Adaptive **Behavior**

Adaptivity: From Metabolism to Behavior

Xabier Barandiaran, Alvaro Moreno
Department of Logic and Philosophy of Science, University of the Basque Country, Spain

In this article, we propose some fundamental requirements for the appearance of adaptivity. We argue that a basic metabolic organization, taken in its minimal sense, may provide the conceptual framework for naturalizing the origin of teleology and normative functionality as it appears in living systems. However, adaptivity also requires the emergence of a regulatory subsystem, which implies a certain form of dynamic decoupling within a globally integrated, autonomous system. Thus, we analyze several forms of minimal adaptivity, including the special case of motility. We go on to explain how an open-ended complexity growth of motility-based adaptive agency, namely, behavior, requires the appearance of the nervous system. Finally, we discuss some implications of these ideas for embodied robotics.

Keywords naturalist approach to normativity · autonomous systems · adaptivity · minimal agency · decoupling of the nervous system · definition of adaptive behavior

1 Introduction

Adaptivity is the capacity that certain systems possess to modify themselves in order to adjust to changes in the environment. As is apparent in many types of system, this capacity is obviously an essential property of living beings that exhibit not only an enormous functional plasticity at the phylogenetic scale, but also a rich repertoire of responses to environmental changes in somatic time. However, many ecological and social organizations are also considered adaptive. Last but not least, a certain number of artificial systems—from virtual computer networks to physical “autonomous” robots—are also designed to show adaptive capacities.

This generic idea of adaptivity obscures two important philosophical and scientific questions: the presupposed distinction between system and environment and the fact that the idea of “adjustment” (and consequently of adaptivity) implies an irreducible normative dimension.

The first problem we call the “problem of identity;” that is, from the set of possible and arbitrary separations between system and environment, which are the ones we choose and why. If we are to attribute adaptive capacities to a system, we must first specify who is going to be adjusted to what. In addition, when talking about adaptivity, we are not just referring to a mere structural adjustment (a liquid adjusts to a container) between a system and its environment, but to a functional one; this brings us to the second problem.

The second problem is the “problem of normativity.” Normativity refers to the value attribution that is given to a process or object (e.g., adaptive or maladaptive to an interaction or structure in an organism, true or false to a cognitive state or belief, beautiful or ugly to a work of art, etc.). Normativity challenges physicalist scientific approaches to the understanding of our world because it introduces a value asymmetry (good/bad, true/false, adapted/maladapted) in the description

Correspondence to: Xabier Barandiaran, Department of Logic and Philosophy of Science, University of the Basque Country, PO Box 1249, 20080 San Sebastian-Donostia, Spain.
E-mail: xabier@barandiaran.net
Tel.: +34 943015516; Fax: +34 943311056

(2008), Vol 16(5): 325–344.
DOI: 10.1177/1059712308093868

of nature. However, although alien to fundamental physics, normativity is an essential component of biology; whether a structure or interaction is adaptive or maladaptive for an organism is a value judgment that a scientist engaged in the analysis and synthesis of adaptive systems must make.

According to these considerations, adaptivity is a capacity that a system possesses to maintain its identity by compensating, through the use of certain internal mechanisms, those environmental changes that jeopardize its identity. However, a system that triggers functional interactions to compensate for dangerous environmental perturbations only because it is externally guided to do so (such as, for example, a teleoperated robot) will never be a truly adaptive system (as the adaptivity lies in the intelligent system that operates it). Thus, a truly adaptive capacity can only occur in systems that define their own goals and their own normative adjustment to the satisfaction of these goals. It is precisely in living systems where the processes triggered are not divorced from their internal organization; they are causally connected to its maintenance. Thus, living systems are true adaptive agents because they generate actions in order to achieve goals that they generate themselves.

Thus, a full understanding of adaptivity is strongly related to the naturalization of concepts such as goal, normativity, identity, and functionality. This obliges us to look at the basic organization of living systems. However, even the simplest organism is a highly complex system. If we try to understand what adaptivity is from a naturalistic perspective, we have to search for the minimal organization that might support autonomous, plastic interactions with its environment.[1] Explaining adaptivity requires us to specify under which basic or minimal organizational conditions it appears, given the best available law-like understanding of the universe and the biological constraints that are observed upon them. Although it is obvious that the appearance of adaptivity requires the existence of a certain level of complexity, we do not know which type of complexity is required for adaptive capacities. In other words, how do we draw the boundaries between adaptive and non-adaptive complex systems? If we are no longer to believe in rigid boundaries, what makes some systems more adaptive than others? In addition, what types of transition determine the increase in complexity of adaptive mechanisms? How is normativity expressed through these transitions?

A naturalistic account of adaptivity should address all these questions from two complementary perspectives: (a) an evolutionary perspective, which should account for the diachronic emergence of adaptivity (what types of evolutionary transition permit the appearance of adaptive capacities); (b) an organizational perspective, which should account for the synchronic emergence of adaptivity from the bottom up (how adaptivity is sustained and enabled by underlying, more fundamental, processes). At the same time, the answer should be grounded on the available scientific knowledge and be capable of providing productive feedback to science both at empirical–analytic and constructive–synthetic (biorobotics and simulation of adaptive behavior) levels.

The structure of this article is as follows. In Section 2, we introduce the notion of basic autonomous systems to characterize living beings (i.e., systems capable of defining and maintaining their own identity). Autonomy provides the naturalized grounding for normative functionality, teleology, agency, and other fundamental concepts for the understanding of adaptivity. In Section 3, we focus on the appearance of the nervous system within basic autonomous systems, the transformations it requires and induces in biological organization leading to the notion of a hierarchically decoupled dynamic domain in control of adaptive behavior. We conclude with some fundamental properties of adaptive behavior that have been closely studied in recent adaptive science: embodiment, situatedness, and emergent functionality.

2 Autonomy and Intrinsic Teleology

It seems intuitive that adaptivity implies the distinction and the specific relationship between a system doing something by itself—an agent—and an environment. This fundamental dichotomy is very often taken for granted and the characterization of adaptivity is reduced to the establishment of the type of relationship between an agent and its environment. However, if we are to proceed from the bottom up, not taking any distinction for granted, we must first ask what types of natural process can constitute a system that produces its own identity by separating itself as different from its environment and establishing some functional interactions with it.

2.1 Self-Maintaining and Far-From-Equilibrium Organization of Minimal Autonomous Systems

A brief examination of the types of process that surround us reveals that the universe has evolved by producing forms of order (such as rocks or solar systems) in some places, whilst in others matter shows no apparent cohesion whatsoever (such is the case of gases,[2] for instance). Ordered matter takes two different forms: in some cases, basic components appear lumped together constituting conservative static structures, and in others they constitute dissipative dynamic structures.

The first type refers to spatially ordered forms of assemblage of material subunits, where this order is temporally instantaneous, such as in rocks or crystals, or temporally unfolded, such as atoms or planetary systems. In both cases, the form exhibited is just an expression of the intrinsic nature of a set of components that interact under certain conditions and that will exist indefinitely once created (i.e., energetically all these systems are conservative).

The other form of order is dissipative, and appears in far-from-equilibrium (FFE) conditions. This form of order is found among so-called "dissipative structures" (Nicolis & Prigogine, 1977). A dissipative structure[3] is a set of interacting elements that generate a cohesive dynamic pattern in FFE conditions. Examples of this type of system are whirlpools, hurricanes, oscillatory chemical reactions, and living beings. What all these different systems have in common is the fact that a huge amount of microscopic elements adopt a global, macroscopic ordered pattern in the presence of a specific flow of matter and energy (i.e., under certain boundary conditions). Interestingly, their internal dynamic cohesion is not only a consequence of the material features of their components but also of the achievement and maintenance (in FFE conditions) of some type of circular causality, as the very macroscopic pattern contributes to the maintenance of dynamic cohesion at the microscopic level. These systems are able to generate and maintain, through recursive dynamics, a type of correlation among their constitutive elements that otherwise would remain disconnected. This recursivity is precisely what provides a minimal form of self-created identity.

Now, it is obvious that whirlpools, hurricanes, and oscillatory chemical reactions are too simple to produce any specific and distinctive form of interactive process with the environment. Thus, in order to discover the origin of adaptivity, we have to look for forms of (self-)organization capable of evolving and generating complex and diverse ways of self-maintenance. Among the wide set of self-maintaining organizations, those based on chemical processes are of particular interest, because they allow the construction of complex recurrent organizations through the creation of local and selective constraints. This type of specific constraint is possible at the chemical level because of the action of the shape and reactive capacities of its constituent components, the molecules.[4] This is the organizational framework of the early prebiotic evolution.

During the process that gave rise to life, a fundamental step was the appearance of systems whose productive activity included the construction of a selective and functionally active membrane. This change led to a progressive takeover of (at least part of) those boundary conditions, thus ensuring the maintenance of the system. In other words, a form of organization appeared in which the conditions of its maintenance were actively controlled by the very organization, thus creating a first and self-sustained separation between system and environment. We call these autonomous (from the Greek *auto-nomos*, self-law) because (some of) the constraints that define the dynamics of the systems are the result of its very organization. In other words, what determines the behavior of the system is not just the physical laws and a set of externally defined boundary conditions but the capacity of the system to redefine (part of) its boundary conditions (Moreno, Etxeberria, & Umerez, 2008).

