. The elements of a(t) are called $a_i(t)$ (for u_i).

Already here, a multiplicity of PDP models occurs: the $a_i(t)$ may be continuous or restricted to a finite number of values. If finite, the number of values may be only (0,1) (inhibited or excited), or (-1,0,+1), or (1,2...9) or any series of natural numbers. If, to the contrary, the activation values are continuous, they may range over all real numbers or be bounded. PDP research explores the consequences of these different possibilities.

3. units *transmit signals* to their neighbors. So with any state of a unit u_i at t, a function $f_i(a_i(t))$ must be associated which maps the present activation state $a_i(t)$ to an output signal $o_i(t)=f_i(a_i(t))$. The sequence of outputs (a unit has many neighbors) is given by a *vector*, o(t). Another classification, on the same lines as the earlier one under (2), may be given for the several o_i's. Superposed to this classification, models may also be distinguished by restrictions on the form of f. (In the simplest case $f=I$, and $o_i=a_i$; in the more general case, it may be a *threshold* function – only becoming active when it exceeds a threshold value – and a *bounded* function, not exceeding certain upper limits.) Finally f may be a deterministic function or a stochastic one, depending only in a probabilistic way on the activation intensities.

4. In the simplest case, each unit provides an additive contribution to its neighbors.(In other words, the total input to a unit is the weighted sum of the separate inputs from the neighbors it is connected to.) The sign of the input is positive when it is excitatory, negative when it inhibits. This is the linear case. However, many more complex rules are possible. (At this stage already, we may consider the case of different *types of connections*, whose inputs are computed in different ways: instead of adding inputs, inputs can also be multiplied. The functional importance of this lies in the fact that for $m \times n = r$, $r=0$ whenever m or n is regardless of the magnitude of the other term; and $r=m$, or $r=n$ if $n=1$ or $m=1$. Such a unit functions as a gate (p.73 and p.425, vol.1). The combination of an additive *and* multiplicative unit is called sigma-pi. Here linearity is left behind.)

Our two volumes do not mention that the cellular network is a graph; it is, and the graph-theoretic properties of the structure should be well worth studying. One of them, the fan-in and the fan-out of a unit, is mentioned (number of ingoing and outgoing lines to a unit).

5. Given two units, u_i and u_j, we need a rule which takes the output vector o(t), representing the output values of the units

and combines it with the connectivity of the matrix: net_{ij} is the net input of unit i to unit j. If many types of connections are introduced the total net_{ij} may get quite complicated. If only excitation or inhibition are distinguished, $net_e-W_eo(t)$ is the excitation for one specific j to i, $net_i=W_io(t)$ has the analogous meaning for inhibition.

6. This next state-function shows *how* $a_i(t+1)=F(net_i(t))$. The possible F's are legion. The linear bias of the book makes quasi-linear $a_i(t+1)=F(\Sigma_j w_{ij} o_j)$ quite frequent (where F is non decreasing). But differentiability (ch. 8) is also used as a constraint on this large family.

7. Modifying patterns of connectivity as a function of experience.
Here lies the aim of the whole enterprise. The network's task is to learn! It can do this by adding new connections, eliminating connections or modifying the strength of the existent ones. The first and the second types have not been studied much, as they appear to be limit cases of the third one.

Hebb's rule has been generalized as follows:
a) without teacher
$\Delta w_{ij} = \eta\, a_i o_j$ (where η stands for rate of learning, a_i is the activation of u_i and o_j is the activation of u_j.

b) with a teacher, the delta rule is the simpler case of
$\Delta w_{ij} = \eta\cdot(t_i(t)-a_i(t))o_j(t)$ (where a_i and o_j are the same as before, and $t_i(t)$ is the teaching input – the target activation provided by the teacher. (This is a direct generalization of Rosenblatt's rule for which the convergence theorem has been proved.)

c) the general formulation (covering the two earlier ones as special cases) is:
$\Delta w_{ij} = g(a_i(t),t_i(t))h(o_j(t),w_{ij})$
Meaning : the change in the connection between u_j and u_i is given by the product of two functions g and h. The function g depends on the activation of u_i, and the teaching input it receives (t_i); the function h depends on the output value of u_j, and the connection strength w_{ij}.

8. Each input is considered as a vector. Input patterns may be sets of orthogonal or linearly independent vectors, or may be sets of arbitrary vectors. (A vector v is linearly independent from a set of vectors $v_1...v_n$ if *no* series of scalars can be found such that $v=c_1v_1+c_2v_2 ... + c_nv_n$; two vectors v_1 and v_2 are orthogonal, if their inner product is zero.) These conditions may be combined with the psychological properties of irrelevance (neither systematically similar nor dissimilar) and of opposition.

