

Arantza ETXEBERRIA, Juan Julian MERELO,
Alvaro MORENO

Studying Organisms with Basic Cognitive Capacities in Artificial Worlds

I. Introduction

One of the most peculiar features of the methodological style of Artificial Life (AL) (Langton, 1989; Langton, Taylor, Farmer & Rasmussen, 1992; Varela & Bourgine, 1992; Fernandez & Moreno, 1992) is the attempt to ground all the processes that concern living systems with the purpose of --sooner or later-- being able to artificially emulate them. After the functionalism which was the main characteristic of AI, now a way is open for a new structuralism that will make possible to study the behavior of focused systems through relations that are closer to what is materially realizable. The AL approach is developed towards the ideal of being able to display genuine evolutions of artificial natures in the computer, so that the simpler will generate the more complex through an emergent evolution. However, an absolute fidelity to that ideal would not make it possible to study systems as complex as the cognitive ones we are interested on. Even basic cognitive phenomena constitute a degree or level of organization of matter where access is difficult (because they present a huge complexity in physico-chemical terms and, therefore, in computational ones). Thus, the only option left in order to study these phenomena is to simplify the underlying materiality by designing certain features.

Then, in this domain we have developed an artificial world model whose main interest is to study the cognitive capacities of the artificial organisms and their evolutionary consequences. This

approach can uncover various types of problems that are often left aside when the cognitive phenomenon is approached only at high levels (Belew, 1991).

The paper is organized in the following way: section 2 presents a review of the properties of the systems that can be in various ways considered autonomous with respect to their environments at different degrees of complexity, in order to establish some criteria to estimate the complexity of cognitive phenomena. In section 3 a characterization of the origin of cognition is presented, as arising from forms of adaptation to the environment that are linked to the genetic specification of the system operation. In section 4 some basic ideas are presented on the modellization of what we consider basic cognitive capacities through models of simulation of artificial worlds. Finally, in section 5 these ideas are briefly discussed and some conclusions are proposed on the advantages of this approach to study cognitive problems.

2. Levels of System-Environment Relationship

A cognitive process only can take place in the frame of a system that maintains some degree of autonomy and self-determination in respect to its environment. However, we cannot consider that all natural systems where some form of autonomy is observed are cognitive. In fact, we can distinguish several degrees of complexity among autonomous systems, according to the kind of interaction they establish with their environments, and their artificial representation should vary according to this degree of complexity. We will try to make this idea clearer through a short course through self-organizing systems in a gradation of increasing complexity so as to establish some criteria on the features that are essential to model artificial cognitive systems.

2.1. Self-maintained minimal systems

The most elemental notion of a system that internally defines its identity in respect to its surroundings can be explored in the formation of systems connected by an operational closure. There

the system properties will critically depend on the components that take part and the relations that are established among them (Kampis, 1991).

In the context of Protobiology those systems are supposed to present a fundamental feature of life: the possibility of forming connected ‘protoorganisms’ whose relationships with the environment depend on the interactions of the molecular components of the system, the system being able to reproduce by fracture and, in some cases, to entirely reconstruct the set of relations that supplies its systemic identity. Even though, such reproduction is not mediated by the existence of ‘template’ components capable of self-replication; thus, reproduction presents a very low degree of reliability, and all system components are at the same hierarchical level.

Models of this kind of systems have been developed by the formation of self-maintaining reaction networks, such as autocatalytic sets. Those are autonomous systems created to grasp the properties of the protobiological organization; they are based on molecular chemical reactions basically developed to pose the problem of the origin of life. Autocatalytic Sets (Kauffmann, 1986; Farmer, Kauffmann & Packard, 1986) are good models to study the minimal conditions of complexity of the systems where some kind of functional emergence will be found. Their main property is that starting from an initial set of components whose interaction is chemical (catalyzed cleavage or condensation of components), it is possible to observe connected sets of components forming a stable network, meaning by stability the capacity to present a coherent behavior in the presence of perturbations and to self-maintain in the continuous flow of energy and materials coming from the network environment. Therefore, an unstable network would lose connectivity, the reactions among components would not allow any global behavior and unconnected parts would appear in the whole of the system.

This capacity of forming systems with an operational closure among components that stand at a single level has been viewed by the Autopoietic Approach (Varela, 1979; Maturana & Varela, 1984) as a basic and sufficient condition for life. The autopoietic organization would appear as the top of the following gradation of

self-organizing processes (Fleischacker, 1988): 1) dynamical systems; such as dissipative structures that do not imply transformation of matter (examples would be whirlwinds or Benard cells); 2) 'oscillating' systems as special cases of crossed autocatalytic reactions (for example, the Belusov-Zabotinsky reaction) that imply transformation of matter; 3) Autopoietic Systems, as networks of component production in watery mediums (for example, living cells). This kind of system is limited by a physical closure constructed by itself (membrane) and maintains its organization by exchange of matter and energy through that structure. The general idea of autopoiesis is self-production (Varela, 1981), as a special form of self-organizing process that constitutes living systems.

