Available online at www.sciencedirect.com

ScienceDirect

BioSystems 91 (2008) 309–319

www.elsevier.com/locate/biosystems

The autonomy of biological individuals and artificial models

Alvaro Moreno*, Arantza Etxeberria, Jon Umerez

Department of Logic and Philosophy of Science, University of the Basque Country
UPV-EHU Tolosa hirib.70//E-20018 Donostia, Spain

Received 7 March 2007; received in revised form 27 April 2007; accepted 11 May 2007

Abstract

This paper aims to offer an overview of the meaning of autonomy for biological individuals and artificial models rooted in a specific perspective that pays attention to the historical and structural aspects of its origins and evolution. Taking autopoiesis and the recursivity characteristic of its circular logic as a starting point, we depart from some of its consequences to claim that the theory of autonomy should also take into account historical and structural features. Autonomy should not be considered only in internal or constitutive terms, the largely neglected interactive aspects stemming from it should be equally addressed. Artificial models contribute to get a better understanding of the role of autonomy for life and the varieties of its organization and phenomenological diversity.

Keywords: Agency; Biological organization; Cognition; Function; Integration; Self-organization

1. Introduction

Autonomy means self-law, to be in charge. In general, this is understood either as the capacity to act according to self-determined principles, or as the duty of recognizing and respecting that aptitude about someone. In an ontological usage it may also mean that a given level or realm is relatively independent with respect to others because it is ruled by its own norms. Yet, according to Maturana and Varela's theory of *autopoiesis*, autonomous capacities stem from self-production and they constitute an identity (Maturana and Varela, 1973, 1980, 1984). Then, not merely autonomous action, but also autonomous being is the subject matter of this last approach.

The theory of *autopoiesis* placed the notion of autonomy at the center of the biological understanding of living beings in a moment (early seventies) when the atmosphere was probably more prepared for systemic developments than inmediately before or after (Wimsatt, 2007; Etxeberria & Umerez, 2006); (as it is well known, the biology of the second half of the 20th century revolved around the concept of the gene to explain most of living phenomenology, including development or evolution). Although for a long time the impact of this notion in main stream biology was not major, at present it has started to draw a lot more attention, especially among those interested in a biology centered in the organism. Examples of this are this early century's return of Systems Biology (Boogerd et al., 2007; Kitano, 2002; O'Malley and Dupré, 2005; Science, 2002), the renewed interest for the Kantian third Critique's view of organisms in the philosophy of biology (Van de Vijver et al., 2003), and the biologically inspired Artificial Life and Robotics (Webb, 2001; Steels and Brooks, 1995; Beer, 1997; Nolfi and Floreano, 2000; Di Paolo, 2003). In all three (biology, philosophy and artificial science), the exploration of autonomy is likely to bring about

* Corresponding author.
E-mail address: alvaro.moreno@ehu.es (A. Moreno).

0303-2647/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2007.05.009

new perspectives for research (Etxeberria et al., 2000). If Modernity[1] thought that autonomy was a desired consequence of the human faculty of reason, contemporary naturalism aims to understand life and cognition as expressions of the autonomy of some material systems. The focus is shifted from the actions deriving from autonomy to the processes able to originate that capacity, to its conditions of possibility.

We want to stress here the significance of autonomy for sciences aiming to understand living systems. In this sense, exploring autonomy should constitute an inevitable axis in the models, artificial systems, etc. involving phenomena such as organization, development, behavior or evolution. In our view, autonomy is the main feature of life, the key notion for any attempt to define it.

Such tight a relation between life and autonomy suggests a complicated status for artificial autonomous systems. On the one hand, if an artificial autonomous system would be produced, then we might call it "alive". The reason is that life may be understood in certain ways as a category beyond the natural/artificial divide. Natural systems generate spontaneously, whereas artificial systems are created by an intelligent agent (generally human)[2] through the use of complex cognitive capacities and resources. But, according to Hans Jonas (1966/2001)), life has some form of primacy for the living, in the sense that it is recognized without the need to analyze it into constituent parts. Although none of the artificial systems so far produced are alive, if a truly autonomous system would come into existence by means of human intervention, then its reality as autonomous or alive would be more significant to us than the fact that it is artificial; its artificiality would be somehow secondary with respect to its aliveness. On the other hand, these systems would be also extensions of ourselves,[3] effects our agency has on the environment, even if their autonomy might prevent us from consider them as mere products (because they would be self-produced). It is doubtless that the exploration of autonomy poses a special kind of challenge for human agency, as it requires the reflexive form of epistemology searched by cyberneticians (Hayles, 1999).

However, we do not contend that (re)producing autonomy should constitute a technological goal for research on life. Concerning our understanding of life, what really matters is the construction of models of autonomy (which, if material, are often called "artificial autonomous systems"). But modelling autonomy is a complex task, a real challenge for an experimental epistemology. Not being exhausted by the ontological production of autonomous system, it may be appealing as an activity in which the purposes and intervention of modellers interact with a dynamic and emergent system, with the result of understanding.

In that context, the goal of this paper is to offer an overview of the meaning of autonomy for biological individuals and artificial models rooted in a specific perspective that pays attention to the historical and structural aspects of its origins and evolution. In the next section we will discuss the question of minimal autonomy, namely, what characterizes autonomy from less complex forms of self-organization and self-maintenance. In Section 3 we analyze the specific features of different levels, degrees and domains of autonomy. Finally, in Section 4 we consider the relation between autonomy and the artificial, mainly focusing on methodological and epistemological aspects.

