THE CONCEPT OF E-MACHINE: ON BRAIN HARDWARE AND THE ALGORITHMS OF THINKING

Victor Eliashberg

Varian Associates, Palo Alto

The goal of this paper is to call in question the popular thesis that the problem of the algorithms performed by the brain (algorithms of thinking) has but little to do with the problem of brain hardware. The paper presents a simple example of a "brain-like" universal associative processor (referred to as E-machine) for which such a thesis would be obviously inadequate. More sophisticated examples of E-machines were studied in Eliashberg (1979).

1. A SIMPLE EXAMPLE OF E-MACHINE

Consider the "neural" network schematically shown in Fig. 1. The big circles with incoming and outgoing lines represent centers, elements which are assigned certain discrete coordinates in the network. The small circles denote couplings, elements whose place in the network is determined by a pair of coordinates of centers. A center may be viewed as a neuron with its dendrites and axon or a neural subsystem that may be treated as a "reduced neuron". A coupling may be interpreted as a synapse or a "reduced synapse", the white and the black circles corresponding to the excitatory and inhibitory synapses respectively. Using the terminology of Nauta and Feirtag (1979), the network of Fig. 1 may be characterized as a three-neuron nervous system (it has three neurons in a path between its input and output). Some simple animals have two- and even one-neuron nervous systems, so the network of Fig. 1 may be viewed as a morphological model corresponding to a rather advanced stage of the evolution of the brain. Accordingly, one may expect to get some interesting information processing characteristics in a neurobiologically reasonable functional model of this network. Therefore to reach the goal of this paper it seems sufficient to show why the traditional brain-hardware-independent approach to the problem of the algorithms of thinking would fail to adequately describe the psychological properties of an animal with such a simple nervous system. So much the more may this approach be inadequate in the case of human brain.

NOTATION.

$N_j(i)$ is the i-th center from the set N_j.

$S_{kj}(i,i')$ is the (i,i')-th coupling from the set S_{kj}.

ν is discrete time (the number of cycle).

$x(\cdot,\nu)$ is the input vector of the model. (The dot substituted for an index implies that the whole set of components corresponding to this index is assumed).

$G^x(\cdot,i)$ is the vector of gains of couplings $S_{21}(i,\cdot)$. This vector will be interpreted as the symbol stored in the i-th location of input long-term memory (ILTM) of the model.

$E(i,\nu)$ is a variable describing a hypothetical state of "residual excitation" (E-state) associated with center $N_2(i)$. Such a state might be interpreted as a certain phenomenological counterpart of the concentration of a chemical participating in a slow reversible reaction. Accordingly, in a more sophisticated model one may introduce several types of E-states. Such states will be viewed as the states of distributed "non-symbolic" short-term memory of the model.

$U_2(i,\nu)$ is the (postsynaptic) potential of center $N_2(i)$.

$J_2(i,\nu)$ is the output signal of $N_2(i)$.

$G^y(\cdot,i)$ is the vector of gains of couplings $S_{32}(\cdot,i)$ interpreted as the symbol stored in the i-th location of output long-term memory (OLTM) of the model. For the sake of simplicity we will assume that the state of long-term memory (ILTM and OLTM) is formed before the first moment of observation and doesn't change later, i.e. we will avoid the problem of learning.

$y(\cdot,\nu)$ is the output vector of the model.

In a rather general form a functional model of the network of Fig. 1 with the above input, state, and output variables may be described as the following machine.

$$y(\cdot,\nu) \leftarrow F_y(x(\cdot,\nu),E(\cdot,\nu),G^*(\cdot,\cdot))$$

$$E(\cdot,\nu+1) \leftarrow F_e(x(\cdot,\nu),E(\cdot,\nu),G^*(\cdot,\cdot)), \text{ where}$$

F_y is the otput procedure, F_e is the next E-state procedure, $G^*=G^x,G^y$. As it was mentioned above in this paper we are not concerned with the next G-state procedure and treat the variable $G^*(\cdot,\cdot)$ as a parameter.

For the goal of this paper it is sufficient to make rather simple assumptions about F_y and F_e. What is important for this goal is the presence of E-states rather than the details of their interaction with the input vector and the G-state and the details of their dynamics. With that in mind let us introduce the following "quasi-neural" description of F_y and F_e. (The reader with an appropriate background will be able to find many other descriptions of these procedures satisfying the requirements of this paper).

OUTPUT PROCEDURE, F_y :

The potential $U_2(i,\nu)$ is the following function of $x(\cdot,\nu),E(i,\nu)$ and $G^x(\cdot,i)$

$$(1) \qquad U_2(i,\nu) \quad S(i,\nu)\cdot[1 + \alpha\cdot E(i,\nu)], \text{ where}$$

$$(2) \qquad S(i,\nu) \quad \sum_{(j)} G^x(j,i)\cdot x(j,\nu) \ , \quad i=1,\ldots n_2$$

The layer of centers N_2 with lateral inhibitory couplings S_{22} performs the random equally probable choice of a center, $N_2(i_o)$, from the subset of centers with the maximum potential (Eliashberg, 1969, 1979). It is assumed that there is some noise.

$$(3) \qquad J_2(i,\nu) = \begin{cases} 1 & \text{if } i=i_o \\ 0 & \text{----} \end{cases}, \text{ where}$$

$$(4) \qquad i_o :\in M(\nu) = \{i / U_2(i,\ ) = \max_{(i)} U_2(i,\nu) > 0\}$$

We are using ALGOL-like notation, $:\in$, to denote the operator of random equally probable choice of an element from a set.