Anyway, even if the criteria of autopoiesis are necessary ones for a definition of life, they are not sufficient (Moreno, Fernandez & Etxeberria, 1990). Important features of living systems such as the possibility of transmitting their organization through self-reproduction and therefore, undergo Darwinian evolution cannot be explained by this approach. In fact, the Autopoietic Approach deliberately leaves those properties aside as consequences of the 'autopoietic structures' that should not be taken into account when defining the 'autopoietic organization' (Varela, 1979). They have important reasons to defend this view; the main could be the necessity to argue against another that attributes the most relevant role in the constitution of life to isolated components like nucleic acids. Instead their aim is to promote a vision of living organization that lies on the whole system, understood as a set of recurrent operations that continuously re-construct the system (Etxeberria, 1992). Therefore, it is the necessity of blurring the role of informational components (templates) as underlying the living and to promote the idea of life as a property of the whole what underlies the autopoietic rejection of these properties (self-reproduction, evolution) from the definition of life. But, in our view, these capacities of living systems denote a degree of system complexity that cannot be given for granted. Minimal autopoietic systems cannot either reproduce with a reasonable degree of reliability nor evolve, therefore our conclusion can be that living systems are necessarily more complex than autopoietic systems.

2.2. *Minimal Living System and environment*

Even if we do not wish to under value the importance of systems, such as the previously described as decisive steps both for the natural origin of life and for an eventual artificial one, from our point of view what in fact characterizes life is the formation of a self-maintaining organization where there exist two entangled levels of components: one of a conservative type (like nucleic acids) and another of a dissipative type (like proteins) forming a functional closure. Pattee (1982) has theorized this organization as forming a ‘semantic closure’.

An organization defined by a ‘semantic closure’ can not only explain the self-maintenance and self-production that characterize the autopoietic organization, but also biological phenomena such as self-reproduction and evolution. Unlike the Informational Paradigm¹ in Theoretical Biology, a paradigm based on the ‘semantic closure’ will not view nucleic acids as trivial informational carriers, because their ‘meaning’ depends on a dynamic interpretation realized by the whole system. Unlike in the Autopoietic Paradigm, living organization does not rely merely on an ‘operational closure’ of a syntactic type, but the history of living systems endows each of them with hereditary structures that act as ‘symbols’ in the frame of the whole system.

According to Pattee, the closure can be understood as an interdependence or complementarity between some dynamic elements, whose change is rate-dependent (proteins) and other symbolic ones, whose change can be described as rate-independent (nucleic acids). Rate-dependent dynamics account for the creative processes of the living, while rate-independent symbols can grasp the historically stabilized features of life. The molecular strings of genes only become symbolic representations when the physical tokens of symbols are directly recognized by the ‘translating’

¹ In other works (see, for example, (Moreno, 1988)), we have referred to the view based on the Central Dogma of Molecular Biology as the Informational Paradigm and have argued that the Autopoietic Paradigm should be considered as an alternative Paradigm in Theoretical Biology.

molecules (aminoacyl-tRNA-synthetase), who exert arbitrary specific matching actions that result in protein synthesis. Then, once folded, proteins finally execute functional actions. Thus, without the rate-dependent functional action of proteins, the ‘meaning’ of the genes would not exist, the semantic closure arises when the ‘translating’ molecules themselves are referents of the gene strings. Therefore, in the cell it is not possible to opt solely for one or the other of the two dimensions of the phenomenon without losing explanatory capacity, as would be the case if we think that all living phenomena arise from the properties of informational molecules (Informational Paradigm) or that all living phenomenology can be explained in terms of dynamic components constituting a coherent whole.

In this organization, due to the replicating capacity of certain components in space and time, the system can ‘construct’ another similar one, it can reproduce itself reliably. This construction takes place in a space that is also constructed from the inside of both systems (original and copy). This capacity ensures: 1) to obtain a fracture plane in space that warrants division into two similar copies; 2) to duplicate certain patterns which are indispensable for the identical operation of original and copy and 3) autonomy of the system, that is the operational closure of the process (Fernandez, 1992).

From this perspective biological information is not independent from the rest of the physico-chemical interactions taking place inside the system or in the system-environment relationship. Instead it is the capacity of certain physical entities of exerting diverse actions in respect to other system components or the whole system; it is not derived from intrinsic properties of system components such as their chemical composition, but from the specific network of interactions where it is exerted. To be able to talk of information there must exist alternative configurations; in the cell information can be stabilized and transmitted due to the existence of template components whose conservative order makes possible their functioning as ‘records’.

In our view this two level organization generates all biological phenomenology. Rosen (1959) developed a series of works to model systems of this kind: M,R (Metabolic Repair) Systems, which are complex reaction networks that can evolve. In this kind of

network there is a reciprocal interaction between the metabolic units (M) and the repairing or genetic ones (R): each R unit of the system depends on the outputs of the M system and M units are controlled by the R units. This kind of interaction of entangled levels is what makes this system far more complex than the canonical reaction networks. Therefore M,R systems are not just models of the emergence of metabolisms (as for example, autocatalytic networks), but a network that grasps the operation of the 'semantic closure'.

2.3. The cognitive specificity

Biological organization can be characterized as a network of component production in which the existence of information is essential to synthesise the specific components (proteins) that control metabolic operation so that the operation of the system can be stable under the fluctuations of the environment. The network is closed by a semipermeable membrane through which there is a selective exchange of reactives and energy (mainly through membrane proteins).

Even if systems pertaining to both of the previously considered categories (protoorganisms, minimal living systems) maintain a relation with their environments (it does not make sense to pose the problem of autonomy for an isolated system), the specific feature of the cognitive level is the appearance of a specialized subsystem for the regulation of that relation, so that the organism is able to develop a structural plasticity that can be modulated in ontogenetic time. Thus, cognitive phenomena bring about the possibility of forming material structures specialized in the maintenance of suitable relations with the environment. The origin of cognitive systems is related to the increase in complexity and selective specificity of that exchange of materials and energy between system and environment (self-organization and selection) and the construction of controls that allow the fixation and reproduction of those paths.