2. Autopoiesis and minimal autonomy

Since the relation between the concepts of autonomy and life is so tight, we may wonder if, in the transition between the non-living and the living there are (were) non-living organizations that can be considered autonomous to some degree. Is the simplest biological form we know also the simplest form of autonomy? Are there reasons to think that non-organic or pre-biotic forms of organization, simpler than known life, could be already autonomous? In other words, this is a question about self-organization, minimal autonomy and the relation between the two of them.

[1] The concept of autonomy has been used in many domains (political philosophy, ethics, biology, robotics...), and it is not evident whether there are clearly traceable genealogical relations among the various usages. In ancient Greece the term was applied to city-states and referred to the political right of self-government. Later, modern philosophy extended the term to the self-determination of persons, both political and ethical. In the sense we claim for in philosophy of biology, artificial life and organismic robotics, the concept plays an important role in the definition of the identity and the interactive capacities of individuals.

[2] It is possible to consider that artefacts or substances produced by animals are equally artificial, and this view has the advantage of situating human cognitive and technical abilities within an evolutionary continuum. It has the problem of blurring some intuitive differences, as both plastic and honey would appear to be equally artificial from this perspective, but we think that the distinction may be recovered if it were needed to argue in a given context (for example, by distinguishing the technological abilities of humans and other animals).

[3] Keller (2007) proposes this understanding. She is inspired by Turner's (2000) work on the way animals use their environmental constructions as extensions of their bodies.

Yet, it is essential to keep in mind that autopoiesis is an a-historical concept: it articulates an organization emerging from the dynamics of components, not from evolution.[4] As formulated by Maturana and Varela, evolution will give rise to a manifold of structures deriving from autopoiesis, but the organization remains the same. Thus, only an evolutionary account implying some increase of complexity in autonomy may justify referring to minimal forms.

The concept of autopoiesis refers to a recursive net of component production that builds up its own physical border. The global net of component relations establishes a self-maintaining dynamics, whose action brings about the constitution of the system as an operational unit. In autopoiesis, components and processes are entangled in a cyclic, recursive production logic.

Although the stress is made in the circular logic of the system, and primarily it is an "internalist" perspective, in other places these authors also consider the relation of the system with its environment. For example, they provide the following definition in their glossary:

> "Autonomy: the condition of subordinating all changes to the maintenance of the organization. Self-asserting capacity of living systems to maintain their identity through the active compensation of deformations". (M&V 1980, p. 135).

This definition contains, at least, two aspects worth stressing in our attempt to establish the lower limits at which autonomy may appear. One is that there has to be an organization toward whose maintenance all changes must be subordinated. The other is that maintenance is active and self-asserting. Thus, phenomena like tornadoes, whirlpools and candle flames, in which self-organizing properties appear to a certain degree, are not autonomous. Indeed (some of) these systems may have "homeostatic" capacities. For instance, a candle flame can compensate a small puff of air by increase in temperature back to steady state and re-establishment of vertical air flow from the flame. But we cannot detect any form of active interaction or "agency" exerted by the system. In that sense, what distinguishes simple forms of self-organization and self-maintenance from autonomy is, as we will explain later in more detail, that the former merely react against external perturbations, not being capable of displaying selective actions.

In a different view of autonomy, that of Pattee (i.e., 1972, 1973, 2001), self-organizing systems like tornadoes are not entitled to be considered autonomous either, but for different reasons. His argument would be that an active compensation requires that the system is capable of some form of selection, as basic as it might be, between different alternatives. If no choice is available, no action can be exerted: the dynamics just results in the advent of the next state of a temporal sequence determined by laws. Is a planet, exerting its gravitational force upon a given object, "active"? What about the water flow of a river eroding the surface of its bed? Or, to come closer to our case, when a tree is bent by the wind and bounces back to its resting position, is it doing an "active compensation" against the "deformation" induced by the wind? According to Pattee, the three of them are examples of systems governed by laws. However, the autonomy of the system depends on the existence of internal constraints able to channel its process dynamics in one of the possible directions. This amounts to some form of additional causality with respect to law-governed dynamics, exerted by the system itself.

Thus, the existence of this type of causal constraints in the system may demarcate the difference between self-organizing systems, as generally understood, and what people like Simon (1996) called "organized complexity", using Weaver's (1948) term. Usually, self-organizing systems are studied as a "one-shot, order-for-free" kind of process whose outcome is the emergence of a global pattern starting from a uniform non-linear dynamics below. This pattern is not able to produce any internal selection within the system. However, organized complexity involves some distinction among parts and their characteristic functions within the system. It often results from processes of composition or self-assembly between heterogeneous subsystems, with the result of some form of hierarchy (as the merging of autonomous identities in the origins of eukaryotic cells). Far from being an immediate process, this kind of organization can only result from iterative processes of self-organization over time (Keller, 2007).

The aim to make some progress concerning this essential distinction is probably on the basis of Kauffman's (2000) approach to autonomy, expressed in thermodynamic terms. Starting from the concept of "autocatalytic set", the main condition required to consider a system as autonomous is that it is "able to perform at least one thermodynamic work cycle" (Kauffman (2000), p. 4). This capacity is implemented through a deep entanglement between work and constraints: "work begets constraints begets work", as some form of closure in the abstract space of catalytic tasks. This insight is based, on the

[4] In this sense, Di Paolo (2005), notes that an autopoietic network of processes reversed in time is still autopoietic. See also (Etxeberria, 2004).

one hand, in Atkins' view (1984) of work as a coordinated, coherent and constrained release of energy, and, on the other hand, in the recognition that work is absolutely necessary to build constraints. In other words, to be autonomous, a system must do work; to be capable of work it requires constraints to channel the flow of energy in an appropriate way, and to build those constraints the system requires appropriately constrained energy flows, that is to say, work. This is the circularity of the "work-constraint cycle".

