Behavioral Adaptive Autonomy. A milestone on the Alife route to AI?

Xabier Barandiaran[1]
[1]Department of Logic and Philosophy of Science, FICE
UPV-EHU University of the Basque Country, Spain.
PO BOX 1249 20080 Donostia - San Sebastian
barandi@sf.ehu.es
http://www.ehu.es/ias-research/barandiaran

Abstract

While central to robotics, biology and cognitive science, the concept of *autonomy* remains still difficult to make operative in the realm of Alife simulation models of cognitive agents. Its deep significance as a transition concept between life and cognition (a milestone on the Alife route to AI) remains obscured in the intricate relation between metabolic/constructive processes and behavioral adaptive processes in living systems. Within a naturalized and biologically inspired dynamical approach to cognition a definition of *behavioral adaptive autonomy* is provided: homeostatic maintenance of essential variables under viability constraints through self-modulating behavioral coupling with the environment, hierarchically decoupled from metabolic (constructive) processes. This definition allows for a naturalized notion of behavioral adaptive functionality (that defines a proper level of modelling within Alife), structurally and interactively emergent: the mapping of the agent-environment system's state space trajectories into the viability subspace of the essential variables of the organism.

Introduction

While central to Alife (Varela and Bourgine, 1992), robotics (Maes, 1991), biology (Varela, 1979) and cognitive science (Christensen and Hooker, 2000), the concept of *autonomy* remains still difficult to make operative in the realm of Alife approaches to cognition[1]. In particular it is not yet completely clear whether an artificial cognitive autonomous agent could be built without the underlaying autopoietic autonomy being implemented; and it is not clear how could the cognitive autonomy of an artificial or simulated agent be measured or implemented. The notion of autonomy and its deep significance remains obscured in the intricate relation between metabolic/constructive processes and behavioral/cognitive processes in living beings and in the highly abstract conceptual framework in which it has been developed by Francisco Varela.

[1]We won't discuss here the notion of autonomy in relation to the origin and synthesis of life, this issue has long being discussed elsewhere (Ruiz-Mirazo and Moreno, 2000). Our main concern is the notion of autonomy in the cognitive domain and its interaction with basic (autopoietic) autonomy.

But far from being a neglectable term, the notion of autonomy has inspired a whole range of research projects within the Alife community and it does in fact capture the core of the conceptual shift behind most Alife research (biological grounding, self-organization, emergence, embodiment and situatedness). In fact, the subtitle of the First European Conference on Artificial Life ("towards a practice of autonomous systems") reflects the significance of the concept.

The main goal of this paper is to provide an specific definition of behavioral adaptive autonomy that can be implemented in Alife scientific practices and used to model adaptive behavior. The main thesis is that the hierarchical decoupling of the nervous system from metabolic constraints specifies the domain of behavioral adaptive dynamics, in this domain autonomy is defined as the capacity of the system to interactively maintain its essential variables under viability constraints.

We start (section 2) by briefly analyzing the variety of uses that the term *autonomy* has had in the literature and compiling (section 3) a set of key notions around autonomous approaches. Section 4 reconstructs the concept of basic (autopoietic) autonomy and the way it relates to functionality. We then move to specify the organization of the nervous system in the context of the whole organism (section 5) to end up defining autonomy and functionality in the dynamical framework of the behavioral adaptive domain (section 6). Finally section 7 discusses some implications of the present approach for Alife simulation models.

A quick overview of the literature

The term *autonomous robotics* has been used since the 90s (Maes, 1991) to refer to a set of engineering constraints on the construction and testing of robots, thus labeling a style of robotic research in cognitive science and engineering. Such constraints include conditions like no remote control of the agent, no external energy supply, mobility in the robot, no human intervention in robot task solving or real-time response in real-world environments. Close to situated robotics (Brooks, 1991), autonomous robotics highlights physicality, embodiment, situatedness and dynamicism versus abstract, virtual and formal approaches artificial intelligence in which agents operate in controlled formal or virtual environments or in toy like worlds without dynamical constraints. As a consequence of the real-world interaction of the robot the emphasis is often put on the *viability* of the robot as a task achieving agent: a self-generated and robust capacity to respond to environmental changes.