Actually, this concept of (minimal) autonomy is very similar to the idea of autopoiesis developed by Maturana and Varela (1980). Our emphasis, however, focuses on the FFE and thermodynamically open nature of these systems, from which a crucial implication follows: interactive dynamics are constitutive of the system and not something to be added, a posteriori, in the form of a structural coupling. For Maturana and Varela, autonomous systems are defined by the abstract property of operational closure, leaving aside material and energetic requirements. As a result, the environment appears only as a source of perturbations due to the concurrent structural coupling between the system and its environment. What matters for autopoiesis is then the conservation of autopoiesis, the compensation environmental perturbations or the accommodation of

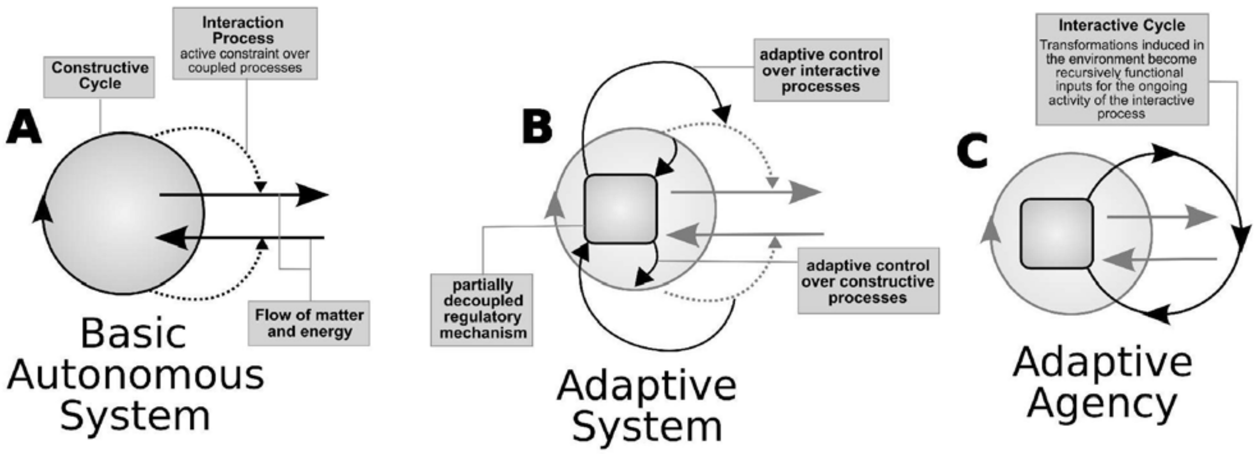

Figure 1 Conceptual diagrams of increasingly complex organizations. (a) Diagram of constitutive processes in basic autonomous systems; the constructive cycle requires a flow of matter and energy that is actively constrained by the system, giving rise to interactive processes and thus bringing about a minimal form of agency. (b) Adaptive systems show a partially decoupled mechanism that regulates interactive and constructive processes adjusting and switching between different alternatives according to external perturbations and conditions. (c) Adaptive agents appear when the interactive processes become a cycle. (Copyright © 2007 Xabier Barandiaran under a *CreativeCommons Attribution – Share Alike* license: http://creativecommons.org/licenses/by-sa/3.0.)

non-disruptive deformations. For a constitutively open notion of autonomy, it is through the flow of matter and energy, required for the appearance of FFE dissipative structures, that the system achieves its unity. Therefore, by definition, the systems appears thermodynamically "hungry," in need of coupling with the environment, which is no longer a mere source of uncomfortable perturbations to be compensated but the source of an essential flow. (The consequences for characterizing adaptive behavior will be crucial as system–environment interactions will become functionally integrated on the very process of becoming an autonomous adaptive system.)

Two constitutive processes can be differentiated within this concept of minimal autonomous organization[5] (see Figure 1a):

1. **Constructive processes** are those that participate in the continuous production of the system (e.g., chemical reactions). Constitutive processes are networked, thus creating a closed organization we call the constructive cycle. The constructive cycle is defined by the production of a set of constraints that recursively regenerate the conditions of their production. The network of reactions that produces the components of the network itself (metabolism) constitutes the most basic example of a constructive cycle in a minimal autonomous system. However, given that this network only exists as a thermodynamically dissipative organization, its maintenance requires interactive processes.
2. **Interactive processes** are the processes generated by the constraining action exerted by the constructive cycle on the flow of matter and energy between the system's boundary and the environment[6] so as to ensure the system's maintenance. At a molecular level, interactive processes require systems endowed with a physical separation (a membrane) between their constitutive organization and the environment. An example of a minimal interactive process is active transport through the membrane, breathing, or adaptive behavior.

Practically all forms of present-day life are capable of performing a wide range of processes in their environments in order to ensure their maintenance. Living organisms at large define their own identity and differentiate themselves from their environments. The way in which they do this is through their metabolic organization: a self-producing network of chemical reactions that controls some of its boundary conditions. Thus, the basic organization of all present-day living beings is essentially what we have elsewhere called basic autonomy[7] (Ruiz-Mirazo & Moreno, 2004).

2.2 Normative Functionality and Intrinsic Teleology Naturalized

Minimal autonomy thus described already provides a naturalized criterion for functionality and normativity in natural systems. The issue of the attribution of normative functionality to natural systems is of major importance for the task of naturalizing adaptivity. As observers, we can define systems arbitrarily. For instance, if we are to use differential equations to build a mathematical model of a system we must (a) choose a set of observables, (b) create some variables representing those observables in our model, and (c) perform operations on the system to establish how variations on an observable produce variations on other observables, so that (d) we can abstract a set of differential equations that specify the rate of change of the variables in the model. If the model adequately predicts the functioning of the system we have an adequate model. At this point, we can attribute a "way of functioning" to the system that corresponds to the functions of our model. Nevertheless, we are faced with the two problems presented in the introduction: (a) how to justify the selection of observables belonging to the system (attribution of identity); (b) how to justify that the system not only functions in a certain way but that, in addition, it "must function" that way and not in another (attribution of normative functionality; see Millikan, 1984 for an influential conception of this problem). For artificial systems, we, as designers or users, can attribute a certain goal to the system according to our intentionality. Under the attributed goal or purpose, we claim that the machine works properly or that it is broken and malfunctions, although the good or bad functioning of the system (the machine) is completely extraneous to the structure of the machine itself. Equally, what belongs to the system and what should be left out (as irrelevant) is determined by the goal we project upon the system; that is, what is an "essential" part of the system (as distinct from its environment) is defined in relation to the desired functionality that we as designers or users expect to achieve in a set of contexts.

In contrast, living systems in general and cognitive systems in particular are capable of defining themselves (as we have explained above) and, more specifically, of determining their own normative functionality; that is, what is good or bad (right or wrong) for them does not depend on an external observer, designer or user but on their own organization. More specifically, in autonomous systems, a process (constructive or interactive) is functional if it contributes to its self-maintenance (Bickhard, 1993; Collier, 1999). A process, in turn, becomes normative if it is dynamically presupposed by other processes in their contribution to the overall self-maintenance of an autonomous system (Christensen & Bickhard, 2002); that is, a behavior or internal component is dynamically coupled with the rest of the components so that the overall maintenance of the whole organization depends on it. Normativity[8] refers to the fact that a set of processes that constitute the system must occur as they do in order for the system itself to exist. A basic example of normative (proper, necessary) functionality is active transportation through the cell membrane. This process becomes normative because the level of chemical concentrations that the membrane's active transport retains within the cell is necessary for some metabolic reactions to maintain the appropriate rate to sustain the network of reactions, which in turn produces the membrane, and so on in a circular and interdependent manner. At a higher level, the normative function of the lung is oxygen intake, because the dynamic-metabolic organization of the rest of the organism relies on this oxygen intake for its functioning and existence. This type of circularity is characteristic of autonomous systems: a set of networked component processes that depend recursively on each other, so that the system, as a whole, is cause and effect of itself.

Because of this circularity in autonomous systems, identity and normative functionality are not observer-dependent but intrinsically causal: The network as a whole (the very system) will not exist in the absence or malfunctioning of the component processes (given its FFE nature and the circular dependency between processes). In other words, in autonomous systems what-the-system-does (the way it functions) and what-the-system-is (its structure) are highly intertwined and they merge together in its organization.

The holistic, integrated, and self-maintaining organization of autonomous systems has some important consequences for the way they are described. For instance, the use of teleological terms to characterize their functioning can be naturalized; unlike its use to describe some artifacts that perform a goal-seeking behavior, such as thermostats or target-seeking missiles. These are artifacts that have been designed to correct their behavior (usually by a negative feedback mechanism) according to an externally defined goal state. Expressions such as "the purpose of the thermostat is to maintain the room temperature at 23 ºC" are used as metaphorical shortcuts to describe the behavior of such systems. However, what the goal state is remains completely extraneous to the mechanism that achieves it; the system is independent of the goal state or set of parameters it controls (which are externally imposed). Autonomous systems are different. Their existence depends on the FFE stability they produce. The stability point or set of points that enable a system to exist are its goal states. These goal states are not just goal states because the system compensates for deviations from them, but because the goal state is the condition of possibility of the system itself. In other words, in autonomous systems, the goal state of the system and the organization that instantiates it are one and the same thing. Autonomous systems have an implicit teleology[9] as their internal causal circularity makes each process of the system a contribution to its global self-maintenance.

