Synthese
DOI 10.1007/s11229-011-9874-z

Autonomy in evolution: from minimal to complex life

Kepa Ruiz-Mirazo · Alvaro Moreno

Received: 24 June 2010 / Accepted: 6 January 2011

Abstract Our aim in the present paper is to approach the nature of life from the perspective of autonomy, showing that this perspective can be helpful for overcoming the traditional Cartesian gap between the physical and cognitive domains. We first argue that, although the phenomenon of life manifests itself as highly complex and multidimensional, requiring various levels of description, individual organisms constitute the core of this multifarious phenomenology. Thereafter, our discussion focuses on the nature of the organization of individual living entities, proposing autonomy as the main concept to grasp it. In the second part of the article we show how autonomy is also fundamental to explaining major evolutionary transitions, in an attempt to rethink evolution from the point of view of the organizational structure of the entities/organisms involved. This gives further support to the idea of autonomy not only as a key to understanding life in general but also the complex expressions of it that we observe on our planet. Finally, we suggest a possible general principle that underlies those evolutionary transitions, which allow for the open-ended redefinition of autonomous systems: namely, the relative dynamic decoupling that must be articulated among distinct parts, modules or modes of operation in these systems.

Keywords Autonomy · Function · Agency · Evolutionary transitions · Dynamic decoupling

K. Ruiz-Mirazo (✉) · A. Moreno
Department of Logic and Philosophy of Science, University of the Basque Country, Avda. Tolosa 70, 20018 Donostia-San Sebastián, Gipuzkoa, Spain
e-mail: kepa.ruiz-mirazo@ehu.es

K. Ruiz-Mirazo
Unidad de Biofisica (CSIC-UPV/EHU), University of the Basque Country, Barrio Sarriena s/n. Leioa, Bizkaia, Spain

Published online: 01 February 2011

The subject I think is going to be most interesting, and which the new biology will open up, is in fact the understanding of ourselves as organisms. For the first time we can now attack the fundamental biology of man, and I see the beginnings of the understanding of our evolution, and our history, and our culture, and our biology as one thing

Sydney Brenner, Nobel Laureate 2002

1 Introduction

The phenomenon of life has always spontaneously attracted the interest of human intellect. Aristotle, possibly the most relevant thinker of antiquity, in fact takes the living as the starting point for developing his metaphysics. However, in modern philosophy many authors, following Descartes's ideas, tried to understand living beings rather by analogy with manmade artifacts, considering their rich functional properties as a result of complex interactions between, ultimately, simple parts. Thus, the Cartesian distinction between *res extensa* and *res cogitans*, which subsumed the biological domain within a global mechanistic vision of nature, facilitated a scientific research program to study living systems. Others, like Kant, were skeptical about this mechanistic view of organisms, and considered them as irreducible complex systems.[1] However, such skeptical and critical views were not able to propose a clear, alternative research program for the study of biological systems and, as a consequence, the Cartesian mechanistic view ended up being much more influential. Therefore, during the 18th and 19th centuries, living beings were mostly regarded as *just* a complicated case of physical or mechanical systems, in an implicit reductionist move from the complex to the simple.

But mechanistic theories attempting to explain the formation of highly organized natural systems faced a fundamental problem: how to achieve this goal without eventually resorting to the teleological idea of some external intelligent design, responsible for that organization. Darwin provided an elegant way out of this problem, by putting forward a plausible mechanism, natural selection, to account for the evolutionary pathways followed by biological systems and, at the same time, by demonstrating that all species, ours included, have a common origin (a cenancestor, or primitive type of organism).

The impact of Darwinian ideas in the philosophical view of human beings was certainly deep. Since the Darwinian revolution human nature has been seen as something in continuity with biology, because it had to be rationally acknowledged that the human species was, after all, generated from other living organisms and through strictly nat-

[1] Kant (1790/1952) argues that an organism is a system comprised of a whole and parts, where the whole is the product of the parts, but the parts, in turn, depend upon the whole for their own proper functioning and existence. He understood a mechanism as a functional unity in which the parts exist for one another in the performance of a particular function (what might today be called the "operational principles" of a machine). Instead, an organism would be a functional and a structural unity in which the parts exist for and by means of one another in the expression of a particular nature. Thus, the emergence of parts in an organism is a result of internal interactions, rather than the assembly of preexisting parts, as in a mechanism or in a machine. For Kant, a machine has only motive force, while an organized being has *formative* force within it, which mechanisms cannot explain. Since the only way to do science for Kant was the Newtonian way, a scientific study of organisms was simply impossible. Therefore, the question of the nature of life and all biological phenomenology could be dealt with only metaphysically.

ural means. As Dennett (1995) has pointed out, Darwinism could be regarded as a corrosive acid, capable of dissolving our earlier belief and forcing a reconsideration of much of sociology and philosophy. Understanding and accepting Darwin's message means rejecting or modifying a good part of our historical intellectual baggage. However, all this is more related with the conception of our own nature as human beings than with a new conception of what living systems are.

Certainly, Darwin provided a new general framework to understand how complex organisms could be generated from simpler ones, but he did not provide a theory of organisms (neither of how they work, nor of how they can originate from physicochemical systems). Therefore, Darwin's theory did not solve all the objections to mechanistic approaches, especially regarding the lack of a complete, satisfactory account of the intrinsic functioning of living organisms. Furthermore, although by the end of the 19th century many scientists were assuming that progress in chemistry and thermodynamics could lead to a fully mechanistic explanation of biological systems (and therefore, would guarantee the success of the Cartesian research program), it remained quite difficult to understand the apparent differences between machines and organisms.

These difficulties in understanding or deciphering the functioning of organisms would explain the rebirth of Kantian ideas in the early 20th century "organicism," which claimed that living systems are irreducible wholes, showing emergent properties, since the intimate processes of life cannot be grasped in terms of our strictly mechanistic, physical or chemical models (Elsasser 1966; Haraway 1976). As a classical example (taken from Driesch), embryological processes would show capacities for self-determination and regulation that no manmade machine could ever achieve. Nevertheless, albeit often correct, the criticisms thrown by the organicist school of thinking to the mechanistic paradigm ended up being rather sterile, because they had no other way out but metaphysics. It seemed that not only the Kantian reasoning about the fundamental differences between mechanisms and the organization of living entities, but also his conclusion (against the possibility of acquiring scientific knowledge, *sensu stricto*, about the latter) had to be assumed.

However, things were going to change, as science (and biology in particular) further advanced. The first seeds for that change were sown during the second part of the 20th century, when, mainly as a result of molecular biology's revolution, the mechanisms underlying the organization of living systems began to be unraveled.[2] Quite interestingly, molecular biologists, along with their major discoveries at the biochemical level, introduced (how consciously would be matter of debate) a whole series of cybernetic/informational concepts, showing that, behind the apparent physicalist or mechanistic character of their approaches and results, a much more intricate and rich epistemological reading had to be done. So, although the scientific practice was utterly reductionist, there was still some hidden, unresolved tension with part of the organicist claims.

At present, in the so-called "post-genomic" era, this tension is somehow being channeled, or released, with the emergence of new fields like *systems biology* (focused on

[2] In parallel, the scientific approach to the problem of the origins of life, previously launched by Oparin (1994/1928) and Haldane (1994/1929), was also obtaining experimental results of great significance (Miller 1953; Oró 1961).

the complexity of biomolecular interaction networks) or *synthetic biology* (in between genetic engineering and artificial life). The spirit of these new scientific approaches is much closer to a holistic or integrative conception of living systems, bringing along novel biotechnological tools—both experimental and theoretical—and giving fresh food for thought and discussion. In a similar way to what artificial life[3] and complexity sciences did at the end of the last century, these emerging fields will surely help to pose old philosophical problems (e.g., the relationship between matter and form, the symbol-grounding problem, the question of emergence, etc.) from different, still barely foreseeable perspectives.

Certain theories in the last century, like autopoiesis, already provided good arguments to search for the roots of normativity, meaning and autonomous agency in our current knowledge of metabolic organization (Weber and Varela 2002). Similarly, other currents of thought in theoretical biology (e.g., Pattee 1982, or the biosemiotic school: Emmeche and Hoffmeyer 1991; Hoffmeyer 1996) argued that notions such as symbols, internal descriptions, codes and messages, would have a first physical implementation in the role played by DNA within a cell. Somehow, the findings of 20th-century biology were starting to make possible the elaboration of a language that could serve as a bridge between molecular mechanisms and system organization (and, thereby, with concepts traditionally full of philosophical connotations, like function, information, etc.).

These possibilities of reformulating old philosophical questions in terms of new scientific insights and tools are being enhanced through the discovery of more and more detail and intricacies involved in the physiology of cells (Harold 2001; de Duve 2002; Bechtel 2006), the development of computational models and network theory as applied to the biological domain (Ravesz et al. 2002; Montoya and Solé 2002; Kitano 2002), the standardization and manipulation of synthetic functional modules ("BioBricks"; see Endy 2005; Andrianantoandro et al. 2006), or other strategies to put together biologically engineered artificial systems (Benner and Sismour 2005; Solé et al. 2007; Rasmussen et al. 2008).

Such a stimulating research context offers a great opportunity to tackle the question of the nature of life under new light. At least, philosophy has more scientific knowledge than ever available to accomplish the task. Our aim in the present paper is to take a step forward in this direction, approaching the nature of life from the perspective of autonomy, and looking into its philosophical implications. We first argue that, although the phenomenon of life manifests itself as highly complex and multidimensional, individual organisms constitute the core of this multifarious phenomenology. Thereafter, our discussion focuses on the nature of the organization of individual living entities, proposing autonomy as the main concept to grasp it. Finally, in the second part of the article, we argue that autonomy is also fundamental to explaining major evolutionary transitions.

[3] Artificial life (also known as "ALife") is an interdisciplinary study of life and lifelike processes that uses a synthetic methodology. According to Bedau (2003), there are three broad and intertwining branches in this research area, corresponding to three different synthetic methods: "soft" artificial life creates simulations or other purely digital constructions that exhibit lifelike behavior; "hard" artificial life produces hardware implementations of lifelike systems; "wet" artificial life tries to synthesize living systems in biochemical media.

In sum, the aim of this paper is to contribute to showing what enormous potential a better understanding of biological organization—of what it really means, and how it was generated—holds to modify our vision of nature and humankind; and, in particular, to realize that the cognitive and material domains are, after all, intrinsically connected. This is probably the strongest challenge ever faced by the dualism of the still-influential Cartesian worldview, which so radically separated the spheres of nature and of mental processes. Today, this Cartesian divide is still influential not so much as an ontological dualism but rather as a kind of explanatory dualism. Thus, what we would like to argue in the following pages is that modern biological sciences are opening up new ways to explain, in fully naturalistic terms, an increasing number of phenomena that until now were linked to the *res cogitans* domain, such as intentionality or purposefulness. Although more and more philosophers and scientists (like Brenner, in our opening quote) already assume and foresee a unified scenario, philosophy at large still needs to take the message seriously on board and contribute to elaborating a new theoretical framework where the connection between human nature and the rest of the natural world can be reassessed, turning the ontological claim into an epistemological achievement.

2 Trying to grasp the core of the living: autonomy in evolution

Perhaps it is not very wise to try to define life because, as with all definitions, any formulation will be limited by the same linguistic terms that have articulated it and, therefore, its validity or application will always be context-dependent (Oliver and Perry 2006). Perhaps it is not very wise to try to define life when there is no general theory of biology yet that provides us with the keywords, in the absence of alternative non-terrestrial examples of it (Cleland and Chyba 2002, 2007; see also Cleland 2011). Or perhaps it is not very wise to try to define life when the nature of life itself is so multifarious and elusive, so full of borderline cases and counter-examples, that one could even doubt if it is, actually, a natural kind (Keller 2008).