The practice of *autonomous robotics* has forced some engineers to go beyond the specification of a list of engineering constraints and to develop a more elaborated notion of autonomy that specifies the kind of interaction process that is established between the robot and its physical environment, the dynamic structure and properties of the controll mechanisms and the underlaying consequences for cognitive science and epistemology. That is the case of engineers like Tim Smithers (1997) or Randall Beer (1995; 1997) who have strongly criticized computational information processing approaches to cognition highlighting dynamicism, embodiment and situatedness or Eric Prem (1997) who has put the emphasis on *epistemic autonomy* "the system's own ability to decide upon the validity of measurements" (Prem, 1997) a process that cannot be reduced to formal aspects, given the physicality of the measuring process, its preformal nature.

This authors have been greatly influenced by the biologist Francisco Varela whose definition of autonomy is much more abstract and encompassing than its robotic application[2]. Autonomy is defined by Varela as an abstract systemic kind of organization; a kind of self-maintained, self-reinforced and self-regulated system dynamics resulting from a highly recursive network of processes that generates and maintains internal invariants in the face of internal and external perturbations. A process that defines its own identity; i.e. its unity as a system distinguishable from the surrounding processes. This abstract notion of autonomy is realized at different biological scales and domains defining different levels, and it is precisely the autonomy of each level what defines its specificity. As a paradigmatic example "life" is defined as a special kind of autonomy: *autopoiesis* or autonomy in the physical space. In turn "adaptive and cognitive behavior" is the result of a higher level of autonomy: that of the nervous system, producing invariant patterns of sensorimotor correlations and defining the behaving organism as a mobile unit in space (Varela, 1979; Varela and Bourgine, 1992; Varela, 1992).

[2] "Autonomous systems are mechanistic (dynamic) systems defined as a unity by their organization. *We shall say that autonomous systems are organizationally closed. That is, their organization is characterized by processes such that (1) the processes are related as a network, so that they recursively depend on each other in the generation and realization of the processes themselves, and (2) they constitute the system as a unity recognizable in the space (domain) in which the processes exist.*" (Varela, 1979, p.55, italics in the original).

But Varela's perspective on autonomy (although highly influential) has been recently criticized by its emphasis on closure[3] and the secondary role that system-environment interactions play in the definition and constitution of autonomous systems. Introducing ideas from complexity theory and thermodynamics authors like Bickhard (2000), Christensen and Hooker (2000), Collier (2002), and Moreno and Ruiz-Mirazo (2000), have defended a more specific notion of autonomy as a recursively self-maintaining far-from-equilibrium and thermodynamically open system. The interactive side of autonomy is essential in the definition: autonomous systems must interact continuously to assure the necessary flow of matter and energy for their self-maintenance. The philosophical consequences derived from the nature of autonomous systems are highlighted by these authors and summarized by Collier in the slogan: "No meaning without intention; no intention without function; no function without autonomy." (Collier, 2002). Autonomy is made the naturalized basis for functionality, intentionality, meaning and normativity. But it is not always clear what the relation is between this basic thermodynamic or constructive autonomy and neurally guided adaptive behavior. It is even argued that dynamical system theory (and thus computational simulation models in Alife) cannot capture the kind of organization that autonomy is (Christensen and Hooker, 2000) or that robots should be self-constructive in order to be "truly" autonomous (Ruiz-Mirazo and Moreno, 2000).

Key notions covered by autonomous approaches to cognition

In general and across the differences between the uses of the term autonomy and the consequences that (more or less explicitly) are derived from it, the notion of autonomy subsumes a set of key notions in cognitive science that have been pushing towards a paradigmatic shift. Among this key notions we find:

- **Biological grounding**: the idea that the understanding of cognition and adaptive behavior must be approached bottom-up at two levels: in terms of the evolution of cognitive capacities in natural history and in terms of the biological mechanisms (neural networks, bodies, neuroendocrine systems, etc.) that produce cognitive behavior.