In our view, Kauffman's account envisages how functionality or utility (implicit in the idea of work) can come out of the causal circularity of the system, where this circularity is not only understood in terms of abstract relations of component production, but also as an energetic logic sustaining the specific chemical recursivity of the system: all the processes ongoing in the system are constrained in order to satisfy the condition of self-maintenance. This is how explanations in terms of functions can and should be introduced on the study of autonomous systems. However, unlike abstract computational functionalism and evolutionary etiological accounts of function, a utilitarian notion of autonomy as differential contribution to self-maintenance (Collier, 1998; Christensen and Bickhard, 2002) uncovers the deep entanglement between structures, functions and stability that is characteristic of autonomous systems.

Yet, in his work, like in autopoiesis, the notion of recursivity remains the conceptual nucleus of autonomy as identity preserving. The system is constituted as a series of causal processes (energy transduction, component production, etc.) converging to a given (initial) state, as an indefinite repetition of the same loop. Thus, although these models admit variations, their essential aspect is the existence of a pattern, which is maintained throughout the self-producing cycles, and constitutes the "identity" of the system.

Thus, from this perspective in order to preserve the system's identity, its actions must counteract possible perturbations from the environment, and further actions of autonomous systems must always have that internal reference only. As a consequence, the actions of an autonomous system in the environment are only side effects of the internal self-production. Although it is sound to consider that the reference of autonomous agency is internal, that is to say, that the self-maintenance will define what is relevant, a broad consideration of agency suggests that the transformation of the environment should be more than a passive consequence of self-production: autonomous agents simultaneously preserve the self-produced identity and transform the environment, not only to make their more suitable to its needs, but also to use it as a control parameter for internal processes or even as a tool for some of their required processes. Then the identity itself should be a process open to "becoming", and not a fixed point.

Then, the question is: what is the role of action in the environment within a theory of autonomy? A broad agential autonomy implies that the organization of the system causes the processes exerted on the environment, whereas those of the environment towards the system are monitored according to internally defined needs. To justify this we would like to distinguish between *constitutive* processes, which produce the identity and largely delimit what the system is, from *interactive* processes, which are not only side effects of the former, but crucial to maintain the identity of the system, with the specific function of controlling the interaction with the environment. We may picture those two, constitutive and interactive, aspects of autonomy as acting in different temporal scales, being the first faster and more fundamental than the second, although both are equally required. It might be clarifying to think of the example of the active transport of ions across the membrane, required to prevent osmotic crisis. This interaction (a form of "work", as it carries ions against gradient) requires an internal sub-organization of different chained reactions. The cell can be maintained due to ion transport interaction, but this can be realized because there is a pre-existent internal system.

Therefore, we are proposing that autonomy requires more than self-organization or self-maintenance. In a similar way, Bickhard (2004) noticed the need of an "infrastructure" in the system, understood as an internal mechanism able to organize and channel energy flows for the system's self-maintenance. This subsystem should be some form of modulation of the self-maintaining processes themselves.

A consequence of this is that autonomy requires functional organization. As it is known, the Kantian approach to organisms describes them as natural purposes and, in his view, this intrinsic teleology makes it difficult to explain them in mechanistic terms. In opposition to this, Maturana and Varela proposed to explain organisms through a circular organization very similar to the Kantian, but they consider organisms as "autopoietic machines", and explicitly aim to avoid teleology.[5] Now, after some extensive exploration of self-organizing properties in the last decades, it is clear that no natural notion of function emerges out of it. But it is also evident that

[5] However, at the end of his life, Varela (Varela and Weber, 2001) adopted a different position, closer to the Kantian approach.

function is required by even the simplest satisfactory theory of life or autonomy.

3. Levels, degrees and domains of autonomy

A key conceptual aspect of autonomy is the distinction of its different kinds and degrees, but it has been largely neglected. Maturana and Varela explained living and cognitive phenomenology starting from the autopoietic organization, and even if they acknowledged that the structures changed in history, they considered that the organization itself was not affected by history. Nevertheless, we think that an adequate theory of autonomy must try to address degrees of autonomy; i.e. the aspect of becoming (more) autonomous.

One of the difficulties for a comparative study of the domains in which autonomy appears (with different degrees) is that all are characterized according to the same type of circular organization. For example, at the biological level the autopoietic organization is:

> "a network of processes of production (transformation and destruction) of components which: (i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in space in which they (the components) exist by specifying the topological domain of its realization as such a network."(Maturana and Varela, 1973, p. 78).

And cognitive autonomy is closure of the dynamics of the nervous system:

> "two main specifying characteristics of cognitive self are given by the nervous system's operational closure which: a) produces invariant sensory-motor patterns b) and specifies the organism as a mobile unit in space" (Bourgine and Varela, 1991, p.).

With this strategy it is difficult to connect these two different domains of autonomy in terms of the complexity of the structures required to present the phenomenological properties associated to each of them. The first one takes into account the production of (chemical) components, and the second of the sensorimotor (or neural) patterns. From these definitions no difference can be drawn related to the structural properties of each one, only the similar circular logic of each system is enhanced. This suggests that the differences between these two kinds of systems are not relevant in what concerns autonomy, but we think they might be crucial.

If the autonomous system is an agent whose identity depends on constitutive and interactive processes, structural changes leading to a greater independence from the environment or to greater control over how it influences the agent become a clue to identify increases in autonomy. This is the idea behind some attempted characterizations. For example, in Cariani's view (1998), a system acts more autonomously if it depends more on internal processes than on external inputs. In a similar way, Boden (1996) considers that:

> "an individual's autonomy is the greater, the more its behavior is directed by self-generated (and idiosyncratic) inner mechanisms, nicely responsive to the specific problem-situation, yet reflexively modifiable by wider concerns." (p. 102).

