Did Artificial Systems Need Random for Learning Strategies ?

Pierre GLIZE

IRIT - Université Paul Sabatier
118, Route de Narbonne - 31062 - Toulouse Cedex
Email : glize@irit.fr

KEY WORDS: cooperativelearning, neural network, genetic algorithm, predictive model, attractors

ABSTRACT: Many analogies found in natural systems give evidence that the role ofnoise in a complex system might well lead to further organization. So,noise seems a good way in order to create novelty or to test the strengthof algorithms.

In this paper, we are going to analyse some artificial learning mechanismssuch as genetic algorithms or neural networks, which may be generallyformulated as an optimization problem by specifying a performancecriterion, and then by using the simple but powerful technique of stochastic hill-climbing along thegradient. In these algorithms, the integration of random is a good way tomaintain the exploration property during searching, useful for avoidinglocal optima or when environment is dynamic.

We claim that artificial learning must overcome their limitations using theexpedient of random search. This is due to attractors always present insidesearch procedures. We discuss in order to find another way to create orderwithout having any presupposed attractors. This is also a central question for anticipatorysystems which must learn about themselves and their environment.

1. The role of random in adaptive systems

1.2. The Role of Noise in Natural Systems

Analogies found in natural systems give evidence that therole of noise in a complex system might well lead to further organization.For example :

- the space positioning mechanism of an ant, has a limitedprecision degree. This leads to some mistakes when it comes back to thenest or to the previous foraging location. But these mistakes are sometimesuseful to discover food,
- in biology, a mutation occurs when the replication mechanism does notgive an exact copy. From the biological viewpoint, replication cannot berealized with zero default, otherwise the genetic material must be veryimportant. Most of mutations are useless, some of them are disadvantageous and only a very few are favourable. Afavourable "mistake" increases the survival of an individual and gives abetter chance to this gene replication.

International Journal of Computing Anticipatory Systems, Volume 1, 1998
Ed. by D. M. Dubois, Publ. by CHAOS, Liège, Belgium. ISSN 1373-5411 ISBN 2-9600179-1-9

Many systems in nature exhibit sophisticated collectiveinformation-processing abilities that emerge from the individual actions ofsimple components interacting via restricted communication pathways. Someoften-cited examples include efficient foraging and intricate nest-building in insect societies, the spontaneous aggregationof a reproductive multicellular organism from individual amoeba in the lifecycle of the Dictyostelium slime mold, the parallel and distributed processing of sensory information by assemblies of neurons in the brain, and theoptimal princing of goods in an economy arising from agents obeying localrules of commerce (Crutchfield,1994). These coherent global activities arerealized by entities having only local view of their environment. Erroneous behaviors or unexpected eventsare not controlled by an individual but must be recovered by its adaptationprocess.

Previous examples show that noise seems a good way in orderto create novelty or to test the strength of algorithms. This was stated by the order-from-noise principleby Heinz von Foerster (von Foerster,1981). This is not random noise, butmore precisely a hidden order which could be discovered by an adaptivesystem. In these cases noise is a good way in order to test the robustness of algorithms and to create novelty.

We want to show in this paper, that the role of random inartificial adaptive systems is very far from what is observed in nature.This has important consequences for the properties of learning process inanticipatory systems.

1.1 The learning process in Anticipatory Systems

In order to be efficient in its environment, an anticipatory system mustinclude predictive models about itself and its environment. This way fordecision-making of an organism is also proposed by some biologists asStewart (Stewart,1997) for which an animal (and a human being) is able to anticipate the local consequences of anhypothetical sequence of innovative actions ; but it does not and cannotpredict the wider consequences of such actions. Adaptability is required tounexpected events, dealing with imperfect and conflicting information frommany sources, and acting before all relevant information is available. Theanticipative reasoning process favours the autonomy of the organism in allowing it many choices of actions in theworld in the near future. But a question remains : how are these modelsacquired ?
- First, the designer of the system could include a prioristatements about itself and the environment in which it will interacts.This is the classical cognitivist approach for artificial systems.
- Second, nothing is presupposed inside the system by the designer. In thiscase he must add it some learning mechanisms in order to acquire thesepredictive models when the system will work.