- **Self-organization, complexity, emergence**: the idea that there is no central processor or homunculi that controlls behavior but a distributed and functionally integrated network of recursive processes from which a coherent behavior emerges as a global product of the system. The notion of autonomy assumes a high degree of complexity in the system introducing constraints on the possible

[3] It is precisely this emphasis on operational closure and its algebraic definition in Varela (1979) what makes controversial its application to dynamical modelling of biological systems.

analysis and functional localization and decomposition of structures.

- **Interactivism, embodiment, situatedness, dynamicism**: Cognition is a process whose development and realization cannot be decoupled from the embodied interaction processes in which it is situated. An autonomous approach assumes a dialectics between independence and structural coupling: an interactive construction of meaning and behavior in which embodiment and situatedness are taken to be essential features of cognition that are best captured by dynamical (rather than traditional computational) notions, thus introducing time and space dependant constraints as essential features on the generation of behavior.

- **Critics to GOFAI**: The use of the notion of autonomy is often associated with a profound critique to what has been the mainstrain paradigm in cognitive science: the view that cognitive processes are logical transformations of computational states bearing a representational relation with observer independent "states of affairs" in the world; where the representational relation is taken to be the mark of the mental and the program-like transformation rules between representational states the causally effective mechanisms in the production of behavior. From autonomous robotics to the philosophy of biology and cognition the approaches focused on autonomy have taken a different starting point, different theoretical primitives from which theories of cognition and adaptation have been built: complex dynamic networks, physically (and thermodynamically) embodied interactions, decentralized control systems, biologically grounded subsymbolic processes, to specifically address some of the problems that GOFAI approaches suffered at both practical-engineering and theoretical levels.

So far so good, the concept of autonomy subsumes a set of new approaches to cognitive science... But what else? Is *autonomy* just an umbrella label to cover an undetermined set of general constraints in robotic and cognitive science? Is it just a heavy-weighted metaphysical concept that only makes sense under the conceptual framework developed by Maturana and Varela? Or can it be conceptually and methodologically tunned in order to be introduced as a scientifically productive concept in empirical and synthetic research? The remaining of the paper will try provide an explicit and positive answer to this question.

Basic autonomy: the root for normative functionality

The origin of the word autonomy comes from the Greek *auto-nomos* (self-law). We can thus provide an intuitive first notion of autonomous systems as those producing their own laws[4]. But this notion requires a previous notion of self: autonomous systems must first produce their own identity; i.e. autonomous systems are primarily those whose basic organization is that of a self-sustaining, self-constructing entity over time and space. Their being is a process of recursive production of their constituting structure: a recursivity that generates a self. It is on top of this sense of *basic autonomy* that other levels of autonomy will appear in natural systems.

Basic Autonomy

Basic autonomy (Ruiz-Mirazo and Moreno, 2000) is the organization by which far from equilibrium and thermodynamically open systems adaptively generate internal and interactive constraints to modulate the flow of matter and energy required for their self-maintenance. Two equally fundamental but distinct aspects of basic autonomy can be distinguished:

a) ***constructive*:** generation of *internal* constraints to control the internal flow of matter and energy for self-maintenance. In this sense the autonomous (autopoietic) system can be understood as a highly recursive network of processes that produces the components that constitute the network itself (Maturana and Varela, 1980). Metabolism is the expression of this constructive aspect.

b) ***interactive*:** the generation of *interactive* constraints modulating the boundary conditions of the system to assure the necessary flow of energy and matter between the system and its environment[5]. Active transport through the membrane of a cell, controll of behavior or breathing are characteristic examples of this interactive constraint generation.

On this basis we can define *constructive closure* as the satisfaction of constructive constraint generation and *interactive closure* as the satisfaction of interactive constraint generation for self maintenance.