Thus, the basic type of autonomy just described is the lower level and most fundamental type of autonomy, that of material and thermodynamic self-construction and self-maintenance, constitutive of all living beings, upon which higher levels of autonomy appear.[10] Basic autonomy generates a cascade of emergent properties such as identity formation, normative functionality, and implicit teleology. By expanding the above analysis and properties to the interactive cycle of autonomous systems, we can naturalize a set of characteristics that are necessary to describe adaptive behavior. These are agency, adaptivity, and explicit teleology.

3 Adaptivity and Agency

From a dynamic point of view, we can abstract a set of boundary conditions and an essential parameter value region to be necessary for the maintenance of a FFE system. Based on the definitions of Ashby (1952), we call these parameters and boundary conditions "essential variables," and the range within which the system's organization can be maintained "viability boundaries." The FFE nature of autonomous systems makes at least one of their essential variables have an intrinsic inertia toward outside the viability boundaries. Some of the essential variables are also non-controlled variables, in the sense that no change of the internal variables of the system can directly control its state. As a consequence, only the coupled system–environment can maintain the essential variables within viability boundaries (hence the importance of interaction processes in autonomous systems).

In general, among the system–environment relationships, and from the point of view of the effect of the process on the system, some processes are functional for the system when they contribute to maintain essential variables within their boundaries of viability. Some others can be dysfunctional when they "push" essential variables outside the boundaries of viability. Many others are neutral (they have no effect on the state of the essential variables). However, from the point of view of the cause of these processes, we can also classify them as active (if the processes are triggered by the system as a whole) or passive (when the interaction is induced from outside or the process is the result of physico-chemical laws without any internally generated constraint acting upon the process; e.g., osmosis). We use the term "agents" for those systems that interact with their environments, so that the changes produced between the system and its environment contribute to its self-maintenance (see Figure 2 for an inclusive classification of the different types of process). The fact that a coupled process can be active, rather than passive, is of fundamental importance for the task of naturalizing agency, and it deserves a more precise definition. By active we mean, in negative terms, that the interaction is produced neither by an external source nor by means of unconstrained physical laws (i.e., spontaneously and independently of the particular organization of the system). In positive terms, an active process is one that makes use of the organization of an autonomous system to produce a constraining effect to ensure its own self-maintenance. Actions require work, and work requires suitable energy. Thus, at the minimal level, active interactions require that the system channels the thermodynamic flow of its FFE organization into the creation of specific constraints.[11] In addition, another crucial factor that characterizes active processes is the complexity asymmetry of the coupled processes, laden to the side of the autonomous system (i.e., the interactive mechanism of the system is, because of its holistic organization, more complex than the coupled process it sustains). "More complex" means that the set of variables internal to the agent are

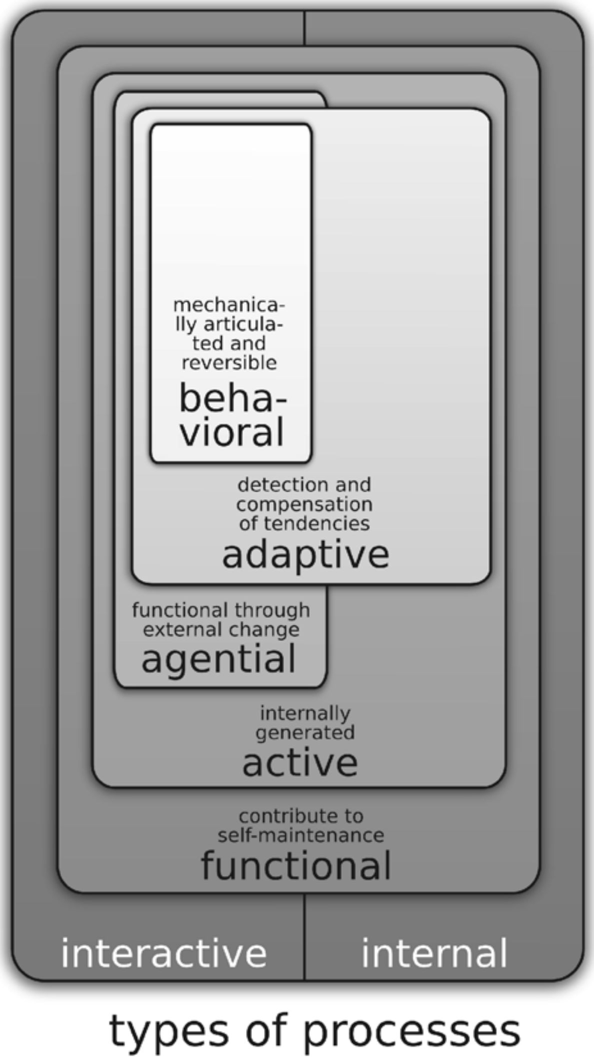

Figure 2 Inclusive classification of different processes in autonomous systems. Functional processes are those that contribute to self-maintenance, and these can be both internal to the organism or interactive with the environment. Active processes are those generated by constraints produced by the system. Within active processes, interactive processes give rise to agency; these are the interactive processes that are functional in virtue of the changes induced in the environment. Adaptive processes are those active and functional processes that are carried out by a specific mechanism, which detects and compensates tendencies of essential variables when they are sufficiently close to the boundaries of viability. Adaptive processes can be internal or coupled. Interactive and adaptive processes give rise to adaptive agency. Finally, behavior appears as a subclass of adaptive agency based on mechanical articulation. (Copyright © 2007 Xabier Barandiaran under a *CreativeCommons Attribution – Share Alike* license: http://creativecommons.org/licenses/by-sa/3.0.)

more functionally differentiated and integrated (Tononi, Edelman, & Sporns, 1998) than those of the environment involved in the dynamic coupling (for a quantitative comparison between internal and behavioral complexity, see Seth & Edelman, 2004).

Robustness is the capacity of a system to maintain its organization in the face of internal and external perturbations. This capacity can be achieved by a number of increasingly complex mechanisms: a simple buffering mechanism, a distributed structural stability or an adaptive regulatory subsystem. In any case, a sense of margin of viability is required; otherwise even the smallest perturbations will destroy the system's organization. The simplest forms of self-maintaining systems, such as hurricanes or candles, cannot really "do" anything in order to compensate for new environmental conditions, they only have a margin of maintenance provided by a buffering or flexible structure. However, more complex forms of self-maintaining systems, such as hypothetical cellular protometabolisms, achieve robustness only by using their internal organization to constrain some of their boundary conditions. In other words, to sustain self-maintenance under different environmental conditions, they must "recruit" their internal organization to regulate the coupled processes, which now become interactive processes. Interactions are thus active and functional processes between the system and its environment. Active ion pumping by the membrane could be an example (Moreno & Barandiaran, 2005; Moreno & Etxeberria, 2005; Ruiz-Mirazo & Moreno, 2004). This interaction becomes functional because, as a result of ion-pumping, it produces an internal change that is necessary for the ongoing activity of the constructive cycle (reduction of ion concentration in the interior of the cell).

The most simple forms of autonomy are structurally stable, in the sense that they can compensate internal and external perturbations by means of a self-regulating mechanism that is integrated and distributed over their constitutive organization; that is, the regulatory mechanisms are embedded in constructive and interactive processes and so regulatory and regulated processes are indistinguishable. The property of structural stability[12] corresponds with the idea of conservation of autopoiesis as originally formulated by Maturana and Varela (1980). However, when FFE systems increase their complexity[13] they become more fragile; noise and environmental perturbations affect their organization, which, given its holistic nature, disintegrates easily. Thus, as complexity increases even more, self-maintenance under relatively wider conditions requires special mechanisms capable of ensuring adequate coupled processes. Structural stability cannot ensure the maintenance of complex systems under different, variable environmental conditions. Hence, in early living systems self-maintenance requires an active control of interactive and constructive processes by measuring different conditions and monitoring its own constitutive processes so as to avoid or prevent dysfunctional situations. In other words, in addition to executing a constructive or interactive process that contributes to self-maintenance, the system is also capable of switching between different alternatives, adjusting them, and so forth, according to external changes. This is the essence of adaptivity recently defined by Di Paolo (2005):

> A system's capacity, in some circumstances, 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, a) tendencies are distinguished and acted upon depending on whether the states will approach or recede from the boundary and, as a consequence, b) 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. (p. 438)

Adaptivity requires the establishment of an explicit normative regulation. As Di Paolo maintains, in non-adaptive self-maintaining systems the natural distinction between self-maintenance and disintegration is not yet accessible to the system, unless it is also able to regulate itself with respect to a norm. Whereas in pre-adaptive systems self-maintenance depends on the range of values that the essential variables can withstand, adaptive systems have the capacity to modulate (internally and interactively) the trajectories of the essential variables of the constitutive processes (unlike structurally stable systems; see Figure 1b). Essentially, adaptivity requires a regulatory control over the basic functioning of the system. Therefore, in order to be adaptive, a system must be organized such that there is a relative decoupling between the dynamics of a regulatory subsystem and that of its basic constitutive organization. This has important consequences, as some of the properties traditionally assigned to autopoietic systems cannot be derived from autopoiesis alone but presuppose adaptive capacities (as argued in detail by Di Paolo, 2005). In addition, in most cases, adaptive regulation takes place not just by transforming outward tendencies into inward trajectories (i.e., not just by avoiding negative tendencies) but by actively seeking to improve the state of essential variables so that regulation takes place not just in reference to the boundary of viability but graded and directed by a "sense of well-being."

