A WORKING CHARACTERIZATION OF INTELLIGENCE AND A NEW MODEL

by

Lev Goldfarb

TR90-062 December 1990

Faculty of Computer Science
University of New Brunswick
P.O. Box 4400
Fredericton, N.B. E3B 5A3

Phone: (506) 453-4566
Fax: (506) 453-3566

This paper was written in January of 1990

A WORKING CHARACTERIZATION OF INTELLIGENCE AND A NEW MODEL

Lev Goldfarb

School of Computer Science and Intelligent Information Systems
University of New Brunswick
Box 4400, Fredericton, N.B.
E3B 5A3
Canada

Intelligent Information Systems
29 Pembroke Crescent
Fredericton, N.B.
E3B 2V1
Canada

> The problem is not what the answer is, the problem is what the question is . . .
>
> A. Poincare

> The most important link in scientific cognition is a scientific problem; without posing a problem there can be no creative activity or discovery.
>
> G. Kirilenko and L. Korshunova

Abstract. A working description of "intelligence" is proposed: intelligence is characterized by autonomously evolving purposeful processing of meaningful signals (events). The purposefulness should be interpreted in terms of orientation in the environment (physical or abstract) from which the "signals" came, and the evolving nature of the process points to the dominant role of learning mechanisms in acquiring the relevant knowledge. A new model for intelligent machines - pattern learning machines, which can be viewed as far reaching symbolic generalizations of the artificial neural nets - is briefly outlined. A longer and more detailed exposition of the model is currently in press.

1. INTRODUCTION

> Artificial Intelligence (AI) is a large and growing field. Graduate students study and perform doctoral research in AI at many universities throughout the world, and scientists and engineers at academic and other research centers contribute to AI's body of concepts and techniques. Industrial organizations are not only beginning to apply AI ideas to manufacturing technology, but are using them in an increasing number of new products. There are AI organizations and societies, AI textbooks, AI journals and magazines, and AI meetings. [1]

Today, 8 years after [1] was written, it would not be an exaggeration to say that AI is one of the most popular areas in computer science. Paradoxically, in spite of this popularity, there is no *working* definition of the subject area of AI. By a working definition I mean a description that, first, can guide and focus one in the research work, and second, can be used by all AI research workers to evaluate the progress made by a certain time. R.C. Schank states in [2, p.3] that AI is "[w]ithout a coherent methodology and clearly defined goals". Some AI researchers "admit that most current AI work fails to meet traditional criteria of scientific theories, and decry the absence of 'competitive argumentation' whereby the power of one simulation can be rigorously compared with the power of another" [3, p.178, where Gardner simply reports the opinion of a group of researchers at Xerox PARC in Palo Alto]. Others note the following, what I consider to be inadmissible situation:

> Some practical success was achieved with symbol manipulation of mathematical equations and pattern recognition, with the result that these fields became independent of the general field of artificial intelligence. [4, p. vii]
>
> The subject matter of artificial intelligence is not therefore fixed, but changes with time. For example, about the end of the 1960's, techniques for reading handwritten letters of the alphabet or numerals were considered as belonging to the field of artificial intelligence. However, when optical character readers were developed, such techniques were no longer considered as artificial intelligence. It appears that it is the fate of artificial intelligence that when techniques in a given field become established and put into practice, they cease to be part of artificial intelligence. [4, p.1]

Now, more than 30 years after the first works on the subject, we still don't even know how far, if at all, we have progressed towards the creation of an intelligent machine. Very often many of us find it difficult even to evaluate the relevance of a specific research work to the above enterprise. This situation has lead some AI researchers to believe that a working description of the subject area will mark a breakthrough in AI. I, personally, believe that until a reasonably clear *initial* working description of the subject area is found, no fundamental progress in AI is possible, simply because such a progress requires *concerted* long-term effort of many scientists.

It is the purpose of this paper, first, to suggest an initial inevitably incomplete but a sufficiently general working description of AI, and, second, to propose a general mathematical model for intelligent machines - pattern learning machines (PLM). Remarkably, both the overall analytical structure of the model and its implications support some relatively recent hypotheses and observations, notably those of J. von Neumann, O. Selfridge, M. Minsky, S. Watanabe, J. Piaget, H. Putnam, S. Kripke, C. Levi-Strauss, E. Rosch and others.