In general autonomy, at any level, will always present a twofold dialectics between internal recursive process and the necessary interactions to maintain them. In autonomous systems internal dynamics are more cohesive and integrated (more complex) than the interactive dynamics it sustains, thus producing a dynamic control asymmetry laden to the side of the autonomous agent.

The origin of functionality and normativity

What defines functionality in autonomous systems is the satisfaction of closure conditions (Collier, 2002) of internal and

[4] Although strictly speaking new physical *laws* will never be created by an organism (or any other macroscopic system) constraints can be generated that specify and govern its behavior.

[5] Unlike dissipative structures which hold their organization only under a restricted set of external conditions that the system cannot modify.

interactive processes. A process (internal or interactive) is functional if it contributes to the global self-maintenance of the system.

In turn functions become *normative*[6] by means of the *dynamic presupposition* of that process in the overall organization of the system (Christensen and Bickhard, 2002) since constructive and interactive functional processes are *the condition of possibility* of autonomous systems (as far from equilibrium and recursively self-maintained systems). The strength of an autonomous perspective resides in the fact that it is the very system who determines and specifies it, not an external observer attributing functions to structures and imposing a normative criteria according to its correspondence with states of affairs in the world nor on the basis of the agents evolutionary history or its structural matching with the environment.

The organization of the nervous system

Following Moreno and Lasa (2003) if an autonomous system needs to recruit the same infrastructure to achieve both constructive and interactive closure then the space of possible biological organization becomes highly constrained. This happens because metabolic reactions (constructive processes) are slower than the reaction times required for available interactive closure opportunities, specially those available for fast body movements (motility) in big organisms (where the relative difference in velocity between metabolic reactions and body movement increases). Thus if a subset of the interactive closure is achieved and controlled by a structure that instantiates processes which are dynamically decoupled from the constructive ones, the space of viable system organization is expanded. That's precisely the origin of the nervous system: the new opportunities for survival offered by the hierarchical decoupling of the nervous system, i.e. behavioral control decoupled from metabolic (constructive) constraints. The relation between metabolic constructive processes (M) and the nervous system (NS) is characterized by four properties that specify the organization[7] of the nervous system in the context of the whole organism:

1. **Hierarchical decoupling of the NS from M:** The NS is hierarchically decoupled from M by the:
   (a) **Bottom-up, local, constructive causation of the NS by M:** constructive/metabolic processes produce and maintain the architecture of the nervous system (neural cells, synapses, myelin, etc.) thus sustaining a new dynamical domain, new variables and relations between variables: the NS. The constructive nature of this causation establishes the *hierarchical* aspect of the decoupling.
   (b) **Dynamic underdetermination of NS by M:** the dynamic state of the NS is underdetermined by metabolic dynamics, i.e. neural dynamics are enabled but not determined by the metabolic production of the neural architecture (this underdetermination specifies the *decoupling* side of the relation).

2. **Downward causal dependency of M on NS:** Because the NS performs interactive functionality for the self-maintenance of the system, M depends on the proper functioning of NS; i.e. the organism's survival depends on neurally controlled behavior.

3. **Global and dynamic meta-regulation of NS by M:** Although dynamically underdetermined by M, because the NS's functionality is defined by its interactive contribution to self-maintenance (and this must ultimately be evaluated by M) M establishes the metaestability condition for the NS. M does not directly evaluate the NS's dynamics but the interactive closure: i.e. the input of matter and energy it gets from the environment. But this meta-regulation, again, underdetermines the dynamics of the NS. Metabolism only indicates if a particular coupling is successful or not in the satisfaction of interactive closure conditions, but does not determine which one of all the possible viable/adaptive couplings should the NS undergo.

4. **Internal cohesive dynamics of the NS:** The other side of the metabolic constructive and meta-regulatory underdetermination of the NS's dynamic state is the recursive capacity of the NS to maintain invariant patterns under internal and external perturbations; i.e. its capacity for self-generated cohesion, the degree in which the system's internal dynamics are more complex than the interactive flow so that the former control the later to compensate for internal and external perturbations[8].