However, the situation in biology at present is not at all like the one of trying to define or find the true nature of water in 17th-century chemistry, when the molecular theory of matter had not yet been developed (Cleland and Chyba 2002, 2007). The amount of accurate scientific data and knowledge about living systems available these days is enormous, so much so that it is becoming hard to assimilate. Of course, the discovery of some system or phenomenon that we unanimously agree to regard as an alternative form of life, when it comes (if it eventually comes, through space exploration or in the lab), will have deep implications and change profoundly our worldviews and our conception of the living. But, meanwhile, we should not abandon efforts to put together explicitly, in a distilled way, what our day-to-day increasing biological knowledge is telling us about the actual concept of life. Quite the contrary, these efforts should be encouraged, as part of what biology is missing right now: more encompassing approaches that contribute to integrating the huge amounts of data and relevant information being continuously generated.

Current scientific knowledge gathered from the diversity of living entities on Earth, and the diversity of investigations carried out on them, enables us to start discerning what is necessary and what is contingent in their organization and, in that way, achieve eventually a more complete or congruent characterization of the phenomenon of life, in its minimal general sense. Almost nobody will nowadays dispute that: (i) living systems are *organized* in a radically different way than nonliving systems and (ii) living systems *change in time* in a radically different way than nonliving systems. The difficulties come as soon as one makes the assumption that *all* living systems are organized and change in time in a common, characteristic way; and, even more, when one tries to specify the terms in which that organization or that evolutionary trend should be accounted for.

These difficulties are related to a third, also widely accepted, feature of the phenomenon of life: the fact that collective-evolutionary and individual-systemic aspects of the living are deeply intertwined. In a certain sense, this tends to support a view of life as a highly adaptive, supple, evolving biosphere (Bedau 1998). Any known living being cannot exist but in the context of a global network of similar systems. And this is clearly reflected in the fact that genetic components (which specify the metabolic machinery and organization of single biological entities), in order to be operational, have to be shaped through a process that involves a great amount of individual systems and many consecutive generations, or reproductive steps. The unfolding of an evolutionary process by natural selection, based on heritable modular templates and genetic mechanisms, allows life to explore many possible combinations and formulas to survive, providing extreme examples of adaptation that make any borderline within the living domain rather diffuse. Actually, evolution, beyond that intrinsic competitive/selective dynamics (which would lead to rather selfish entities; Dawkins 1976), weaves a collective network of increasingly complex and entangled cooperative relations among entities at different phenomenological levels and with different cohesive strength (Dupré and O’Malley 2009).

Yet, acknowledging the collective dimension of the phenomenon of life does not mean blurring the key role of individuality. However embedded in evolutionary and ecological webs living systems might appear, their metabolic functioning still points to a basic organizational core that should be properly characterized. If the essence of biological organization is conceived as a web of diverse interactions *among* rather than *within* current living systems (extending the idea of living system to molecular replicators, viruses, prions, organelles, parasites, etc.), it will not be possible to determine whether organisms should be taken as a basic starting point (i.e., as a highly integrated and cohesive type of organization that gets progressively complex) or just as some occasional result of an ongoing dynamics of loose cooperative relations among different kinds of biological entities. This is an important issue, because it could turn out that without such a highly integrated and cohesive individual organization, living systems would not be able to build a wider and more complex historic-collective meta-network (which, in turn, provides the necessary conditions for their long-term maintenance and evolution). Furthermore, without a strong idea of individual metabolic organization, it would be very difficult to provide a naturalized account of concepts like functionality, agency, unit of selection, etc., or to make a clear-cut distinction between organisms and other forms of cooperative or “ecological” networks (Ruiz-Mirazo et al. 2000). After all, life seems to demonstrate, in the course of evolution, that increasingly complex forms of individual agents have emerged and developed, bringing forth progressively sophisticated constructive, interactive and cognitive capacities.[4]

But, how can an "individualistic" perspective on the phenomenon of life be, at the same time, congruent with the historic and collective dimension? Elsewhere (Ruiz-Mirazo et al. 2010) we have argued that these two dimensions are covered by defining life as "a complex network of self-reproducing *autonomous* agents whose basic organization is instructed by material records generated through the *open-ended,* historical process in which that collective network evolves," a statement that actually derives from the realization that collective-evolutionary and individual-systemic aspects of the living are so deeply intertwined that they cannot be set apart.

The two fundamental concepts in this definition are autonomy and open-ended evolution. By 'autonomy' we mean the property of a system that builds and actively maintains the rules that define itself, as well as the way it behaves in the world. So autonomy covers the main properties shown by any living system at the individual level: (i) self-construction (i.e., the fact that life is continuously building, through cellular *metabolisms*, the components which are directly responsible for its behavior) and (ii) functional action on and through the environment (i.e., the fact that organisms are *agents*, because they necessarily modify their boundary conditions in order to ensure their own maintenance as far-from-equilibrium, dissipative systems).[5] Open-ended evolution, in turn, covers the properties of life as a collective-historical phenomenon, i.e., as an intricate network of interacting individuals (organisms), bringing about other similar individuals, and undergoing a long-term process of change which allows for an indefinite increase in their complexity (always under the constraints of a finite physical-chemical world).

More importantly (although arguing for this would require another full paper; see Ruiz-Mirazo et al. 2008), a process of open-ended evolution cannot occur except in the context of a population of autonomous systems. And, conversely, the unfolding of autonomous systems and their long-term maintenance depend on their insertion in an open-ended evolutionary route.[6] So it is really the integration of these two main ideas, autonomy and open-ended evolution, that provides a complete, rich enough picture of the phenomenon of life. Such a synthetic conception is, in addition, an attempt to bring together the main results and theoretical legacy of two different traditions in biology: the physiological-biochemical tradition, focused on immediate or "proximate" causes, and the evolutionary-historical tradition, interested in the final or "ultimate" ones (Mayr 1982).

Therefore, this definition implies a view of the phenomenon of life in which individuality and agency cannot be severed from a wider collective organization: as the

[4] In fact, an ever more influential school of thinking claims that the study of human mental capacities cannot be achieved without embodied, strongly integrated strategies (Varela et al. 1991; Clark 1997; Gibas 2005; Thomson 2007; Calvo and Gomila 2008) and, especially, without a "biogenic" approach to understanding the appearance and subsequent development of those capacities (Hooker 1995; Christensen and Hooker 2000; Lyon 2006; Lyon and Keijzer 2007).

[5] We come back to this point later in the paper.

[6] Later in this paper, we provide further support for this claim, analyzing how different kinds of autonomy were actually involved in the main evolutionary transitions in the history of biological systems on Earth.

individual organization unfolds, it creates and supports a more encompassing historic and collective network, which in turn sustains and facilitates its evolution in a changeful environment. Furthermore, as Bedau (1996) has pointed out, a significant aspect of the environment to which any given organism must adapt is the set of all other organisms with which it interacts. "So, when a given organism adapts and changes, the evolutionary context of all the other organisms changes. Thus, even without an externally changing environment, adaptation can be a co-evolutionary process that internally changes the selection pressures which shape adaptation, thus making open-ended adaptive evolution an intrinsic property of the system" (Bedau 1996, p. 339).

Nevertheless, the focus of analysis in the present paper will be on biological complexity at the level of individuals, i.e., organisms. Apart from a particularly cohesive organization, organisms display a particularly marked impulse or urge to persist in their state of being (Spinoza's *conatus*). It is, therefore, important to understand life at that individual level, and analyze carefully the implications of the emergence in the natural world of systems with that capacity to act for their own benefit, to constitute identities that distinguish themselves from the environment (at the same time as they keep interacting with it as open, far-from-equilibrium systems). Actually, in those conditions, the environment becomes a world full of significance, an *umwelt* (Uexküll 1982/1940): facts that externally could appear as purely physical or chemical develop into positive, negative, or neutral influences on the system, depending on whether they contribute to, hinder, or have no effect on the maintenance of its dynamic identity. Thus, even the simplest living organism creates a set of preferential partitions of the world, converting interactions with their surrounding media into elementary norms or values, as we will explain more extensively below. And here is where the nature of living systems as autonomous agents, as inventors of worlds with meaning, becomes manifest (Hoffmeyer 1996).

So, from the origins of life to the origins of humanity we envision a complex series of transitions in which autonomous systems are, one way or another, involved. The goal would be, then, to investigate those transitions and try to extract conclusions at a more general scale. In other words, as we suggested in the introduction, we would like to investigate in which way autonomy contributes to a naturalization project that comes to terms with the presently overruling scientific worldview about the biological and human domains, or in which way it could help to look at traditional problems in philosophy (specifically, in philosophy of biology) through a new lens. The following pages explore these issues.

3 Minimal biological complexity: the autonomous roots of functionality and agency

Autonomy may initially appear too heavy a word to be part of a general definition of life. Originally used in the context of law and sociology (in the sense of self-government, from the Greek polis) or human cognition and rationality (in the sense of a cognitive agent that acts according to rationally self-generated rules, cf. Kant), for many it will sound like a high-level concept, with too many non-strictly-biological connotations. Broadly speaking, autonomy is understood as the capacity to act according to self-determined principles. It might also mean that a given ontological or phenomenological level is relatively independent with respect to others, because it is ruled by its own norms (Moreno et al. 2008). However, the idea of autonomy can adopt a more specific, minimal sense ["basic autonomy," as we have called it (Ruiz-Mirazo and Moreno 2000, 2004)] related to the capacity of a system to self-define, to construct its own identity. It is in this more basic sense that autonomy proves relevant for the definition of life, since it provides the necessary explanatory power to account for the complex material organization underlying any living organism; namely, its *metabolism.*[7]

The essence of the idea of metabolism is that of a cyclic, self-maintaining network of reactions by means of which the components of a system are continually produced, in far-from-equilibrium (global matter-energy flow) conditions.[8] Actually, it was a simplified and rather abstract version of this cyclic idea of metabolism that inspired Maturana and Varela (1973, Varela et al. 1974) to define living beings as autopoietic (self-producing) systems. So we are not the first ones to use the concept of autonomy in this minimalist, biologically significant way (Varela 1979). However, in contrast to the autopoietic school, we are particularly interested in finding the possible material and thermodynamic roots of the concept (see also Kauffman 2000, 2003).[9]

The main motivation for this search is to provide a link with physics and chemistry, so that the idea of autonomy itself is naturalized and can serve as a bridge from the nonliving to the living domain (Ruiz-Mirazo et al. 2004). And the main reason why autonomy is applicable to this gap between chemistry and biology, beyond standard self-organization, is because there is a strong sense in which component production relationships in a system (chemical transformations, reaction feedbacks, mutual interactions, catalytic effects, etc.) can be interpreted as *self-constraining* processes, i.e., as the generation by a system of the local rules (constraints) that govern its dynamic behavior (Pattee 1972). So it is through this self-constraining action that the system actually defines itself, *constructs an identity of its own*.

