A Comparison of Natural and Artificial Intelligence

A. Harry Klopf
Air Force Avionics Laboratory
Wright-Patterson Airforce Base, Ohio 45433

This paper was motivated, in part, by Price's recent (1975) tutorial comparing human and computer visual systems. While I found his comparisons interesting, a fundamental question continued to trouble me, as it has now for a number of years. Namely, does artificial intelligence (AI) really bear any fundamentally important relationship to natural intelligence? Or are the two sufficiently different so that a comparison of human and computer visual systems, for example, might be of the same significance as a comparison of locomotion in four-legged animals and four-wheeled vehicles?

My own conclusion is that artificial intelligence doesn't today bear much of a relationship to natural intelligence. I suggest this not because I think intelligence can't be mechanized but because I think the required mechanisms may be quite different from those AI researchers are considering. For two decades now, artificial intelligence research has sought increasingly powerful information processing systems through the development of new forms of software, sensors, and effectors for the digital computer. In recent years, this research has frequently been assessed in terms of such questions as whether the research has accomplished much, what in principle can be accomplished, who should accomplish it, and whether AI is dangerous (see, for example, Dreyfus, 1965, 1972; Lighthill, 1973; Roszak, 1972 and Weizenbaum, 1972). I would like to address a different set of questions that arise out of a new view of natural intelligence. The purpose of this paper is to briefly describe this new view and to examine its implications for the AI paradigm.

An Adaptive Network Approach to Intelligence

Disenchantment with the adaptive network approach to intelligence has grown over the years to the point where now the AI paradigm virtually excludes such an approach. Simple perceptrons, perhaps the best known class of adaptive networks, have been shown to be of very limited capability (Minsky and Papert, 1969). While the need to better understand adaptive networks is still acknowledged, few AI researchers are any longer excited by the possibilities. In this paper, I want to suggest that this virtual rejection of the adaptive network approach is premature; that, in fact, the adaptive networks of the 50's and 60's failed because of a fundamental deficiency that has been carried over intact into the AI systems of the 70's.

The deficiency to which I refer may be simply stated: *we have been trying to build goal-seeking systems out of non-goalseeking components*. This has been the implicit strategy of both adaptive network and AI researchers. An alternative strategy for the design of intelligent systems would be to *build goal-seeking systems out of goal-seeking components*. More specifically, it is suggested that specialized system goals be built upon a foundation of generalized component goals. What I mean by specialized goals are such goals as theorem proving, speech recognition and game playing. What I mean by generalized goals are such goals as maximal positive reinforcemtne or minimal negative reinforcement where *all* inputs are reinforcers. The theory described below should help to clarify these definitions.

To illustrate the approach I am advocating and to explain why I think it will work, I will briefly describe a new view of natural intelligence. This view suggests that natural intellignece, with its specialized global goals, is built upon a foundation of generalized goals implemented at a local level. Such a theory has been described in detail in Klopf (1972) and in summary form in Klopf (1974). I will do little more in this section than present the abstract from the latter reference. The approach I have proposed is one I refer to as a heterostatic theory to distinguish it from the homeostatic theories that have been widely considered beginning with Claude Bernard's work in 1895. Whereas homeostatic theories suggest that the goal os living systems is a steady-state condition, the heterostatic theory to be summarized below postulates that the goal of living systems is a *maximal* condition. Specifically, the heterostatic theory suggests that the following two assumptions are central to an understanding of brain functions:

1. The individual neuron is a goal-seeking system.
2. The goal of the neuron is to maximize the amount of excitation and minimize the amount of inhibition that it is receiving. (It is proposed that the neuron seeks this goal with a selfcontained adaptive mechanism that is a simple embodiment of B. F. Skinner's (1938) operant conditioning procedure.)

These assumptions lead to the view that a brain, in essence, is a collection of neurons that are striving to obtain one type of signal (excitation) and to avoid another type (inhibition). It is hypothesized that the collective behavior of these neurons emerges at the psychological level as a striving to obtain pleasure and to avoid pain. At the sociological level, a similar pattern appears to be repeated. Brains and societies are seen to be fundamentally similar adaptive systems.

The proposed neuronal model yields a theory that is consistent with the experimental data of neurophysiology and psychology. Neuronal and cortical polarization studies, the mirror focus and epileptic foci, habituation, dishabituation, classical and operant conditioning, and extinction appear to be understandable in light of the proposed model (see Klopf, 1972).