Adaptivity is a capacity that all present-day living organisms possess.[14] The simplest mechanisms of adaptive regulation fall into two different categories. One is exemplified in the Operon activation and deactivation of genes as a switch between metabolic pathways according to certain environmental conditions. The other is constituted by a whole subsystem of biochemical pathways not directly involved in the basic self-constructing metabolic network (as is the case of chemotactic agency in *Escherichia coli*; see examples at the end of this section). However, the common characteristic of both cases is that some degree of dynamic decoupling from the basic constitutive processes is required. In the first case, metabolically off-line gene-strings act as instructive switches between different metabolic pathways. In the second case, chemical pathways that are independent of the basic metabolic-constructive cycle sustain the interactive loop. This decoupling of genetic regulatory mechanisms from the basic metabolic network allows a selective choice among a large amount of not yet functional dynamical states of the constitutive self-maintaining metabolic network. These decoupled systems open the possibility to consistently speak in terms of an internally generated mechanism for normative regulation. The capacity to differentiate between, and compensate for, tendencies requires that whatever makes a distinction and generates a compensation be dynamically differentiated from what it distinguishes and acts upon (which presupposes operational mechanisms to distinguish between the different implications of equally viable paths of encounters with the environment).

Thus, we are talking about two dynamic "levels" in the system: the constitutive level, which ensures ongoing self-construction, and the (now decoupled) interactive subsystem, which regulates boundary conditions of the former (e.g., concentration gradient of metabolites across the membrane through displacement to richer environments). It is clear that decoupling permits a specific phenomenon: independence of low-level functionality (constructive processes) from high-level variation (state of regulatory mechanism). Thus, free from the lower-level constraints, higher level variation can be left to "spontaneous" dynamics, provided that a further coupling is established linking higher-level to lower-level states in a functional way so that this higher-level variation (or freedom) is in turn functionally recruited to serve lower-level functions. This way, certain states of the low-level network will be stabilized when contributing to the creation of new forms of self-maintenance.

We can naturalize the claim that some interaction or process is perceived as bad or good by, and for, the very system (and not only by, and for, the external observer); that is, this good or bad functioning for the system is objective because it is detected and compensated by the system, in an effective, functionally integrated way. Thus, adaptive systems are an instance of explicit teleology as, in addition to having an intrinsic goal (because of their basic autonomous organization), they also act according to this goal, generating global constraints, over their minimal basic organization, so that a meta-regulatory process emerges.

Adaptivity takes two basic forms depending on whether the mechanisms of regulation take place at the constructive or interactive level. In the first case, internal or external perturbations are compensated by adjusting or transforming constructive processes (such is the case of the Lac operon mechanism). The second form of adaptivity turns out to be of particular interest because it gives rise to adaptive agency: adaptation to perturbations is achieved through recursive interactions with the environment so that interactive processes become a cycle (see Figure 1c). Interactions become functional in virtue of the changes induced outside the system or, more specifically, in the relationship between the system and its environment. Motility-based interactions are the most clear examples of this type of functional feedback through the environment. A paradigmatic case is bacterial chemotaxis: in order to change its current environment to another that contains more nutrients, bacteria engage on a sensorimotor cycle. The set of interactive processes performed by the bacteria become cyclic because it is realized by modulating effector processes according to the detected conditions of the environment in a recursive manner. However, there are also other forms of adaptive agency not based on motility. A plant, for instance, can become an adaptive agent by secreting and detecting chemical substances, engaging on a defense–attack coupling with certain fungi on its proximal environment. In these cases, the transformations induced on the system–environment relationship also become functional inputs for achieving adaptive regulation; hence the notion of "interactive cycle" as a distinctive character of adaptive agency (unlike cases of internal adaptive processes or single functional actions into the environment, such as ion-pumping).

The appearance of adaptive agency implies the emergence of detection and effector mechanisms by which the adaptive regulation is linked with the environment and the constitutive organization of the autonomous systems. In addition, two regulatory processes appear linked to each other: the adaptive regulation of the essential variables through recursive interactions with the environment on the one hand, and the regulation of this interactive cycle according to its effects on the essential variables on the other. This intertwined regulation along with the detection–response coupling with the environment brings forth a genuine type of teleological agency, in which explicit teleology is extended into environmental interactions.

We can now introduce the concept of motility as the capacity of an agent to move by its own means, so that it is able to execute fast (relative to its size) directional movements to change the environment looking for preferred conditions. In the case of adaptive motility, detection of and functional response to environment-relevant changes becomes a "sensorimotor" cycle, the viability of which is strongly affected by size–time constraints. It is this high size–time (speed) constraint that characterizes sensorimotor adaptability from other forms of adaptability.

Let us summarize the proposed distinctions by providing some characteristic examples.

1. **Simplified bacteria.** *Buchnera aphidicola* (Moya, Peretó, Gil, & Latorre, 2008; Shigenobu, Watanabe, Hattori, Sakaki, & Ishikawa, 2000) can be considered to be a living system in its own right as the machinery of self-(re)production and energetic and material exchange with the environment is at work, but should be considered to be neither an adaptive nor an interactive agent. This bacterium is an extremely degenerated organism. It lives as an endosymbiont inside the *bacteriome* of certain insects where the local environment is highly homogeneous and co-adapted to its meta-

bolic organization. It obtains the necessary energetic and chemical input for self-maintenance from its medium with almost no change of environmental conditions. As a consequence, adaptive mechanisms in *Buchnera aphidicola* are highly degenerated and these organisms are considered to be close to a structurally stable minimal-autonomous system (to the extent that they are currently used as models organisms for the study and synthesis of minimal life forms).

2. **Lac operon mechanism in *E. coli***. The normal metabolism of *E. coli* depends on the presence of glucose in its environment. However, when the levels of glucose in its environment become very low and another sugar (lactose) is abundant in the environment, a mechanism called lac-operon is activated: The detection of lactose triggers the expression of certain dormant genes,[15] which in turn instruct a new metabolic pathway that metabolizes lactose. This metabolic mechanism is adaptive because it implies a (meta)regulation of the internal constructive processes according to the detection of a certain environmental condition (the presence of lactose and the absence of glucose), which jeopardizes the self-maintenance of the system. Nonetheless, this form of adaptivity does not imply agency as the changes produced for self-maintenance are mainly internal.
3. **Chemotaxis in *E. coli***. *E. coli* alternates tumbling (random rotation) and forward displacements, generating a net displacement toward environments with increasing concentrations of metabolites. This is achieved through a "two component signal transduction" mechanism, capable of measuring the temporal difference of attractant concentrations in the environment and changing the frequency of the flagellar rotation accordingly (see van Duijn, Keijzer, & Franken, 2006 for a more detailed analysis). In this form of adaptivity, the system interacts functionally with its environment (moves up-gradient until the necessary level of sugar is encountered). The interaction is functional in virtue of the transformations induced in the environment (the concentration of sugar increases in relation to the system) and the system operates recursively on these environmental changes (sugar-detection and frequency of flagellar rotation are correlated). Among an open set of environmental variables, the sugar concentration gradient shapes a functional world of interactions for the bacteria's self-maintaining capacity.

In short, some fundamental elements, such as functionality, normativity, and teleology, are the consequence of the autonomous organization present in the simplest biological systems. Adaptive agents (those autonomous systems capable of interacting with their environments by detecting and compensating for tendencies of their essential variables) extend an explicit teleology to their environments. As seen in the next section, biological organizations that support this type of agency through biochemical mechanisms are severely limited in their capacity for open-ended agential complexity. Thus, without a qualitatively different internal organization of autonomous systems, adaptive agency eventually exhausts its capacity for complexity growth. The appearance of the nervous system (NS) solves this problem, leading to a whole set of bodily changes and to a qualitatively different organization of adaptive agency.

4 Nervous System and Behavioral Agency

4.1 Limitations of Agential Mechanisms at the Metabolic Level

As we have seen, the appearance of adaptivity implies a certain degree of decoupling within the organization of the agent; there should be mechanisms that support the interactive processes whose functioning is relatively independent of those involved in the constructive processes. This internal organization of adaptive agency therefore requires a relatively complex metabolism, instructed and regulated in somatic time by an off-line conservative structure (DNA) or by a different and relatively independent subsystem of chemical reactions (van Duijn et al., 2006). These early adaptive mechanisms are context-specific regulatory mechanisms and mostly genetically specified.