Nevertheless, construction requires material and energetic resources; here comes the thermodynamic dimension of the problem. Kauffman (2000) elegantly solves this

[7] Metabolism is currently seen as the set of processes that allow cells to build and replace their structures, grow and reproduce, and respond to their environments. Biochemists typically think of metabolism at the cellular level and describe it as a network of chemical reaction pathways in which each step is kinetically controlled by an enzyme, where the characteristic feature is that enzymes are, themselves, materially fabricated within the system. In other words, a metabolic organization is "enzymatically closed" or, more precisely, closed to efficient (enzymatic) causation (Rosen 1991). In a derived way, in multicellular organisms, metabolism is seen as a set of intercellular relations that ensure the processes of building and replacement of the structures of the whole organism, so as to ensure its maintenance and functional interaction with its environment.

[8] The notion of metabolism as a chemical process, as a transforming "force" of energy and matter is quite old (e.g., Schwann's ideas on metabolic forces; Schwann 1939). Modern studies of metabolism have focused on the linear or cyclic transformation of diverse chemicals, but always with a nutrition-bioenergetic orientation. Only recently has the importance of the global circular nature of metabolism (and its connection with ideas like self-maintenance, self-production, self-repair, etc.) been recognized (see, for instance: Cornish-Bowden et al. 2007) following last century's theoretical contributions of authors like Rosen (1958, 1991), Gánti (1971) or Maturana and Varela (1973).

[9] The main reason why energetical and thermodynamic requirements should be taken into account when defining an autonomous organization stems from the impossibility of understanding the actual *origin* of self-constraining/self-maintaining/self-producing systems unless we take these requirements into account.

problem by postulating a cyclic connection between constraints and work (fully profitable energy) at the core of any chemical autonomous system: "[W]ork begets constraints begets work." The capacity to generate work requires constraints, but the production of constraints also involves some useful energy input. So autonomous systems would be those in which this loop is established and robustly maintained.[10]

Now, there is an important part of the job that Kauffman leaves unfinished: determining the amount and type of constraints (and associated work conversion processes) required to achieve minimal autonomy. Ultimately, this question will have to be addressed empirically, but a preliminary theoretical analysis of it would be convenient to guide that empirical research. Without getting into the detailed underlying chemical reasoning (Ruiz-Mirazo and Moreno 2004; Morowitz et al. 1988; Harold 1986; de Duve 1991), a minimal self-constructing system must include, at least, a *membrane* (i.e., a semipermeable compartment, which is fundamental for non-isolating individualization), a set of *catalysts* (for the adjustment and coordination of reaction rates, i.e., the most elementary type of kinetic regulation) and some *energy currencies* (for thermodynamic viability, i.e., endergonic and exergonic couplings that involve inter-convertible forms of work). Regardless of the chemical specificities, what is more significant for the purposes of the present article is to notice that, even in this minimal case, (i) a *variety* of constraints (and work forms), of completely different nature, must come together and (ii) these constraints are not just internal but include the production, maintenance and modulation of boundary conditions. In other words, they involve some control (i.e., an asymmetric influence, exerted by the system) on the domain of interactions with the environment.[11] These two features are central to understanding why autonomy, in this basic, chemical sense, already involves both a constitutive and interactive dimension, and may hold the key to naturalizing the ideas of *function* and *agency*, as we argue more extensively below.

3.1 Naturalizing function

There has been a long-standing debate in philosophy of science about the idea of function and how to develop a scientifically well-grounded account of it. This is particularly relevant for philosophy of biology, since explanations in terms of functions

[10] An autonomous system involves therefore a material-organizational apparatus of energy management that can implement an operationally closed constructive-relational system, in such a way that the component production relationships it contains continuously renew the aforementioned apparatus. Accordingly, autonomy is not an abstract, material-independent property, as it was defined by the autopoietic school. An autonomous organization should implement specific control mechanisms upon the energy flows necessary for the physical realization of an operationally closed component production system. This deep entanglement between the constructive-relational and energetic-thermodynamic dimensions of the problem discards the possibility of a purely computational autonomous system (namely, virtual systems confined to the cyberspace, like the creatures of the so-called "strong AL"; see Moreno and Ruiz-Mirazo 1999; Boden 1999).

[11] We are speaking here about a hypothetical set of conditions for autonomy, pushing things to a strict minimalist context. In the case of full-fledged living organisms (e.g., bacterial cells, Kauffman's favorite example when discussing his ideas on autonomy) it is more obvious that these conditions are met, surely together with other more complex ones, which ensure an autonomous behavior.

are characteristic of our models of living systems and remain essentially irreducible to physics or chemistry. Although that irreducibility might also apply to human-made machines or artifacts (Polanyi 1968), here we will restrict our discussion to the biological case, in which functional relationships are not imposed by an external designer but must be endogenously produced.

Among the different philosophical proposals made in recent years to characterize and naturalize the concept of biological function, the predominant approach has been the so-called "etiological" (Wright 1973), according to which functional attributions should be based on the causal history (i.e., the *etiology*) of the particular phenotypic trait under analysis. Among these, the most widely embraced are "selected effects theories" (SET), which try to naturalize functions through the idea of evolution by natural selection (Millikan 1989; Neander 1991; Godfrey-Smith 1994). These theories provide a naturalistic account of biological functions, but at the expense of falling into epiphenomenalism (Christensen and Bickhard 2002), in the sense that they do not take into consideration the actual, current properties of the system under analysis. For SET, functional attributions are not related to the way a system is made or behaves in the present, but to its previous evolutionary history. In addition, it is difficult to explain, by means of this type of approach, how functional relationships originate in the natural world. Since the causal history supporting the functionality of a biological trait is based on the action of natural selection, from the point of view of a strict naturalization program they face the problem of providing an account of the origin of the process or mechanism of natural selection itself.

It is quite problematic to naturalize biological functions in terms of natural selection, because it is not possible to resort to the mechanism of natural selection without assuming the previous existence of systems with some phenotypic-functional diversity. Contrary to the view of SET defenders, natural selection requires the presence of functions and not the other way around. In fact, natural selection is a rather complex and indirect mechanism and, in order for it to start operating a population of systems that reproduce and bring about some sort of heritable phenotypic variation is required, as Lewontin (1970) and Maynard Smith (1986) already highlighted. Expressed in such general terms, these conditions might appear relatively trivial to fulfill, but they are not, as a more thorough examination of prebiotic candidate systems shows.[12] In particular, the question of how phenotypic variation is generated in the first place deserves attention, as it is directly related to the problem of the origin of functional diversity in a system. In this context, an account in terms of (basic) autonomy, like the one we present in this paper, offers an obvious advantage because, in contrast to a population of "selfish" replicating molecules (e.g., RNA quasi-species; see Eigen and Schuster (1979), a population of autonomous protocells[13] would provide a wide enough

[12] Indeed, the concrete properties/strategies of reproduction, heredity, variability, etc. displayed by those systems affect the resulting selective dynamics of the population (Szathmary 2006).

[13] By "protocells" we mean systems that possess a self-made topologically closed boundary, similar to present day biological cells, but whose level of complexity is still far below the living threshold (i.e., they represent hypothetical "infrabiological systems," which are currently being tested experimentally, both *in vitro* and *in silico*: see Solé et al. 2007; Mansy 2008; Rasmussen et al. 2008; Ruiz-Mirazo and Mavelli 2008).

functional space for the implementation of a process of natural selection "without dead ends" (Wicken 1987; Moreno and Ruiz Mirazo 2009).

Some other authors have proposed a radically different way to naturalize functions, also within the general framework provided by well-established biological knowledge. These alternative approaches can be regarded as "systemic" or "dispositional" (Cummins 1975; Nagel 1977; Craver 2001; Davies 2001) because they conceive functional relationships in terms of the current properties of a system, typically a full-fledged living organism. Following this perspective, in the last years a new approach is taking shape that attempts to define functional relationships in terms of the current organization and dynamic behavior of the system under analysis (Schlosser 1998; McLaughlin 2001; Christensen and Bickhard 2002; Delancey 2006; Edin 2008; Mossio et al. 2009). Bickhard (2000) and Christensen (Christensen and Bickhard 2002), for instance, conceive functions as contributions to the self-maintenance of an autonomous system. More recently, Mossio et al. (2009) defend a similar position, but grounded in the context of (presumably previous) self-organizing phenomena. In any case, all of these latter types of naturalizing approaches, implicitly or explicitly, convey the idea that functions are specific causal relations attributed to differentiated parts of a self-maintaining system, whose organizational homeostasis is thus preserved.

This idea fits very well with our approach to the origins of life in terms of autonomy. We were just pointing out in previous paragraphs that a variety of material constraints (and associated work conversion processes) would be required to bridge the gap between self-organization and self-construction (i.e., proper autonomy, however minimal). And this precisely corresponds with one of the main requirements ("distinguishability" in the part-whole relationship) that are necessary to demarcate, within the general class of far-from-equilibrium dissipative systems, those showing functional features, like living systems. Indeed, autonomy would involve the endogenous production of distinguishable parts (membrane, catalysts, energy currencies, etc.) that contribute in different ways (i.e., through remarkably different constraining actions) to the constitution and maintenance of a whole, integrated entity (a protocell), whose organizational homeostasis would reinforce the conditions for stability of those very component parts. Furthermore, this emergent autonomous organization would become the reference to ground *normativity* in the system, in the sense that all the molecular processes involved (self-assembly, chemical reactions, auto/cross catalysis, etc.) can be somehow "internally evaluated," according to their contribution—or lack thereof—to the maintenance of that global organization. (We will come back to the question of normativity below.)

3.2 Naturalizing agency

The concept of agency, again, seems to belong to a high-level phenomenological sphere, associated traditionally with cognitive or behavioral sciences and artificial intelligence (AI). Like Barandiaran et al. (2009) recently remarked, it has been used in those fields often in a rather intuitive and uncontroversial way, as an alternative to other, more heavily loaded terms, such as "intentionality," "willingness" or "purposeful action." However, agency can also be interpreted in a more elementary or basic sense: simply as the activity or dynamics of a system through which it gets engaged in an "interactive loop" with the environment (Smithers 1997). This type of characterization, even though coming from the field of (situated-embodied) robotics, is more amenable to naturalization than the inherited traditional conception from AI.[14]

In fact, if one looks into the problem of how natural self-organizing systems could develop into more complex self-maintaining systems, and eventually into autonomous (self-constructing) systems, from a thermodynamic perspective, it appears rather obvious that these systems must engage in such an interactive loop with their respective environments (although the way to achieve this, in practice, is not so obvious). All open, far-from-equilibrium systems, from the simplest dissipative structures to living systems, strongly depend on boundary conditions (gradients, influx/outflux of different compounds, energy transduction mechanisms, etc.) in order to sustain the processes of generation of internal "order," in accordance with the generalized second law of thermodynamics. Now, what can autonomy mean in such a context, in which there is always an ultimate dependence on material-energetic resources provided by the environment? One of the keys is to analyze the situation in terms of the stability or robustness in the maintenance of those far from equilibrium conditions that precisely allow the emergence of a system with a distinct dynamic behavior. Unlike physical or chemical dissipative structures, in which patterns of dynamic order form spontaneously, but whose stability relies almost completely on externally-imposed boundary conditions, autonomous systems build and actively maintain most of their own boundary conditions, making possible a robust far-from-equilibrium dynamic behavior.

Thus, the heart of the issue is: how does a system develop the capacity to *channel* the flow of matter and energy through itself, so as to achieve robust self-construction (i.e., self-construction that includes regulation loops with its immediate environment)? The importance of this capacity can be readily confirmed if we turn to concrete biological examples. The stability/homeostasis of all living cells depends on the synthesis and maintenance of sophisticated molecular mechanisms, like those making possible an active control of electrochemical gradients or chemotactic behavior.[15] One of the main points to highlight here is that this involves a clear *asymmetric* system-environment relationship: The system, through the implementation of these mechanisms (typically embedded in its boundary) has the final responsibility for coupling with the environment, i.e., the responsibility for defining the terms in which it relates to the surrounding milieu. Such an asymmetry is important because it is where we find the roots of agent behavior.