If we consider that the system will be in a dynamicenvironment or it will be able to acquire new competences during itslifetime, predefined models can not be given during the design phase : theymust be acquired by a learning process. Thus, if an anticipatory mechanism seems to be interestingfor an organism, a major key problem remains : how does the system acquiremodels about itself and about the environment ? This is a non trivialproblem, even if we consider having many learning algorithms in our disposal to this goal. This learning,based on the past and present state, is difficult in its general processbecause the system must be able to learn these models even if the worldchanges, and also itself evolves. Thus, it must learn without any presupposition about the world : this is the mainpurpose of our paper.

In the next chapter, we analyse some artificial learning mechanisms such asgenetic algorithms or neural networks, which may be generally formulatedas an optimization problem by specifying a performance criterion, and thenby using the simple but powerful technique of stochastic hill-climbing along the gradient.

2 Random Strategies in some Artificial Systems

When we want to modelize some natural behavior (such as in ethology) in using an artificial system, we are unable toknow precisely all the underlying conditions of an action but onlyprobabilistic behaviors. Some general laws about these probabilisticfunctions could be founded by observations of natural systems. In order to obtain a behavior of a virtual individualclosed to these observations, a designer employs generally distributionfunctions associated to a random function. This randomization process isnot the purpose of our work.

We are interested here by random internalized in learning artificial algorithms. Inthis use, random maintains the exploration property during searching foravoiding local optima (funtions having many optima) or when environment isdynamic (i.e. optima evolving during time). This use is very far from hazard occuring in natural environmentbecause no individual can know all the consequences of his acts.

2.1 Genetic algorithms

Genetic algorithms (GAs) is a member of the class ofstochastic optimization procedures called evolutionnary algorithms (EAs). It also includes evolutionaryprogramming (EP) and evolution strategies (ESs) (De Jong, 1993). Acomparison of these different methods can be found in (Bäck and Schwefel 1993). The general process of a genetic algorithm is thefollowing (see figure 1) :

- The first generation is composed of a randomly generatedpopulation of chromosomes (e.g. candidate solutions to some problem). Eachchromosome is a string of 1's and 0's in the simplest form.
- The fitness of each chromosome in the population iscalculated with a given evaluation function.
- A subset of the population is then selected depending on their fitnessand the crossover genetic operator is applied between tem to create a newpopulation.
- On this new population the mutation genetic operator is applied on eachchromosome in order to obtain the new generation. Go to step 2.

For Goldberg (Goldberg,1989), mutation plays a secondaryrole in the operation of genetic algorithms. "Mutation is needed because,even though reproduction and crossover effectively search and recombine extant notions, occasionallythey may become overzealous and lose some potentially useful geneticmaterial [...]. In artificial genetic systems, the mutation operatorprotects against such an irrecoverable loss. [...] We note that the frequency of mutation to obtain good results inempirical genetic algorithm studies is on the order of one mutation perthousand bit (position) transfers».