The above is fairly specific statement of the nature of the proposed theory. It is instructive to consider the theory in more general terms, especially for the purpose of contrasting it with the currently accepted paradigm. Ever since the discovery of the neuron, most basic research into the mechanisms of intelligence has rested on two (sometimes implicit) premises:

1. Complex goal-seeking systems are composed of simple non-goal-seeking components.
2. The goal of living systems is homeostasis (a steady-state condition).

The writer believes both premises are false and that this in part explains why our hard won and vast accumulation of data has not yet yielded a unifying theory. Today much is known about the functioning of nervous systems and neurons but the adaptive properties of brains remain mysterious. Such a condition is suggestive of the need for a new framework within which the current accumulation of data can be freshly examined. Generally, new frameworks are obtained only by considering new assumptions.

1. Complex goal-seeking systems are composed of simple *goal-seeking* components.
2. The goal of living systems is *heterostasis* (a maximal condition).

This pair of premises appears to yield a theory of brain function (and more generally a theory of adaptive systems) consistent with the available evidence (see Klopf, 1972).

Historically, viewing the neuron as a goal-seeking element goes back to 1895 and the work of Frued. Frued proposed that CNS neurons, in their "primary function," seek to minimize tha amount of excitation received. Frued also suggested that a "secondary process" occurred that maintained a low, steady-state level of excitation required for basic processes such as respiration.

NATURAL vs ARTIFICIAL

Cybernetic research into adaptive networks grew out of the early work of investigators such as Rashevsky (1938) and McCulloch and Pitts (1943). Adaptive network research by Minsky (1954), Farley and Clark (1945) and Rosenblatt (1957, 1960, 1962) moved toward local elemental goals when they initiated investigations into networks of elements that received positive and negative reinforcement signals. However, a global reinforcer was always the source of the reinforcement signals and only one or two of each element's inputs were reinforcing. A true generalized elemental goal was considered by Griffith (1962, 1963) who proposed a network element that sought to minimize a "power function" of its inputs and outputs. As an alternative to Griffith's neuronal model, Wilkins (1970) has proposed a generalized goal-seeking element that seeks to minimize the "total information exchange during a computation... consistent with the computation." Neuronal models like those proposed by Griffith and Wilkins will have to be thoroughly understood if we are to develop a general theory of adaptive networks. However, the apparent appetite of the single neuron and the whole organism for positive reinforcement (see Klopf, 1972) is not consistent with Griffith"s and Wilkins' views of living systems as minimizers.

Implications of the Heterostatic Theory for the AI Paradigm

From the perspective of the heterostatic theory a particular assumption of AI researchers emerges as being of crucial importance. It is a negative assumption that relates to a number of AI attributes. In Nilsson's (1974) survey of AI research, the assumption is stated as follows:

> ...knowledge about the structure and function of the neuron--or any other basic component of the brain--is irrelevant to the kind of understanding of intelligence that we are seeking. So long as these components can perform some very simple logical operations, then it doesn't really matter whether they are neurons, relays, vacuum-tubes, transistors, or whatever.

Implicit in this statement is the assumption that complex goal-seeking systems consist of simple nongoal-seeking components. Note that neurons are plced in a class with "relays, vacuum tubes, transistors, or whatever." Having unknowingly downgraded the neuron, AI is then able to dismiss it as irrelevant. If the neuron is in fact a goal-seeking system in its own right, then Nilsson's (1974) introductory statement about AI needs to be considered in a new light:

> The field of Artificial Intelligence (AI) has as its main tenet that there are indeed common processes that underlie thinking and perceiving, and furthermore that these processes can be understood and studied scientifically.

The adaptive network theory outlined above suggests that the heterostatic neuron represents the fundamental process underlying thinking and perceiving. If this is so, then the brain can no longer be envisioned as a collection of logic gates. Rather it must be seen as an organized population of billions of goal-seeking "creatures," each "talking" to some thousands of others and each adapting its behavior in accordance with whether or not it is getting what it wants. Such a view of brain function calls into question certain aspects of the AI paradigm.

Knowledge acquisition or learning is a central property of intelligence as we know it in living systems. It is however a process largely ignored in AI research. AI research has tended to factor intelligence into three components, these being knowledge acquisition, organization and utilization. In living systems these three aspects of intelligence appear to be integral parts of a single process. This is also true for the proposed heterostatic neuron. However, in AI research, where intelligence has be fractionated and the focus is on knowledge organization and use, learning has come to assume an unnatural role, as indicated by Nilsson's (1974) statement:

> ...we cannot have a program learn a fact before we know how to tell it that fact and before the program knows how to use that fact. We have been busy with telling and using facts. Learning them is still in the future....