So the interactions between an autonomous system and its environment are not merely physical interactions, because the viability of the system itself is at play. There are, of course, material and energetic exchanges occurring all the time, but these are

[14] For instance, Russell and Norvig (1995) definition: "[A]n agent is anything that can be viewed as *perceiving* its environment *through sensors* and *acting* upon that environment *through effectors*" (our italics; quoted from Barandiaran et al. 2009).

[15] Bioenergetic studies, like Harold (1986) or Skulatchev (1992), show that this active control of boundary conditions (this "vectorial metabolism") is fundamental for all cells. But also experimental models, like complex bioreactor networks embedded in simple vesicles, prove that compartmented biochemical reactions cannot proceed for very long unless the system has a boundary that mediates adequately the exchanges with the environment, harvesting the necessary "fuel" (Noireaux and Libchaber 2004).

intrinsically biased by the rules that ensure the system's viability. An autonomous system is continuously performing actions, doing things that contribute to its own maintenance. A big stone in the river holds water from flowing, and some bacteria ferment milk to produce yoghourt. Although both systems do something, we do not call the stone an agent. The difference between the two cases is not in the degree of change operated by one or the other, but in the consequence of that change: only in the latter case does the change contribute to the maintenance of the performer of the action. Thus, for the autonomous case one could say that the interactions with the environment are, at the same time, a result of the constitutive processes of the system and a necessary condition for their continuity. In other words, there is a reciprocal dependence between what defines the "self" (or the subject) and the actions derived from its existence, because it is not really possible to separate the system's *doing* from its *being* (Moreno and Etxeberria 2005).

It was philosopher H. Jonas (1966) who first defended, against the Cartesian view, that metabolism should ground intrinsic teleology and natural agency, because of its inherent self-maintaining nature. Our approach, articulated around the notion of autonomy, serves to re-formulate his basic insights from a somewhat different, scientifically up-to-date perspective. Since the very existence of an autonomous system depends on the effects of its own activity, this activity has an intrinsic relevance for the system itself. Such intrinsic relevance generates a naturalized criterion to determine what the system is *supposed to do*, as Barandiaran and Moreno (2008) and Mossio et al. (2009) have pointed out. In fact, the whole system (and its constitutive processes) *must* behave in a specific way; otherwise it would cease to exist. That is why the activity of the system becomes its own norm; or, more precisely, the conditions of existence of its constitutive processes and organization are the norms of its own activity, both inwards and outwards.

Therefore, normative function and agency are two sides of a single coin: autonomy. Internally, an autonomous system is an organization of functional relationships among distinguishable components whose causal effect is their cohesive or cooperative integration within that self-maintaining and self-producing organization. Agency, in turn, is the external observable expression of that ongoing activity: it covers those relationships among distinguishable components whose causal effect, mediated through the environment, is to contribute to the robust self-maintenance and self-production of the system. But robustness leads us to the problem of how these relationships are regulated, and how it is possible to move from homeostasis to adaptive behavior.

3.3 Regulation and adaptivity

So far we have analyzed autonomous agency in its minimal form. Nevertheless, in current living systems autonomous agency is also adaptive. "Adaptive agency" would be, as Di Paolo (2005) has defined it, the capacity of a(n autonomous) system

> to regulate its states and its relation to the environment with the result that, if the states are sufficiently close to the boundary of viability, 1. tendencies are distinguished and acted upon depending on whether the states will approach or

> recede from the boundary and, as a consequence, 2. tendencies of the first kind are moved closer to or transformed into tendencies of the second and so future states are prevented from reaching the boundary with an outward velocity. (Di Paolo 2005, p. 468)

Adaptive systems have, therefore, the capacity to modulate (internally and interactively) the trajectories of the essential variables of their constitutive organization. This capacity requires that metabolic paths be, in turn, modulated and specifically harnessed so as to compensate for external perturbations before the system reaches a critical point. In other words, adaptive regulatory constraints perform their activity by *appropriately* changing metabolic fluxes. For example, bacteria are able to monitor and regulate their internal processes so that they generate the necessary responses, anticipating internal tendencies and being able to evaluate graded differences in the outcomes of otherwise equally viable states. This implies some sort of control mechanism (composed by receptors, effectors and a transformation function between them) by which the system distinguishes and compensates tendencies that, if no compensation were carried out, would bring the system too far from its optimal state. So adaptive agency permits introduction of a minimal form of functional detection.

Therefore, we can move now beyond the implicit normativity of minimal autonomy and naturalize the claim that some interaction or process is detected as "bad" or "good" *by* and *for* the very system (not *by* and *for* an external observer). This good or bad functioning *for* the system is objective because it is detected and compensated for *by* the system, in an effective, functionally integrated way. Thus, adaptive systems are an instance of explicit normativity. In addition to having an intrinsic norm due to their basic autonomous organization, they also have the capacity to *modify their actions* to fulfill this norm better and more robustly, generating global constraints on their minimal, elementary way of functioning. Thus a regulatory form of control emerges, operating upon the basic self-maintaining and self-producing infrastructure (Barandiaran and Moreno 2008). But this poses a new problem, namely, where does such an explicit normativity come from? In order to answer this question we shall analyze the origin of adaptive agency in the context of prebiotic evolution.

This remarkable capacity was presumably absent in prebiotic autonomous systems, whose stability in time would rely on internal redundancies and feedback loops. Before the complex and indirect process of natural selection started to operate, primitive autonomous systems could probably ensure stability against perturbations through reorganization processes that involved higher levels of modularity and feedback-regulatory relationships among those modules. However, this scenario is bound to face a strong bottleneck, since control relationships can be established only between modules whose characteristic/operational time scales are clearly different.[16] In other words, there has

[16] As Bechtel (2007) has pointed out, if control is to involve more than strict linkage between components, a property that varies independently of the basic operations is required. The manipulation of this property by one component can then be coordinated with a response to it by another component, so that one component can exert control over the operation of the other. Therefore, what is required for an effective control system is a property that is sufficiently independent of the processes of material and energy flow that it can be varied without disrupting these basic processes, but still able to be linked to parts of the mechanism so as to be able to modulate their operations.

to be some sort of *dynamic decoupling* between the controlled and controller subsystems.

Such a bottleneck could be overcome when proto-metabolic systems incorporated heritable modular components, which allowed a completely different and indirect exploration of control relationships. The key point here would be that the search in the sequential space of modular components is "stoichiometrically free" (Griesemer and Szathmáry 2009) from the basic constructive organization of the system, so these components can drive new chemical paths that are dynamically decoupled from low-level, constructive processes, and thus perform regulatory tasks (e.g., maintaining functionally adjusted concentrations of internal metabolites in the face of variations of the metabolic flux).

So when primitive autonomous systems incorporated heritable modular components, a completely different, freer and more indirect exploration of control relationships was possible. In this way, the "appropriateness," the norm that specifies what the system adaptively has to do, comes from "outside," from a more encompassing meta-network of processes that take place at a very different characteristic time scale. Relatively free from direct metabolic constraints, the incorporation of heritable modular components allows an indefinite exploration of a huge space of possibilities, finding (through a selective process of retention, taking place at large space-temporal scales) new functional patterns of organization, thus leading to embodied normative regulatory mechanisms in the individual organization.[17] This means that only those autonomous organizations driven by catalysts and functional components whose specificity is determined through trans-generational processes (i.e., through a dynamically decoupled set of events)[18] can achieve adaptive agency.

In conclusion, as primitive autonomous systems increased in complexity, they constructed a spatially and temporally wider organization (an evolutionary framework) in which a digital-type of memory is transmitted (i.e., modular templates acting as records; see Pattee 1977), so that a huge sequential space could be collectively explored and only those sequences functional for the individual systems could be retained. Accordingly, as the minimal core of autonomy unfolds, it creates a wider, collective organization, which involves many similar systems and their respective environments. At the same time, the organization of these individuals becomes internally more complex (enlarging their functional diversity), as well as their interactive, agential capacity (e.g., anticipatory adaptivity).

[17] The introduction of sequential hereditary components had further consequences, when the first systems with reliable heredity (hypothetical RNA-based metabolisms) led to the invention of a second strong decoupling (DNA-RNA-protein metabolisms), according to which autonomous systems, once again, completely reorganized their way of functioning. This is achieved through a new cyclic causal correlation established between *two operational modes* in the system, so that it can "self-interpret" the sequences of the genetic components (Pattee 1977, 1982). The new form of organization, usually described in informational terms, is what allows the access to an unlimited number of metabolic organization designs, and, therefore, to open-ended evolution (Ruiz-Mirazo et al. 2008).

[18] This is what a natural selection process, in fact, involves.

4 The development of biological complexity in time: the open-ended redefinition of autonomy

In the previous section we analyzed some of the implications of choosing autonomy as the fundamental concept to grasp the core of the organization of living systems at its basic, cellular level. But is the idea of autonomy in any sense also helpful for understanding evolutionary transitions, i.e., the appearance of new, more complex forms of biological organization in time? Since the beginning of the history of life on Earth, organisms have grouped together, constituting more or less cohesive aggregates that increase, under certain circumstances, the possibilities of survival of the individual systems.[19] Bacteria, for example, show a huge variety of this type of collective—and often just temporary or *ad hoc*—associations, based on intercellular self-organization processes that tend to expand their overall adaptive capacity. These communities show features, and occasionally seem to even behave, like multicellular organisms (O'Malley and Dupré 2007). This would be the case, for example, of bio-films (which build a common physical border, a polymeric matrix that keeps the cells together and attached to a surface), myxobacteria (body-like colonies that have developed their own "life cycles"), or the so-called "magnetotactic multicellular prokaryote" (a bacterial aggregate that exhibits an unusual "ping-pong" motility in magnetic fields; see Keim et al. 2004).

However, given the limited degree of functional differentiation within these collective associations, the cohesion and interactive capacity of the global system, as such, is rather weak (at least, weaker than the one displayed by its constitutive parts, the cells). Thus, a closer examination does not allow one to consider them as proper autonomous agents. Actually, the constitution of new composite forms of autonomy, endowed with their own agency, was a much more difficult process than the formation of more or less cohesive colonial aggregates. This is because the creation of a full-fledged autonomous entity is not possible without a stronger subordination of the constitutive elements to the new functional requirements of the emerging global autonomy. The issue is quite tricky, as O'Malley and Dupré (2007) demonstrate, and probably can be solved only in relative or comparative terms, since the constitutive elements of these new forms of autonomy are, themselves, living cells (with their own metabolic organization, etc.). Thus, there is an obvious tension between the autonomous logic of those units and the conditions that they must satisfy to constitute the new, potentially "higher-order" autonomy.[20]

When or how would it be possible to discern whether a group of autonomous cells is just gathering together to improve their overall fitness or becoming part of a higher-order autonomous entity? The key is in the comparative degree of cohesion and

[19] According to Bonner (2000), there is another reason why evolution tends to favor collective associations: because they produce an increase in size and the uppermost size niche is never filled. Therefore, there must have been some selection pressure for integration and coordination, even if it has been fruitful only when combined with the appearance of mechanisms that help to accommodate the newly created, larger system.

