

Towards a Practice of Autonomous Systems

Paul Bourguine

Laboratoire d'Intelligence Artificielle
CEMAGREF,
BP 121
92185 Antony Cedex

Francisco J. Varela

CREA,
Ecole Polytechnique,
1 rue Descartes
75005 Paris

1 NEW TRENDS

Artificial Life is the most recent expression of a relatively long tradition of thought which searches the core of basic cognitive and intelligent abilities in the very capacity for being alive. Metaphorically, this current of thought would see in a modest insect, rather than in the symbolic abilities of an expert, the best prototype for intelligence.

This recent surge of interest in 'artificial' life has to be understood in the background of the tradition inaugurated in earnest with cybernetics, seeking explicitly common grounds for the living and the artificial. In the early 50's, pioneering work of people such as Grey Walter and his turtles and W.Ross Ashby and his ultra-stable machine, were not given all the attention they deserved, as research efforts swung towards symbolic computation as the prime mover for research leading to what is still today the dominant form of artificial intelligence and cognitive science. After 30 years of a research program emphasizing symbolic computations and abstract representations, we can now benefit from this accumulated experience and the more recent re-discovery of connectionist models and neural networks which have come to pose important challenges to the dominant computationalist tradition (Meyer and Wilson, 1991).

The guiding intuition to both cybernetic forerunners and current proponents of artificial life is quite similar: the need to understand the class of processes that endow living creatures with their characteristic autonomy. Autonomy in this context refers to their basic and fundamental capacity to be, to

assert their existence and to bring forth a world that is significant and pertinent without be pre-digested in advance. Thus the autonomy of the living is understood here both in regards to its actions and to the way it shapes a world into significance. This conceptual exploration goes hand in hand with the design and construction of autonomous agents and suggest an enormous range of applications at all scales, from cells to societies.

From what we have said already, it is clear that in our view, AL has to go further in its self-definition. According to a recent formulation: "Artificial Life is a field of study devoted to understanding life by attempting to abstract the fundamental dynamical principles underlying biological phenomena, and recreating these dynamics in other physical media - such as computers - making them accessible to new kinds of experimental manipulation and testing" (Langton et al. 1991:xiv). Rather we think artificial life can be better defined as a research program concerned with autonomous systems, their characterization and specific modes of viability. This view is not in contradiction with the previously quoted one; rather, we claim that the foregoing definition needs to be made more precise, by focusing on those dynamical processes that assure the key features of autonomy more than any other dynamical principles present in living systems. Furthermore, it is by focusing on living autonomy that one can naturally go beyond the tempting route of characterizing living phenomena entirely by disembodied abstractions, since the autonomy of the living naturally brings with it the situated nature of its cognitive performances.

As white light seen through a prism, the autonomy of the living is articulated by a number of constitutive capacities such as viability, abduction and adaptability concepts on which research can actually advance. This first European Meeting on Artificial Life was intentionally focused on such key points of research, in an attempt to reveal common trends and to examine conceptual foundations. Whence the title: towards a practice of autonomous systems. Needless to say, not every contributor shares our viewpoint on the matter, but they share it to a sufficient extent to respond enthusiastically to our call for papers under this description.

This is the view that animates this book. The rest of this Introduction is our attempt to clarify further the cluster of key notions: autonomy, viability, abduction and adaptation. These notions form the conceptual scaffolding within which the individual contribution contained in this volume can be placed. Hopefully, these global concepts represent fundamental signposts for future research that can spare us a mere flurry of modelling and simulations into which this new field could fall.

2 AUTONOMY AND OPERATIONAL CLOSURE

At first glance, when studying a (natural or artificial) behaving system, one has two alternative ways of addressing it. If the system is considered as an heteronomous device, it is addressed as an input/output device, essentially defined by a set of instructions and the related control that will act upon it. In contradistinction, if a system is considered as an autonomous device, the center of attention is placed on emergent behaviors and internal self-organizing processes which define what counts as relevant interactions.

Obviously, living systems are the prime example of autonomous devices, and the most important source of examples of generalizable insights. It is on this basis that Varela (1979) proposed the following:

Hypothesis 1 (Closure Thesis):

Every autonomous system is operationally closed.

The notion of closure here is intended in its algebraic sense: a domain K has closure if all operations defined in it remain within the same domain. The operation of a system has therefore closure, if the results of its action remain within the system itself. This notion of closure has nothing in common with the idea of a closed system or closedness, which means an incapacity to interact. Obviously we are interested in interacting systems; the main point is the nature of internal dynamics of the systems which determine how the arriving interactions are interpreted, rather than being pre-given as information-rich inputs. One of the most paradigmatic cases of operational closure is the very origin of life as the emergent unit of minimal cellular organization, where the biochemical closure of membrane constitution and metabolic repair make the cell a viable self-distinguishing autopoietic unit (as discussed more extensively elsewhere in Varela *et al* .1974; Varela, 1991).