The cognitive system accomplishes an integrating role for the organism, it is superimposed on the biological ground forming a hierarchical functional network. The characteristic behaviors of this

kind of organization appear in the upper level to which corresponds co-ordinates the lower ones, but its function cannot be isolated from the operation of lower levels. The cognitive system adapts the behavior of the organism to a changing environment acquiring knowledge in the course of its life and being open to the relation with the world. This relation is epistemic or informational, because it implies the detection of relevant aspects of the environment, encoding some physical patterns into informational ones (symbols) that finally trigger functional actions. The relation of the cognitive system with the biological one is pragmatic and, even if pragmatics cannot fully explain semantics (as syntax cannot either), it makes possible to understand the cognitive system as one of self-determination for the organism. As a consequence, by the insertion of the cognitive system in that global context we can understand or analyze its representations as symbol systems characterized by a triple dimension of syntax, semantics and pragmatics. In fact the reduction of cognition to a single one of them has originated an intense debate in cognitive science around the idea of representation (Cariani, 1989); if the pragmatic aspect of biological functionality is left aside the semantic relation of referentiality of representations becomes intractable.

3. From Adaptation to Learning

According to the presentation of the last section, there seems to be a gap between what we considered Minimal Living Systems and Cognitive Systems. In order to find a link, we should first of all answer the following question: how can this internal system that correlates the behavior of the organism and the characteristics of the environment originate and evolve? The constitution of a minimal living system able to reliably self-reproduce implies the possibility of phylogenetic adaptation of populations of organisms to changing environments through processes of genetic change (what inspires Genetic Algorithms, a computational procedure of problem solving that we will see later). However, the capacities we are interested in do not depend only on phylogenetic adaptability, but require structures that are variable and modulable in somatic time.

3.1. Adaptation and the origin of the sensori-motor loop

Besides phylogenetic adaptation, all existent living systems possess some mechanism of ontogenetic adaptation. Basically, ontogenetic adaptation consists in a mechanism of functional self-modulation of the metabolic network. In its simplest form, adaptation is achieved through the selective activation of the pertinent genes when certain environmental conditions are detected. This kind of adaptation can be understood as ways of connecting detection mechanisms with those that regulate the genetic repertoire, producing changes in metabolic paths that will not have reproductive consequences, but can enhance the production of components that trigger precise functional actions.

These detection mechanisms constitute the most elemental version of perception. Several authors (Pattee, 1982; Conrad, 1988) have proposed that the classifying capacity of the substrate-recognition by enzymes is the most elemental form of a detection process. This hypothesis is supported by the fact that all the increase of complexity of epistemic processes that arises in biological evolution (including the functioning of the nervous system) is grounded on mechanisms of enzyme recognition (Koshland, Goldbeter & Stock, 1982). But a process of perception entails more than the enzyme pattern recognition capacity by the enzyme. Pattee (1992) exposes more accurately his first position, by stating that the 'detection' (or perception) occurs when pattern recognition is arbitrary, repeatable, stable and with operative consequences. Besides this process is rate-independent, because it must be distinguishable from merely dynamical processes and certain record or memory is necessary which, finally, can be reduced to that separation of rates or scales. From Pattee's point of view, perception is an intermediate process between dynamics (physical laws) and computation (the process of symbol manipulation by rules): even if its result were not symbolic and computable, it must be some kind of record and must be distinguished from dynamics. Intuitively that distinction brings forth the idea of some kind of store that keeps the result of perception to deliver it operatively later.

The problem of this position is that in order to consider something as perception, it is indispensable to be able to functionally recognize/evaluate the discrete output as a significant event/structure in respect to what it presumably 'detects'. As an example, the enzyme recognition of the substratum is not in itself an act of perception (of the substratum) unless there is an operationally closed network that interprets the aforementioned change of enzyme configuration (for example, by a change of the metabolic path that synthesises certain product). In the case of the immune system, components that do not belong to the operationally closed system are molecularly recognized and evaluated by the network itself, which can distinguish between what is or is not its own. But a phenomenon of this nature takes place also in the cellular domain, because it is the cell itself --or, better, the network that defines the reproductive identity of the cell-- which evaluates or recognizes the enzyme changes according to certain events/structures of the environment. Thus, when certain membrane proteins or a specific set of such molecules (Kremen, 1992) receive specific physical patterns, they undertake a configurational change that triggers metabolic-motor reactions; these ones, in their turn guide subsequent perceptions, so that a new functional closed loop is formed.

So, a system with mechanisms of perception must be essentially functional, while the converse is not true, functionality is not a sufficient condition to speak of perception. In the case of a minimal biological system like the one presented in 2.2., only able to maintain and reproduce itself, the enzymes recognizing genetic information should not be considered as 'sensors', but only as generically functional components, because there is not any previous or more basic mechanism for the functional evaluation of the different metastable states of those enzymes. As a consequence, there must pre-exist a system able to self-define its identity in order to be able to conceive even the most elemental process of perception that refers to something external. This way the changes triggered by the perceptive process could modulate the behavior of the system in correlation with the external or environmental changes that have been 'recognized'.