[20] In general, associations among autonomous systems can 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).

agency of the cell with regard to the collection of cells. In bio-films and other types of prokaryotic communities we can observe very interesting coordinated dynamics, but what we should consider is whether or not the *inter*cellular relations are more complex, functionally diversified and cohesive than *intra*cellular ones. For example, at the end of the last section we were highlighting the role of regulation (dynamic decoupling, etc.) in explaining the organization of all present-day cells, their agency and strong functional integration. Prokaryotic communities, however (as far as we know, at least), do not show those kinds of hierarchical or modular regulatory mechanisms, neither inwards (internally), nor outwards (in their relations with the environment). They are very complex, dynamically coordinated organizations, with great plasticity, since most of the changes that occur at the level of each cell are reversible. However, these associations seem to be fundamentally based on self-organizing principles that generate patterns of global coherence, emerging out the local interactions among cells and the development of their signaling or "communication" capacities.

In order to go further, both in functional integration (internal cohesion) and agency at the level of the global association, it might be necessary for the units, the cells, to give up a good part of their individual plasticity, becoming more irreversibly diverse and interdependent. Only from that point can the functional integration, agency and plasticity of the emerging level become greater than that of its constitutive units. Nevertheless, the tension between those two different levels of organization is there and it probably took very long to be solved. In fact, such difficulties were surely the reason why the first form of a composite autonomy, the eukaryotic cell, can still be considered as an intermediate or transition case, in the sense that its current organization is not itself constituted by truly autonomous cells.

4.1 Eukaryotic autonomy

According to the most widely accepted theories these days, eukaryotes appeared as a result of a long process of symbiotic relationships among prokaryotic cells, probably of different types (Margulis 1991; Margulis and Sagan 2002; López-García and Moreira 2004), leading to an altogether new class of individuals, in which the former autonomous prokaryotic cells became semi-autonomous organelles.[21] Without getting into the details, this is a clear example of an evolutionary transition that cannot be explained just by the gradual accumulation of small random mutations: Other kinds of mechanisms must operate in such a revolutionary change. Watson and Pollack (2003) have suggested, along those lines, that the composition of pre-adapted extant entities into a new whole should be an important, alternative source of phenotypic variation, with interesting consequences for their *evolvability*[22]. Very significantly, this new,

[21] These organelles (mitochondria, chloroplasts, nuclei, etc.) have lost irreversibly not only their capacity to live independently from one another, but also their ancient metabolic organization, even though they still keep traces of their former single-cell nature.

[22] It is now quite unanimously accepted that the mechanisms of evolution (especially as far as phenotypic variation or plasticity is concerned, i.e., adaptability, generation of new functionalities, etc.) have themselves evolved (Conrad 1979; Wagner and Altenberg 1996; Kirschner and Gerhart 1998). The reason for these evolutionary changes seems to be the robustness and flexibility of the processes involved, which make them

highly integrated organization (thanks, in particular, to the nucleation of DNA) has the possibility for more elaborate regulation and processing of genetic information and, thereby, for a much more complex internal organization than in prokaryotes. As J. Mattick (2004) remarks:

> Because bacteria lack a nucleus, transcription and translation occur together: RNA is translated into protein almost as fast as it is transcribed from DNA. There is no time for intronic RNA to splice itself out of the protein coding RNA in which it sits, so an intron would in most cases disable the gene it inhabits, with harmful consequences for the host bacterium. In eukaryotes, transcription occurs in the nucleus and translation in the cytoplasm, a separation that opens a window of opportunity for the intron RNA to excise itself. Introns can thus be more easily tolerated in eukaryotes. (Mattick 2004, p. 63).

In other words, the *decoupling* between transcription and translation permitted a much higher level of genetic regulatory control, which, in turn, would be required to increase the organizational complexity and plasticity of the whole cell (Taft et al. 2007). Actually, RNA has a clear advantage over proteins for transmitting information and regulating activities that involve the genome itself: RNAs can encode for short, sequence-specific signals as a kind of bit string or zip code. These embedded codes can direct RNA molecules precisely to receptive targets in other RNAs and DNA. The RNA-RNA and RNA-DNA interactions could in turn create structures that recruit proteins to convert the signals to actions. But it is just in the last few years that we are becoming aware of the relevance and potential of these post-transcriptional regulatory mechanisms, thanks to the advances in comparative genomics and the realization that non-coding DNA (or "junk DNA," as it was initially called) is pivotal precisely for understanding eukaryotic plasticity and later evolutionary transitions.

Thus, unlike colonies of prokaryotic cells, eukaryotes were able to put together, within their new autonomous organization, the necessary ingredients to generate a radically new (and richer) level of functional and morphological diversity. This involved other internal innovations (e.g., development of the cytoskeleton) but also more sophisticated mechanisms of interaction with the environment and other similar cells (e.g., new intercellular communication strategies). Only this richness in functional and agent capacities can explain why eukaryotic cells can constitute the building blocks for even more complex and integrated forms of associative units: multicellular organisms.

4.2 Multicellular autonomy

Multicellular organisms are the next step in the evolutionary generation of new forms of autonomy. But in what sense is a multicellular organism different from (a) associations of individual cells or (b) the eukaryotic cells that we have just analyzed?

Footnote 22 continued
particularly suitable for complex development and physiology. According to Nehaniv (2003), properties such as the facilitation of extradimensional bypass and robustness to genetic variability (Conrad 1990), heritability of fitness (Michod and Roze 1999), modularity, and robustness to developmental variation (Kirschner and Gerhart 1998) play important roles in this respect.

Regarding (a), it is true that certain forms of eukaryotic multicellular individuality are not qualitatively different from many prokaryotic multicellular associations; nevertheless, eukaryotic multicellular organisms can generate new forms of autonomous individuality, which show an incomparable degree of integration and complexity. As for (b), unlike eukaryotic unicellular organisms, multicellular organisms generate a form of individuality in which their constitutive parts remain, themselves, autonomous (in a significant, metabolic sense).

Admittedly, many forms of multicellular organisms show a relatively weak integrative cohesion, and their agent-interactive capacities do not involve great innovations with regard to those of their constitutive parts. This is the case not only with early multicellular metazoans, but also with most plants and fungi. In these cases it is not easy to define what the term 'organism' means. As Sterelny and Griffiths have pointed out

> "the organism" turns out to be a highly contestable notion. ... If there is a common-sense view of the organism, it is the idea that organisms are complex, coadapted, and physically integrated. They have differentiated parts. They are physically cohesive, with an inside and an outside. Since many metabolic processes depend on the existence of this inside/outside distinction, organisms are often equipped with homeostatic mechanisms to ensure that the inside remains stable despite variation outside. One major problem with this definition is that it fits plants badly. (Stelreny and Griffiths 1999, p. 173)

Then, the decision to consider many of these multicellular systems as full-fledged individual organisms is a matter of degree rather than of clear conceptual differences. The difficulties lie precisely in the case of less complex forms of multicellular integrated systems. As we have already argued, a possible solution is to compare the degree of cohesion and agency of the new integrated unit with that of its constitutive cells. If the emergent, global entity shows a more complex degree of functional diversification and cohesion at the level of its *inter*cellular relations than at the level of its *intra*cellular ones, and the interactive agential processes of the whole also become more complex and integrated than those happening at the level of the parts, it is legitimate to consider it as a full fledged (multicellular) organism.

However, in these highly integrated multicellular organisms, the more clearly autonomous their collective behavior becomes, the more difficult it is to consider also as autonomous their constitutive units (namely, their unicellular parts), because they are, indeed, very heavily constrained. In a multicellular organism cells need to adapt their autonomy[23] to serve that of the entire organism. Therefore, they become strongly dependent on other cells to obtain the material and energetic resources required to carry out their own metabolism. In these conditions, the main reason to see the constitutive cells of a multicellular organism still as autonomous entities is, precisely, because they do not lose their fundamental metabolic organization. Furthermore, their aptitude to participate in a highly complex global organization lies in their potential to achieve a large diversity of highly specific (but, at the same time, versatile and adaptive)

[23] This occurs through irreversible differentiation processes that make them apt to live only in a very specific environment, tightly surrounded by other cells.

functions. And these specialized-versatile functions require in turn a type of entity, which, from the point of view of internal organization, remains autonomous, even if it becomes capable of living only in a collectively built and continuously maintained environment. This is why Maturana and Varela (1992) characterized multicellular organisms as "second-order autopoietic systems."

The transition to multicellular organisms is very interesting and intriguing because it is a process in which, again, bigger and more complex individuals[24] are generated. Regarding gene regulation networks in the system, important innovations had to take place, on top of the natural expansion of RNA editing strategies. In particular, new regulatory controls must be implemented, so that the expression of the diverse genes occurs in the right place, and at the right time, during the process of differentiation (or proper development). The role of the so-called "homeotic genes" has been recognized since long ago in that respect. But now their true significance has to be reassessed, or reinterpreted, in conjunction with discoveries about other high-order regulators (e.g., "microRNAs"), which have been proven to control, as well, the timing of processes that occur during development (like stem cell maintenance, cell proliferation or apoptosis; see Mattick 2004). In any case, in parallel to these new genetic control mechanisms, many other structural and organizational innovations must have happened (in terms of intercellular communication and adhesion devices, sexual reproduction and germ-soma distinction and integration, etc.). Such a complex transition, of course, could not occur out of the blue, and many different attempts must have been made by the developing life on Earth before a solution was found. As a consequence, on the way to the emergence of full-fledged multicellular organisms numerous intermediate types of collective integration surely took shape, some of which have survived and still appear in nature today.[25]

From an organizational perspective, the process of the cells' progressive integration required that what previously were interactive processes among autonomous systems be transformed into internal or constitutive processes of a more encompassing system. At the same time, completely new agent-interactive capacities had to be developed at that emerging global-collective scale, continuously feeding back (and forward again) with the environment, to ensure the viability of the new autonomous entity. The great diversification potential of eukaryotes was surely crucial for that. As a reminder of the

[24] The assumption is that organisms—unlike other categories—are or tend to be individuals, in the sense that they are unique, not only genetically speaking but also as a result of their ontogenetic history of interactions (Ruiz-Mirazo et al. 2000).

[25] Multicellular organisms with differentiated cells have evolved independently on at least three occasions: animals, higher plants and fungi, thus conforming to different developmental schemes. The most ancient forms of multicellular individuality were reversible systems, like *Physarum*, or the "organism" called *Dyctiostelium discoideum*, an association of unicellular amoeba that show a life cycle with processes of cellular differentiation and germ-soma separation, but also disaggregate and live independently as soon as environmental conditions are favorable (i.e., enough food is available). A subsequent step would be individuals whose constitutive parts can no longer disintegrate and continue living independently. An extremely interesting case, as it represents an obvious transition step, is *Physalia physalis*, more commonly know as the Portuguese man o' war (Gould 1985). *Physalia physalis* is an individual composed of four kinds of specialized polyps and medusoids with different genomes, that have gone through a joint evolutionary process, as a result of which a manifest, irreversible distribution of tasks was established (digestion, prey detection, defense and navigation).

importance of the interactive dimension of the process, it is quite appropriate to quote Hooker (2009), who contrasts his position with the autopoietic way of thinking:

> The emergence of multicellular organisms represents a massive expansion in both interactive capacity and eventually self-regulation of that capacity and in this lies their rich adaptabilities that make them so successful. The focus of understanding such evolutionary functional change should thus be, not on finding repeated levels of the same stringency of closed cellular autopoietic organization, but instead on the effective mastery of increased interactive openness. That mastery is achieved through increased self-regulatory capacity to modify, in situation-dependent ways, both the internal metabolic and external environmental cycles. (Hooker 2009, pp. 521–522)

In addition, from an evolutionary perspective, the emergence of multicellular organisms had to solve many conflicts. On the one hand, as Bonner (2000, p. 50) remarks, multicellular organisms must solve two opposite selective pressures, since "natural selection is simultaneously pushing for a large stage in the life cycle that can compete for food and for a minute single-cell stage that is essential for sex reproduction (and often asexual as well). The result is that all multicellular organisms (…) have a unicellular stage and a larger stage of varying dimensions in their life cycle." On the other hand, as Buss (1987) and Michod (1999) have argued, much of the evolution of development in metazoans can be explained as trade-off solutions of the conflict between selective pressures acting at the level of the cell and those acting at the level of the multicellular individual.[26] When the survival of a certain unit depends so much on other cooperating units that the dynamics of selection scales up to the next level (i.e., a new unit of selection appears) it is because the interconnections of the lower-level units are very strong and the system-environment interactive dimension, where selective forces actually operate, is transferred to the integrated whole (Wimsatt 1980). In other words, in multicellular organisms higher-level, global fitness is significantly decoupled from the fitness of the constitutive cells.