Boden's point of view here is that the degree of autonomy is linked to the system's capacity for self-modification of behavior producing mechanisms[6] (see also Boden, this issue).

The problem with these approaches lies on the difficulty to determine what higher degrees of self-modification are. It is plain that many unicellular organisms already show high capacities for self-modification, even higher than those of complex multicellulars in terms of self-repair, morphological self-modification and alike. However, autonomy is a capacity that may increase in history, in ontogenetic and phylogenetic terms. This becoming might be seen as an access to a wider interactive functional universe,[7] a higher degree of autonomy appears to be related with the creation of more complex, hierarchically organized, functional constraints on the environment. In evolutionary terms the process of becoming more autonomous sometimes consists in the integration of already existing autonomous systems via appropriate controls.[8] This process may involve that what previously were interactive processes among autonomous systems result in internal or constitutive processes of a more encompassing system. In the case of the origins of eukaryotic cells, the new internal relations involve the loss of the autonomy of the incorporated system. In multicellular organization, the new constituent parts, previously independent, do not

[6] In particular, she claims that representational re-description mechanisms are necessary to produce deliberate self-control.

[7] By a wider functional universe we mean an increase in the number and variety of the interactive processes for ensuring the maintenance of the system, along with an increase in the number of hierarchical regulatory controls.

[8] Complex interactions among autonomous systems can also be seen as trade-offs of mutual benefits (Christensen and Bickhard, 2002). When collective benefits are limited, the autonomy of the more encompassing system is weak, whereas parts retain significant autonomy. See also Buss (1987).

loose metabolic autonomy, but their interactive processes become heavily constrained. In what respects colonies and societies the cohesion of the collective has different grounds in each of them. The first depends mainly on a self-organizing dynamics producing an encompassing agential capacity that raises the adaptability that isolated parts would have; occasionally they may almost appear as an integrated multicellular organism.[9] But colony autonomy does not become a lot more complex. In the case of animal societies, individual organisms remain autonomous to a large extent. This is probably linked to the fact that societies are constituted by organisms endowed with cognitive systems, which show more intricate forms of agency than the integrants of colonies. Societies are organized according to complex hierarchies of regulatory controls that provide suitable environments for complex forms of agency, like cognitive communication. However, although in certain cases it may appear that the society as a whole behaves as an agent, the identity causally grounding the actions in these kinds of systems is by far more complex at the level of the individuals than at the level of the society.

The case of multicellular organisms is quite different. Here new and more complex forms of agency appear in the evolution of animals, grounded on the stabilization of self-regulated functions through mechanisms of internalization able to protect them against, environmental perturbations. The way for evolution to achieve highly integrated multicellular systems seems to require something more than forms of self-organization among an increasing number of constitutive systems, like in colonies and certain societies. Multicellular organisms, specially in the case of animals, require the creation of complex regulatory mechanisms in order to generate an integrated functional unit from the relations among the constitutive cells. All these processes contribute to an increase of autonomy, at least they generate phenomenologies that only metaphorically adjust to the autopoietic model of autonomy: while they are still cases of self-production, a lot more than that can be said about them. In evolution other processes of autonomization[10] have been proposed to illustrate some form of progress towards more independence from the environment (Rosslenbroich, 2005, 2006). For instance, the extracellular matrix appearing common in the development of metazoan, which appeared early in evolution, allows intracellular conditions to regulate and protect internal cells from the external environment. This independence is relative because it is built at the same time that “many interconnections with and dependencies upon” the environment are retained and perhaps elaborated. Some of the elements are: spatial separation from the environment (membrane, walls, intergumens etc.), establishment of homeostatic functions, internalization of morphological structures or function from an external position.

The creation of regulatory mechanisms requires a hierarchical organization, in which the degree of autonomy of the constituents is restrained. As we have pointed out, the way for evolution to achieve highly complex biological systems does not seem to be to increase the number and variety of the constituent parts organized in a distributed manner, but to generate new, nested forms of regulatory control.[11] For example, the increasing process of autonomization in the evolution of vertebrates goes together with the fact that their metabolic organization is fully and precisely controlled by their brain. Their characteristic agency has been made possible by the development of the nervous system, which evolved as a powerful regulatory mechanism to control and integrate complex underlying processes. There is a strong association between the evolution of highly integrated and complex bodies and the evolution of cognitive autonomy (Moreno and Lasa, 2003; Moreno and Etxeberria,

[9] This is the case, for example, of the so-called “magnetotactic multicellular prokaryote”, a bacterial aggregate that exhibits an unusual “ping-pong” motility in magnetic fields (Keim et al., 2004).

[10] “Increasing autonomy is defined as the evolutionary shift in the evolutionary system/environment relationship, so that the direct influences of the environment are gradually reduced and stability and flexibility of self-referential, intrinsic functions within the system is generated” Rosslenbroich, 2006, p. 61. The term autonomization was used by Schmalhausen (1949).

[11] Self-organization alone cannot explain this process. An explanation for it can go along the following lines: “(. . .) whilst ‘self-organization’ is celebrated for its capacity to generate global patterns it has significant limitations as a means of resolving the problems presented by integration pressure. The most important of these are slow action and poor targeting capacity. Precisely because achieving the global state depends on propagating state changes through many local interactions the time taken to achieve the final state can be long, and increases with the size of the system. Moreover, since there is no regulation of global state, the ability of the system to find the appropriate collective pattern depends on the fidelity of these interactions. Here there is a tension: if the self-organization process is robust against variations in specific conditions the process will be reliable, but it will be difficult for the system to generate multiple finely differentiated global states. Alternatively, if the dynamics are sensitive to specific conditions it will be easy for the system to generate multiple finely differentiated global states, but difficult to reliably reach a specific state. (. . .) Consequently the most effective means for achieving the type of global coherence required for functional complexity is through regulation, including feedback mechanisms and instructive signals operating at both local and larger scales”. (Christensen, 2007).