2. A WORKING DEFINITION OF INTELLIGENCE

What is intelligence? We may agree that intelligence is (or characterized by) an *intelligent processing* of information. In this description one can immediately single out *two interacting elements*: intelligent agent (machine) and the information itself (that "comes" from the environment).

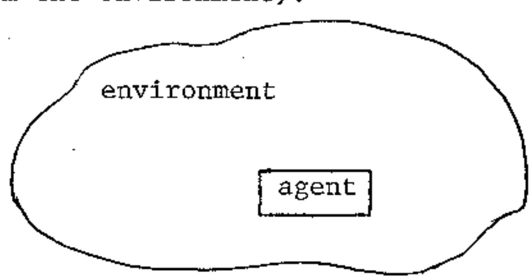

Figure 1.

Above all, it should be stressed that the intelligent nature of the *process* can in general be determined by observing responses of the agent to a *sequence of inputs* and not necessarily to a *single input*, since the intelligent agent may not respond "intelligently" to a single input, especially if this agent evolved in an environment very different from that of the observing agent. Thus, the intelligence should be viewed (and examined) as a purposeful *process*. This purposefulness can be established sooner or later depending on whether the difference in the "architectures" and "training" (maturation) environments of the two communicating agents are smaller or larger. In this respect the well known Turing test is not quite satisfactory, since it tacitly assumes that the intelligent agent (computer) evolved in essentially the same environment as we did.

It appears that the verb "evolve" used above is, in fact, the key word in the description of an intelligent process. It is doubtful if any process that has not evolved in some environment over any period of time can be called intelligent, since whatever "intelligence" it possesses must have been "produced" by some other intelligent agent - its designer. Besides, as we shall discuss next, without a direct bond between the events in the environment and their representation inside the agent (semantic bond) - the bond that can be established only during the agent's learning periods - these events cannot be meaningfully (intelligently) interpreted by the agent. At the same time, from a purely pragmatic point of view, the acquisition of knowledge in a "rich" environment must be an ongoing process: the knowledge must be constantly updated and compacted, because of the restrictions on the memory size and the processing speed.

In this connection one should also note that a "true" offspring of an intelligent agent may be considered to possess the intelligence but only as a potentiality, which can be realized during a maturation period. All neurophysiological evidence collected so far clearly indicates that even the neuronal "architectures" simply cannot and are not completely specified genetically, but are "finalized" during various "critical" maturation periods [5, pp. 186-196]. Moreover, it is well known that all natural (biological) intelligent agents themselves evolved over very long evolutionary periods.

A vivid illustration of what I for the moment will call a pseudo-intelligence, i.e., intelligence that has not evolved over any period of time, one can find in the "Chinese room" example proposed by modern philosopher J. Searle (see, for example, [6, Ch.2], or [3, p. 172]). In this hypothetical example, which has acquired some notoriety in the AI circles and which was inspired by the Turing test, an agent that knows absolutely no Chinese is locked in a room which has all the books that provide complete instructions (in the language of the agent) as to how to assemble correct answers in Chinese to any question in Chinese received from the outside. To the outside interrogating agent who knows Chinese, the answers that are received electronically from the "Chinese room" are indistinguishable from those that would have been given by an agent that knows Chinese. "What is the difference between the Chinese [room] case and the English case? . . . You understand the questions in English because they are expressed in symbols whose meanings are known to you. Similarly, when you give the answers in English you are producing symbols which are meaningful to you. But in the case of the Chinese [room], you have none of that. In the case of the Chinese [room], you simply manipulate formal symbols according to [instructions] . . . and you attach no meaning to any of the elements." [6, pp. 33-34]

From the analysis of this example Searle draws conclusions that "there is no way that the system can get from the syntax to the semantics" [6, p.34] and that "the project of trying to create minds solely by designing programs is doomed from the start" [6, p.39].