4.3 Toward complex multicellular autonomy

Finally, in metazoan evolution we can envision also the appearance of new, highly integrated systems as modifications or redefinitions of an autonomous organization. These increasingly complex forms of autonomy are based on deep changes at the developmental, body-plan level. In turn, developmental plasticity is possible, among other things, because the regulatory possibilities offered by homeotic genes combined with RNA editing processes expand enormously at these stages. There is evidence that the "microRNA repertoire" continues enlarging during metazoan evolution (with very clear signs at the transitions to bilaterians, vertebrates and placental mammals; see Hertel et al. 2006). In relative terms, the part of the genome that is responsible for these regulatory and epigenetic mechanisms keeps growing in importance, whereas

[26] This perspective can shed new light on phenomena as diverse as cancer, gastrulation and germ line sequestration.

the part responsible for core metabolic functions remains basically the same. The most recent surprise in that sense is that vertebrate genomes contain thousands of noncoding sequences that have persisted virtually unaltered for many millions of years. And these sequences are much more highly conserved than those coding for proteins, which was totally unexpected.

But what is particularly interesting here is how this potential for phenotypic diversification can be channeled to redefine radically both the constitutive and the interactive organization of the multicellular individuals, generating higher forms of complex integration. For example, this is the case when, in the development of some metazoans, a new kind of cell (the neuron) started to differentiate. Neurons differentiated as cells capable of forming branches, interconnected through plastic electrochemical pathways and capable of propagating and modulating electric potential variability. In fact, these interconnected cells led to the establishment (about 600 million years ago) of a dynamic network capable of managing an efficient coordination between sensor and motor/effector structures in multicellular organisms.

Since the very beginning of its evolution, neural organization appeared as an extended network capable of producing a *recurrent* dynamics of specific patterns independent of the underlying metabolic transformations that the organism undergoes. Unlike chemical signals circulating within the body, which directly interact with metabolic processes due to their diffusive nature, the electrochemical interactions between neurons make open-ended recurrent interactions within the very nervous system possible. What made neural interconnections so special is that they created an incredibly rich and plastic internal world of patterns of fast connections, *dynamically decoupled* from the metabolic processes (Moreno and Lasa 2003). Thanks to this decoupling, that new dynamical domain could be recruited for highly complex regulatory functions, concerning both interactive processes (behavior) and metabolic processes (internal regulation). Furthermore, the evolution of the nervous system shows an increase in regulatory controls. Indeed, as nervous systems evolve, multiple modulation relationships across a wider range of temporal and spatial scales become manifest (Bickhard 2009).

Interestingly, the increase in hierarchical regulatory controls is not in opposition with the opening of richer domains of variability, which in turn is a fundamental source for evolvability (Nehaniv 2003; Sniegowski and Murphy 2006). As Kirschner and Gerhart (1998) describe, there seems to be a process of progressive "deconstraining," in the sense that all these regulatory controls on core functions somehow loosen their originally much tighter metabolic links, conferring higher plasticity and evolvability on the system. So the impressive morphological and physiological diversification of metazoan lineages is based, again, on the evolution of various regulatory processes controlling the time, place, and conditions of use of the conserved core processes, which have modified their capacity to produce heritable phenotypic variation.[27]

[27] These regulatory controls, and certain core processes, have special properties relevant to evolutionary change. The properties of versatile protein elements, weak linkage, compartmentation, redundancy, and exploratory behavior reduce the interdependence of components and confer robustness and flexibility on processes during embryonic development and in adult physiology. They also confer evolvability to the organism by reducing constraints on change and allowing the accumulation of non-lethal variation (Kirschner and Gerhart 1998).

For example, the appearance of the vertebrate body plan might well be linked to an increase in developmental degrees of freedom, attributed to the appearance of the neural crest (Moss 2006): a set of cells that *detach* and migrate freely from the neural tube during development, giving rise to the central and autonomic nervous system, hormone-producing cells and derivatives, bones and cartilage, etc.

> The neural crest is the (…) sine qua non of the possibility space that opens up with the transition to vertebrate life (…) As the neural plate folds in the embryo (…) a radically new developmental event in the evolutionary history of multicellular life-forms takes place. The cells of the neural crest systematically break away from the tissue-sheet pavement of a tightly connected epithelium, transforming into *detached*, individuated, stromal-type, migratory cells. As individual cells finding their way through the interior space of the embryo, guided by an ongoing dialogue of signal exchange between extracellular matrix and cell surface, they take up residence in, and contribute to, the further development of every tissue and every organ of the body (…) [The neural crest] entails (…) a capacity for cells to step back from the manifold of ambient stimulus and to be prepared to pick and choose which stimulus to make salient and thus in so doing *a capacity to enjoy an unprecedented level of internal autonomy,* it entails an ability to be receptive to multiple modalities of stimulus and to make nuanced distinctions between them, and it entails an ability to transmit and transduce sensory signals so as to elicit rapid as well as protracted adjustments of its internal state of the respective cells. The kind of enterprise that vertebrate existence brought into being was one characterized by two-way dialogues between the sensory surface of the organism and its ambient surround and between the sensory surface of the organism and its viscera, its interior. It is the advent of the neural crest that makes this transition to new levels of contingent responsiveness possible. (Moss 2006, pp. 932–934)

So evolution shows that the appearance of increasingly complex systems goes together with the invention of different ways of hierarchical regulation to manage internal functional variety and develop agent plasticity. What really matters in biological evolution, as Mattick (2004) already highlighted, is not so much the generation of complexity, but its functional and selective control.

4.4 The nature of complex biological autonomy

From the perspective of autonomy that we present in this paper, evolution can be seen as a process in which, starting from basic autonomous systems as the main building blocks, reorganization processes (in the spheres of both internal-metabolic cycles and interactive loops with the environment) allow for strong cooperative types of behavior. Some of these new types of behavior, provided that differentiation and coordination of the building blocks are developed in a balanced and robust way, can lead to a novel, globally integrated autonomous systems. By incorporating new hierarchies of dynamically detached domains and regulatory controls this trend can proceed even further, producing new forms of autonomy capable of creating or taking over more complex, flexible and diversified functional interactions with the environment.

This is what makes a big difference with respect to colonies, societies, and even primitive multicellular organisms, whose cohesion relies more on self-organization than on specific regulatory control mechanisms. Whereas the increase in complexity of associations of self-organized, distributed systems hits apparent ceilings or bottlenecks, regulatory control development seems to allow for an open-ended increase in the complexity of autonomous organizations.

Starting from forms of collective associations where the constitutive, autonomous units are more integrated and cohesive than the collectivity, evolutionary transitions show the appearance of increasingly integrated systems, leading to new forms of autonomous agents. The organization of those agents, then, becomes much more complex, functionally diversified and cohesive than that of their constitutive units. This evolutionary trend has two types of important consequences: one is related with the way constitutive processes of complex organisms become progressively autonomous from the environmental conditions; the other concerns the interactive processes, opening new domains of autonomous identity.

Rosslenbroich (2005, 2006, 2009) has argued that the most innovative evolutionary transitions constitute a process of "autonomization," which he understands in the sense that the direct influences of the environment are gradually reduced and a stabilization of self-referential, intrinsic functions within the system is generated. For instance, the extracellular matrix common in the development of metazoan organisms, which appeared relatively early in their evolution, makes possible intracellular conditions to control and protect internal cells from the external environment. This increase in independence is relative, because it is built at the same time as "many interconnections with and dependencies upon" the environment are retained, and perhaps redefined. Some of the ingredients for autonomization, according to Rosslenbroich, are: spatial separation from the environment (membrane, walls, integuments etc.), establishment of homeostatic functions, and internalization of morphological structures or functions from an external position. Gerhart and Kirschner (1997) have also argued that one of the advantages of multicellular organisms is that they can more effectively protect themselves from environmental changes by producing their own internal environment.

But the consequences of the appearance of more complex forms of autonomy are even more important at the interactive level, as Hooker (2009) and other authors (Christensen and Bickhard 2002) have emphasized. From that perspective autonomy is not merely the development of intrinsic, normative functions, but more importantly of new domains of interactive capacities. This is confirmed by our study of systems with higher degrees of autonomy, which certainly show an increase in the capacity to create or take over complex and larger environmental interactions (namely, new forms of agency): e.g., flexible behavior, anticipatory strategies or self-directed anticipatory learning (Christensen and Hooker 2000). Ultimately, this leads to the creation of new domains of autonomous agency, such as cognitive or mental autonomy (Barandiaran and Moreno 2006; Barandiaran 2008).

In sum, this analysis of several of the major evolutionary transitions in the history of life on Earth reinforces the key role that the concept of autonomy must play in the characterization of the living, from the most elementary (or even preliminary) forms of it to the most complex. Thus, if any general theory of biological systems ever gets developed, according to our view, autonomy ought to be one of its central axes, if not *the* central one.

5 Conclusion

In conclusion, if we understand the phenomenon of life as a complex network of processes that take shape and propagate both at an individual and a historical-collective dimension—in our terms, as the history of the proliferation of various forms of autonomous organization (from chemical to unicellular, multicellular, developmental, cognitive, and, only recently, to rational autonomy)—the radical Cartesian separation between nature and mind simply disappears. The capacity of a system to determine itself, to establish its own rules and norms of behavior, and to create meaningful environments no longer belongs exclusively to the realm of rationality. At the same time, the natural world cannot be regarded as a universe where blind forces, acting without any sense or purpose, operate: the study of the fundamental mechanisms underlying biological organization, with all their intricacies, has clearly refuted that possibility.

Instead, the overall picture coming out of a careful, interdisciplinary analysis of diverse biological systems with diverse levels of complexity is that they self-constrain or self-regulate in increasingly sophisticated ways. In other words, as living beings originate and evolve, they create supra- and macromolecular material mechanisms that act as additional "local rules." These rules or constraints affect the underlying molecular processes (in ways that are compatible, of course, with the general interaction rules of physical laws), and take them to *cyclic*, far-from-equilibrium, dynamic states where there are still open degrees of freedom to be explored. Or, perhaps more accurately, in their own self-constraining, these systems, through the new rules, are able to *generate* completely new degrees of freedom, new possibilities of dynamic behavior, which were inexistent until that moment. In any case, according to this picture, if matter finds the right conditions, like those on Earth a few thousand million years ago, and later on, in the evolution of the planet, it can locally organize and reorganize itself many times, giving as a result emerging patterns of dynamic-cyclic order and complex behavior.