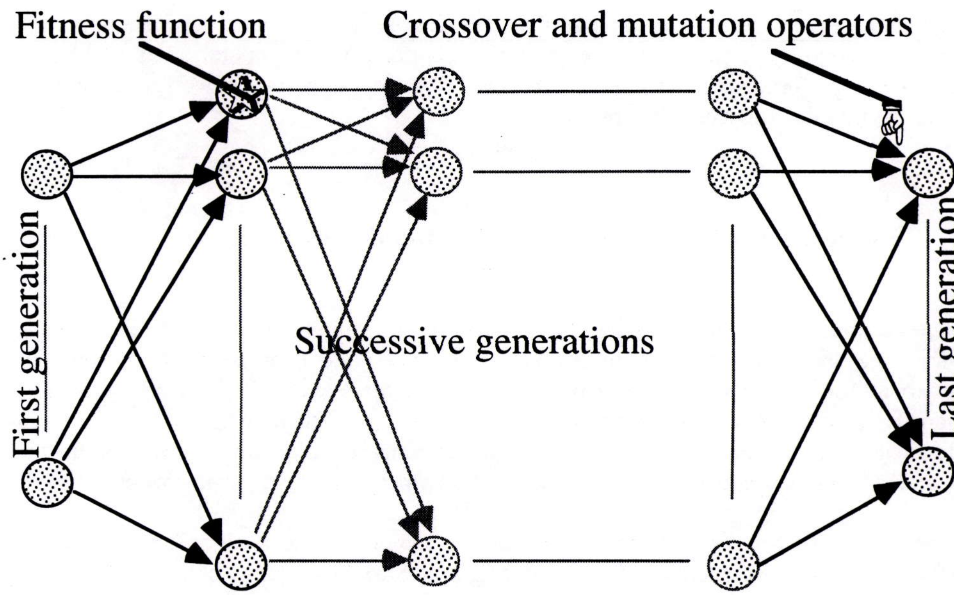

Figure 1 - Population evolution in a Genetic algorithm

When the parents are distributed around the global optimum(Fogel 1995a) , it is an evidence that recombination is sufficient to attain thisoptimum. But, this is not the general case. Salomon (Salomon 1996)demonstrates that mutation alone is sufficient to find the global optimumof separable, multimodal functions within O(n ln n) time, whereas crossover alone is not sufficient for this goal.

2.2 Neural Networks

A neural network model is characterized by three basic components (seefigure 2) :
- The network is a set of interrelated nodes (the neurons) by orientedweighted links.
- The activation rule is a local procedure used by eachnode to evaluate its activation level depending on the surroundingnodes.
- The learning rule used locally to modify the weights of links in order toadapt the network behavior. The basic learning process, initially proposedby Hebb (Hebb,1949), is based on the observation of biological brain inwhich changing occurs between neurons having a high degree of correlated activity.

The method used for finding the correct adaptation is knownas gradient descent. The process consists in minimizing the «error-surface» by descending this surface downhill, i.e., in the direction of thenegative gradient; we will finally reach at the bottom of the surface. Atthat point, the error can no longer be decreased and the procedurefinishes. The existence of local minima can very easily lead to a failure of thegradient descent search. If such a situation occurs one could try startingfrom a different initial weight setting. Fortunately, it seems that theerror surface of a network with many weights has very few local minima. Apparently, in such networksit is always possible to slip out the local minimum by some otherdimension. A more reliable method for escaping from local minima in agradient search is called simulated annealing (Kirckpatrick,1983).