For our purposes here, we wish to focus on the operational closure of sensory-motor action. We claim that contemporary neurosciences - like cell biology for the case of the cellular organization - gives enough elements to conceive the basic organization of a cognitive self in terms of the operational closure of the nervous system (Maturana and Varela, 1980; Varela, 1979, 1991). Again this is not interactional closedness; we speak of closure to highlight the self-referential quality of the interneuron network and the perceptuo-motor surfaces whose correlations it subserves. More specifically, the nervous system is the operational closure of reciprocally related modular sub-networks giving rise to ensembles of coherent activity in such a way that :

(i) they continuously mediate invariant patterns of sensory-motor correlation of the sensory and effector surfaces;

(ii) they give rise to a behavior for the total organism as a mobile unit in space.

The operational closure of the nervous system thus brings forth a specific mode of coherence, which is embedded in the organism. This coherence is a cognitive self: a unit of perception/motion in space, sensory-motor invariances mediated through the interneuron network. The passage to cognition occurs at the level of a behavioral entity, and not, as in the basic cellular self, as a spatially bounded entity. The key to this cognitive process is the nervous system through its distinctive operational closure. In other words, the cognitive self is the manner in which the organism, through its own self-produced activity, becomes a distinct entity in space, though always coupled to its corresponding environment from which it remains nevertheless distinct. A distinct coherent self which, by the very same process of constituting itself, configures an external world of perception and action.

These two cases, the cellular unit constituted via the metabolic network and the constitution of a cognitive agent via the nervous system, are two prime examples of the way in which the Closure Thesis is validated. Its intent is to make the admittedly intuitive notion of autonomy into a more pragmatic one, in the same spirit of, say, how Church's Thesis works by identifying calculability with recursivity on the basis of accumulated experience. Similarly, our experience of autonomy as manifested in living systems makes it plausible that we can generalize : autonomy is always obtained when we endow the system with a rich enough closure of its constitutive process.

3 ABDUCTION AND VIABILITY

Autonomous systems must be, in harmony with Hypothesis 1, endowed with the capacity to be viable in the face of an unpredictable or unspecified environment. The concept of viability can be made quite precise, as discussed by J.P.Aubin (1991). The key idea is that

one considers the dynamical description f of a system's closure as giving rise, not to an unique solution, but to an ensemble of possible solutions. One works with differential inclusions (Aubin and Cellina, 1989) rather than with the more familiar differential equations. From amongst the ensemble of possible trajectories, the system must "choose" one so as not to depart from a domain of constraints which guarantee its continuity, the viability subspace K .

Aubin has developed these ideas mostly in reference on how an heteronomous system is kept viable by an observer introducing the appropriate control parameters, which are different from the state parameters. This is clearly inadequate for autonomous systems since they are not endowed with an external controller giving well defined inputs. The system is defined by its closure, the embodiment of sensori-motor cycles configuring what counts as perception and action. Every state change $s(t)$ is the basis for the new state given by the closure dynamics $f(s(t))$ eventually modulated by the coupling with independent and unspecified perturbations. If we consider for clarity the discrete case, this can be written as:

$$s(t+1) \in f(s(t))$$

The system will cease to operate when there are no more accessible states : its domain of dis-organization is thus defined. The viability domain can only be defined relative to f . In the classical theory of viability and control, one assumes K known and the observer or the system chooses the next state if one exists within K :

$$\forall s \in K \quad f(s) \cap K \neq \emptyset$$

This point of view is too demanding for an appropriate theory of autonomous systems, since it supposes an omniscient control from outside, even if the controlling agent is supposed to be not human but environmental optimizing parameters. For an autonomous unit we must modify the definition as follows :

$$\forall s \in K \quad f(s) \subset K$$

This says that the function f for an autonomous system must guess, at any moment, a set of solutions that are all viable by eliminating all others. This capacity is best expressed in the notion of abduction. C.S. Peirce states abduction as "the mind's capacity to guess the hypothesis with which experience must be confronted, leaving aside the vast majority of possible hypothesis without examination" (Peirce, 6.530). This is the reverse of the usual logical implication arrow, which is close to what we need for an autonomous behavior. To fix ideas let us give a logical characterization of abduction as follows:

Definition: An abductive machine is a function $f: A \rightarrow 2^B$ which interprets the indices in a space A to produce a restricted domain of hypothesis in a space B .