Therefore, we consider that reorganization processes are the basis of all major evolutionary transitions. Starting from "basic autonomy," which can be considered as a reorganization of natural "self-organization" phenomena, all subsequent forms of autonomy derive from deep structural rearrangements of the way the system operates internally or in relation to its environment (or both). In that sense, perhaps the most important idea we want to transmit in this paper is the role of regulatory controls and the implications that these have in terms of the modular, hierarchical relationships to be established in the system. In particular, we highlighted a rather general principle underneath those transitions, allowing for the open-ended reorganization of autonomous systems: the relative *dynamic decoupling* that distinct parts, modules or modes of operation in the system must show with regard to one another. As a result of many subsequent transitions, the decoupling between the last "upper-level" or regulatory set of mechanisms and the underlying "lower-level" dynamics can appear very strong (making understandable, to a certain extent, a Cartesian type of dualistic view). Deeper analyses, however, should eventually reveal the intricacies of the hierarchical control relationships that sustain and have made the whole process possible.

This picture is, therefore, very different from the inherited view, based on the general theoretical framework of physics with its main weight on fundamental interaction laws. Philosophy of biology and particularly work about the nature of life, in light of the new discoveries and theories bound to appear in the near future to explain living phenomena and their origin, should contribute to doing away with that physicalist tradition in general philosophy and push, with renewed force, the naturalization program. Humankind will become increasingly aware that the unfolding of ever more complex forms of autonomy, as it has occurred on Earth, brings with it the origin and evolution of functionality, symbols and information, agency, and, thereby, normativity, teleology, cognition, intentionality and, eventually, consciousness and morality. Thus, in a nutshell, the understanding of life (which, even in its minimal expression, will contain an irreducible element of autonomy) will reveal itself, philosophically speaking, to be of greater and greater importance.

On similar but parallel lines, another legacy of Cartesian dualism was a dichotomy established in the sphere of values. In the past, the life of all species other than humans was considered to have a purely utilitarian value.[28] In this sense, the significance of the understanding of life, of how life has evolved and keeps evolving, has another, no less important, aspect. Indeed, we can already perceive that human behavior starts to be more clearly linked to a value system that provides different living beings with a gradual scale of value weights. This significant—though, globally speaking, still preliminary—change derives from the new evolutionary worldview on the phenomenon of life, which makes obvious the incongruence of drawing a strict line of separation between the value of human life and that of other living species. Thus, biological diversity is becoming a value in itself, not only in utilitarian, practical terms, but also as a consequence of understanding the profound interdependence among all forms of life, particularly the more complex ones.

The underlying and more acute problem is whether the human species is going to be able to modify or get rid of its traditional value system, to view and locate itself in a world of axiologically interdependent beings, and lead an ecologically sustainable way of life. We know that the achievement of this goal is an enormously difficult enterprise, which involves in practice all fields of knowledge, technology and political and economical decision-making; but what is sure is that a better understanding of the nature of life is also important for that task.

[28] Let us recall Spinoza, again, to illustrate this traditional view: "Save men we do not know any individual thing in nature in whose mind we may rejoice or which we may join to us in bonds of friendship or any other kind of habit: and therefore whatever exists in nature besides man, reason does not postulate that we should preserve for our advantage, but teaches us that we should preserve or destroy it according to our various needs, or adapt it in any manner we please to suit ourselves" (Spinoza, *Ethics*, part IV, chapter-appendix XXVI).

Acknowledgements Funding for this work was provided by grant IT 505-10 from the Basque Governement, grants, FFI2008-06348-C02-01 and FFU2009-12895-CO2-O2 from the Spanish Ministry of Science and Education (MICINN). Kepa Ruiz-Mirazo holds a *Ramón y Cajal* Research Fellowship. Finally, we would like to acknowledge the review and editorial work of Mark Bedau and Emily Parke on the original text, as well as the suggestions for improvement provided by Cristian Saborido, who read several parts of the first manuscript.

References

Andrianantoandro, E., Basu, S., Karig, D., & Weiss, R. (2006). Synthetic biology: New engineering rules for an emerging discipline. *Molecular Systems Biology*. doi:10.1038/msb4100073.

Barandiaran, X. (2008). *Mental life: A naturalized approach to the autonomy of cognitive agents*. PhD Dissertation, University of the Basque Country (UPV/EHU).

Barandiaran, X., & Moreno, A. (2006). On what makes certain dynamical systems cognitive. *Adaptive Behavior, 14*(2), 171–185.

Barandiaran, X., & Moreno, A. (2008). Adaptivity: From metabolism to behavior. *Adaptive Behavior, 16*(5), 325–344.

Barandiaran, X., Rohde, M., & Di Paolo, E. (2009). Defining agency: Individuality, normativity, asymmetry and spatiotemporality in action. *Journal of Adaptive Behavior, 17*(5), 367–386.

Bedau, M. (1996). The nature of life. In M. Boden (Ed.), *The philosophy of artificial life* (pp. 332–357). New York: Oxford University Press.

Bedau, M. (1998). Four puzzles about life. *Artificial Life, 4*, 125–140.

Bedau, M. (2003). Artificial life: Organization, adaptation, and complexity from the bottom up. *Trends in Cognitive Science, 7*, 505–512.

Bechtel, W. (2006). *Discovering cell mechanisms: The creation of modern cell biology*. Cambridge: Cambridge University Press.

Bechtel, W. (2007). Biological mechanisms: Organized to maintain autonomy. In F. Boogerd, F. Bruggeman, J. H. Hofmeyr, & H. V. Westerhoff (Eds.), *Systems biology: Philosophical Foundations* (pp. 269–302). Amsterdam: Elsevier.

Benner, S. A., & Sismour, A. M. (2005). Synthetic biology. *Nature Reviews Genetics, 6*, 533–543.

Bickhard, M. H. (2000). Autonomy, function, and representation. *Communication and Cognition: Artificial Intelligence, 17*(3–4), 111–131.

Bickhard, M. H. (2009). The biological foundations of cognitive science. *New Ideas in Psychology, 27*, 75–84.

Boden, M. A. (1999). Is metabolism necessary? *British Journal for the Philosophy of Science, 50*, 231–248.

Bonner, J. T. (2000). *First signals: The evolution of multicellular development*. Princeton: Princeton University Press.

Buss, L. (1987). *The evolution of individuality*. Princeton: Princeton University Press.

Calvo, P., & Gomila, T. (Eds.). (2008). *Handbook of cognitive science: An embodied approach*. Amsterdam: Elsevier.

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

Christensen, W. D., & Hooker, C. A. (2000). An interactivist-constructivist approach to intelligence: Self-directed anticipative learning. *Philosophical Psychology, 13*, 5–45.

Clark, A. (1997). *Being there: Putting brain, body and world together again*. Cambridge, MA: MIT Press.

Cleland, C. E. (2011). Life without definitions. *Synthese*. doi:10.1007/s11229-011-9879-7.

Cleland, C. E., & Chyba, C. F. (2002). Defining 'life'. *Origins of Life, 32*, 387–393.

Cleland, C. E., & Chyba, C. F. (2007). Does 'life' have a definition? In W. T. Sullivan, III & J. A. Baross (Eds.), *Planets and life: The emerging science of astrobiology*. Cambridge: Cambridge University Press.

Conrad, M. (1979). Bootstrapping on the adaptive landscape. *BioSystems, 11*, 167–182.

Conrad, M. (1990). The geometry of evolution. *BioSystems, 24*, 61–81.

Cornish-Bowden, A., Cárdenas, M. L., Letelier, J.C., & Soto-Andrade, J. (2007). Beyond reductionism: metabolic circularity as a guiding vision for a real biology of system. *Proteomics 7*, 839–845.

Craver, C. F. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of Science, 68*, 53–74.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy, 72*, 741–765 (Reprinted in D. J. Buller (Ed.). (1999). *Function, selection, and design* (pp. 57–83). Albany, NY: SUNY Press.).

Davies, P. S. (2001). *Norms of nature: Naturalism and the nature of functions*. Cambridge, MA: MIT Press.

Dawkins, R. (1976). *The selfish gene*. New York: Oxford University Press.

De Duve, C. (1991). *Blueprint for a cell: The nature and origin of life*. Burlington, NC: Neil Patterson Publishers.

De Duve, C. (2002). *Life evolving*. New York: Oxford University Press.

Delancey, C. (2006). Ontology and teleofunctions: A defense and revision of the systematic account of teleological explanation. *Synthese, 150*, 69–98.

Dennett, D. (1995). *Darwin's dangerous idea*. New York, NY: Simon and Schuster.

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

Dupré, J., & O'Malley, M. A. (2009). Varieties of living things: Life at the intersection of lineage and metabolism. *Philosophy and Theory in Biology*, 1 (e003), 1–25.

Edin, B. (2008). Assigning biological functions: Making sense of causal chains. *Synthese, 161*, 203–218.

Eigen, M., & Schuster, P. (1979). *The hypercycle: A principle of natural self-organization*. Berlin: Springer.

Elsasser, W. M. (1966). *Atom and organism*. Princeton, NJ: Princeton University Press.

Emmeche, C., & Hoffmeyer, J. (1991). From language to nature: The semiotic metaphor in biology. *Semiotica, 84*(1/2), 1–42.

Endy, D. (2005). Foundations for engineering biology. *Nature, 438*, 449–453.

Gánti, T. (1971). *The principle of life* (1st ed.). Budapest: Gondolat.

Gerhart, J., & Kirschner, M. (1997). *Cells, embryos, and evolution*. Malden, MA: Blackwell Science. (in Hungarian).

Gibas, R. W. (2005). *Embodiment and cognitive science*. Cambridge: Cambridge University Press.

Godfrey-Smith, P. (1994). A modern history theory of functions. *Noûs, 28*, 344–362.

Gould, S. J. (1985). *The flamingo's smile* (Chapt. 1.5). New York: W. W. Norton.

Griesemer, J., & Szathmáry, E. (2009). Ganti's chemoton model and life criteria. In S. Rasmussen, M. A. Bedau, L. Chen, D. Deamer, D. C. Krakauer, N. H. Packard (Eds.), *Protocells: Bridging nonliving and living matter* (pp. 481–512). Cambridge: MIT Press.

Haldane, J. B. S. (1994/1929). The origin of life. In D. W. Deamer & G. R. Fleischaker (Eds.), *Origins of life: The central concepts* (pp. 73–81). Boston: Jones and Barlett.

Haraway, D. (1976). *Crystals, fabrics, and fields*. Baltimore: Johns Hopkins University Press.

Harold, F. (1986). *The vital force: A study of bioenergetics*. New York: Freeman.

Harold, H. (2001). *The way of the cell*. Oxford: Oxford University Press.

Hertel, J., Lindemeyer, M., Missal, K., Fried, C., Tanzer, A., Flamm, C., Hofacker, I. L., & Stadler, P. F. (2006). The expansion of the metazoan microRNA repertoire. *BMC Genomics, 7*(1), 25.

Hoffmeyer, J. (1996). *Signs of meaning in the universe: The natural history of signification*. Bloomington, IN: Indiana University Press.

Hooker, C. A. (1995). *Reason, regulation and realism: Toward a regulatory systems theory of reason and evolutionary epistemology*. Albany: State University of New York Press.

Hooker, C. (2009). Interaction and bio-cognitive order. *Synthese, 166*, 513–546.