Thus, a perceptive process starts with the detection of certain changes taking place in the boundaries of the organism. It is basically a selective process of pattern recognition linked with certain functional consequences for the system which performs it. To fix the sensorimotor loop, living systems should selectively discard (either phylogenetically or/and ontogenetically) a great amount of components and metabolic paths. In this way, epistemic coupling is achieved through a recursive interaction with the environment (producing its modification), as a mutual and progressive organism/environment adjustment until certain stability points are reached. The organism/environment relation can be seen as a closed correlation between perceptions of the relevant properties of the environment --its 'affordances' in terms of Gibson (1979)-- and motor actions on it. Both processes are complementary in the sense that perception must be active (the organism moves towards its goal object, acts to perceive it) and action must be guided by perception. Perception is a requisite for optimum action, but both are entangled in a closed loop.

3.2. Cognition and the development of adaptive sensors and effectors

From the viewpoint of its origins cognition and learning arise as a result of greater complexity of the sensorimotor loop. Functionally speaking this increase in complexity is directed to the control, integration and hierarchical organization of an increasing number of biological activities. Although cognition does not define the set of biological needs, it is directed to the optimization of their realization. Therefore, even if cognition cannot be studied apart from biological functions, it has a different specificity: their global integration through mechanisms that imply informational processes.

While in purely adaptive organisms perception is, as we said before, the direct cause of certain metabolic-motor actions, in cognitive organisms the physical patterns impinging on sensors are transformed in trains of discrete sequences (which constitute information) that modify the state and dynamics of a network of connections where sensory information is processed. Unlike in metabolic networks, where there is no distinction between units and

connections (Minch, 1993), in neural networks the stress is made on the variability of connections and on the control (by/of the very network, throughout other layers or global patterns) upon such structural variations. Therefore in the former case structural changes take place only in the frame of phylogenetic evolution, while in the latter this kind of process can also take place in somatic time (learning). That is why the concept of (epistemic) information processing needs the development of a system of channelling as rich and modifiable as possible.

When high level cognitive functions are being considered, most research strategies emphasize essentially the increase in complexity of the intermediate net connecting sensors and effectors. Even if, no doubt, this is a fundamental factor, one should not forget that evolution toward more complex forms in the system processing sensory information is correlative to the complexity of sensors and effectors. Often, when we face the task of building artificial models of cognitive systems, this is left aside or under valued, mainly because cognition is not approached from a radically evolutionary perspective, that is, as a development of the sensorimotor loop. At higher levels of cognition the increasing complexity of the different elements of the system makes them appear as nearly autonomous subsystems. But it is an empirically verifiable fact that in natural cognitive systems there is a closed and tangled correlation between the development of sensors, the information processing network and effectors and cognitive science should take it into account.

This is why we think that if the model of artificial cognitive systems we are going to develop presents a neural network to allow learning, its sensors and effectors must be also adaptive. By adaptive sensors we understand those able to change the mapping between the type of output signals and their functional consequences through learning, so that the meaning assigned to the result of the sensor can vary according to the different circumstances of the environment in which it is produced and, thus, it is modifiable by learning. Given that this principle in a converse order can be applied to effectors, in this work we will focus on the study of sensors.

4. Basic Ideas to model Cognitive Systems in Artificial Worlds

In order to place the problem of modelling a cognitive system in its biological ground, we should move from the domain of Artificial Intelligence (AI) to the one of Artificial Life (AL), where evolutionary modelling of autonomous systems (Maes, 1990; Brooks, 1992; Harvey, Husbands & Cliff, 1992) focuses on the necessity of devising ways to approach their study as physical embodied systems, either biological or artificial (Varela, Thompson & Rosch, 1991). Contrary to the high-level, disembodied perspective of Classical Artificial Intelligence, its goal is to understand intelligence as a form of adaptivity that has evolved phylogenetically and is ontogenetically developed. The strategy to study is not to start modelling high level intelligent behavior such as theorem proving or playing chess as processes developed independently of their biological background, but to develop a lifelike structure (real or simulated) for the cognitive agent.

Some of the issues emphasized by the new approach are: 1) the importance and pre-eminence of low level capacities for an accurate notion of intelligence, therefore 2) the necessity of studying sensorimotor loops underlying behavior, so that 3) action as a control of perception, arising from the situatedness of an agent in its environment or ecological niche.

Work in this field is based both on robots (realizations) and simulations, as far as simulations are computational models of physical interactions underlying adaptivity and cognition. Realizations have the difficulty of how to implement evolution connected with reproduction, so it is a usual practice to combine both methods; it is not currently possible to create a building procedure robust enough to allow genetic variations, and simple enough to be implemented on a small machine. A review of the simulation/realization controversy in this field can be found for example in Brooks (1992) and Harvey, Husbands & Cliff (1992). A theoretical discussion of the problem in Pattee (1989).

In order to construct the structure of the artificial organisms, evolutionary modelling uses a procedure to design cognitive

architectures inspired by the operation of Darwinian evolution: Genetic Algorithms (GA). GAs provide an automatic method for structure development that permits to obtain interesting architectures from a population of random possible ones. This population undergoes an evolutionary process inspired by the genetic recombination of sexual reproduction (mutation and crossover) and selection is exerted upon it depending on a fitness function designed according to the desired behavior.

4.1. Evolution of the Artificial World

Thus, our effort is directed toward the creation of a simulated Artificial World (AW) where artificial organisms (AOs) can be found. All processes of the AW, either physical or epistemic (such as perception, learning or anticipatory behavior) take place in this artificial world, so they are simulations.