This abductive capacity is both an eliminating capacity and a hermeneutic (i.e interpretive) capacity. We can evaluate these capacities in probabilistic terms by assuming a uniform measure μ on B , so that

- the eliminating abductive

$$\text{capacity is } 1 - \frac{\mu(f(s))}{\mu(B)}$$

- the hermeneutic abductive

$$\text{capacity is } \frac{\mu(f(s) \cap K)}{\mu(f(s))}$$

We are less interested here in the eliminating capacity than in the hermeneutic capacity. Let us remark that the hermeneutic capacity of all states of K is one, but, clearly, the above definition of a viability domain does not give the system a finite horizon of life. Once the system steps into its viability domain, it will remain for ever within it. For a more realistic description, we consider domains where the hermeneutic capacity is only close to one.

We will now utilize the hermeneutic abductive capacity as a key notion for the understanding of autonomous devices with a large adaptability, since they do not know *a priori* their viability domain (Bourguine, 91):

Hypothesis 2:

Every autonomous system behaves as an abductive machine with a hermeneutic acquired capacity close to unity along its trajectory of states.

It is clear that in living systems the hermeneutic capacity is almost always close to unity, since they realize an enormous numbers of sensori-motor cycles and yet remain viable. Everything happens as if the living system had the capacity to find viable behavior within a given life horizon, and thus as if both Hypothesis 1 and 2 were valid. Notice how operational closure and viability domain are closely linked although not equivalent. The first refers to the algebraic closure of the system on the basis of its previous state and its coupling; the second to the fact that it remains within bounds so that its operation may continue and hence within the viability domain. Hypothesis 1 and 2 represent, so to speak, the two complementary sides of the same process.

This situation has been the result of millions of years of natural evolution, and it thus differs strongly from the case of artificial devices. But it makes it clear that one central topic in artificial life is how to define and implement, as part of the systems closure, effective abduction processes. This demands a further clarification of what such abduction capacities might be, which is the theme of the next section.

4 VIABILITY AND METADYNAMICS

One important key to abduction capacities, proper to living systems endowed with a nervous system, has been widely explored recently in the form of distributed processes within a very well connected network. To the extent that such studies are based on

network architectures such as those found in all unsupervised learning schemes, their operational closure is guaranteed.

In general we wish to characterize an adapting process as a metadynamical procedure, that is, a procedure g which changes the dynamics f of the system after each time step. Thus the unit is expressed by the triplet:

$$Unit(g, f, s)(t) \rightarrow Unit(g, f + \delta f, s)(t + 1)$$

There are at least three important manners in which such metadynamics have been conceived so far, and it useful to see them as variants of the same research as just described:

1. Neuronal strategies: the nodes of the networks are fixed (and considered as some form of idealized neurons), but their connections are allowed to vary according to various algorithms, such as Hebb's rule or its variants (see e.g. Amit, (1989) for review of this large literature). The constraints for an autonomous system is that learning should be either by reinforcement (various models) or by unsupervised learning (again various models available) (Kohonen, 1984). Metadynamics which assume a direct supervised learning such as back-propagation, are less directly applicable for implementation of autonomous devices for obvious reasons.

2. Genetic strategies: here the emphasis of the metadynamics is not so much on a fixed set of nodes in a network (neuronal or not), but on the way in which one can apply rules for the replacement and updating of the participating nodes. A classical approach is via recombination of the previously existing nodes, much like in genetic recombination of chromosomal DNA. This approach was originally developed by Holland (1975, 1987) and is in full expansion (Grefenstette and Goldberg, 1991).

3. Immunitary strategies: like the previous case, connections between cells are not modified *per se*, but the list of active or participating agents changes continuously. The difference with genetic

algorithms is that the new participating nodes are not recombinations of previously existing ones, but newly recruited nodes from a large potential pool. The core of this most recent of metadynamical processes is inspired by the amazing adaptability of the immune system (Bersini and Varela, 1991).

In all these cases, the operational closure of the system is respected by a trajectory, not only in state space but in continual change in the defining dynamics, which maximizes the chances that the choice of the next state will remain in the viability domain. Thus the observation of natural systems teaches us yet another lesson: metadynamical capacities are essential for their capacity for viability, and, by the same token, for their adaptability in vastly different and unpredictable environments.

5 ARTIFICIAL LIFE: PERSPECTIVES

In this introductory essay we have tried to be explicit about some key notions which are shared by various researchers in artificial life. We reiterate that in order to avoid falling into the trap of a mere fashionable buzz word, or a fascination with technological wizardry without direction, it is important not to lose sight of the deep issues that animate this re-surgence of research. Our view, as we stated at the outset, is that AL finds its *elan* because it (re-) discovers the central role of the basic abilities of living system as the key to any form of knowledge from simple sensori-motor capacities all the way to symbolic interaction. Therefore, the key is the identification of such basic living properties and our stance here is that autonomy is the emblematic quality which needs to be unfolded into clear and practical concepts as expressed in our two Hypotheses.