2005). In other words, systems with higher degrees of autonomy show an increase in the capacity to create and/or take over complex and larger environmental interactions, because of a more intricate organization of their constitutive identity. Their autonomy is also based on a circular, recursive organization, but this also includes many hierarchical levels and many embedded regulatory controls.

4. Autonomy and the artificial

As already said, autonomy is a challenge for artificial systems: in the ontological sense, because the autonomy of artificial systems is limited; in the epistemological, the question is whether we can learn about autonomy by constructing artificial models, and what kind of models are required.

The challenge is the creation of systems that are, at the same time, artificial and not fully designed. Though the goal to create a true autonomous system should be differentiated from the epistemological aim of constructing artificial systems as models, these two objectives have been often linked in the sciences of the artificial. In the field of Artificial life, for example, the study of living organisms was faced "by attempting to synthesize life-like behaviors within computers and other artificial media" (Langton, 1989, p. 1). Here the very concept of autonomy becomes both the end and the way; if a system is made by strict design, it can hardly be autonomous: either its actions would not reconfigure its identity, or, if they did, the system would not be anymore the result of direct design.

As autonomy implies that the identity of the system is self-created, a hypothetical artificial production of an autonomous system would require an indirect form of human action, so that by partially "get(ting) the human being out of the loop" as Langton described (Boden, 1996; Risan, 1997), some of the human creativity turns out to be externalized to the dynamical behavior of the artifact itself.

The originality of this project stems from a bottom-up, "emergent" methodology and the mimesis of some of the natural ways to generate complex processes (i.e., evolution, development or learning). In computer simulations, the result does not have to be – at least in its specific form – analytically inferred and it must be more complex[12] than what has generated it, in the sense that the dynamic unfolding of the simulation is far richer than the simple aggregation of the local rules that generate it. The design of the model should be simple and similar to the conditions by which natural processes give rise to something more complex. With this aim, a set of techniques, known under the common terms of "emergent computation" (Forrest, 1991), "artificial evolution" (Harvey et al., 2005), etc. have been developed to bring into existence (along with progresses in hardware) an explosion of computational "lifelike creatures".

Two broad kinds of techniques have been developed to study living phenomenology using artificial systems. One follows a part/whole strategy aiming to replicate the complexity of the system taking into account the relation of constituents with the totalities they form. In general, the method used is bottom-up and tries to study the self-organizing properties of the system. Examples of this kind of research line are the formation of chemical self-maintaining networks, protocells, structures in neural networks, ant colonies, the organization of the immune system or of robot societies. The main conceptual and technical problem of this approach is how to capture the circular causality that characterizes autonomy, which should be both top-down and bottom-up, so that the macroscopic behavior emerges from microscopic local rules of simple parts, whereas the macro emergent effects feed back to modify and control the bottom level. There are already good models that capture some of the circularity and self-organizing properties characteristic of autonomous systems (Varela et al., 1974; McMullin and Varela, 1997; Fontana, 1992). The challenge now lies on developing models in which a richer repertoire of functional behavior is integrated into the model and emergent from its local interactions, together with the exploration of spontaneous (i.e., not externally imposed) formation of a hierarchical regulatory organization.

The other strategy is artificial evolution, which tries to substitute the human design of artificial agents by an evolutionary process. In what respects the evolution of autonomous systems, most of the first models of artificial evolution relied on a relatively adaptationist version of Darwinian evolution, partly obliged by the use of fitness functions to guide the search process, and mainly because the evolving structures were represented as idealized genotypes, instead of being self-reproducing autopoietic systems (Griesemer, 2000; Etxeberria, 2000, 2004). The role attributed to artificial evolution regarding the modeling of autonomous systems has been generally limited to optimize the system towards a unique desired functionality (which was, in addition, generally externally defined). As a result, the system was generally very limited on its behav-

[12] In the sense of less trivial, because it can be – and usually it is – a "simplified" pattern with respect of the low level interactions that led to it.

ioral repertoire and, most importantly, its behavior was uncoupled to internal (self-maintaining or stability) conditions. Recent attempts to overcome these limitations (in particular those focused on selection for internal stability coupled to adaptive behavior (Di Paolo, 2003, this issue) might be able to generate more complex forms of autonomous organization, making room for functional diversity and the necessary hierarchical regulatory organization required to manage it. It is within this context where the previously raised questions related to the interplay between levels, degrees and domains of autonomy could be addressed. But evolution itself, that is to say, the problem of the role of autonomy in evolutionary processes, is still largely neglected, and the proposal of "natural drift" as the main evolutionary phenomenon (Maturana and Mpodozis, 2000) has not been pursued further in artificial systems.

Until very recently, most of the models more directly aiming to reproduce the constitutive organization of autonomy ignored or neglected thermodynamic and/or energetic constraints. However, since autonomous systems cannot be but far from equilibrium dissipative organizations (where the flow of matter and energy across the system makes interactive and constitutive processes inherently interdependent), the abstraction of these aspects in the design of models hides the strong interconnection between the interactive and the constitutive dimensions of real autonomous systems. This limitation is currently being overcome in some recent work on more realistically simulated chemical networks that could lead to minimal autonomous systems (Daley et al., 2002; Kauffman, 2003; Olasagasti et al., 2007; Mavelli and Ruiz-Mirazo, 2007), although still in a very preliminary stage of research.