Following the wise advice of nature, we will use algorithms that try to mimic the way Nature works: Genetic algorithms. In them several solutions to a problem (in our case, the problem of surviving in an AW) are encoded in a 'genome', and they compete to be the best solution. Genetic algorithms will allow us to implement 'evolution', by coding all somatic characteristics of the AO (its 'body') in a genome. Only the fittest are allowed to reproduce, by mixing their genetic information (a string of 0s and 1s) with other good solutions.

Until now most of the effort has been channelled in the direction of modelling the cognitive structure of the autonomous agent, that is, its nervous system. The search of forms of development of neural networks (phenotypes) starting from the genotypes specified by the symbol string optimized/evolved by the genetic algorithm is a very difficult one. The use of genetic algorithms to encode neural networks has had an important development in the literature of the area --see, for example, Ackley & Littman (1992)--, nevertheless until now no biologically plausible form of representation has been proposed to understand in constructive terms the relation between genotype and phenotype for the artificial living systems.

Anyway, the field seems to be sensitive to the necessity of coevolving agent morphologies as well as neural networks (Brooks, 1992; Harvey, Husbands & Cliff, 1992) so that 1) the biological neural structure has the function of producing behavior, 2) behavior becomes adapted to a certain ecological niche and 3) the structure allowing adaptation has been conformed by evolution and can evolve and 4) the nervous system coevolves with the rest of the agents morphological traits.

The main problem to achieve it is the use of variable-length genes in the genetic algorithm, that would allow for open-ended evolution (usually by a enlargement of the genome, as shown in Harvey (1992). Nevertheless, this problem has been solved in several ways, as has been proposed by Koza (1990) and other authors (for instance Goldberg, Deb & Kob (1991) with their mGA paradigm). A system with open-ended evolution should use one of these algorithms.

In our model, the genome contains a complete description of the AO, including some 'metabolic' and 'somatic' characteristics (speed and energy consumption, for instance). The neural net of each AO will allow it to learn departing from its genetic information, changing the values of the connections between sensors and effectors continuously, thus, changing structurally. Neural nets combine several units (threshold-logic units), connected with each other, in the same way as biological neurones. The value of the connections can vary, making neural nets learn, or associate inputs with outputs in a meaningful way (working, for instance, as associative memories). These neural networks are correlated to features of their environments and react toward changes by varying their configuration in a proper way as different couplings take place. The adaptive sensors are also implemented by neural nets, and effectors are adaptively connected between them and the neural net.

The genetic description is compiled (in the computer language sense) at the time of birth. In principle, this process is deterministic, i.e., compiled structures follow necessarily from its encoded form. Nevertheless, as we said before, new forms of representation should be sought in which development were similar to the biological, so that the phenotypic structures were described in the genome only loosely, while other dynamical properties would follow the genome

in a functional way, (trying to simulate the duality between informational and dynamical levels that we have emphasized in subsection 2.2.) or could arise from interaction and competition between neurones, as proposed by Edelman (1987). Anyway, even if it is very difficult to think of ways of achieving this, the genome complexity and the coding of the neural net structures are positive steps towards it.

In order to evaluate the efficiency of these programs, besides watching the AO perform its duties within the world, several off-line tools are built for the program edition, execution, and alteration. For instance, a graphical representation of the cognitive subsystem is useful to evaluate differences among subsequent generations, or differences between different 'tribes' in the world, if something of the sort emerges.

4.2. Cognition of Artificial Organisms

Right at this moment, research in the field of AL has not developed any truly cognitive system, in the sense that it is really embedded in an organismic morphology and all its structure varies in a phylogenetic as well as an ontogenetic scale. The vast majority of artificial organisms --see for example, Ackley & Littman (1992)-- show only a moderate grade of adaptation, usually through the change of the neural net weights, being change at the level of the neural net structure far less common; most systems are currently lacking adaptive sensors and effectors.

A cognitive AO must obviously learn inside the AW. Learning does not mean only change of internal rules, it implies also the possibility of assigning new 'meanings' to the features of the environment that are detected (inputs) as the interaction of the AO with the environment takes place in different circumstances. This would mean that learning is mainly directed toward the realization of different biological needs or functions, that can perhaps change or develop in ontogenetic time (escaping from predators, hunting, mating, etc. are functions that will have to be fulfilled at different stages of the development of the AO). It requires the development of adaptive sensors, capable of growing and varying the realized

mappings according to learning (and similarly, the possibility of developing new effectors).

In AL, a ‘tabula rasa’ approach does not usually give good results. The organisms that populate the artificial world in the first place should have some innate capabilities (for instance, they should have at least one sensor and one effector that would allow them to move, and some kind of reflex or motivation to move unless they find food). Evolution requires many AOs present in the world at the same time because the ‘fitness landscape’ to be explored is huge and it is the only way of being sure that a sufficient amount of viable organisms will be created.

In our model the neural net structure remains frozen as it has been created, and only weights change during its lifetime, accounting for learning. These weights, that reveal the ontogenetic learning of the organism, are not inherited by the next generations. If new connections are created as a consequence of evolution, their values are randomly set.

4.3. Adaptive Cognitive Subsystems

We can analyze our AOs in terms of their sensitive, processing and effector subsystems. These must evolve and adapt phylogenetically, therefore the algorithms used must be adaptive in the ontogenetic and phylogenetic scales. New sensors should be developed during evolution, and the processing subsystem should make new associations between sensors and effectors, giving new meaning to inputs, while the AO is ‘living’ (that is, during the finite time it is allowed to function inside the simulated world).

4.4. Adaptive Sensors

Sensors cannot be understood if they are not related to a world, their function is to react to certain characteristics of the world that surrounds the AO, and process physical data in order to extract high-level sensory information: size of the object in front, odour, distance and so on.