Such prerequisite structuring of information before it can be learned is not the case for living systems, although it is true that prestructuring more nearly becomes a requirement as the information to be learned becomes more abstract.

The unnatural role of learning in AI systems can be related to the lack of a generalized goal. Highly specific goals do not lend themselves to the kind of general interaction with an environment that widely scoped learning probably requires. A generalized goal that can "drive" all of the activities of a system is needed. To obtain a generalized goal, the heterostatic theory suggests the following approach: translate all information for transmission into a pair of binary codes and then employ subsystems that seek to maximize the difference between the amounts of each of the two types of codes being received. One code is to be sought by the subsystems and the other to be avoided. In the case of the brain, it appears that the code that is sought consists of excitatory impulses and that to be avoided consists of inhibitory impulses. Such codes represent a generalization of the roles played by positive and negative feedback loops in conventional control systems. Perhaps greater coordination of subprograms and system generality can be achieved in AI systems with a similar approach. To some extent, this has been done in robotics research [for example, see Andreae (1963) and Doran (1968)] where a robot's problems have been formulated in terms of variables similar to those of pleasure and pain (see review by Ernst, 1970).

Among AI researchers, the need for learning mechanisms is being increasingly felt. Nilsson (1974):

> Today, the knowledge in a program must be put in "by hand" by the programmer although there are beginning attempts at getting programs to acquire knowledge through on-line interaction with skilled humans. To build really large knowledgeable systems, we will have to educate existing programs rather than attempt the almost impossible feat of giving birth to already competent ones.

To date, AI researchers have employed both procedural and assertional representations for knowledge. Learning, if it is based on a generalized goal-seeking mechanism, will by the very nature of the mechanism favor procedural representations. Also, a change in attitude will be required on the part of those who structure the environments for and train such systems. In the case of a system possessing a generalized goal, it will become important to ask not only what the machine can do for us but also what we can do for the machine.

To summarize, we see that the AI approach treats intelligence as an emergent phenomenon whereas in the heterostatic theory intelligence is viewed as a "grass-roots" phenomenon. Perhaps AI research will begin to move in this direction, as might be concluded from the kinds of networks discussed in Minsky's (1975) recent work on a theory of frames.

AI research must continue if we are to establish the limits of the digital computer's capabilities. However, there appears to be no basis for expecting that this research can ever yield systems of an intelligence comparable to that of man. AI research, in employing a centralized, serial, nonadaptive substrate (a conventional digital computer) appears to be 180 degrees out of phase with the one architecture that is known to work, namely that of the living brain. For highly intelligent systems, it seems that research into

decentralized, parallel, pervasively adaptive architectures may be required.

References

Andreae, J. H. (1963) Stella, a scheme for a learning machine, *Proc. IFAC.*

Bernard, Claude (1859) Lecons sur les proprietes physiologigues et les alterations pathologigues des liquides de l'organisme, *Bailliere* 1, Paris, France.

Doran, J. E. (1968) Experiments with a pleasure-seeking automation, *Machine Intelligence* III.

Dreyfus, H. (1965) "Alchemy and Artificial Intelligence," RAND Corporation Paper P3244 (AD 625 719), Dec. 1965.

Dreyfus, H. L. (1972) *What Computers Can't Do*, Harper and Row, New York.

Ernst, H. A. (1970) "Computer-Controlled Robots", IBM Research Report RC 2781 (No. 13043).

Farley, B. G., and Clark, W. A. (1954) Simulation of a self organizing system by a digital computer, *IRE Trans. on Information Theory*, Vol. PGIT-4, pps. 76-84.

Freud, Sigmund (1895) Unpublished, untitled paper subsequently published in Freud, Sigmund (1964-) *Standard Edition of the Complete Psychological Works of Freud*, (ed. by James Strachey), Macmillan, New York, 1 :281-387.

Griffith, V. V. (1962) A Mathematical Model of the Plastic Neuron, Goodyear Aircraft Corporation, Akron, Ohio, Report No. GER-10589.

Griffith, V. V. (1963) "A Model of the Plastic Neuron," *IEEE Trans. on Military Electronics*, MIL-7:243-253.

Klopf, A. Harry (1972) *Brain Function and Adaptive Systems - A Heterostatic Theory*, Air Force Cambridge Research Laboratories Research

Klopf, A. Harry (1974) Brain function and adaptive systems - a heterostatic theory, *Proceedings of the 1974 International Conference on Systems, Man and Cybernetics*, IEEE Systems, Man and Cybernetics Society, October 2-4, 1974, Dallas, Texas.