Another, complementary, way to study minimal autonomy is "synthetic biology". Synthetic biology can be considered as a redefinition and expansion of biotechnology, with the ultimate goals of being able to build engineered biological system (Benner and Sismour, 2005). Thus, it goes further than classical genetic engineering and seeks the complete fabrication of living beings starting from the same (or very similar) natural materials. On these lines there are currently different research programs under development that involve the fabrication of some sort of minimal artificial organism or protocell (Solé et al., 2007). On the one hand, following a 'top-down' strategy, different researchers are trying to find out the simplest form of a living cell by modifying extant unicellular organisms with genetic engineering techniques. The aim is to find an artificial cell with the minimal genome able to sustain their most basic vital functions (Hutchinson et al., 1999; Cho et al., 1999; Luisi et al., 2002). On the other hand, from a 'bottom-up' approach, artificial life *in vitro* (Szostak et al., 2001; Pohorille and Deamer, 2002) aims to synthesize minimal systems with capacity for self-maintenance and/or reproduction, starting from the most basic molecular components.

Finally, we have to mention the important research line of biologically inspired robotics (Steels and Brooks, 1995; Beer, 1997; Webb, 2001). This research is mainly focused on self-organized sensorimotor interactions, often neglecting the relationship between interactive and constitutive aspects of autonomy by focusing exclusively on the former. However in the complex interactive autonomy the guiding motives for behavior cannot be reduced to self-maintenance or survival, and that is why some researchers have proposed that the sensorimotor domain itself may generate autonomous forms of constitutive organization (Di Paolo, 2003; Barandiaran and Moreno, 2006). Particularly important for research in this direction is how the notions of value, intentionality and emotions relate to the concept of autonomy and its dynamic organization (Di Paolo and Iizuke, this issue; Ziemke, this issue).

To sum up, current approaches to artificial modeling are largely biased by a partial view of autonomy. As we have seen in Section 2, for a system to be autonomous, in its minimal sense, it has to fulfill certain requirements. And in Section 3, we have developed this concept, showing that it allows to deal with the varieties and degrees that biological evolution has unfolded. These requirements are more demanding than many of the standard views on autonomy developed by the theory of autopoiesis or by current biologically inspired robotics. On the one hand, most of the popular research methodologies for modelling autonomy (i.e., evolutionary robotics, dynamical systems approach, etc.) are almost exclusively focused on the study of the interactive dimension of autonomy, neglecting the constitutive one. On the other hand, the – relatively few – models dealing with the constitutive dimension tend to ignore the interactive aspect. This is quite understandable, since it is easier to study limited aspects of autonomy than its whole complexity. However, we consider that further advances will require the development of new models (some of which are already on their way) that take into account both the constitutive and the interactive dimension of autonomy together, as some of the key features of autonomy cannot be understood but as arising from the intimate coupling between the two mentioned dimensions. Ruiz-Mirazo and Mavelli (this issue) and Ikegami and Suzuki (this issue) constitute preliminary attempts to achieve this interconnection between constitutive and interactive aspects of autonomy in the metabolic domain, while Ziemke (this issue) and Di Paolo and Iizuke (this issue) show complementary approaches to deal with interactive and constitutive aspects at the sensorimotor level.

5. Conclusions

We have taken into account that the theory of autopoiesis started a research line on autonomy that extends classical philosophical approaches in that autonomous organization is not only behind the capacity to behave autonomously, but it also defines the system ontologically. This ontological account assimilates autonomy to the kind of systems that are alive. However, our approach to autonomy deviates from autopoiesis in that we try to discuss and contribute towards a material (and not merely organizational) and historical concept. In this sense, we have tried to explain why we consider that autonomy should not be considered only in internal constitutive terms, but that both the interactive and constitutive aspects stemming from autonomy should be equally taken into account. The main consequences of our focus are the need to explain function and hierarchical integration of parts within the theory of autonomy, starting from self-organization but in such a way that the explanation of autonomy is not exhausted by it.

We think that this approach makes it possible to develop better suggestions for work on artificial models. By attempting to (re)produce autonomous systems, we can at the same time learn how and why they are the way they are. As the fields of Cybernetics before and both Artificial Intelligence and Artificial Life later have showed, progress in our understanding about the property of autonomy in general and the autonomy of living and cognitive systems in particular requires an intensive use of computational simulations, robotic models and even synthetic biochemistry and biotechnology. However, the main goal of this activity should not be to produce artificial autonomous systems, but to get a better understanding of the role of autonomy for life and the varieties of its organization and phenomenological diversity.

Acknowledgements

Funding for this work was provided by grants 9/UPV00003.230-15840/2004 from the University of the Basque Country-Euskal Herriko Unibertsitatea, and HUM2005-02449 from the Ministry of Science and Education (MEC) and Feder funds from the E.C. We thank the valuable suggestions of two anonymous referees and of the editors of this issue (X. Barandiaran and K. Ruiz-Mirazo) to a previous version.

References

Atkins, P.W., 1984. The Second Law. Freeman, New York.

Barandiaran, X., Moreno, A., 2006. On what makes certain dynamical systems cognitive. a minimally cognitive organization program. J. Adapt. Behavior 14 (2), 171–185.

Beer, R., 1997. The dynamics of adaptive behavior: A research program. Robot. Auton. Syst. 20 (2–4), 257–289.

Benner, S.A., Sismour, A.M., 2005. Synthetic biology. Nat. Rev. Genet. 6, 533–543.

Bickhard, M.H., 2004. The dynamic emergence of representation. In: Clapin, H., Staines, P., Slezak, P. (Eds.), Representation in Mind: New Approaches to Mental Representation. Elsevier, Amsterdam, pp. 71–90.