Although I disagree with the Searle's conclusions, when interpreted broadly, it is hard to disagree with the following point, which, unfortunately, has not been articulated well: machines whose responses to the external events are *completely* preprogrammed (i.e., when no fundamental learning, or internal restructuring, occurs) cannot be called intelligent, because, besides the inflexibility, they have no *intelligent* way of interpreting the external events. In fact, it is easy to see that the realization of the above hypothetical example is impossible, simply because the number of potential questions is practically infinite, so that no complete collection of books for the Chinese room example can exist. Since our brain is also of finite capacity, we can draw the crucial conclusion that an intelligent interpretation of the external events must involve, as was mentioned above, an ongoing context-dependent reduction of a potentially infinite number of external events to a finite number of internal "symbols" representing the events. This, as well as other evidence, strongly suggest that, contrary to what is often assumed in AI (see, for example, [1, pp.2-3]), the "transduction", or "symbol" formation, processes must form a major part of what is usually called "central" or "symbol" manipulating cognitive processes.

As a result of the development over the last 10 years of the new model for intelligent machines [7], which is very briefly outlined in the second half of this paper, it became reasonably clear to me that the most *economical* route towards the context-dependent intelligent interpretation, including dynamic representation, of the external events is a taxonomic one, i.e., via dynamic, or reconfigurable, pattern learning mechanisms that contrary to the present architectures, evolve in an *irreversible* manner. It appears to be the only route that by-passes the wall between the syntax and the semantics, the wall that prompted Searle to conclude that AI is not possible. The agent then would possess the dynamic capacity to associate with a group of external events an internal "symbol", i.e., to generate the mapping by means of which the semantics is induced. Moreover, it turns out that the learning processes are absolutely necessary in order to make the more conventional propositional mechanisms of AI computationally feasible in reasonably complex environments.

It goes without saying that I am not alone in choosing the "taxonomic" route as the basis. In fact, this assumption was instrumental in the formation of a large and successful "sister" area, Pattern Recognition, which includes Artificial Neural Nets and which formed at about the same time as AI did. As I have mentioned elsewhere, I believe that the artificial separation of the two areas is responsible to a considerable degree for the current principal difficulties of the two areas. What is also of importance is that some of the leading neuroscientists point to the taxonomic basis as well: "The essential requirement for learning, logic, and other mental functions that are the usual subjects of AI research is the prior ability to categorize objects and events based on sensory signals reaching the brain". [8, p.155]

It is not difficult to see that all non-taxonomic, i.e., propositional, models for a context-dependent interpretation of the external events cannot evolve autonomously, but, again, must be preprogrammed and therefore are not sufficiently flexible, particularly for "rich", or natural, environments.

To summarize, we come to the proposed characterization of an intelligent process as *autonomously evolving purposeful processing of meaningful signals (events)*. The purposefulness should be interpreted in terms of orientation in the chosen environment (physical or abstract) from which the "signals" come. The most basic understanding of the orientation is in terms of categorization. The evolving nature of the processes points to the dominant role of learning mechanisms in acquiring the relevent knowledge. It appears that such processes, contrary to the present computing processes, *must* evolve in an *irreversible* manner, i.e., the structure of the machine must also evolve.

The progress in the area can then be judged by the complexity of the environment in which the system can function and the degree to which the process can autonomously evolve. Autonomous maturation of the system does not, of course, exclude the presence of a teacher. The assessment of the progress, then, in addition to the above, has to take into consideration the achieved degree of the system's independence from the teacher. In case of biological agents, the role of the "teacher" plays the environment itself in the form of food, preditors, adverse chemical reactions, etc.