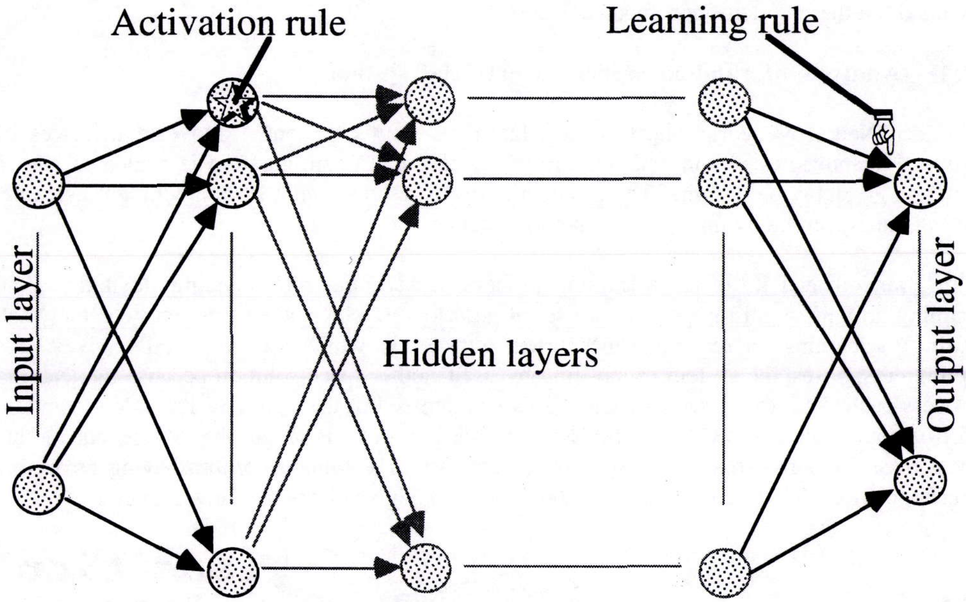

Figure 2 - Representation of a multi-layer perceptron

Normally, it is not possible to go uphill in a gradientdescent. When applying simulated annealing every adaptation is made with acertain probability. This introduces the possibility of going uphill,enabling an escape from local minima. Since it is more probable of getting out of a less deep minimum bychance the system is most likely to end in a global minimum instead of alocal minimum. In simulated annealing this process converges by slowly«freezing» the system, i.e., by decreasing the probability of adaptation. A similarstrategy is applied in the Boltzman neural network.

The learning algorithm may be formulated as an optimizationproblem by specifying a performance criterion, and then by using the simplebut powerful technique of stochastic hill-climbing along the gradient.Importantly, a such procedure is locally implementable. Learning is guided by a "teacher" or bya "critic" using a finite set of "exemplars". The nature of the feedbackprovided by the external trainer needs different weight adjustmentprocedures such as the various versions of the back propagation algorithm, or the reinforcement methods outlinedbefore.

3 Alternatives to Random in Learning Strategies

Adami (Adami,1994) claims that "In almost all cases of learning in natural systems, the fitness of a certain configuration (or"hypothesis") is determined within the system.[...] We shall call systemsthat can perform this feat "auto-adaptive", to emphasize the fact that wedo not provide a fitness -or error- function. For example, all adaptive natural systems are "auto-adaptive"in following the previous meaning and noise is outside of the learningalgorithm : it cannot be avoided but these systems accomodate itspresence.

3.1 Analysis of random search in artificial systems

Artificial Neural Networks and Genetic Algorithms shortly presented aboveare instances of artificial adaptive systems in which the fitness of thecurrent configuration is evaluated with a function given by the designer.The problem is that a system cannot learn anything outside the boundaries specified by this given evaluationfunction.

Thus, random activity inside a learning process has acentral role : adding flexibility. Any learning artificial algorithmpossesses a set of attractors in which a system can potentially fall during its learning phase. When the current attractor in which the systemfalls, gives sub-optimal responses, the system cannot find by itselfanother space solution because the learning process pushes it into the sameattractor. This is expressed in the figure 3. The role of random (for example mutationin GA or simulated annealing in NN) is to get the system out of this localspace. To summarize : random is only necessary in learning algorithmshaving erroneous presuppositions (some teleogical goal) about the world in which the systemwill interact.

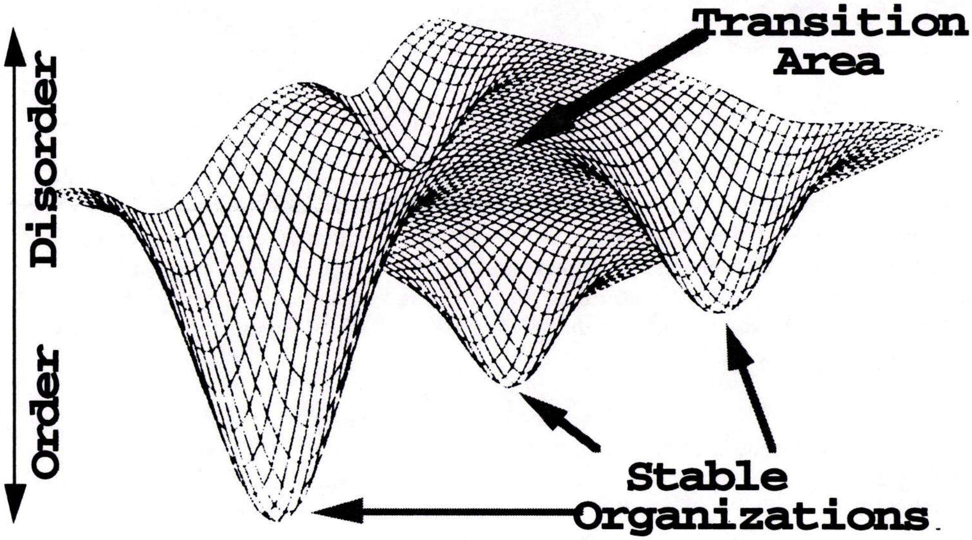

Figure 3 - Representation of predefined attractors by artificial learningalgorithms