The output vector is determined as follows

$$(5) \qquad y(\cdot,\nu) \quad \sum_{(i)} G^y(\cdot,i)\cdot J_2(i,\nu)$$

NEXT E-STATE PROCEDURE , F_e :

The dynamics of E-state is described by the first order difference equation, the time constant, $\tau(i)$, of this equation depending on whether $E(i,\nu)$ increases, $\tau(i)=\tau^+$, or decreases, $\tau(i)=\tau^-$.

(6) $\tau(i)\cdot[E(i,\nu)-E(i,\nu)] = S(i,\nu)-E(i,\nu)$,

where

(7) $\tau(i) = \begin{cases} \tau^+ & \text{if } S(i,\nu) > E(i,\nu) \\ \tau^- & \text{if } S(i,\nu) \leq E(i,\nu) \end{cases}$

In what follows the system described by Exps (1) - (7) will be referred to as Model 1.

2. SOME GENERAL INFORMATION PROCESSING CHARACTERISTICS OF MODEL 1

2.1. Universality with respect to the Class of Combinatorial Machines

Let $x(\cdot,\nu), G^x(\cdot,i) \in X$, where X is a finite set of positive normalized vectors. Let $y(\cdot,\nu)$, $G^y(\cdot,\nu) \in Y$, where Y is a finite set of positive vectors. Let the number of locations of LTM be as big as required ($n_2 \to \infty$), so any desired software , $G^x(\cdot,\cdot), G^y(\cdot,\cdot)$, may be put into the LTM of Model 1 before the beginning of observation. Let $\alpha = 0$ (the mechanism of STM of Model 1 is "turned off").

It can be shown that in this case Model 1 can be programmed to simulate an arbitrary probabilistic combinatorial machine (with rational probabilities) with the input alphabet X and the output alphabet Y.

2.2. Universality with respect to the Class of Finite-State Machines

Let us split the input and output vectors of Model 1 each into two subvectors $x(\cdot,\nu) = (x_1(\cdot,\nu), x_2(\cdot,\nu))$, $y(\cdot,\nu)$ $(y_1(\cdot,\nu), y(\cdot,\nu))$, and let us introduce the delayed feedback $x_2(\cdot,\nu+1)$ $y_2(\cdot,\nu)$.

Let $x_1(\cdot,\nu) \in X$, $x_2(\cdot,\nu), y_2(\cdot,\nu) \in Q$, $y_1(\cdot,\nu) \in Y$, where X and Q are finite sets of positive normalized vectors, Y is a finite set of positive vectors. Let as before $\alpha=0$.

It can be verified that this modification of Model 1 can be programmed to simulate any probabilistic finite state machine (with rational probabilities) with input alphabet X, state set Q, and output alphabet Y.

2.3. E-States as the Mechanism of Mental Set

Let $\alpha>0$ (the mechanism of STM of Model 1 is "turned on"). Let us assume at first that the time constants are very big ($\tau^+ \to \infty$, $\tau \to \infty$) so the E-state of Model 1 does not change considerably during the interval of observation. Let other conditions be as in section 2.1.

Suppose the LTM of Model 1 contains all possible input/output pairs (associations) from $X \times Y$, i.e., for all $(a,b) \in X \times Y$ there exists $i \in \{1,..n_2\}$ such that $G^x(\cdot,i)=a$ and $G^y(\cdot,i)=b$. It can be shown that for any (deterministic) combinatorial machine with the input alphabet X and the output alphabet Y there exists an initial E-state, $E(\cdot,0)$, of Model 1 such that this model in this state simulates the above machine.

Thus being observed as a black box, Model 1 with a fixed state of its LTM may appear to an experimenter as any of m^l deterministic combinatorial machines ($l=|X|$, $m=|Y|$) depending on the "state of mind" ("mental state"), $E(\cdot,\nu)$, of this model. This result can be naturally extended to the class of (deterministic) finite-state machines in the case of the Model 1 with the feedback of section 2.1.

Let us remove the condition of infinite time constants but still assume that these constants are rather big. In this case the E-state of Model 1 will change slowly ("adiabatically") so this model will gradually change into different combinatorial machines. Thus the "non-symbolic" expressions (6),(7),determined at a hardware level, controll the "personality" of Model 1 as a symbolic machine.