As we have already said, the AO sensors can be adaptive in two senses:

- in a phylogenetic scale, they can develop the capacity of detecting new features of the environment. New sensors can be developed, according to the ‘physical’ characteristics of the world.
- in an ontogenetic scale, they can vary the internal meaning assigned to previously detected inputs, as the functioning of the sensorimotor loop acquires new experience of its surroundings. This is achieved by variations in the values of the connections among sensors, sensors and effectors, and sensors and the neural net.

How can these adaptive sensors be simulated? There are several problems to develop them from scratch, that is, from an initial system with random neural connections, for it would take aeons of computer time, and evolution surely evolved sensorimotor loops from systems that are already very complex, as was explained in sections 2 and 3. Research on the properties of cognitive systems from an evolutionary approach is still in a rather primitive stage. Indeed, most researchers working in the modellization of sensors usually create neural networks with fixed weights, with no learning, and connectivity patterns and strengths taken from experimental data. Even these simple sensors are of such a computational complexity that we cannot imagine to develop a population of beings, each with a 16x16 retina. Raw information picked up by this retina should be then processed to obtain high-level information. This is a problem that falls into the domain of artificial vision, but AL prefers minimalist versions that can help understand more general features of life (or lifelikeness), such as adaptation to an environment rather than detailed biological structures. Perhaps the computational power and the development of the field will make more complex models possible in a few years, but not in the current state of affairs.

For the moment we take what we call a ‘toolbox approach’. We can think that we already have all these sensors (distance, colour, size, shape) potentially developed, we put then in a ‘genetic toolbox’, and thus they can evolutionary exchange into another in reproduction or develop new ones, as a response to the presence of a new stimulus in a world. Each sensor is genetically specified to be

sensitive to a certain type of physical properties of the world. This includes a range in which it is effective and some information about the features it is able to detect.

Obviously, sensors must be sensitive to all the characteristics of objects of the world. One way of doing that is to define an Object Description Language (ODL) to perform couplings between relevant features of the AW and an OA sensor. Couplings are realized by some adjustment or matching rules between sensor specificities and features of the world described by this ODL. Due to this object-sensor interaction the AO can act according to the inference of the object characteristics realized by its sensors.

Our world would be composed of a potentially unlimited number of objects, all possible programs written in this language. The matching can help the OA act in the AW because the ODL includes qualities like whether the object is movable, whether it can be decomposed in smaller units, and so on. In this way, the AO can interact with the world, and in turn, the latter can interact with it.

Every AO has an adaptive sensor, or a set of them, only sensitive to some aspects of the objects of the world. A sensor fires if the rule it contains (for instance, BLACK and BIG) is met by the object/objects in front. Every time a rule fires, its firing is passed on to the neural network that constitutes the information processing subsystem of our AO. This way, adaptive sensors are not only sensitive to intrinsic object characteristics, but also to relational ones, like distance, orientation and speed.

Obviously, and taking into account what has been previously said about the coevolution of cognitive subsystem (entangled in the whole organism, that is, in its biological structure), it does not make sense to study the evolution of sensory organs as parts that are isolated from the rest of the cognitive system. Sensors coevolve with the neural net that processes their information, and the effectors that allow the AO to move to the perceived object, change it or get away from it.

4.3.2. Neural Net Subsystem

As we have already said, the task of implementing a neural network of variable structure and size, and to code it in a gene is

difficult, because the goal of simulating a complete world restricts the use of the available computational resources. Once again, we cannot pretend to evolve complicated learning algorithms from simple rules. Besides, each of these algorithms would have such a huge set of inputs (present and previous states and weights of the network) and outputs (variation of all weights), that even a simple set of rules would be computationally cumbersome. Thus, we can only hope to label each weight or directed connection as Hebbian or anti-Hebbian (in fact all learning rules can be reduced to this one), and let the structure change genetically. Each neurone is then labelled as input, output, or pass through, and information cascades from inputs to outputs, every discrete step going from one neurone to the next one. Information from several time steps is then concentrated in the output neurones.

The genetic coding of the neural network will include:

- A connection map, that tells how many neurones there are, and its connections.
- A neurone labelling, that classifies each neurone as input, output or pass-through.
- A connection or weight labelling, possibly mixed with the first, that tells whether the connection is anti-Hebbian or Hebbian.
- Initial values of weights.

The length of this part of the genome and its meaning can vary phylogenetically, making possible the development of a potentially infinite amount of neural networks.

4.3.3. Adaptive Effectors

In order to simulate effectors, we will take the same approach than for sensors. Effectors manipulate the world, affect some characteristics of the objects of the environment or change the spatio-temporal relation of the AO with respect to the world. Features of the objects of the world change as events take place, but this variation will only be appreciated in the following step, because in an AW time is discrete.

Effectors are also genetically coded. For instance, we are not concerned with how deambulatory mechanisms are developed and, therefore, we will use a toolbox approach too. There can be a finite amount of effectors (like walk, eat, mate, emit sound), which can appear or not in an AO. These effectors are connected between them, to the sensors and to the neural net, so that sensorimotor loops can emerge.