The reader will find these 8 points the essential content of

vol. 1, ch. 2. In order to convey the excitement the authors felt during this study, it seemed important to this reviewer to describe the immense field of possibilities that the study of these networks yields by pointing out the many divergent models that can be constructed on the basis of the seemingly simple neuron-style idea. It also explains why there is no super-model available at the moment which approximates both the structure *and* the function of the human brain. Each model has its positive and negative features, and is put forward for specific and special tasks.

Progress in this uncharted field is slow. Obviously, networks that indulge in associative learning (after n times, the input units and the output units having been submitted to similar *pairs* of patterns, the network *acquires* the capability of producing the output when confronted with the input) are *less* important than networks that learn to respond to interesting patterns in the environment; yet both are needed (the first store relations, the second detect features). Equally obviously, a pattern association program with teacher, in which only certain weights can be modified, is less interesting than an auto-association process in which all connections can be modified, and where, without the teacher, portions of the input will -if learning occurs- evoke the remainder. Three unsupervised models of the second type are presented in ch. 5,6,7.

Strength and limitations.

The book is eminently readable and clear. Mathematical concepts are clearly explained. In the second volume, the authors show both their results and their modesty.

1. By including two chapters about what is known and what is not about the neo-cortex in the neurosciences, an attempt is made to show how far the presented models still are from the real neo-cortex. Not only that. They also indicate possible directions for future inquiry. F. Crick, the co-discoverer of DNA is the coauthor of one of them; and T.J. Sejnowski, author of the other, emphasizes those open areas in which computational questions might be asked about the neo-cortex. Both authors are aware that they are far from their goal.

D.A. Norman, working on cognitive science outside of the PDP group, concludes the book with an evaluation of its weaknesses and its strengths.

As one would expect, the learning processes come under fire. How *do* the learning laws arise? The noteworthy absence of any reference to *the developmental* origin (Piaget's problem) is highlighted; the type-token relation, although studied several times

(ch. 2, ch. 3, ch. 16), is not easily solved while it is easy do
so using semantic networks and frame based classical approaches;
symbol processing systems are fundamentally variable interpre-
tation processes (p. 540, vol. 2), but PDP networks lack vari-
ables. Furthermore, at any one moment the system reaches *one* glo-
bal equilibrium: it seems necessary to introduce a multiplicity of
PDP systems; and the *need* for an external trainer seems to require
either other persons *or* an extra, evaluative, (eventually sequen-
tial) system. Norman calls it deliberate conscious control (DCC).
These two last objections can be turned into major strengths how-
ever. Both the importance of social contact and of self-awareness
by introspection has been underlined recently. If they could be
combined with the smooth instantaneous functioning of PDP (res-
ponsible for perception, memorization, automatic behavior), the
interaction of subconscious with conscious mechanisms, of proces-
sing and evaluative structures appear to follow naturally. Indeed,
"a nice fallout"(p. 546, vol.2; Norman's words!).

2. The strength of the approach is shown, together with its
limitations, by applying it to amnesia (vol.2, ch. 25), place
recognition and goal location (vol. 1, ch. 23), sentence pro-
cessing (vol.2, ch. 19) learning the past tense (vol.2,ch. 18),
a distributed model of memory (vol.2,ch. 17), the trace model for
phoneme identification (vol.2, ch. 25).

To this philosopher, there are other attractions which caused
him to be interested in it in the first place. Recently, a move-
ment called 'evolutionary epistemology', initiated by J. Piaget,
K. Lorenz and D. Campbell, caused a minor revolution in epistemo-
logy. Knowledge as biological adaptation seemed a fine idea, but
a biologically evolved model of the knowing subject was lacking.
This gap −according to me− seems to be on the verge of being fil-
led. *I suggest a synthesis of evolutionary neurology with PDP* as a
natural next step for the evolutionary epistemologist. As an enti-
cement to do so, he could find in vol.2 (ch. 14) a PDP interpre-
tation of "schema" − an ubiquitous idea starting with Kant, essen-
tial for Bartlett, central in Piaget and the general idea behind
scripts and frames, explicitly taken up again, since 1975, by
Bobrow, Norman and Rumelhart. Its implementation here is that a
schema is a global set of constraints on the connection which,
under given context and input, maximizes the goodness of fit of
the internal network to the external environment. Schemata *are*
stored in memory as 'sets of connection strengths' (p. 21, vol.2);
an example of a room−schema is worked out; sub−schemata exist
(small configurations of units embedded in a schema); schemata
represent knowledge on all levels, they are active processes,
having slots or default values. I suggest a study of the degree to