In contrast with classical models in AI and cognitive science, AL if understood as we have presented it here, leaves behind the notion of knowledge as being a form of problem-solving of problems already posed; knowledge becomes instead the capacity to pose the

relevant problems which are solved in order to preserve viability (Bourgine & Le Moigne, 87). These two positions are sufficiently different to become a "nouvelle AI" in the words R.Brooks (Steels, 1992). Inside cognitive science there has also an historical progression, which began with classical cognitivism (or computationalism), continued with connectionism which opened the way to a full appreciation of issues dear to AL. The last step in this progression has been characterized by one of us as an enactive view of cognitive processes , which also places the autonomy of the system as its center and is thus naturally close to AL (Varela, 1989; 1991).

The comparative table below is an attempt to evoke a progressive shift of

research trends and their corresponding emphasis in some dimensions of analysis (Varela, 1989; Bourguine, 1991).

To conclude, this first ECAL is a living proof that the issues just discussed are being addressed by a growing community of researchers from all disciplinary fields, and that theoretical, conceptual and engineering progress is quite possible in notions that until recently were dismissed as merely metaphorical. The practice of autonomous systems is not any longer a matter of mere vague speculation in contrast to a well developed theory of control systems. It is the necessary enlargement of the field of science to encompass what is most interesting in life and knowledge.

	<i>Cognitivism --></i>	<i>Connectionism --></i>	<i>Enaction</i>
<i>problems:</i>	well defined		not representable
<i>resolution:</i>	heuristic methods		adaptive methods
<i>reasoning:</i>	deduction		abduction
<i>behavior:</i>	given		emergent
<i>knowledge:</i>	symbolic		know-how
<i>characterization</i>	controlled		autonomous
<i>success criteria:</i>	validity		viability

6 REFERENCES

Amit,D. (1989), Neural Networks, Cambridge U.Press,1989.

Aubin, J.P.(1991) Viability Theory, Birhäuser, Berlin

Aubin,J.P. and C.Celina, (1989) Differential Inclusions, Springer-Verlag, Berlin.

Bersini,H. and F.Varela, (1991), Hints for adaptive problem solving gleaned from immune networks, in: H.-P.Schwefel and R.Männer (Eds.), Parallel Problem Solving from Nature, Lecture Notes in Computer Science N°496, Springer Verlag, Berlin, pp.343-354.

Bourgine P., Le Moigne J.L. (1987), The intelligence of economics and the economics of intelligence,in: Economics and Artificial Intelligence, Ed. J.L. Roos, Pergamon Press, 1987.

Bourgine P. (1991) , Heuristic and Abduction, Report n° 65, Dept. of Computer Science, University of Caen .

Grefenstette, J .J. and D .E. Goldberg (1991) (Eds.) Genetic Algorithms, Morgan Kauffman, 1991

Holland, J.H.(1975), Adaptation in natural and artificial systems. Ann Arbor : the university of Michigan Press.

Holland, J.H., (1980), Adaptive algorithms for discovering and using general patterns in growing knowledge-bases, Int.J.Policy Analysis Inform. Syst. 4 : 245-268.

Holland J.H. (1987), Genetic algorithms and classifier systems : foundations and future directions, Proc. Second ICGA .

Kohonen T.(1984) Self-Organization and Associative Memory. Springer Verlag, Berlin.

Langton, C., Taylor, C., Farmer, J.D., Rasmussen, S. (1991) Artificial Life II, Addison-Wesley, New Jersey.

Meyer J.A., Wilson S.W. (1991), From animals to animats, M.I.T./Bradford Books, Cambridge, MA.

Maturana, H. and Varela, F. (1980) Autopoiesis and Cognition: The realization of the Living, D.Reidel, Boston.

Steels, L. (1992) (Ed.), Situated Cognition, Emergent Functionality and Symbol Grounding, Artificial Intelligence Journal, (in press)

Varela, F., Maturana, H. and Uribe, R. (1974) Autopoiesis: the organization of the living, its characterization and a model, BioSystems 5:187-195.

Varela, F., (1979) Principles of Biological Autonomy, Elsevier/North-Holland, New York.

Varela, F. (1989) Connaître: Les sciences cognitives, Eds. du Seuil. Paris.

Varela, F., (1991) Organism: A meshwork of selfless selves, in: A. Tauber (Ed.) Organism and the Origin of Self, Kluwer Associates, Dordrecht, pp.79-107.