Lighthill, J. (1973) "Artificial Intelligence: A General Survey," *Artificial Intelligence: A Paper Symposium*, Science Research Council Pamphlet, Science Research Council, State House, High Holburn, London, April 1973.

McCulloch, W. S., and Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity, Bull. Math. Biophys. 5:115-137 [reprinted in McCulloch, W. S. (1965) *Embodiments of Mind*, M.I.T. Press, Cambridge, Mass., pp. 19-19].

Minsky, M. (1954) "Neural Nets and the Brain-model Problem," doctoral dissertation, Princeton University, Princeton, New Jersey (available from University Microfilms, Ann Arbor, Michigan).

Minsky, M., and Papert, S. (1969) *Perceptrons: An Introduction to Computational Geometry*, M.I.T. Press, Cambridge, Mass.

Minsky, Marvin (1975) "A Framework for Representing Knowledge," *The Psychology of Computer Vision*, McGraw-Hill, New York.

Nilsson, Nils J. (1975) "Artificial Intelligence," Artificial Intelligence Center Technical Note 89, Stanford Research Institute, Menlo Park, Californis; also presented at IFIP Congress 74, Stockholm, Sweden, August 5-10, 1974.

Price, Keith (1975) A comparison of human and computer vision systems, *ACM SIGART Newsletter*, No. 50, Feb. 1975, pp. 5-10.

Rashevsky, N. (1938) *Mathematical Biophysics*, University of Chicago Press, Chicago, Illinois.

Rosenblatt, F. (1957) "The Perceptron: A Perceiving and Recognizing Automaton, Project PARA," Cornell Aeronautical Laboratory Report 85-460-1.

NATURAL vs ARTIFICIAL

Rosenblatt, F. (1960) "On the Convergence of Reinforcement Procedures in Simple Perceptrons," Cornell Aeronautical Laboratory Report VG-1196-G-4, Buffalo, New York.

Rosenblatt, F. (1962) *Principles of Neurodynamics*, Spartan Books, New York.

Roszak, T. (1972) *Where the Wasteland Ends: Politics and Transcendence in Post-Industrial Society*, Doubleday, 1972.

Skinner, B. F. (1938) *The Behavior of Organisms: An Experimental Analysis*, Appleton-Century, New York.

Weizenbaum, J. (1972) "On the Impact of the Computer on Society," *Science*, Vol. 176, No. 609.

Wilkins, Michael G. (1970) "Neural Modelling: Methodology, Techniques and a Multilinear Model for Information Processing," Biological Computer Laboratory Technical Report No. 19, University of Illinois, Urbana, Illinois.

Automatic Proof of Correctness of a Binary Addition Algorithm

J Strother Moore
Computer Sciences Laboratory
Xerox Palo Alto Research Center
3333 Coyote Hill Road Palo Alto, CA 94304
(MOORE @ PARC-MAXC)

Abstract

The Boyer-Moore Pure LISP Theorem Prover ([1], [2], and [3]) has recently proved the correctness of a program implementing binary addition with a carry flag. Because of the relative complexity of the binary addition algorithm and the automatic nature of the proof, this result was thought interesting enough to warrant this technical note.

Representations

The Pure LISP Theorem Prover uses lists of NILs to represent integers. Thus, the integer 6 is represented by a list with six NILs in it: (NIL NIL NIL NIL NIL NIL). ADD1 is defined to CONS a NIL onto such an integer. As in Peano arithmetic, the addition of two such integers is defined recursively in terms of ADD1. This is the function PLUS exhibited in DEFINITIONS below.

An alternative representation of integers is as binary numbers implemented as lists of 1's and 0's (where 1 is (NIL) and 0 is NIL). Because the least significant bits are processed first, it is convenient to reverse the usual order of the bits. Thus, 6 can be represented as the "binary number" (0 1 1).

Binary Addition

It is possible to write a LISP function which adds two such binary numbers and produces a third. This function uses a flag to effect carries into the high-order bits. The definition is exhibited as BINADD below. The definition is complicated somewhat since two binary numbers are not necessarily lists of the same length[1]. If the carry flag is on and one of the binary numbers is exhausted before the other, it is necessary to treat the shorter number as though it had the proper number of high-order zeroes.

The correctness of BINADD can be stated by defining a function (MKBIN) which maps from integers to binary numbers, and its inverse (MKINTEGER) which maps from binary numbers to integers. The statement of correctness is then:

[1] For example, 6 is (0 1 1) but 3 is (1 1). Since insignificant high-order zeroes may be present, 6 can also be represented by (0 1 1 0), (0 1 1 0 0), etc.