which the most worked out schema dynamics (Piaget, 1976) is compatible with this concept of schema. The idea of interiorization (=mental models), essential for Piaget's schema, is anyway already suggested (p. 41, vol.2): instead of one PDP, the authors introduce two (p. 40-42, vol.2): the first receives inputs and relaxes to an appropriate equilibrium state which will, in turn, change the inputs to the system. The other piece of the system is similar in nature, except that it is a model of the world on which we are acting. This consists of a relaxation network, which takes as input some specification of actions to be carried out, and produces an interpretation of "what would happen if we did that"(=how the inputs would change). This first philosophical suggestion induces one lingering doubt: Will the so defined schemata not be *too* fluctuating and dependent on the inputs to represent our *relatively* independent intellectual life? To doubt is, however, to ask; not to reject.

The second philosophical attraction regarding PDP's offer is a step forward in the study of the mind–body relation. Its core is described in vol.2, ch. 22. Precisely by viewing PDP models as dynamical systems (p.397,vol.2), one transcends the computational perspective to reach the systems–theoretical one.

P. Smolensky, starting out from the central PDP equation :
$u_v(t+1)=F[\Sigma_\mu W_{v\mu}G(U_\mu(t))]$ (where F is a nonlinear sigmoid function from inflowing activation to a unit to the new activation of the unit, and G is a nonlinear threshold function) states (p.398) that the *knowledge* contained in the PDP model is stored in the connection strengths $\{W_{v\mu}\}$ or the weight matrix W. For any dynamical system, the set of its possible states (state space), and the set of its possible trajectories (pathspace) with eventual stable paths at the limit, have to be considered. Abstracting from the nonlinearities F and G, the purely linear
$u_v(t+1)=\Sigma_\mu W_{v\mu}u_\mu(t)$.

Smolensky studies the kinematics (geometry of the state space as a vector space) u(t+1)= W u(t), that can be parametrized both by pattern coordinates, and by unit coordinates. He proves that "the evolution equation for the linear model has exactly the same form in the pattern coordinate system as it has in the unit coordinates"(p.411,vol.2). This result is connected with the mind–body problem in two *ways*: 1. If mind is to be viewed as a higher level description of the brain, one possibility is that the higher level describes *collective* activity of the lower level. (p. 429 vol.2). The isomorphism of levels hypothesis constitutes a mathematically analyzable question about mind/brain duality (p.431, vol.2). 2. More importantly, there is a second kind of isomorphism involved as the following table makes it clear : (p. 393, vol.2).

The Mappings from the Mathematical World to the Neural and Conceptual Worlds

Neural	Mathematical	Conceptual
neurons	units	hypotheses
spiking frequency	activation	degree of confidence
spread of depolarization	spread of activation	propagation of confidence: inference
synaptic contact	connection	conceptual-inferential interrelations
excitation/inhibition	positive/negative weights	positive/negative inferential relations
approximate additivity of depolarizations	summation of inputs	approximate additivity of evidence
spiking thresholds	activation spread threshold G	independence from irrelevant information
limited dynamic range	sigmoidal function F	limited range of processing strength

To be sure, as Smolensky with characteristic modesty shows, linearity is only approximate. Until now, no way is found to preserve the isomorphism for nonlinearity. (This itself is essentially connected to the necessary competition of distributed representations.) But —at long last— the patterns of functioning of the connection matrix are *both* suitable objects for the predicates of self-awareness, *and* intelligibly linked to the activity of neurons.

And all of this is achieved within the limits of general dynamic systems theory close to a model that suggests the application of evolutionary laws.

To insist on the philosophical promise of this collective inquiry is a suitable conclusion for a review in a philosophical journal.

Bibliography

A. Burks (Ed.), 1970,   *Essays on Cellular Automata*.
   Urbana:University   of Illinois Press.
D.O. Hebb, 1949, *The Organization of Behavior*.
   New York: Wiley.
McCulloch, W.S. and Pitts, W., 1943 "A Logical Calculus of the
   Ideas Immanent in Nervous Activity."
   Bulletin of Mathematical Biophysics, 5, 115–133.
M. Minsky and S. Papert, 1968, *Perceptrons*.
   Cambridge, Ma.: M.I.T. Press.
J. Piaget, 1975,   *L'équilibration des Structures Cognitives* :
   Problème Central du Développement.
   Paris: P.U.F.
F. Rosenblatt, 1962, *Principles of Neurodynamics*.
   New York: Sparton.

Leo Apostel