3. TRANSFORMATION SYSTEMS

In some sense the transformation systems, which are introduced in [7], have emerged as a result of the search for the model that integrates into a coherent whole presently separate areas of AI and Pattern Recognition (PR). Unfortunately, some leading AI researchers, who have had decisive influence on shaping the agenda in AI, have insisted that the numerical (vector) representations and the techniques dominating PR are fundamentally different from the "symbolic" representations and techniques employed in AI. Within PR, already in the late 60s, it became clear that the nature of the chosen object representation should not have a drastic effect on the PR model, but since no unified framework has been found, a new large subarea, syntactic, or structural, PR, emerged. In fact, the model proposed in [7] is also an outcome of the work directed towards the unification of the two principal models for PR, vector space and syntactic. In view of the space limitations, I shall only outline the basic theoretical features of the model. A more detailed exposition together with an example, without which, unfortunately, the model is not easily accessible, can be found in paper [7] and its sequel.

A transformation system (TS) is a pair $(O,\mathcal{S})$, where O is a set of structured objects (strings, labelled trees, labelled graphs, etc.) and $\mathcal{S}$ is a finite set of (reversible) substitution operations. Each substitution operation (rule) $S_i \in \mathcal{S}$ specifies allowable local (i.e., related to a part of the object) rules for object transformation, which are generalizations of the familiar substring substitution operations to more general structured objects. The substitution rule says, in effect, that if a certain structural element is found in the structured object, then it can be (reversibly) replaced by another structural element, thus transforming the original object into another one.

At any given moment a TS has a fixed set of the substitution operations, which can be considered as the basic machine operations defining the machine configuration. The machine (PLM) has the capacity during the learning process to expand, if necessary, its set of operations by introducing some new operations that are formed according to some fixed set of composition rules from the existing machine operations. The composition rules may include, for example, formation of a new substitution operation that represents parallel application of the existing operations, i.e., the new operation works as a specified parallel group of the existing operations. One can show that the class of PLM with parallel composition rules are more powerful than those with only sequential composition rules. In case of strings, for example, if deletion-insertion of 'a' and 'b' are in the set of operations, the sequential composition rules allow addition of operation deletion-insertion of 'ab' or 'ba'.

It is useful to assume that the set of initial substitution operations is complete, i.e., every object $o_1 \varepsilon O$ can be transformed into any other object $o_2 \varepsilon O$ by means of the operations from the set (without forming any new operations).

So far we are still in the classical computational setting, which is a generalization of a Thue system [9, p.287]. The next definitions, particularly the second one, introduces into the classical *discrete* setting a *continuous* metric geometry (or rather a family of such geometries), which play the decisive role in the learning process.

The intrinsic distance function in the TS $(O,\mathcal{S})$, $\Delta: O \times O \to R_+$, is defined as the minimum number of the substitution operations necessary to transform one object into another. Representing objects as vertices of a graph and the operations as edges of the graph, one can view $\Delta(o_1,o_2)$ as the length of the shortest path between objects o_1 and o_2. It is important to note, that the addition of operations can only reduce the distances.

The intrinsic distance function gives rise to the very central concept of the parametric family of distance functions $\{\Delta_\omega\}_{\omega\varepsilon\Omega}$, $\Omega \subset R_+$, where the definition of Δ_ω is obtained from that of Δ by assigning weighting

$$\omega = (w^1,w^2,\ldots,w^m), \quad w^i \geq 0, \quad \sum_{i=1}^{m} w^i = 1, \quad \text{to the set } \mathcal{S} =$$

$\{s_i\}_{1\leq i\leq m}$ of substitutions in the transformation system $T = (O,\mathcal{S})$. In other words, the shortest path is replaced by the shortest weighted path.

The choice of a specific weighting scheme $\bar{\omega}$, and therefore of the corresponding distance function $\Delta_{\bar{\omega}}$ (or metric geometry), occurs during the learning process.

4. LEARNING IN TRANSFORMATION SYSTEMS

For simplicity we shall consider only non-sequential learning. Modifications for the sequential learning readily follow from the latter.

During learning, the pattern learning machine (PLM) is presented with a finite set of positive and negative examples of a class O_1 of objects, $O_1 \subset O$. As was mentioned above, during learning in the simplest case the PLM tries to find an optimal weighting scheme $\omega^* \varepsilon \Omega$ that would allow an (optimal) recognition of the learning class O_1. Once a satisfactory ω^*, and therefore Δ_{ω^*}, is found, the future recognition of objects from O_1 may proceed along the lines outlined in monograph [10] (see also [11]), which is concerned with the mathematical model for efficient algorithms for object recognition in the metric model. In essence, all such recognition algorithms, which are not considered here, try to assign a new pattern to the closest class using a minimum number of the distance computations (between the new pattern and the training, or learning, patterns).