Now, the main theoretical question remains : can we design artificialsystems without having an external evaluation function ? This question isassociated with our ability to isolate universal characteristics of thelearning process. In fact, the very existence of a universal learning process has yet to be established(Adami,1994). The consequence of a positive answer is the ability tosuppress any random process for learning strategy.

3.2 Examples of alternative

Salomon (Salomon,1997) suggests that the set of functions on which GAsyield optimal performance converges to a rather "polar" set of functions.To overcome this problem, he proposes an algorithm for derandomizing a GA.The goal is to substitute the stochastic application of mutations by a deterministic mechanism that yieldsoptimal performance while it does not require an exponential memory size. He proposes a deterministic GA which can berecursively constructed in a bottom-up way in order to solve n-hard problemin O(n) time, n^2-hard problems in $O(n^2)$ time and soforth.

Salome (Salome,1994) has developed a self-structuring neural net classifierin which learning is not based on an error function but on the values ofsynaptic weights. A settled neuron presents a connection strength above acertain settling threshold fixed initially. A settled neuron will sharply truncate its input space and it is a candidate for duplication. A useless neuron hasa connection strength below the uselessness threshold parameter, here againinitially fixed. It will be suppressed of the network. Thisself-structuring process is guided by the utility of each neuron i.e. an implicit analysis of cooperative activityinside the network.

Hogg and Huberman (Hogg,1992) showed that when agentscooperate in a distributed search problem, they can solve it faster thanany agent working in isolation. A similar result was obtained by Mataric (Mataric,1994) with a set of mobilerobots fourraging in order to bring back tiles to "home". She has observedthat when the number of individualist robots increases, the globalperformance decreases due to the interferring activities. For her, the ideal result will be obtained with robotshaving altruistic behaviors.

Multi-agents systems are composed of several agents capableof mutual and environmental interactions. Each agent has a local view ofthe environment, generally specific goals and is unable to solve alone the global task devoted tothe system. For most application tasks, it is extremely difficult or evenimpossible to correctly determine the behavioral repertoire and concretactivities of a multi-agent system a priori, that is, at the time of its design and prior to its use. Thiswould require, for instance, that it is known a priori which environmentalrequirements will emerge in the future, which agents will be available atthe time of emergence, and how the available agents will have to interact in response to these requirements.This kind of problems resulting from the complexity of multi-agent systemscan be avoided or at least reduced by endowing the agents with the abilityto adapt and to learn, that is, with the ability to improve the future performance of the total system,of a part of it, or of a single agent (Weiß,1996). Multi-agent learning relies on or even requires the presence ofmultiple agents and their interactions. Many authors in this domain(Goldman,1994), (Sekaran,1995), (Sen,1995), (Weiß,1993) have studied in order to analyze the role of social behavior ofagents on the global performance. They found that cooperation betweenagents improves the results. If we consider each agent of the system as a piece of knowledge, these works mean that knowledge is well learnwhen it is organize in a cooperative manner. This is a criterionindependant of the meaning (the semantic), and thus could be a goodapproach for a general learning theory.

3.3 The cooperative process in an anticipatory system

When a system (a living being or an artificial system) isfunctionally adequate, it realizes the "right" function in its environment.The primary consequence of a functional adequacy, is the system ability to"survive" even in a changing world. At a given time, the next action of an anticipatorysystem is based on an expectation about future events. When this system isfunctionally adequate in its environment, it is able to predict very frequentlythe behavior of its environment. We can observe sequences of events donealternatively by the system and the environment. The process seems to beorderd in a cooperative fashion, even if it is not realized intentionally. A contrario, when ananticipatory system is functionally inadequate, many unexpected events willoccur implying conflictual situations and the system will be unable toobtain the desired state of the world. It is also a strong evidence that an anticipatory system will be unable toact in an unpredictible (random) world.