However, the use of biochemical mechanisms supporting interactive tasks severely limits the capacity to achieve increasingly complex forms of agency. The reason is that there is a serious bottleneck in the evolution of movement-based agency supported by biochemical mechanisms. First, the bottleneck appears because the level of complexity that the adaptive subsystem can achieve (within the biochemical medium) without severe interference with metabolic processes is very limited. Second, as the size of the organism increases, the fast and plastic correlation between sensor and effector surfaces becomes harder (or even impossible in multicellular organisms) as a result of the slow velocity of diffusion processes. Third, there is also the problem of achieving unified body coordination for displacement. The type of rather sophisticated motile agency displayed by *Paramecium* illustrates the tension produced by the combination of these three factors: epithelial conduction (through Ca channels) is used to solve fast and coordinated beating of cilia because, unlike the case of *E. coli,* it could not be achieved by mere diffusive mechanisms. However, the complexity (in terms of functional diversity and integration) that homogeneously spreading epithelial conduction can achieve is very limited.

The appearance of multicellularity posed an important challenge in the evolution of agency, as at this size biochemical mechanisms cannot support fast and versatile motility. There are two causes of this problem: the enlarged internal distance between parts of the body, which needs to be connected in short delays (so that the organism can move fast and coordinately), and the need to modulate the organization of connections selectively (to obtain the adequate sensorimotor correlations) for versatile, plastic agency. Hence, if biochemical network plasticity were the only mechanism for accomplishing adaptive interaction and self-maintenance, the forms of movement-based agency would probably be very limited at the multicellular size.[16] However, when in the development of some metazoans a new type of cell (the neuron) started to differentiate itself, this limitation could be overcome. Neurons differentiate as cells capable of forming branches, interconnected through plastic electrochemical pathways and capable of propagating and modulating electric potential variability.[17] 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 (Llinás, 2001).

4.2 Hierarchical Decoupling of the Nervous System

Since the very beginning of its evolution, neural organization appeared as an extended network capable of producing a recurrent dynamic 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 because of their diffusive nature, the electrochemical interactions between neurons make open-ended recurrent interactions within the NS itself possible. The NS constitutes a cellular infrastructure that converts metabolic energy into finely modulable electrodynamic processes, thus creating a new dynamic level almost free from the thermodynamic constraints characteristic of the biochemical level of metabolic-constructive processes. What makes neural interconnections so special is that they create an incredibly rich and plastic internal world of patterns of fast connections, hierarchically decoupled from the metabolic processes.

As we have pointed out elsewhere (Barandiaran, 2004; Barandiaran & Moreno, 2006; Moreno & Etxeberria, 2005; Moreno & Lasa, 2003), the hierarchical decoupling of the NS from metabolism means that metabolism generates and sustains a dynamical system (the NS) minimizing its local interference with it. The term "hierarchical" refers to the fact that metabolism produces and maintains the architecture of the NS by providing the necessary energy to feed its dynamics. However, the term "decoupling" means both (a) that neurons minimize interference in their local metabolic processes with their ion-channeling capacities and (b) that the metabolic-constructive organization of the organism (digestion, circulation, etc.) under-determines the activity of the NS, which depends on its internal dynamics and its embodied sensorimotor coupling with the environment. Operationally speaking, if we are to predict the state of the NS, hierarchical decoupling means that neither local states of cell metabolism nor the state of metabolic organs alone are going to be very useful; on the contrary, the electrochemical states of other neurons and their embodied sensorimotor coupling with the environment might provide a much better model for prediction. In other words, the biophysical specificity, high connectivity, embodiment, and situatedness of neural electrochemical dynamics make it irreducible to the metabolic substrate of its constituent components (the neurons) and the organismic processes of self-construction and repair (state of other body organs and processes involved).[18]

Hierarchical decoupling permits the specification of a set of operational primitives (the lowest level dynamical observables necessary and sufficient to model nervous activity) and their relationships, which constitute the control mechanisms of animal agency. It is commonly accepted that the primary operational primitives are the change of membrane action potentials of neurons over time (that generally take the form of spikes), which conserve dynamic variability in terms of spike frequencies and time distance between spikes. Synaptic connections, however, specify a connectivity matrix (the transformation functions between primary operational primitives) while neural modulators (local and global synaptic modulators and action potential threshold modulators) become secondary operational primitives (as they become operational primitives in virtue of their effect on the spikes). The search for these dynamic primitives and their functional higher level causal organization constitutes the search for a neural "code:" the set of primitive variables and relationships that constitutes the dynamic domain capable of modeling adaptive behavior. In other words, the type of local operational differences that can make a systematic global difference in behavior (spike rates, interespike intervals, time of arrival, gas-net modulation, synaptic modulator, axonal growth, etc.).

The decoupling of neural processes from the underlying metabolic processes raises the question of how to characterize this specific dynamic domain. As we have seen, the active electrochemical conductivity of the components of the NS (the neurons) is organized in spikes or action potentials, thus generating a new dynamic domain that is built up tangentially to the metabolic processes of the organism, although realized and sustained by it. At the same time, the all-or-nothing characteristic of neural spikes allows for a stable combination of these, which, added to the network structure of the NS and the action of neural modulators, generates a high-dimensional, non-linear, recurrent, and recursive domain. Nonlinearity allows distinctiveness of states, while recurrency, provided by the structure of the network, allows circularity or reentry (Edelman, 1987). Recursivity, however, takes place because spikes can affect themselves through the neural modulators they activate. As a result, the effective dimensionality of the system is constantly being redefined by its own activity. Thus, unlike the aforementioned forms of decoupling within the biochemical organization, which allow primitive forms of adaptive agency, the hierarchical decoupling of the NS permits an open-ended growth of complexity in the forms of agency (impossible to achieve by previous agential mechanisms).

In addition, there is a causal link between the neural domain and certain external processes that belong to other dynamical levels (e.g., metabolic processes in the muscles). This causal connection is largely independent of energetic or material aspects, as neural states produce changes in body states by formal rather than energetic means in a type of lock-and-key causality (what we call formal causality). This causal link is established through the pattern of spikes and not through the energetically determined causality through which these very patterns of spikes propagate. Thus, for example, the motor action caused by neural spikes is not determined by the electrochemical energy that constitutes action potentials but by their form or pattern, which muscle cells "interpret" (i.e., the process by which the neurotransmitters that neurons generate act by selecting metabolic energy to produce movement). This process is similar to the electric patterns that travel along wires and connect two computers; these patterns produce changes in the terminal not by virtue of the electric energy they convey but by virtue of the sequence of changes in amplitude and frequency. In other words, the neurotransmitters that neurons generate (when a given pattern of spikes from further neurons arrives) trigger a cascade of chemical processes in the muscles that convert patterns of spikes into mechanical work.

Thus, the NS, hierarchically decoupled and endowed with the capacity for formal causation, justifies the characterization of the neural domain as properly informational.[19] Neural primitives can be considered as non-dynamical elements in relation to the underlying metabolic processes in the sense that, from the point of view of the modeling of metabolic-constructive dynamics of the organism, the NS appears decoupled. From the point of view of musculoskeletal dynamics, the NS acts as a formal control system, independent from the particular energetic details of how movement is achieved in the muscles.

4.3 Behavioral Agency

Interestingly, the appearance of multicellular organisms endowed with a (sub)system allowing fast, efficient, and plastic agency, was necessarily accompanied by other important changes in their internal structure. The appearance and evolution of the NS brought along with it changes in the organization of internal circulation, the system of fixation, and even the body shape. Thus, the unfolding of multicellular organisms whose way of life is based on motility requires radically different internal organization, namely a new bodyplan that allows for a whole set of transformations at both the internal and interactive levels (Moreno & Lasa, 2003). First, as we have already mentioned, there is the appearance of a decoupled non-biochemically based adaptive subsystem: the NS. Second, there are specific tissues and body structures capable of channeling metabolic energy into efficient mechanical energy (muscles, skeleton, etc.). Fast movement, in multicellular organisms, is only possible through specialized organs, which directly convert metabolic energy into mechanical energy independently of the continuous process of metabolic self-maintenance and morphological transformations that the organism undergoes by means of cell growth and reproduction. We denominate the new form of adaptive agency based on motility of those multicellulars endowed with a NS controlling a mechanical body as properly behavioral agency.[20] Thus, behavior is actually fast adaptive motility decoupled from morphological and, in general, metabolic-constructive processes.[21]

It is precisely the hierarchical decoupling of the NS and its sensorimotor coupling with the environment that makes it possible to study adaptive behavior in terms of sensorimotor dynamics (as is the case in several fields, such as robotics, cognitive neuroscience, or embodied psychology) and qualifies behavior as a specific phenomenon distinct from generic biology. In contrast, the explanation of the interactions of plants with their environments would require the introduction of additional (non-sensorimotor constraints) such as the rate of growth through cell replication according to exposure to light, availability of water in the immediate surroundings and a host of similar agency–metabolism interdependences. None the less, hierarchical decoupling from metabolism and sensorimotor coupling with the environment does not mean that the metabolic substrate of behaving organisms is irrelevant. In contrast to the local decoupling from metabolic processes, a global coupling follows, so that adaptive behavior will ultimately have to satisfy the demands of metabolism.