[6]Normativity refers to the value attribution 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 believe, 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 of nature, an asymmetry that is not present in any of the fundamental laws of physics. But, although alien to fundamental physics, *normativity* is an essential component of biology and cognitive science (and consequently for Alife and AI): whether an structure or interaction is adaptive or maladaptive for an organism is a value judgment that a scientist engaged in the study of living and intelligent systems must do. A judgment that must be justified in naturalistic terms; i.e. in the very organization of the system under study and not from a set of value preferences in the observer scientist.

[7]The identity characterizing properties of a system, i.e. the set of properties that identify a system as being a member of class.

[8]This is close to what Varela refers to as "operational closure" although we believe that internal cohesion is achieved through interaction processes rather than through internal recursivity alone: i.e. closed sensorimotor loops are integrated in the recursive functioning of neural dynamics

We can now abstract a second domain in biological systems (hierarchically decoupled from basic autonomy): *the domain of the organism's behavioral adaptive dynamics*, specified by the organization of the nervous system. This new dynamic domain, decoupled from local metabolic processes, provides a qualitative lower level boundary for the characterization of the specificity of cognition and allows for specific dynamical modelling of adaptive behavior. It is in this modelling domain that we will be able to define behavioral adaptive autonomy and thus a new level of functionality (properly cognitive but still biologically grounded).

Dynamical modelling of autonomy and functionality in the behavioral domain

Dynamically considered metabolism only acts as a set of control parameters for the nervous system; the behavioral domain is dynamically blind to metabolism's constructive functioning (although it has to be sensible to global metabolic conditions). Thus the constructive processes of basic autonomy can be modelled as a set of essential variables which tend to stay away from equilibrium; representing the cohesive limits of constructive processes and their interactive closure conditions. A similar approach was already taken by Ashby (1952) half a century ago (from whom we have taken the term essential variables) and recently recovered by Beer (1997) and Di Paolo (2003) in (evolutionary) simulation modelling of adaptive behavior. The dynamical autonomy of the behavioral domain allows for a naturalistically justified assumption of dynamical system theory (DST) as the proper conceptual framework to think about autonomy and cognition in this domain. If we model: a) the agent's NS and the environment as coupled dynamical systems (situatedness), b) coupled through sensory and motor transfer functions (embodiment), and c) the metabolic processes as essential (far from equilibrium) variables only controllable from the environment and signalling the NS; we get that functionality and autonomy can be redefined in the behavioral domain.

Behavioral adaptive autonomy

In the behavioral domain thus considered, a new level of autonomy can be described, hierarchically decoupled but interlocked with basic (metabolic) autonomy: *behavioral adaptive autonomy*.

We can now, in dynamical terms, explicitly define *behavioral adaptive autonomy* as:

> homeostatic maintenance of essential variables under viability constraints [**adaptivity**] through a self-modulating behavioral coupling with the environment [**agency**], hierarchically decoupled from metabolic (constructive) processes [**domain specificity**].

This definition highlights three main aspects of behavioral adaptive autonomy:

Adaptivity: The "homeostatic maintenance of essential variables under viability constraints" condition assures a naturalized and autonomous criteria for (adaptive) functionality. Adaptivity is thus defined from the perspective of the maintenance of the organism, not from the perspective of structural adequation between the organism and the environment. It is not the organism that matches the environment in a given prespecified way, but on the contrary it is through the particular way in which the agent satisfies the homeostatic maintenance of essential variables that an adaptive environment (a world) is specified cut out from a background of unspecific physical surroundings. Next section will further analyze the the nature of behavioral adaptive functionality thus considered.