More specifically, during the learning process all events occur around the following function g: $R^m \to R_+$ (m is the number of operations in S)

$$g(\omega) = \frac{f_1(\omega)}{1 + f_2(\omega)} ,$$

where $f_1(\omega)$ is the smallest Δ_ω-distance from the positive learning (training) objects (in class O_1) to the negative ones, and $f_2(\omega)$ is the average Δ_ω-distance between the positive learning objects. One can show that f_1 and f_2 are concave and piecewise linear.

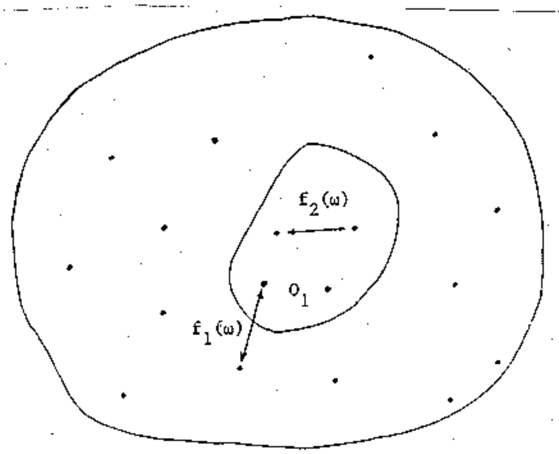

Figure 2

It is easy to see that the subset $\overline{\Omega}$ of the simplex

$$\Omega=\{\omega = (w^1, w^2, \ldots, w^m) \mid \sum_{i=1}^{m} w^i = 1, \quad w^i \geq 0 \quad \forall i\}$$

corresponding to maximum(s) of the above function over Ω represent the best choice for the above optimal weighting scheme ω^* for the current configuration of the PLM ($S=\{S_i\}_{1\leq i\leq m}$). Both the maximum value of g as well as the <u>relative complexity of the class</u> O_1 w.r.t. S,

$$\varepsilon = \min_{\omega\in\overline{\Omega}} \mathcal{E}(\omega) ,$$

where $\mathcal{E}(\omega) = -\sum_{i=1}^{m} w^i \log w^i$, play the key role in the decision about the adequacy of the current set of operations, S, i.e., the current configuration of the PLM. If necessary, the set S is modified and the optimization step is repeated.

The computed optimal weighting scheme not only allows one to introduce for the first time a non-probabilistic concept of the relative class (object) complexity, desirability of which was mentioned by Kolmogorov in [12], but it also provides a bridge between the perception (recognition) mechanisms and the more conventional, propositional, mechanisms in the PLM: the largest weights w^i in the optimal weighting scheme point to the contextually "important" features (operations) in the object representation. This bridge (between perception and propositional mechanisms) is established by assigning operations from S to the variables of the predicate language, or of the frames, used in the propositional object (class) descriptions. In other words, given a specific class F of predicate formulas (wff),e.g., conjunctive descriptions, the set $\overline{F}$ of <u>optimal propositional descriptions</u> from F is obtained by <u>limiting the set of variable used</u> in formulas from F to those corresponding to the largest weights w^i in ω^*. Since the cardinality of F depends exponentially on the number of the chosen variable, the importance of the above learning stage for the computational feasibility of the propositional stage becomes obvious. The propositional descriptions themselves are "read off" from those obtained by substituting the corresponding operations for free variables in the wff's applied to the positive objects.