4.4 Neurodynamic Constraints, Self-Organization and Adaptive Behavior

The function of the neural domain in the overall organization of the organism is to achieve behavioral adaptivity (i.e., adaptive maintenance of essential variables under viability boundaries through the neural sensorimotor control of the interactive coupling with the environment; Barandiaran, 2004). However, the fact that the global dynamics of NS–body–environment produce an adaptive maintenance of essential variables under viability boundaries does not specify how this functionality is achieved (i.e., what the dynamic organization of the NS is like and how it is related to behavioral adaptivity). Because it is impossible for metabolic needs alone to instruct functionally such a high-dimensional space, the understanding of behavioral agency requires us to explicitly define what types of constraint act on the NS generating functional order.

Dynamically speaking, the activity of the NS is defined by internal and external constraints. External constraints are those that are not the result of the activity of the NS itself. We can distinguish two main types of external or innate[22] constraints on the NS. Architectural constraints look like a first candidate for the "instruction" and generation of order in the NS, and can, in fact, generate highly constrained structures and dynamics in some primitive NSs.[23] These constraints are the result of genetically triggered anatomical-developmental processes selected in relation to the adaptive sensorimotor interactions they produced when coupled to body–environment systems. Body signals, however, act as strong perturbations of neural dynamics through specific signals (pain, hunger, pleasure, etc.) that operate by modulating the overall activity of the NS (generally through specific neural modulator pathways such as the dopaminergic system). The role of body signals is to regulate the behavior sustained by neural activity in relation to the adaptive needs of the organism. These signals shape neural dynamics toward the satisfaction of certain goals; they are generally originated on metabolic needs and can be considered innate from the point of view of the early activity of the NS.

So far, then, neural dynamics can be captured through the specification of (a) neural primitives, (b) a set of innate (architectural and body signal) constraints, and (c) sensorimotor interactions. However, in most of the known neural systems the complexity of the possible neural dynamics appears underdetermined by these constraints. Nonetheless, we still find functional order. Therefore, different (non-external) principles of order are required to explain them. More specifically, the dynamics of the NS enter a process of local and interactive self-organization through the recursive activity of neural dynamics and sensorimotor interactions. The hierarchical decoupling achieved through the electrochemical functioning of neural interactions and their capacity to establish a highly connected and non-linear network of interactions provides a dynamic domain with open-ended potentialities, not limited by the possibility of interference with basic metabolic processes (unlike diffusion processes in unicellular systems and plants). It is precisely the open-ended capacity of this high-dimensional domain that opens the door to spatial and temporal self-organization in neural dynamics and generates an extremely rich dynamic domain mediating the interactive cycle, overcoming some limitations of previous sensorimotor control systems.

A paradigmatic example of self-organizing patterns in the NS is given by central pattern generators (CPGs) where the interaction between neurons on a local circuit generates robust oscillatory patterns. However, self-organization also appears at the level of the coupling between NS, body, and environment. In fact, the effect of sensory perturbations propagates recurrently through the network generating muscle contractions, which in turn feeds back to sensory neurons both through the changes that movement induces in the immediate sensory environment and through proprioceptive feedback. The recurrent embodied coupling of the NS to the environment results in adaptive behavioral patterns whose functional stability is the result of the dynamic integration of neural, body, and environmental features. Cases of interactive self-organization have been reproduced over more than two decades of robotic research[24] and span across almost all the domains of living behavior (except for a few rigid and cue-bound reflex-like reactive responses).

So let us recapitulate. In large (multicellular) living systems, sensorimotor agency requires a dynamic domain fully decoupled from local metabolic constraints, namely, the NS. This system, embedded in a sensorimotor architecture, is organized in terms of internal and interactive self-organized processes constrained by innate architectures and body signals. The functionality of the system is defined by the maintenance of the system within viability boundaries, which gives rise to the unfolding of adaptive behavior.

5 Situated and Embodied Nature of the Sensorimotor World

At this level of sensorimotor adaptive behavior, we can fully identify some more fundamental properties of natural behavioral processes that are found to be at the basis of all cognitive processes. Some of these properties might already appear in earlier (even unicellular) agents, especially those based on motility and endowed with sensory and motor mechanisms (van Duijn et al., 2006), but their full significance shows up in neurally guided behaving systems.

Behaving systems are situated systems, and their relation with the environment is relative to their situation in it in a non-trivial manner. The behaving organism is not coupled to the environment as a Watt Governor might be coupled to a water flow. Sensory input is not only a function of the environment and the transformations that the agent induces in it but a function of the controlled relative position of the agent in its environment (spatially and geometrically structured through motility). This is a fundamental property that has been ignored in most of the literature on artificial intelligence. For instance, when functional behavior is taken to be the result of extracting statistical properties or patterns from a string of predefined inputs, the consequences are non-trivial. For example, a non-situated system that is reactive (i.e., whose output is determined by the instantaneous input and a historically non-modifiable internal structure) cannot solve a non-Markovian task (i.e., it cannot successfully classify an environmental condition if detecting this condition requires it to extract a sequential order). In contrast, a situated system with a reactive controller can transform non-Markovian tasks into Markovian tasks just by means of exploiting its relative position in its environment (Izquierdo-Torres & Di Paolo, 2005).

In addition, functional behavior is defined by the body in two different dimensions, which we call sensorimotor and biological embodiment, respectively (for a detailed discussion, see Ziemke, 2003). Adaptive robotics has devoted more attention to sensorimotor embodiment, which is a function of bodily properties in relation to the environmental sensorimotor coupling of the agent. The body's physical and mechanical properties shape possible interactions and relative positions through enabling biomechanical constraints. The space of motor outputs to be instructed by the organism is not a uniform multidimensional space defined by a number of degrees of freedom. In contrast, motor embodiment defines a biased "landscape" within that space determined by the shape, elasticity of joints, relative orientation, and a host of similar body constraints. An extreme case of motor embodiment is given by dynamic walking models (McGeer, 1990), in which, even in the absence of neural control, a mechanical system (a pair of legs) determines a well-structured environmental coupling with its environmental surface giving rise to coherent and robust walking behavior. Instead, in the same situation, a disembodied approach would have required an exhaustive control motor output anticipating trajectories and a host of feedback control mechanisms. However, embodied sensory surfaces define a range of sensory inputs and particular transformations and the filtering of them. Finally, both embodied sensory and motor embodied surfaces appear highly intertwined because of the circular and recursive nature of sensorimotor interactions that have evolved together. We can call these enabling constraints because they bias the potential dimensionality of the sensorimotor coupling so as to enable or facilitate self-organized developmental and adaptive interactions (cross-modal sensorimotor spaces, developmental scaffolding by bodily changes, structural adaptation with certain object size and shape, etc.). In addition, and from a computational perspective, embodiment also means that much of the cognitive processing is carried out as embedded in the structure and mechanical functioning of sensorimotor processes.

A less commonly emphasized type of embodiment is biological embodiment, which defines the ecological network of interactive necessities of the agent in order to satisfy its basic biological conditions of possibility.[25] Biological embodiment is in continuous feedback with the sensorimotor flow. In fact, the world of a behaving organism is not so much an independent, physical world but the coupling of this external world with the "internal" world: the dynamics of the constructive cycle "expressed" through body signals harnessing neural dynamics so as to satisfy metabolic needs. The primary function of sensorimotor dynamics is therefore to maintain the essential variables under viability boundaries. So the primary sensorimotor correlations in the organization of behavior are defined by the effect of the sensorimotor coupling on the dynamics of its biological embodiment. In fact, the world that comes about through biological embodiment can be understood as a mapping between the sensorimotor environment and the basic autonomous viability conditions. This is, properly speaking, the adaptive environment that a natural agent defines.

Artificial agents built according to the so-called "autonomous situated robotics principles" show emergent behavior arising from real (or realistically simulated) perception–action cycles; they are (or tend to be) able to measure the relevant parameters of the environment to control certain degrees of freedom of the system from the very situation in which the system finds itself (and not from the point of view of an external observer), and to physically act in it. However, what is lacking in most (if not all) allegedly adaptive or autonomous artificial systems is, as Di Paolo (2003) has pointed out, a self-concern of the processes they undergo "because the desired goal is not designed by the robot but by the designer." In other words, these robots lack their own normativity, which is a consequence of their lack of biological embodiment. This means that artificially created adaptive agents are not fully embodied. While artificial embodiment does not preclude a useful study of adaptive behavior, it severely limits the attribution of genuine agential or adaptive capacities to these systems. An alternative research program would require us to model autonomy at the level of behavior itself, synthesizing agents capable of both maintaining their own behavioral organization through interactions and generating a new level of identity and self-maintenance at the sensorimotor level, in analogy with basic or metabolic autonomy (Barandiaran & Moreno, 2006; Di Paolo, 2003).

6 Conclusions

We have argued that adaptivity requires autonomy. Artificial adaptive systems are, in fact, possible only because normative criteria are externally imposed by human beings, who, as living organisms, are autonomous systems. Autonomous systems create their own identity and differentiate themselves from the environment. However, autonomy is a necessary, but not a sufficient, requirement for adaptivity. As we have seen, the appearance of adaptivity also requires a regulatory subsystem within an autonomous system, partially decoupled from the dynamical organization of the constitutive processes of the system.