2.4. Why Model 1 Jeopardizes a Brain-Hardware-Independent Approach to the Problem of the Algorithms of Thinking

Imagine a researcher trying to develop a theory of the behavior of Model 1 per se (Model 1 treated as a black box) without being concerned with the hardware phenomena in this model. Let us assume that the researcher is used to work with a hardware-independent higher level language, say LISP, and deal with traditional symbol manipulation concepts. It is very likely that Model 1 would fool such a researcher by pretending to behave as a "classical" symbolic machine. The researcher with the above mentioned methodological mental set and the knowledge base associated with the "classical" symbolic information processing paradigm would have little chance to think about the "non-symbolic" E-states of Model 1, much less to find the hardware expressions (6),(7) describing the transformations of these states. It is not difficult to imagine how a sophisticated chemistry of the brain (see, e.g., Iversen, 1979, Kandel, 1979) may lead to non-symbolic brain hardware expressions much more complex than Exps (6),(7). Thus the above mentioned researcher would hardly have a better chance to come up with an adequate theory of human mental states than he (she) does in the case of Model 1.

3. METHODOLOGICAL REMARKS

3.1. On the Whole Brain and the Parts of its Behavior

Let (A,s_o) be a hypothetical machine corresponding to the human brain, A, in its initial (roughly newborn) state, s_o. After several years of learning (A,s_o) is changed into an intelligent system (A,s_n). Model 1 and the more sophisticated examples of E-machines studied in the Eliashberg (1979) manuscript give reasons to believe in the following methodological thesis.

There may exist a relatively simple description of (A,s_o) in terms of a machine with the state set similar to that of the brain. There may be a good possibility to find this description by trying to answer in a single context a large enough set of specially selected basic neurobiological and psychological questions. At the same time it may be hopeless a strategy to try to find adequate descriptions of some nontrivial "parts" of the behavior of (A,s_n) without looking for a description of the "whole" system (A,s_o).

3.2. On the General Relation between the Theory of the Brain and the Information Processing Psychology

To clearly understand the implied methodological meaning of the above mentioned thesis it is useful to compare the general relation between (A,s_o) and the information processing psychology with the general relation between the basic equations of a traditional physical theory and this theory. As an example of the latter relation let us take the Maxwell equations and the classical electrodynamics. (Don't take this metaphor too literally!).

The relatively simple Maxwell equations allow one to describe all variety of arbitrary complex classical electromagnetic phenomena as a result of interaction of these equations with the correspon-

ding variety of "external worlds" represented in the form of various boundary conditions, media, and sources. To find the Maxwell equations did not mean to solve all the problems of classical electrodynamics, the latter having developed (and continuing to do so) a number of specific problem-oriented models and concepts. It would hardly be possible, however, to adequately formulate all these specific problems, let alone solve them, without the Maxwell equations.

3.3. On Skipping "Simple" Problems in order to Solve Complex Ones Faster

For the goal of this paper it is especially important to emphasize that the Maxwell equations were found in an attempt to adequately formalize and extrapolate some "simple" basic knowledge about electromagnetic phenomena (Faradey law,etc.). This formalization and extrapolation created a powerful mathematical tool allowing to adequately approach complex problems of classical electrodynamics.

Imagine a physicist trying to develop a computer program for simulating the behavior of electromagnetic field in a complex microwave device without being concerned with the equations describing the fundamental properties of this field. A researcher trying to skip "simple" basic neurobiological and psychological questions, in order to solve complex problems of human information processing faster, may well be in a situation similar to that of such a physicist.

ACKNOWLEDGEMENTS

I want to express my gratitude to Prof. H. Landahl, Prof. J. McCarthy, Prof. H. Martinez, Prof. I. Sobel, Prof. L. Stark, and Prof. L. Zadeh for useful comments.

REFERENCES

Eliashberg, V.M. On a class of learning machines. Leningrad, USSR, Proc. of VNIIB, 54, 1969 (in Russian).

Eliashberg, V. The Concept of E-machine and the Problem of Context-Dependent Behavior. Palo Alto, Ca., 1979, Copyright 1980, by V. Eliashberg, TXu 40-320, US Copyright Off.

Iversen, L.L. The chemistry of the brain. Sc. American, Sept. 1979.

Kandel, E.R. Small systems of neurons. Sc. American, Sept. 1979.

Nauta, W. & Feirtag, M. The organization of the brain. Sc. American, Sept. 1979.

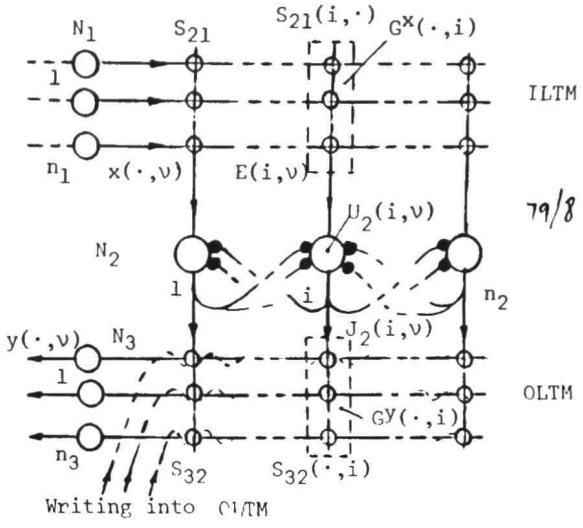

Fig. 1