Boden, M., 1996. Autonomy and artificiality. In: Boden, M. (Ed.), The Philosophy of Artificial Life. Oxford University Press, Oxford, pp. 95–108.

Boden, M., this issue. Autonomy: What is it?

Boogerd, F., Bruggeman, F., Hofmeyr, J., Westerhof, H. (Eds.), 2007. Systems Biology: Philosophical Foundations. Elsevier, Amsterdam.

Bourgine, P., Varela, F.J., 1991. Towards a practice of autonomous systems. In: Bourgine P., Varela, F.J. (Eds.), Towards a Practice of Autonomous Systems. Proceedings of the First European Conference on Artificial Life. MIT Press, Cambridge, MA. pp. xi–xvii.

Buss, L.W., 1987. The Evolution of Individuality. Princeton University Press, Princeton, NJ.

Cariani, P., 1998. Epistemic autonomy through adaptive sensing. In: Proceedings of the 1998 IEEE ISIC/CRA/ISAS Joint Conference, Gaithersburg, MD, September 14–16, 1998, pp. 718–723.

Cho, M.K., Magnus, D., Caplan, A.L., McGee, D., The Ethics of Genomics Group, 1999. Ethical considerations in synthesizing a minimal genome. Science 286 (5447), 2087–2090.

Christensen, W., 2007. The evolutionary origins of volition. In: Spurrett, D., Kincaid, H., Ross, D., Stephens, L. (Eds.), Distributed Cognition and the Will: Individual Volition and Social Context. MIT Press, Cambridge MA.

Christensen, W., Bickhard, M., 2002. The process dynamics of normative function. Monist 85 (1), 3–28.

Daley, A.J., Girvin, A., Kauffman, S.A., Wills, P.R., Yamins, D., 2002. Simulation of chemical autonomous agents. Z. Phys. Chem. 216, 41–49.

Di Paolo, E.A., 2003. Organismically-inspired robotics: homeostatic adaptation and natural teleology beyond the closed sensorimotor loop. In: Murase, K., Asakura, T. (Eds.), Dynamical Systems Approach to Embodiment and Sociality. Advanced Knowledge International, Adelaide, Australia, pp. 19–42.

Di Paolo, E.A., 2005. Autopoiesis, adaptivity, teleology, agency. Phenomenology and the Cognitive Sciences 4 (4), 429–452.

Di Paolo, E.A. this issue. How (not) to model autonomous behaviour.

Etxeberria, A., 2000. Artificial evolution: creativity and the possible. In: Bedau, M., McCaskill, J., Norman Packard, N., Rasmussen, S. (Eds.), Artificial Life V. MIT Press, Cambridge, MA, pp. 555–562.

Etxeberria, A., 2004. Autopoiesis and natural drift: genetic information, reproduction, and evolution revisited. Artif. Life 10 (3), 347–360.

Etxeberria, A. & Umerez, J. (2006). Organización y organismo en la Biología Teórica ¿Vuelta al organicismo? Ludus Vitalis 26.

Etxeberria, A., Moreno, A., Umerez, J. (Eds.), 2000. Special issue on the contribution of artificial life and the sciences of complexity to the understanding of autonomous systems. CCAI: Commun. Cogn. Artif. Intell. 17 (3/4), 111–254.

Fontana, W., 1992. Algorithmic chemistry. In: Langton, C.G., Taylor, C., Farmer, J.D., Rasmussen, S. (Eds.), Artificial Life II. Addison-Wesley, Redwood City, MA, pp. 159–209.

Forrest, S. (Ed.), 1991. Self-Organizing, Collective, and Cooperative Phenomena in Natural and Artificial Computing Networks. MIT Press, Cambridge, MA.

Griesemer, J., 2000. Reproduction and the reduction of genetics. In: Beurton, P., Falk, R., Rheinberger, H.-J. (Eds.), The Concept of the Gene in Development and Evolution. Historical and epistemological perspectives. Cambridge University Press, Cambridge, pp. 240–285.

Harvey, I., Di Paolo, E.A., Tuci, E., Wood, R., Quinn, M., 2005. Evolutionary robotics: a new scientific tool for studying cognition. Artif. Life 11, 79–98.

Hayles, N.K., 1999. How we became posthuman: virtual bodies in cybernetics. In: Literature and Informatics. The University of Chicago Press, Chicago.

Hutchinson III, C.A., Peterson, S.N., Gill, S.R., Cline, R.T., White, O., Fraser, C.M., Smith, H.O., Venter, J.C., 1999. Global transposon mutagenesis and a minimal mycoplasma genome. Science 286 (5447), 2165–2169.

Ikegami, T., Suzuki, K., this issue. From homeostatic to homeodynamic self.

Jonas, H., 2001. The phenomenon of life. Toward a philosophical biology. Northwestern University Press, Chicago, ILL.

Kauffman, S.A., 2003. Molecular autonomous agents. Phil. Trans. R. Soc. London A 361, 1089–1099.

Keim, C.N., Martins, J.L., Abreu, F., Rosado, A.S., Lins de Barros, H., Borojevic, R., Lins, U., Farina, M., 2004. Multicellular life cycle of magnetotactic prokaryotes. FEMS Microbiol. Lett. 240 (2), 203–208.

Kauffman, S., 2000. Investigations. Oxford University Press, Oxford.

Keller, E.F., 2007. The disappearance of function from 'self-organizing systems'. In: Boogerd, F., Bruggeman, F., Hofmeyr, J.-H., Westerhoff, H.V. (Eds.), Systems Biology. Philosophical Foundations. Elsevier, Amsterdam.