The goals of the learning process may vary. If the object (class) recognition is the only objective of a specific learning process, then the process may be terminated when the value of $g(\omega)$ is greater than 1 and (preferably) a large number of w^i's are zeros (the relative complexity ε is small). In this case a small number of the nonzero weights ensures efficiency of the Δ_ω-distance computation, while the corresponding value of $g(\omega)$ ensures a reasonable class perception. In case of the string representation, the known dynamic programming Wagner-Fisher algorithm for computing the Levenshtein, or edit, distance between two strings of lengths n each is of order $O(n^2)$ on a sequential machine [13] and of order $O(n)$ on a parallel machine with n processors. It goes without saying, that having chosen a concrete environment, the use of the corresponding special purpose chip for the distance computations in the PLM is quite prudent.

As the set S of operations, the configuration of the PLM, is modified dynamically, the corresponding set $\overline{F}$ of optimal description is also modified dynamically. Thus, if the objective of the learning is a special type of the propositional object description, the stopping criteria are different than those mentioned above, when the objective is purely "perceptual", i.e., object (class) recognition.

Thus, the PLMs for the first time offer models of truly reconfigurable learning machines, whose configuration may change during learning processes. It is important to note that the change in the configuration of a PLM usually results in changes in the perception of the classes (objects) in the environment, since the former change results in the change of the family $\{\Delta_\omega\}_{\omega\varepsilon\Omega}$ of distance functions from which the optimum distance $\Delta_{\omega*}$, responsible for the calss (object) perception, is selected. The learning, then, represents a truly irreversible process.

The following thesis appears to be true. If $(O,\mathcal{S})$ is a complete transformation system with a sufficiently general family of the composition rules, then the PLM can learn any *class* O_1 of objects and can produce any corresponding propositional class description. In the future it is important to prove the latter for some restricted definitions of the class of objects.

Finally, it is also important to note that the above model allows the PLM to store in memory not the entire object representations but only those context-dependent "important" features that were detected during the learning processes. Thus, the amount of information that is to be stored is not proportional to the "size" of the object representation, but only to the complexity of the contexts in which the objects have been perceived. This feature of the PLM deserves to be called intelligent memory.

5. CONCLUSION

In this section I will very briefly mention some relatively recent hypotheses and experimental observations that support the proposed model. First, I must mention the brilliant predictions by J. von Neumann, who suggested in [14] that the standard (logical) theory of automata by itself is inadequate for modeling intelligence, since the formal logic "is one of the technically most refractory parts of mathematics" and, therefore, a successful model would have to make more "contact with the continuous concept of the real or of the complex number", in order to improve considerably the efficiency of the automata.

> All of this will lead to theories which are much less rigidly of an all-or-none nature than past and present formal logic. They will be of a much less combinatorial, and much more analytical, character. In fact, there are numerous indications to make us believe that this new system of formal

> logic will move closer to another discipline which has been little linked in the past with logic. This is thermodynamics, . . . and is that part of theoretical physics which comes nearest in some of its aspects to manipulating and measuring information. Its techniques are indeed much more analytical than combinatorial, which again illustrates the point that I have been trying to make above. [14, p. 101]

Second, it is important to mention the pioneering work of Selfridge in PR [15, §II.F], which, incidentally, inspired some early well known work in AI [15, §II.F]. Selfridge proposed to apply some sequences of transformations to the images, but was not sure how the machine can generate the necessary sequences of transformations.

Third, I want to mention the work of S. Watanabe [16, §§4.1, 4.2], who insists (with his Theorem of the Ugly Duckling) on the crucial role of the weighting scheme in producing the context-dependent similarity (distance) measure.

Fourth, it is interesting to remember the following passage from a well known paper by Minsky:

> We certainly need (and use) something like syllogistic deducation; but I expect mechanisms for doing such things to emerge in any case from processes for "matching" and "instantiation" required for other functions. Traditional formal logic is a technical tool for discussing either *everything that can be deduced from some data* or *whether a certain consequence can be so deduced*; it cannot discuss at all what *ought* to be deduced under ordinary circumstances. Like the abstract theory of Syntax, formal Logic without a powerful procedural semantics cannot deal with meaningful situations. [17, p. 262]

Fifth, the proposed model for the first time clarifies the nature of the "assimilation", "accommodation", and "equilibration processes introduced by one of the leading cognitive scientists of the century J. Piaget [18, Ch.1]. The concept of a reversible operation, often mentioned but not defined analytically, also lies at the foundation of the Piaget's theory of cognitive development [19, Ch.4].