The most interesting aspect of motility is the appearance of new types of interactive process not directly involved in the maintenance of the constitutive (metabolic) organization of the system. These new interactive processes give rise to the sensorimotor domain. Therefore, sensorimotor processes are neither mere interactive processes performed by an autonomous system, nor just adaptive agency, but specific, metabolically quasi-independent interactive dynamics of movement-based adaptivity.

However, biochemically supported adaptive motility becomes a serious organizational problem as size and complexity increases. Thus, full-fledged behavior only appears when adaptive motility is supported by a metabolically decoupled regulatory subsystem (the NS) embodied on a musculoskeletal architecture, allowing fast and plastic functional sensorimotor coordination at the multicellular scale without interference with other constructive processes (metabolism, growth, replication, etc.). Adaptive behavior is therefore body movement through neutrally controlled sensorimotor interactions that satisfies biological constraints: behavior oriented towards survival and reproduction. Although self-organizing patterns appear in the neural domain, as a result of internal and external recurrent interactions, the biological embodiment still defines the normative framework and teleology of adaptive behavior. The organism as a whole is an autonomous agent but, from the point of view of sensorimotor dynamics, viability boundaries are externally defined (by biological or basic autonomous needs). In this sense, adaptive behavior lies within the biological domain.

Finally, the creation, through the NS, of a highly rich sensorimotor domain decoupled from the functioning of the metabolic organization not only allows an open growth of the complexity of adaptive processes, but also has a collateral consequence: the progressive takeover of both the organization of behavior and of the body itself by the neural system, opening the way to the appearance of the cognitive domain. The study of how this new domain has appeared, its capacity to provide non-metabolic normativity and teleology, and its relation with the biological domain has been explored elsewhere (Barandiaran, 2007; Barandiaran & Moreno, 2006).

Acknowledgments

Funding for this work was provided by grants 9/UPV 00003.230–13707/2001 from the University of the Basque Country, and BMC2000–0764 and HUM2005-02449/FISO from the Ministry of Science and Technology. In addition, X. Barandiaran has the support of the doctoral fellowship BFI03371-AE from the Basque Government.

Notes

1 The search for a minimal organization is meant to be a methodological strategy to find the "essence" of a process that appears vastly complex in nature. In physics, the Bohr model of the hydrogen atom may be taken as a paradigmatic example. The search for minimalism does not rule out other approaches (e.g., the modeling of a component subsystem, such as the visual system of a cat in cognition or the Kreb cycle in metabolism). What a minimalist approach permits is a model template to which other pieces of modeling work can be attached or integrated.

2 Needless to say, at very large scales gases might achieve forms of cohesion on the basis of gravitational forces such as stellar gas clouds. For a smaller case, however, gases are states of matter characterized by their lack of cohesive unity outside containers, or contingent forms that they can acquire for very short periods of time (such as clouds). This is not to say that a gas cannot take any cohesive form under any condition. As we see, dissipative structures can appear, structuring non-cohesive material substrates under very specific conditions.

3 A dissipative structure is in fact a self-organizing system. By self-organization we mean that local non-linear interactions between components generate a global behavior, which is maintained through a certain number of constraints of which at least one is a product of the global pattern (Ruiz-Mirazo, 2001). Note that the global pattern is not instructed (dynamically specified) from outside, nor can it be reduced to or predicted from the activity of any of its local components.

4 Most "purely physical" self-organized systems (such as lasers, spin glasses or Benard cells) are unable to create local and selective constraints to generate a proper organization (understood as a disposition of parts generating a set of integrated but differentiable functions). The chemical domain, however, can potentially combine dissipative self-organization with the power of self-assembly that the structural stability of complex molecules permits (e.g., proteins). It is, therefore, in the (bio-)chemical domain where more complex forms of organization appear, in which a dissipative global organization generates and sustains functionally differentiated parts (membranes, catalysts, membrane gates, flagella, etc.).

5 Although conceptually separable interactive and constructive processes might be (and originally are) instantiated by the same organization.

6 When using the term "environment" we are not referring only to a physical environment but also to an ecological and social environment that is the result of evolutionary–ecological relationships, intraspecies social relationships, and the very recursive effect of the agent on these socio-ecological environments.

7 Although necessary, this type of organization is not sufficient for a minimal characterization of living beings. Life requires that a set of hereditary and non-reactive components be coupled with the dynamic internal medium of the basic autonomous organization (i.e., metabolism). These hereditary components support an open-ended capacity to conserve, instruct, modify, and reproduce organizational complexity; in other words, inserting basic autonomous organizations in a historical and collective process of evolution (Ruiz-Mirazo, Peretó, & Moreno, 2004).

8 We distinguish between normativity, regulation, and normative regulation. From an organizational and naturalist perspective, normativity is a somehow transcendental property defined by the organizational conditions of possibility for a dissipative organization. Regulation, however, refers to the control or active compensation of perturbations according to a given goal state or rule. An organism might or might not be able to regulate itself according to its autonomously defined norms. For instance, it could be regulating its temperature to 42 ºC, maintaining this temperature invariant in the face of perturbations, although this temperature might be harmful. Thus, it could happen (and it often happens) that an organism is regulating itself badly, but still regulating. So there must be a principle by which, independently of the actual regulatory functioning of the organism, we can justify the claim that the organism is doing it wrong. This right/wrong, good/bad, value attribution must be naturalized in the normativity that the FFE organization that an autonomous system brings forth. We call normative regulation the regulation that is carried out according to the normativity of the organism.

9 The term "teleology" here should be devoid of any intentional or representational character, it simply denotes the participation or contribution of a process into a self-maintaining organization. The term "teleology" was introduced by Kant, in his *Critique of Judgement*, to refer to such a process type as parts are means and ends of themselves through their participation on a holistic organization; see Weber and Varela (2002) for a further development of this issue.

10 When referring to basic autonomy in living beings (unicellular, multicellular, with or without specialized organs, etc.) we mean the network of processes that constitute the self-maintaining and self-constructing organization of the system (excluding other biological functions, such as reproduction, immune defense, etc.). Minimal autonomy, in contrast, refers to the minimal organization capable of recursive self-maintenance and construction. In the most simple autonomous systems, interactive processes do not achieve the status of cycles. Functional actions exerted on the environment, such as ion-pumping, do not feed back through the environment, to generate additional transformation that in turn give rise to subsequent actions.

11 The term "constraint" is here used to mean that some physical or chemical processes are limited or shaped, channeled toward directions that do not follow from the effect of physical laws or principles without the presence of that limitation that is in turn produced by the systemic organization (Pattee, 1972). It is here that the conceptual model by Kauffman (2000) of autonomous systems as those capable of instantiating work–constraint cycles provides an insightful approach. To perform work (i.e., useful directed release of energy) constraints need to be built by the system and, conversely, to create constraints work is required. It is the characteristic coupling between exergonic (free-energy releasing and thermodynamically spontaneous) and endergonic (free-energy requiring and non-spontaneous) chemical reactions that provides the means to achieve a work–constraint closure in basic autonomous systems. It is in this precise sense in which the system can be said to be the active source of a functional interaction, when these work–constraint cycles are recruited or mobilized to regulate or direct system–environment interactions.

12 Structural stability can be thought of as a mechanism to achieve robustness that typically involves the recovery of a stable state that was characteristic of the system before the perturbation took place. Jen (2002) proposes the following working definition: "Loosely speaking, a solution (meaning an equilibrium state) of a dynamical system is said to be stable if small perturbations to the solution result in a new solution that stays 'close' to the original solution for all time. Perturbations can be viewed as small differences effected in the actual state of the system: the crux of stability is that these differences remain small for all time. … A dynamical system is said to be structurally stable if small perturbations to the system itself result in a new dynamical system with qualitatively the same dynamics. Structural stability will typically involve a system configuration such that small changes in parameters leave the system behavior almost unaffected and/or such that changes in variables are compensated through negative feedback loops that bring the system to the original stable attractor (i.e., the spontaneous compensation of perturbations when these fall into the basins of attraction)." (p. 2)

13 In terms of number and variety of components and interrelations, as well as in the degree of integration of the different parts of the system.

14 Structurally stable autonomous systems could have appeared on Earth before the invention of organisms endowed with an instructed metabolism (a genetic code). However, it is difficult to imagine how an adaptive control mechanism could spontaneously arise and adequately function in relation to metabolic needs without natural selection assuring that the coupling between the control mechanism and the controlled processes be functional for the system as a whole.

15 This is of course a simplified version of the real mechanism. More than activated, the expression of Lac operon gene complex (which encodes three enzymes required for metabolizing lactose) is repressed. When lactose is present in the environment, it binds to the repressor (inducing a conformational change that unlocks is repressive capacity) and transcription of Lac operon starts (but at a very low rate). It is under the absence of glucose when an activator protein complex (CAP–CAMP) increases the transcription of an activator protein that the transcription of the Lac operon increases.