Kitano, H., 2002. Computational Systems Biology. Nature 420, 206–210.

Langton, C.G., 1989. Artificial life. In: Langton, C. (Ed.), Artificial Life I. Addison Wesley, Redwood City, CA, pp. 1–47.

Luisi, P.L., Oberholzer, T., Lazcano, A., 2002. The notion of a DNA minimal cell: A general discourse and some guidelines for an experimental approach. Helvetica Chim. Acta 85, 1759–1777.

Maturana, H.R., Varela, F.J., 1973, De máquinas y Seres Vivos. Autopoiesis: La organización de lo Vivo. Editorial Universitaria. Santiago (1994, 3rd edition including new prefaces by each author).

Maturana, H., Varela, F., 1980. Autopoiesis and Cognition: the Realization of the Living. Reidel, Dordecht.

Maturana, H., Varela, F., 1984. El árbol del conocimiento. Editorial Universitaria, Santiago de Chile.

Mavelli, F., Ruiz-Mirazo, K., 2007. Stochastic simulations of minimal self-reproducing cellular systems. Philo. Trans. Roy. Soc. London B. 362, 1789–1802.

McMullin, B., Varela, F., 1997. Rediscovering Computational Autopoiesis. In: Husbands, P., Harvey, I., (Eds.), Fourth European Conference on Artificial Life. MIT Press, Cambridge, MA, pp. 38–47.

Moreno, A., Lasa, A., 2003. From basic cognition to early mind. Evol. Cogn. 9 (1), 12–30.

Moreno, A., Etxeberria, A., 2005. Agency in natural and artificial systems. Artif. Life 11 (1–2), 161–175.

Maturana, H.R., Mpodozis, J., 2000. The origin of species by means of natural drift. Revista Chilena de Historia Natural 73, 261–310.

Nolfi, S., Floreano, D., 2000. Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines. MIT Press/Bradford Books, Cambridge, MA.

Olasagasti, F., Moreno, A., Peretó, J., Morán, F., 2007. Energetically plausible model of a self-maintaining protocellular system. Bull Math Biol. 69 (4), 1423–1445.

O'Malley, M.A., Dupré, J., 2005. Fundamental issues in systems biology. BioEssays 27, 1270–1276.

Pattee, H.H., 1972. Laws and constraints, symbols and languages. In: Waddington, C.H. (Ed.), Towards a Theoretical Biology 4, Essays. Edinburgh University Press, Edinburgh, pp. 248–258.

Pattee, H.H., 1973. The physical basis and origin of hierarchical control. In: Pattee, H.H. (Ed.), Hierarchy Theory: The Challenge from Complex Systems. Braziller, New York, pp. 71–108.

Pattee, H.H., 2001. The physics of symbols: bridging the epistemic cut. BioSystems 60 (1/3), 5–21.

Pohorille, A., Deamer, D.W., 2002. Artificial cells: prospects for biotechnology. Trends Biotechnol. 20 (3), 123–128.

Risan, L., 1997. Why are there so few biologists there? Artificial life as a theoretical biology or artistry. In: Husbands, P., Harvey, I., (Eds.), Fourth European Conference on Artificial Life, MIT Press, Cambridge, MA, pp. 28–35.

Rosslenbroich, B., 2005. The evolution of multicellularity in animals as a shift in biological autonomy. Theory Biosci. 123, 243–262.

Rosslenbroich, B., 2006. The notion of progress in evolutionary biology. The unresolved problem and an empirical suggestion. Biol. Philo. 21, 41–70.

Ruiz-Mirazo, K. Mavelli, F., this issue. On the way towards 'basic autonomous agents': stochastic simulations of minimal lipid-peptide cells.

Schmalhausen, I., 1949. Factors of Evolution. The University of Chicago Press, Chicago.

Science, 2002. Special issue on systems biology. Science 295, 1661–1682.

Simon, H.A., 1996. [1969,3rd ed.] The Sciences of the Artificial. MIT Press, Cambridge, MA.

Solé, R.V., Munteanu, A., Rodriguez-Caso, C., Macía, J., 2007. Synthetic Protocell Biology: from reproduction to computation. Philo. Trans. Roy. Soc. London B 362, 1789–1802.

Steels, L., Brooks, R.A. (Eds.), 1995. The Artificial Life Route to Artificial Intelligence: Building Embodied Situated Agents. Lawrence Erlbaum, Hillsdale, NJ.

Szostak, J.W., Bartel, D.P., Luisi, P.L., 2001. Synthesizing life. Nature 409, 387–390.

Turner, J.S., 2000. The Extended Organism: the Physiology of Animal-built Structures. Harvard University Press, Cambridge, MA.

Varela, F.J., Maturana, H., Uribe, R., 1974. Autopoiesis: the organization of living systems, its characterization and a model. BioSystems 5, 187–196.

Van de Vijver, G., Van Speybroeck, L., Vandevyvere, W., 2003. Reflecting on complexity of biological systems. Kant and beyond? Acta Biotheoretica 51 (2), 101–140.

Weaver, W., 1948. Science and complexity. Am. Sci. 36 (4), 536–544.

Webb, B., 2001. Can robots make good models of biological behaviour? Behav. Brain Sci. 24, 1033–1050.

Wimsatt, W.C., 2007. On building reliable pictures with unreliable data: An evolutionary and developmental coda for the new systems biology? In: Boorgerd, F., Bruggerman, F., Hofmeyr, J.-H., Westerhoff, H.V. (Eds.), Systems Biology. Philosophical Foundations. Elsevier, Amsterdam, pp. 103–120.

Ziemke, T., this issue. The role of emotion on biological and robotic autonomy.