Thus, anticipation is another way to express thecooperative process between the system and its surroundings. The underlyingassumption of anticipatory systems could be formulated as follows : If the interactions between an anticipatorysystem and its environment are cooperative then the system is functionallyadequate. It is exactly the same assumption which seems to be used in the alternatives presented in the previous paragraphs.

The large number of applications using the principle of anticipatory systempresented during this conference is a proof that it is right. A derivatedconsequence is that cooperation (the underlying form of interaction ofthese systems) leads to efficient activity. This is a way proposed by some authors (Hogg,1992),(Piquemal-Baluard,1996). At this stage the designer can give to the system another "evaluation function" : becooperative in its environment, which is not context dependent. Thislearning process is studied in this volume by (Camps,97).

4. Conclusion

In this paper, we have tried to point out some underlyinghypothesis essential in anticipatory systems :
- First, in order to work adequately an anticipatory system must be ableto learn rights models about itself and its environment.
- Second, from our viewpoint, the integration of randomfunction in the most of the artificial learning algorithms reveals theirad-hoc conception and limits consequently the generality of anticipatorysystems.
- Third, the form of interactions between an anticipatory system and theenvironment is similar to a cooperative process which could be used as avery general assumption for artificial learning.

If experiments in learning become more and more varied and diverse, this isdue to the lack of general theory of learning. This is the real reason forthe presence of random search in artificial learning algorithms. For Adami(Adami, 1994), "One of the most elusive tasks associated with formulating a theory of learning is theisolation of universal characteristics of the learning process. In fact,the very existence of a universal learning process has yet to beestablished".

We think that this problem is central for anticipatory systems because theycannot work autonomously if they are unable to learn good models aboutitself and the environment. In fact, we have indicated works in which thisopen question is studied and some directions are indicated.

5. Bibliography

Adami Chris (1994) Learning and complexity in genetic auto-adaptivesystems - Physica D

Andrew Alex M. (1979) Autopoiesis and self-organization, Journal ofcybernetics, Vol9 N°4

Andrew Alex M. (1989) Self-organizing systems - NewYork : Gordon &Breach

Aubin J.P. (1991) Viability theory - Birkhauser

Bäck T., Schwefel H.P. (1993) An overview ofevolutionary algorithms for parameter optimization Evolutionarycomputation Vol1 N°1 - MIT Press

Bigelow Julian, Wiener Norbert, Rosenblueth Arturo (1943)"Behavior, Purpose and Teleology", Philosophy of Science 10

Camps Valérie, Gleizes Marie-Pierre,Trouilhet Sylvie (1997) Properties Analysis of a Learning Method forAdaptive Systems - First Conference on Anticipatoty Systems - Liege,Belgium

Crutchfield James P., Mitchell Melanie (1994) The evolutionof emergent computation - Santa Fe Institute - Technical Report N°94-03-012

Forrest Stephanie (1991) Emergent computation :Self-organizing, Collective, and cooperative phenomena in Natural andArtificial Computing networks - Special issue of Physica D - MIT Press /North-Holland

Forrest S., Mitchell M. (1992) What makes a problem hardfor a genetic algorithm ? Some anomalous results and their explanation.Machine Learning

Goldberg David E. (1991) Genetic algorithms - Addison-Wesley USA

Goldman Claudia V., Rosenschein Jeffrey S.(1994)Emergent Coordination through the Use of Cooperative State-Changing Rule -AAAI

Henseler J. (1995) Back propagation - In Artificial neuralnetworks - Braspenning P.J., Thuijsman F., Weijters Editors - LectureNotes in Computer Science - N°931

Hogg Tad, Huberman Bernardo A. (1992) Better than the best: The power of cooperation - Lectures in complex systems - Addison Wesley

Huberman Bernardo A. (1991) The performance of cooperative processes - InEmergent computation : Self-organizing, Collective, and cooperativephenomena in Natural and Artificial Computing networks - Edited byStephanie Forrest - Special issue of Physica D - MIT Press

Kauffman S. (1993) The origins of order: Self-organizationand selection in evolution. Oxford University Press.