16 It has been argued that plants possess epithelial cells, which can be sensitive to local chemical or tactile stimuli triggering a change of electric potential capable, in principle, of producing fast agential responses (Simons, 1981). However, plant intercellular communication is not based on epithelial cell communication, which lacks directional and selective propagation and is unable to organize modulation and regeneration of signals. Instead, the communication system of plants is based on a type of channel called plasmodesmata, which works by transporting (either passive or actively) a large variety of chemical signals. However, this mechanism is far from showing the speed, plasticity, and recursive modulation of signals of neural networks. Not surprisingly, plasmodesmatal connections seem to be limited to adjacent cells (Trewavas, 2003). In addition, the bodyplan of plants does not allow them to develop musculoskeletal structures, which by virtue of their capacity to channel energy into reversible mechanical motion, are of fundamental importance for behavioral agency.

17 The neuron is a cell specialized in connecting sensorimotor surfaces in a plastic, fast, and (metabolically speaking) cheap way, coordination systems based on epithelial conduction being a limited precursor of neural coordination.

18 None the less, it must be noted that to the hierarchical decoupling of the NS follows a global coupling to some metabolic states of the organism in order to satisfy its adaptive/metabolic needs.

19 Not in the sense of representational or semantic but in the sense of propagation of dynamic variability as measured by information theory.

20 What one of the authors has formerly called “neural agency” (Moreno & Etxeberria, 2005).

21 Certain forms of unicellular motility (such as the case of chemotaxis in *E. coli*) can also be included under this category, despite their limited capacity to achieve relatively complex forms because of the intrinsic bottleneck we have previously described.

22 By innate, we do not mean here that there is a genetically determined architecture of neural pathways in the NS but that, given an evolutionarily stable environment, a developmental process triggered by environmental and genetic factors, certain anatomical structures are stabilized as a result of recurrent interactions in the developmental process.

23 For instance, the nematode *Caenorhabditis elegans* contains precisely 302 neurons and about 5,000 synapses with highly stereotyped connections whose complete wiring diagram is already well known and equivalent among individuals of the same species.

24 Situated and autonomous robotics (Brooks, 1991; Clark, 1997; Maes, 1990; Matarić, 2002; Pfeifer & Scheier, 1999) has provided a set of insightful models of embodied and interactively self-organized behavior: obstacle avoidance (Brooks, 1990), wall following (Steels, 1991), behavioral categorization (Cliff, Husbands, & Harvey, 1993) and a number of other interactive behavioral phenomena that exploit recurrent interactions with the environment (Ziemke, 2003).

25 By this term we mean not only metabolic self-maintenance but a number of organizational constraints derived from the evolutionary dimension of living beings (reproduction, kin caring, sexual selection, etc.).

References

Ashby, W. R. (1952). *Design for a brain. The origin of adaptive behavior*. London: Chapman & Hall.

Barandiaran, X. (2004). Behavioral adaptive autonomy. A milestone in the Alife route to AI? In J. Pollack, M. Bedau, P. Husbands, T. Ikegami, & R. Watson (Eds.), *Proceedings of the 9th International Conference on Artificial Life* (pp. 514–521). Cambridge, MA: MIT Press.

Barandiaran, X. (2007). Mental life: Conceptual models and synthetic methodologies for a post-cognitivist psychology. In B. Wallace, A. Ross, J. Davies, & T. Anderson (Eds.), *The world, the mind and the body: Psychology after cognitivism* (pp. 49–90). London: Imprint Academic.

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

Bickhard, M. H. (1993). Representational content in humans and machines. *Journal of Experimental and Theoretical Artificial Intelligence*, *5*, 285–333.

Brooks, R. A. (1990). Elephants don't play chess. In P. Maes (Ed.), *Designing autonomous agents* (pp. 3–15). Cambridge, MA: MIT Press.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence Journal*, *47*, 139–160.

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

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

Cliff, D., Husbands, P., & Harvey, I. (1993). Evolving visually guided robots. In J.-A. Meyer, H. Roitblat, & S. Wilson (Eds.), *From Animals to Animats 2: Proceedings of the 2nd International Conference on Simulation of Adaptive Behavior (SAB92)* (pp. 374–383). Cambridge, MA: MIT Press/Bradford Books.

Collier, J. (1999). Autonomy and process closure as the basis for functionality. In J. L. R. Chandler & G. van der Vijver (Eds.), *Closure: Emergent organizations and their dynamics. Annals of the New York Academy of Sciences*, *901*(1), 280–290.

Di Paolo, E. (2003). Organismically inspired robotics. In K. Murase & T. Asakura (Eds.), *Dynamical systems approach to embodiment and sociality* (pp. 19–42). Adelaide, Australia: Advanced Knowledge International.

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

Edelman, G. M. (1987). *Neural Darwinism: The theory of group neuronal selection*. New York: Basic Books.

Izquierdo-Torres, E., & Di Paolo, E. A. (2005). Is an embodied system ever purely reactive? In M. S. Capcarrere, A. A. Freitas, B. J. Bentley, C. G. Johnson, & J. Timmis (Eds.), *Advances in Artificial Life. Proceedings of the Eighth European Conference (ECAL 2005), Lecture Notes in Computer Science series* (pp. 252–261). Berlin: Springer Verlag.

Jen, E. (2002). Stable or robust? What is the difference? Retrieved 22 April 2008, from http://www.santafe.edu/~erica/stable.pdf

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

Llinás, R. R. (2001). *The I of the vortex. From neuron to self.* Cambridge, MA: MIT Press.

Maes, P. (Ed.). (1990). *Designing autonomous agents. Theory and practice from biology to engineering and back.* Cambridge, MA: MIT Press.

Matarić, M. J. (2002). Situated robotics. In *Encyclopedia of Cognitive Science* (Vol. 4, pp. 25–30). London: Nature Publishing Group, Macmillan Reference Limited.

Maturana, H., & Varela, F. (1980). *Autopoiesis and cognition: The realization of the living*. Dordrecht: Reidel.

McGeer, T. (1990). Passive dynamic walking. *International Journal of Robotics Research*, *9*(2), 62–82.

Millikan, R. G. (1984). *Language, thought and other biological categories*. Cambridge, MA: MIT Press.

Moreno, A., & Barandiaran, X. (2005). A naturalized account of the inside–outside dichotomy. *Philosophica*, *73*, 11–26.

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. *Evolution and Cognition*, *9*(1), 12–30.

Moya, A., Peretó, J., Gil, R., & Latorre, A. (2008). Learning how to live together: Genomic insights into prokaryote–animal symbiosis. *Nature Reviews Genetics*, *9*, 218–229.

Nicolis, G., & Prigogine, Y. (1977). *Self-organization in nonequilibrium systems*. New York: Wiley.

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.

Pfeifer, R., & Scheier, C. (1999). *Understanding intelligence*. Cambridge, MA: MIT Press.

Ruiz-Mirazo, K. (2001). *Physical conditions for the appearance of autonomous systems with open-ended evolutionary capacities*. Ph.D. dissertation, University of the Basque Country (complete version only available in Spanish at http://www.sc.ehu.es/PhBioGroup).

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

Ruiz-Mirazo, K., Peretó, 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.

Seth, A. K., & Edelman, G. M. (2004). Environment and behavior influence the complexity of evolved neural networks. *Adaptive Behavior*, *12*(1), 5–21.

Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., & Ishikawa, H. (2000). Genome sequence of the endocellular bacterial symbiont of pahids Buchenera sp. APS. *Nature*, *407*, 81–86.

Simons, P. J. (1981). The role of electricity in plant movements. *New Phytologist*, *87*, 11–37.

Steels, L. (1991). Towards a theory of emergent functionality. In J. Meyer & R. Wilson (Eds.), *Simulation of adaptive behavior* (pp. 451–461). Cambridge, MA: MIT Press.

Tononi, G., Edelman, G. M., & Sporns, O. (1998). Complexity and coherency: Integrating information in the brain. *Trends in Cognitive Science*, *2*(12), 474–484.

Trewavas, A. (2003). Aspects of plant intelligence. *Annals of Botany*, *92*, 1–20.

van Duijn, M., Keijzer, F. A., & Franken, D. (2006). Principles of minimal cognition. *Journal of Adaptive Behavior*, *14*(2), 157–170.

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

Ziemke, T. (2003). Embodied AI as science: Models of embodied cognition, embodied models of cognition, or both? In F. Iida, R. Pfeifer, L. Steels, & Y. Kuniyoshi (Eds.), *Embodied artificial intelligence* (pp. 27–36). Heidelberg: Springer-Verlag.

About the Authors

Xabier Barandiaran is a Ph.D. student at the University of the Basque Country. He is a graduate student in philosophy from the University of Deusto (Bilbao, Spain) and obtained an M.Sc. in evolutionary and adaptive systems at the University of Sussex (Brighton, UK). His main research areas are the philosophy of artificial life, neurophilosophy, enactive cognitive science, and naturalized epistemology.

Alvaro Moreno is Professor of Philosophy of Science at the University of the Basque Country, in Spain, where he has created a research group specialized in complex systems, philosophy of biology, and artificial life. He is the author of over a hundred publications and organizer of several international workshops on the relationship between biology and cognition. *Address*: Departamento de Logica y Filosofia de la Ciencia, Universidad del Pais Vasco UPV-EHU, Apartado 1249, 20080 San Sebastian, Spain.
E-mail: alvaro.moreno@ehu.es