4.4. Current Model and Work in Progress

A system previous to this has been implemented with the aim of studying several problems of population evolution in function of the cognitive capacities of the AOs (Merelo, Paton *et al.*, 1992; Merelo, Moreno & Moran, 1992; Merelo, Moreno & Etxeberria, 1993). Now we intend to improve this model by applying to it the ideas that have been previously exposed on adaptive sensors. Even if our ‘toolbox approach’ does not make possible to study physical perception, it is useful to see other emergent features of the relation of the cognitive system and biological function in evolution. Several improvements of this model are currently being studied: development of organisms, relation between innate and developed structures, differential evolution of sensors, coevolution of organisms and environments, etc. so that the model will hopefully cover other theoretical issues apart from those presented in sections 2 and 3.

5. Conclusion

The attempt to pose the problem of the appearance of new cognitive capacities in relation to the biological structure of organisms, so that it is possible to observe the evolution of biological functions (feeding, reproducing, etc.) and their fulfilment via cognitive processes in artificial worlds is a great challenge for cognitive science. Traditionally it has focused on the study of high level phenomena and has considered that the underlying biological structure played a small if not insignificant role in the realization of the different cognitive tasks: perception, learning and memory.

This way, our model could overcome some epistemological limitations of current connectionist approaches of cognition. In such approaches, the cognitive systems are not able to autonomously find solutions for certain tasks, nor to determine their goals by themselves or change the ones specified from the outside (Van der Vijver, 1991). As a consequence, the (relative) self-organization occurring in the cognitive process is external and not linked to the constructive self-organization of the very cognitive system. In our opinion, the root of this unsatisfactory situation lies in the fact that the cognitive process is not considered as related in its origin to the self-reproductive one (Moreno & Etxeberria, 1992), and this has some consequences in the debate on the problem of representation in cognitive science.

Critics of classical AI maintain that the knowledge an organism has of its environment does not rely on a symbolic representation that can be specified from the outside, but it is not easy to explain how structures that are functional for the organism can originate and accomplish an epistemic function in relation to the environment. Very often this problem has been taken so far as to the adoption of anti-representationalist positions of different sorts (Bersini, 1992), for example, when it is defended that most of behavior is based on sensorimotor automatisms that do not require internal representational models. This case is usually argued defending either that knowledge depends on the real or detailed structure of the environment whose perception guides action without the need of forming internal structures (Brooks, 1991), or that representation is the result of the structure of the cognitive organism itself and information can be considered embodied in the internal constraints

of the organization of the subject, which can make sense out of certain perturbations coming from the outside (Varela *et al.*, 1991).

The first position reduces the problem of cognition to a mere reactivity towards the environment and, even if it can lead to an interesting engineering strategy that is biologically more realist than the one of previous AI, epistemologically it erases the problem of cognition, for there is no cognitive subject left. The second one underestimates the problem of cognition in a similar way, because the transformations undertaken by the subject in relation to the environment cannot be considered as knowledge of anything, as there is no environment to be known.

A study of cognitive processes grounded in the biological structures of organisms like the one we have proposed here makes it possible to re-settle the problem of cognition as a phenomenon of construction of a cognitive system in the interaction with its relevant environment, a process through which hierarchical representational structures are created with a functional value associated to the biological survival of the AO.

Arantza ETXEBERRIA

School of Cognitive and Computing Sciences
University of Sussex, UK

Juan Julian MERELO

Department of Electronics and Computer Technology
University of Granada, Espagne

Alvaro MORENO

Department of Logic and Philosophy of Science
University of the Basque Country

Acknowledgements

We want to thank Inman Harvey and Gertrudis van der Vijver for their comments and criticisms on an earlier draft of this paper. A DIGCYT Project Number PB92-0456 from the Ministry of Education made this research possible. Arantza Etxeberria and Alvaro Moreno also were supported by the Research Project Number 003.320-HA137/92 of the University of the Basque Country. Arantza Etxeberria

has a Postdoctoral Fellowship from the Department of Education, Universities and Research of the Basque Government to work at the University of Sussex (U.K.).

References

- ACKLEY D., LITTMAN M. (1992) Interactions between learning and evolution. *Artificial Life II*. C.G. Langton, C. Taylor, J.D. Farmer, S. Rasmussen (éds), Redwood City (CA): Addison-Wesley, pp. 487-509.
- BELEW R. K. (1991) Artificial Life. A constructive Lower Bound for Artificial Intelligence. *IEEE Expert*. February, pp. 8-15.
- BERSINI H. (1992) Animat's I. *Towards a Practice of Autonomous Systems*. F. Varela, P. Bourguine (éds.), Proc. of the first European Conference on Artificial Life, MIT Press/Bradford Books, pp. 456-474.
- BROOKS R. (1991) Intelligence without representation. *Artificial Intelligence*., 47, pp. 139-159.
- BROOKS R. (1992) Artificial Life and Real Robots. *Towards a Practice of Autonomous Systems*. F. Varela, P. Bourguine (éds.), Proc. of the first European Conference on Artificial Life, MIT Press/Bradford Books, pp. 3-10.
- CARIANI P. (1992) Emergence and Artificial Life. *Artificial Life II*. C.G. Langton, C. Taylor, J.D. Farmer, S. Rasmussen (éds), Redwood City (CA) : Addison-Wesley, pp. 775-797.
- CONRAD M. (1988) Quantum Mechanics and Molecular Computing: Mutual Implications. *Int. J. of Quantum Chemistry: Quantum Biology Symposium*, 15, pp. 287-301.
- EDELMAN G. M. (1987) *Neural Darwinism. The Theory of Neuronal Group Selection*. New York : Basic Books.
- ETXEBERRIA A. (1992) *El origen de la cognicion en los sistemas biologicos y el enfoque conexionista*. Thèse. Université du Pays Basque.
- FARMER J.D., KAUFFMAN S.A., PACKARD N.H. (1986) Autocatalytic Replication of Polymers. *Physica D*, 22, pp. 50-67.
- FERNANDEZ J. (1992) *Vida Artificial. Un estudio epistemológico*, Thèse. Université du Pays Basque.
- FERNANDEZ J., MORENO A. (1992) *Vida Artificial*, Madrid : Eudema.
- FLEISCHAKER G. (1988) *Autopoiesis: The system logic and origin of life*. Thèse, Boston University.
- GIBSON J.J. (1979) *The Ecological Approach to Visual Perception*. Hillsdale (NJ): Lawrence Erlbaum.
- GOLDBERG D.E., DEB K., KORB B. (1991) Messy Genetic Algorithms Revisited: Studies in Mixed Size and Scale. *Complex Systems*, 4, pp. 415-444.
- HARVEY I, HUSBANDS P., CLIFF D.(1992) Issues in Evolutionary Robotics. *From animats to animas!: Proceedings of the Second International Conference on Simulation of Adaptive Behavior*. J.A. Meyer, H. L. Roitblatt, J.W. Wilson (éds), Cambridge: MIT Press, pp. 364-373.