Sixth, the model remarkably well explains the results of observations by a number of leading psychologists, anthropologists, and philosophers, including E. Rosch, C. Levi-Strauss, S. Kripke, H. Putman, about the fundamental role of the categorization processes [3, p. 238 and Ch. 12].

Finally, it is important to stress, that formally the new model does not contradict existing models, but rather extends and encompasses them. In addition, it explains and predicts many more features than various known models and therefore can easily be tested. Above all, it suggests that the unbounded power of the intellect resides mainly in its *dynamic operational* structure that can evolve in an economical manner through the constant interactions with the environment, rather than in the limited propositional mechanisms that are fundamentally separated (cut off) from the environment and serve only as languages for communicating our perceptual experiences.

REFERENCES

[1] N.J. Nilsson, Artificial Intelligence: Engineering, Science or Slogan?, *AI Magazine*, Winter 1981-82, pp. 2-9.

[2] R. C. Schank, The Current State of AI: One Man's Opinion, *AI Magazine*, Winter/Spring 1983, pp. 3-8.

[3] H. Gardner, *The Minds New Science*, Basic Books, New York, 1985.

[4] Y. Shirai and J. Tsujii, *Artificial Intelligence: Concepts, Techniques and Applications*, J. Wiley & Sons, Chichester, 1986.

[5] B. Bridgeman, *The Biology of Behaviour and Mind*, J. Wiley & Sons, New York, 1988.

[6] J. Searle, *Minds, Brains and Science*, Harvard University Press, Cambridge, Massachusetts, 1984.

[7] L. Goldfarb, On the Foundations of Intelligent Processes I: An Evolving Model for Pattern Learning, to appear in *Pattern Recognition*, 1990.

[8] G. N. Reeke, Jr., and G.M. Edelman, Real Brains and Artificial Intelligence, in *The Artificial Intelligence Debate: False Starts, Real Foundations*, ed. S.R. Graubard, The MIT Press, Cambridge, Mass., 1988.

[9] H. R. Lewis, and C. H. Papadimitriou, *Elements of the Theory of Computation*, Prentice-Hall, Englewood Cliffs, New Jersey, 1981.

[10] L. Goldfarb, A New Approach to Pattern Recognition, in *Progress in Pattern Recognition 2*, eds. L.N. Kanal and A. Rosenfeld, North-Holland, Amsterdam, 1985, pp. 241-402.

[11] L. Goldfarb, A Unified Approach to Pattern Recognition, *Pattern Recognition*, Vol. 17, No. 5, 1984, pp. 575-582.

[12] A.N. Kolmogoroff, Logical Basis for Information Theory, *IEEE Trans. Inform. Theory*, Vol. IT-14, 1968, pp. 662-664,.

[13] J. B. Kruskal, An Overview of Sequence Comparison, in *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparisons*, eds. D. Sankoff and J.B. Kruskal, Addison-Wesley, Reading, Massachusetts, 1983.

[14] J. von Neumann, The General and Logical Theory of Automata, in *Perspectives on the Computer Revolution*, ed. Z.W. Pylyshin, Prentice-Hall, Englewood Cliffs, N.J., 1970.

[15] M. Minsky, Steps Toward Artificial Intelligence, in *Computers and Thought*, eds. E.A. Feigenbaum and J. Feldman, McGraw-Hill, New York, 1963.

[16] S. Watanabe, *Pattern Recognition: Human and Mechanical*, J. Wiley & Sons, New York, 1985.

[17] M. Minsky, A Framework for Representing Knowledge, in *Readings in Knowledge Representation*, eds. R. Brachman and H. Levesque, M. Kaufmann, Los Altos, California, 1985, pp. 245-262.

[18] B. J. Wadsworth, *Piaget's Theory of Cognitive and Affective Development*, Fourth Edition, Longman, New York, 1989.

[19] H. G. Furth, *Piaget & Knowledge: Theoretical Foundations*, Second Edition, The University of Chicago Press, Chicago, 1981.