Jonas, H. (1966). *The phenomenon of life: Toward a philosophical biology*. New York: Harper and Row.

Kant, I. (1790/1952). *Critique of judgment*. Oxford: Oxford University Press.

Kauffman, S. (2000). *Investigations*. Oxford: Oxford University Press.

Kauffman, S. (2003). Molecular autonomous agents. *Philosophical Transactions of the Royal Society of London A, 361*, 1089–1099.

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

Keller, E. F., et al. (2008). What is wrong with the question, What is Life?. In P. Marrati (Ed.), *Concepts of life*. Stanford: Stanford University Press.

Kirschner, M., & Gerhart, J. (1998). Evolvability. *PNAS, 95*(15), 8420–8427.

Kitano, H. (2002). Computational systems biology. *Nature, 420*, 206–210.

Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics, 1*, 1–18.

López-García, P., & Moreira, D. (2004). The synthrophy hypothesis for the origin of eukaryotes. In J. Seckbach (Ed.), *Symbiosis: Mechanisms and model systems* (pp. 133–147). Dordrecht, Boston, London: Kluwer Academia Publishers.

Lyon, P. (2006). The biogenic approach to cognition. *Cognitive Processes, 7*, 11–29.

Lyon, P., & Keijzer, F., et al. (2007). The human stain: Why cognitivism can't tell us what cognition is and what it does. In B. Wallace (Ed.), *The mind, the body and the world: Psychology after cognitivism?* (pp. 132–165). London: Imprint Academic.

Mansy, S., et al. (2008). Template directed synthesis of a genetic polymer in a model protocell. *Nature, 454*, 122–126.

Margulis, L. (1991). *Symbiosis as a source of evolutionary innovation: Speciation and morphogenesis*. Cambridge, London: MIT Press.

Margulis, L., & Sagan, D. (2002). *Acquiring genomes: A theory of the origins of species*. New York: Basic Books.

Mattick, J. (2004). The hidden genetic program of complex organisms. *Scientific American, 291*(4), 60–67.

Maturana, H., & Varela, F. J. (1973). *De máquinas y seres vivos: Una teoría sobre la organización biológica*. Santiago de Chile: Editorial Universitaria S.A.

Maturana, H., & Varela, F. (1992). *The tree of knowledge*. Boston: Shambala.

Maynard Smith, J. (1986). *The problems of biology*. Oxford: Oxford University Press.

Mayr, E. (1982). *The growth of biological thought*. Cambridge, MA: Harvard University Press.

McLaughlin, P. (2001). *What functions explain: Functional explanation and self-reproducing systems*. Cambridge: Cambridge University Press.

Michod, R. E. (1999). *Darwinian dynamics: Evolutionary transitions in fitness and individuality*. Princeton, NJ: Princeton University Press.

Michod, R. E., & Roze, D. (1999). Cooperation and conflict in the evolution of individuality. Part III. Transitions in the unit of fitness. In C. L. Nehaniv (Ed.), *Mathematical and computational biology: Digital evolution, hierachical complexity, and computational morphogenesis. American Mathematical Society Series: Lectures on Mathematics in the Life Sciences* (Vol. 26, pp. 47–91).

Miller, S. L. (1953). A production of amino acids under possible primitive Earth conditions. *Science, 117*, 528–529.

Millikan, R. G. (1989). In defense of proper functions. *Philosophy of Science, 56*, 288–302.

Montoya, J. M., & Solé, R. V. (2002). Small world patterns in food webs. *Journal of Theoretical Biology, 214*, 405–412.

Moreno, A., & Etxeberria, A. (2005). Agency in natural and artificial systems. *Artificial Life, 11*(1–2), 161–176.

Moreno, A., Etxeberria, A., & Umerez, J. (2008). The autonomy of biological individuals and artificial models. *BioSystems, 91*(2), 309–319.

Moreno, A., & Lasa, A. (2003). From basic adaptivity to early mind: The origin and evolution of cognitive capacities. *Evolution and Cognition, 9*(1), 12–24.

Moreno, A., & Ruiz-Mirazo, K. (1999). Metabolism and the problem of its universalization. *BioSystems, 49*(1), 45–61.

Moreno, A., & Ruiz Mirazo, K. (2009). The problem of the emergence of functional diversity in prebiotic evolution. *Biology and Philosophy, 24*(5), 585–605.

Morowitz, H. J., Heinz, B., & Deamer, D. W. (1988). The chemical logic of a minimum protocell. *Origins of Life and Evolution of the Biosphere, 18*, 281–287.

Moss, L. (2006). Redundancy, plasticity, and detachment: The implications of comparative genomics for evolutionary thinking. *Philosophy of Science, 73*, 930–946.

Mossio, M., Saborido, C., & Moreno, A. (2009). An organizational account for biological functions. *British Journal for the Philosophy of Science, 60*(4), 813–841.

Nagel, E. (1977). Teleology revisited. *Journal of Philosophy, 74*, 261–301.

Neander, K. (1991). Function as selected effects: The conceptual analyst's defense. *Philosophy of Science, 58*, 168–184.

Nehaniv, C. L. (2003). Evolvability (editorial, special issue on evolvability, dedicated to the memory of Professor Michael Conrad). *BioSystems, 69*(2–3), 77–81.

Noireaux, V., & Libchaber, A. (2004). A vesicle bioreactor as a step toward an artificial cell assembly. *Proceedings of the National Academy of Sciences USA, 101*, 17669–17674.

Oliver, J. D., & Perry, R. S. (2006). Definitely life but not definitively. *Origins of Life and Evolution of the Biosphere, 36*, 515–521.

O'Malley, M. A., & Dupré, J. (2007). Size doesn't matter: Towards a more inclusive philosophy of biology. *Biology and Philosophy, 22*, 155–191.

Oparin, A. I. (1994/1928). The origin of life. In D. W. Deamer & G. R. Fleischaker (Eds.), *Origins of life: The central concepts* (pp. 31–71). Boston: Jones and Barlett.

Oró, J. (1961). Mechanism of synthesis of adenine from hydrogen cyanide under possible primitive earth conditions. *Nature, 191*, 1193–1194.

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

Pattee, H. H. (1977). Dynamic and linguistic modes of complex systems. *International Journal of General Systems, 3*, 259–266.

Pattee, H. H. (1982). Cell psychology: An evolutionary approach to the symbol-matter problem. *Cognition and Brain Theory, 5*(4), 325–341.

Polanyi, M. (1968). Life's irreducible structure. *Science, 160*, 1308–1312.

Rasmussen, S., Bedau, M. A., Chen, L., Deamer, D., Krakauer, D. C., Packard, N. H. (Eds.). (2008). *Protocells: Bridging nonliving and living matter*. Cambridge: MIT Press.

Ravesz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabassi, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *Science, 297*, 1551–1555.

Rosen, R. (1958). A relational theory of biological systems. *Bulletin of Mathematical Biophysics, 20*, 245–260.

Rosen, R. (1991). *Life itself: A comprehensive inquiry into the nature, origin and fabrication of life*. New York: Columbia University Press.

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

Rosslenbroich, B. (2006). The notion of progress in evolutionary biology. The unresolved problem and an empirical suggestion. *Biology and Philosophy, 21*, 41–70.

Rosslenbroich, B. (2009). The theory of increasing autonomy in evolution—a new proposal for understanding macroevolutionary innovations. *Biology and Philosophy, 24*, 623–644.

Ruiz-Mirazo, K., Etxeberria, A., Moreno, A., & Ibáñez, J. (2000). Organisms and their place in biology. *Theory in Biosciences, 119*, 43–67.

Ruiz-Mirazo, K., & Mavelli, F. (2008). On the way towards 'basic autonomous agents': Stochastic simulations of minimal lipid-peptide cells. *BioSystems, 91*, 374–387.

Ruiz-Mirazo, K., & Moreno, A. (2000). Searching for the roots of autonomy: The natural and artificial paradigms revisited. *CCAI (Special Issue on Autonomy), 17*(3–4), 209–228.

Ruiz-Mirazo, K., & Moreno, A. (2004). Basic autonomy as a fundamental step in the synthesis of life. *Artificial Life, 10*(3), 235–259.

Ruiz-Mirazo, K., Pereto, J., & Moreno, A. (2004). A universal definition of life: Autonomy and open-ended evolution. *Origins of Life and Evolution of the Biosphere, 34*, 323–346.

Ruiz-Mirazo, K., Peretó, J., & Moreno, A. (2010). Defining life or bringing biology to life. *Origins of Life and Evolution of the Biosphere 40*, 203–213.

Ruiz-Mirazo, K., Umerez, J., & Moreno, A. (2008). Enabling conditions for open-ended evolution. *Biology and Philosophy, 23*(1), 67–85.

Russell, S. J., & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Englewood Cliffs, NJ: Prentice Hall.

Schlosser, G. (1998). Self-re-production and functionality: A systems-theoretical approach to teleological explanation. *Synthese, 116*, 303–354.

Schwann, T. (1839). Mikroskopische Untersuchungen über die Übereinstimmung in der Struktur und dem Wachstum der Thiere und Pflanzen. Berlin [1847, Microscopical researches into the accordance in the structure and growth of animals and plants]. London: Sydenham Society.

Skulatchev, V. P. (1992). The laws of cell energetics. *European Journal of Biochemistry, 208*, 203–209.

Smithers, T. (1997). Autonomy in robots and other agents. *Brain and Cognition, 34*, 88–106.

Sniegowski, P. D., & Murphy, H. A. (2006). Evolvability. *Current Biology, 16*, R831–R834.

Solé, R. V., Munteanu, A., Rodriguez-Caso, C., & Macía, J. (2007). Synthetic protocell biology: From reproduction to computation. *Philosophical Transactions of the Royal Society of London B, 362*, 1727–1739.

Stelreny, K., & Griffiths, P. E. (1999). *Sex and death: An introduction to philosophy of biology*. Chicago: The University of Chicago Press.

Szathmary, E. (2006). The origin of replicators and reproducers. *Philosophical Transactions of the Royal Society of London B, 361*, 1761–1776.

Taft, R. J., Pheasant, M., & Mattick, J. S. (2007). The relationship between non-protein-coding DNA and eukaryotic complexity. *BioEssays, 29*, 288–297.

Thomson, E. (2007). *Mind in life*. Cambridge, MA: Harvard University Press.

Varela, F. J. (1979). *Principles of biological autonomy*. New York: North Holland.

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

Varela, F. J., Thomson, E., & Rosch, E. (1991). *The embodied mind*. Cambridge, MA: MIT Press.

von Uexküll, J. (1982/1940). The theory of meaning. *Semiotica, 42*(1), 25–87.

Wagner, G. P., & Altenberg, L. (1996). Complex adaptations and the evolution of evolvability. *Evolution, 50*(3), 967–976.

Watson, R. A., & Pollack, J. B. (2003). A computational model of symbiotic composition in evolutionary transitions. *Biosystems, 69*(2–3), 187–209.

Weber, A., & Varela, F. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences, 1*, 97–125.

Wicken, J. S. (1987). *Evolution, thermodynamics and information: Extending the Darwinian program*. Oxford: Oxford University Press.

Wimsatt, W. (1980) Reductionistic research strategies and their biases in the units of selection controversy. In T. Nickles (Ed.), *Scientific discovery: Case studies* (Vol. II, pp. 213–259). Dordrecht: Reidel.

Wright, L. (1973). Functions. *Philosophical Review, 82*, 139–168.