- HARVEY I (1992) Species Adaptation Genetic Algorithms: A basis for a Continuing SAGA. *Towards a Practice of Autonomous Systems*. F. Varela, P. Bourguine (éds.), Proc. of the first European Conference on Artificial Life, MIT Press/Bradford Books, pp. 346-354.
- KAMPIS G. (1991) *Self-modifying systems in Biology and Cognitive Science*. Oxford: Pergamon Press.
- KAUFFMAN S.A.(1986) Autocatalytic sets of proteins. *J. Theor. Biol.* 119, pp. 1-24.
- KOSHLAND D.; GOLDBETER A., STOCK, J.B. (1982) Amplification and Adaptation in Regulatory and Sensory Systems. *Science* 217, pp. 220-225.
- KOZA J.R. (1990) '*Genetic Programming*': A paradigm for genetically breeding population of computer programs to solve problems. Tech. Rep. Stan-CS-90-1314, Stanford University.
- KREMEN A. (1992) Biological molecular energy machines as measuring devices *J. Theor. Biol.* 154. pp.405-413.
- LANGTON C.G. Ed. (1989) *Artificial Life*. Redwood City (CA): Addison-Wesley.
- LANGTON C.G., TAYLOR C., FARMER D., RASSMUSEN, S. Eds. (1992) *Artificial Life II*. Redwood City (CA): Addison-Wesley.
- MAES, P. Ed. (1990) *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back*. MIT/Elsevier.
- MATURANA H., VARELA F. (1984) *El árbol del conocimiento*. Santiago du Chili : Editorial Universitaria.
- MERELO J.J., MORENO A., MORAN F. (1992) Cognition and perception in Artificial Life. Poster présenté à SAB-92, Hawaii.
- MERELO J.J., PATON M., COSANO J., CIGUERO M., BERMUDEZ A. MOLINA J.A., MORAN F. (1992) Combination of hebbian NN and GA in Artificial Life organisms. Poster présenté à la *III Artificial Life Conference*, Santa Fe (NM).
- MERELO J.J., MORENO A., ETXEBERRIA A. (1993) Artificial Organisms with adaptive sensors. In *Proceedings of the II European Conference on Artificial Life*, Bruxelles.
- MINCH E. (1992) Anima y Animus: una perspectiva sobre la vida artificial y la inteligencia artificial. *Revista de la Real Academia de Ciencias*, Tomo LXXXVI, 4.
- MORENO A. (1988) Paradigmas en Biología Teórica. *Gagavai* (III), pp. 85-98.
- MORENO A., FERNANDEZ J., ETXEBERRIA A. (1990) Cybernetics, Autopoiesis and the Definition of Life. *Cybernetics and Systems*. R. Trappl (éd.) 90, pp. 357-364.
- MORENO A. and ETXEBERRIA A. (1992) Self-reproduction and representation. The continuity between biological and cognitive phenomena. *Uroboros II*. (1), pp.131-151.

- PATTEE H. (1982) Cell Psychology: An evolutionary Approach to the Symbol-Matter Problem. *Cognition and Brain Theory*. 5 (4), pp. 325-341.
- PATTEE, H. (1989) Simulations, Realizations and Theories of Life in Langton (éd.) *Artificial Life*. Redwood City (CA): Addison-Wesley, pp. 63-77.
- ROSEN R. (1959) Some Realizations of (M,R)-Systems and their Interpretation. *Bull. Math. Biophysics*. 21, pp. 109-128.
- VAN DE VIJVER G. (1991) The emergence of meaning and the antinomy of naturalism. *Uroboros*. I (2), pp.153-175.
- VARELA F. (1979) *Principles of Biological Autonomy*, New York: Elsevier North Holland.

- VARELA F. (1981) Describing the logic of the living. The adequacy and limitations of the idea of Autopoiesis. In *Autopoiesis: a Theory of Living Organization*. M. Zeleny (éd.) New York: Elsevier North Holland, pp. 36-48.
- VARELA F., THOMPSON E., ROSCH E. (1991) *The embodied mind*. Cognitive Science and Human Experience, Cambridge (MA): MIT Press.
- VARELA F., BOURGINE, P. (éds.) (1992) *Towards a Practice of Autonomous Systems*. Proc. of the first European Conference on Artificial Life, MIT Press/Bradford Books.