Kirckpatrick S., Gelatt C.D., Vecchi M.P. (1983) Optimization by simulated annealing - Science Vol220 N°4598

Macready William G., Wolpert David H. (1995) What makes anoptimization problem hard ? Santa Fe Institute Technical Report N°95-05-046

Matàric Maja J (1994) Interaction And IntelligentBehavior - PHD of Philosophy Massachussetts Institute of Technology

Maturana Humberto R. , Varela Franscisco J. (1980)Autopoiesis and cognition - The realization of the living - D. ReidelPublishing Company - North Holland

Maturana Humberto R., Francisco Varela (1994) L'arbre de la connaissance -Addison Wesley

Mitchell M. , S. Forrest (1994) Genetic algorithms andartificial life - Artificial life Vol1 N°3 - The MIT Press

Piquemal-Baluard C., Gleizes M.P., Camps V., Glize P.(1996)Une thèse pour l'auto-organisation des systèmes artificiels -Congrès européen de Systémique - Rome

Rasmussen Steen , Carsten Knudsen, Rasmus Feldberg, MortenHindsholm (1991) The coreworld : Emergence and evolution of cooperative structures in acomputational chemistry - In Emergent computation : Self-organizing,Collective, and cooperative phenomena in Natural and Artificial Computingnetworks - Edited by Stephanie Forrest - Special issue of Physica D - MIT Press / North-Holland

Rosen Robert (1985) Anticipatory systems - Birkhauser

Rosen Robert (1991) Life itself : A comprehensive inquiryinto the nature, origin, and fabrication of life - New York : ColumbiaUniversity Press

Salome Tristan , Bersini Hughes (1994) An algorithm forself-structuring neural net classifiers - 2th IEEE conference on neuralnetworks

Salomon Ralph (1996) Neural networks in the context ofautonomous agents : important concepts revisited - Artificial NeuralNetworks in Engineering

Salomon Ralph (1997) Derandomization of genetic algorithm -Fifth congress on intelligent techniques and soft computing EUFIT'97

Sekaran Mahendra, SEN Sandip (1995) To help or not to help- Seventeenth Annual Cognitive Sciences Conference - PitsburgPennsylvannia

Sen Sandip, SEKARAN Mahendra (1995) Using reciprocity to adapt to others -IJCAI

Stewart John (1997) On the evolutionnary origins ofrepresentations and intentions - In Auto-organisation et comportement - GuyTheraulaz et François Spitz Coordonnateurs - Editions Hermès

Tate David M., Smith Alice E. (1993) Expected allelecoverage and the role of mutation in genetic algorithms - In 1stInternational Conference on Genetic Algorithms - Stephanie Forrest (Ed) -Morgan Kaufmann Publishers

Thornton Chris (1995) Why GAs are hard to use -http://www.cogs.susx.ac.uk/users/christ/ index.html

Varela Francisco (1975) A calculus of self-reference -International journal of general systems, Vol2

Varela F., Thompson E., Rosch E. (1991) The embodied mind:Cognitive science and human experience. Cambridge MA: MIT Press.

von Bertalanffy Ludwig (1993) Théorie générale des systèmes -Editions Dunod

von Foerster H. (1981) Observing Systems: Selected Papersof Heinz von Foerster. Seaside, CA: Intersystems Publications

Von Foerster Heinz, Zopf G. W. (1962) Editors - Principlesof self-organization - Pergamon Press

Weiß Gerhard (1993) Learning To Coordinate Actions InMulti-Agent Systems - in Proceedings of the International Joint Conferenceon Artificial Intelligence

Weiss Gerhard (1996). Adaptation and Learningin Multi-Agent Systems. In Lecture Notes in Artificial Intelligence 1042.Springer Verlag.