Agency: The "self-modulating behavioral coupling" condition for behavioral adaptive autonomy specifies the *agency* of the organism in the adaptive process. "Self-modulation" is the consequence of the cohesive dynamics of the nervous system by which the it's dynamics are more complex than the interactive ones in the generation of the internal invariants (the homeostatic maintenance of essential variables under viability constraints). The notion of self-modulation refers to this controll asymmetry in the production of behavior and that's precisely what we call agency. It can't be otherwise, if the state of essential variables is only accessible for the agent (through internal sensors: level of glucose, feeling of hot, pain, etc.) the homeostatic regulation must be guided by the agent's nervous system and not by the environment. Thus the NS needs to evaluate it's structural coupling through value signals from the essential variables. This way a *value system* guided by the state of essential variables and acting as metaestability condition for structural plasticity of sensorimotor transformations becomes a fundamental component of behavioral autonomy, and a defining component of agency. The higher the agent's capacity for adaptively guided self re-structuring (plasticity) the higher it's behavioral adaptive autonomy and hence its agency[9].

Domain specificity: The hierarchical decoupling of the nervous system from metabolic processes provides a naturalized criteria for the domain specificity of behavioral autonomy, distinct from other adaptive domains in nature (bacterian networks, plants, etc.). This domain specificity should not be considered as independency but as hierarchical decoupling (explained above), which allows for a justified specific modelling of behavioral autonomy separated from local constructive aspects. Nonetheless it should be noted that two kinds of autonomy are interlocked here: basic autonomy and behavioral autonomy. Both domains are mutually required, the behavioral domain satisfies interactive closure of basic autonomy and basic autonomy constructs the body and neu-

[9] By this condition external contributions to adaptation (such as parents or artificially induced constraints in toy-like worlds), although functional for the agent, would be excluded from the domain of autonomous adaptation.

ral variables defining the NS's architecture. At the same time basic autonomy acts as a meta-regulator of the NS's dynamics.

Behavioral adaptive functionality

Functionality, in the behavioral domain thus considered, can be defined as the homeostatic effect of an interaction process on the maintenance of essential variables under viability constraints and, more specifically, as the mapping of the interactive trajectories (in the agent-environment coupled dynamic space) into the state space of the essential variables. Normativity is transitive from basic autonomy to the behavioral domain through the maintenance of essential variables under viability constraints. Thus normative functionality (adaptation) is *the mapping of the agent-environment coupled system's state space trajectories into the viability subspace of the essential variables*.

Because this definition of function does not compromise any structural decomposition in functional primitives (unlike traditional functionalism), a dynamical approach to behavioral functionality can hold two kinds of emergence[10]:

a) *Internal emergence*: It appears when the agent's internal structure is causally integrated (and the NS often is), i.e. interactions between components are non-linear and components are highly inter-connected, functional decomposition of components (localization) is not possible. The functionality of the system *emerges* from local non-linear and recursive interactions between components.

b) *Interactive emergence*: Because essential variables are non-controlled variables for the agent, functionality is interactively emergent (Steels, 1991; Hendriks-Jansen, 1996), not in the trivial sense that essential variables need external input, but in the sense that achieving this often requires closed sensorimotor loops for the agent to enact the necessary sensorimotor invariants to control essential variables.[11]

What this double emergent condition shows is that the way the specific adaptive function is achieved involves a dynamic coupling between agent and environment where no particular decomposable structure of the agent can be mapped into functional components: functionality is the outcome of an interaction *process* (that can be modified by the cognitive agent according to its perceived satisfaction of closure conditions).

[10]We are here talking of weak emergence in the sense of an holistic, recursive and distributed causal structure that produces a global ordered/invariant pattern. We are not arguing for a strong or ontological emergence that defends the appearance of a new property or object non reducible to the underlying processes.

[11]Very often interactive emergence reinforces structural emergence because "interactions between separate sub-systems are not limited to directly visible connecting links between them, but also include interactions mediated via the environment" (Harvey et al., 1997, p.205)

Discussion

Now, the problem with behavioral adaptive autonomy is the problem of a higher characterization and development of its understanding, specially in relation to its self-regulating, emergent and complex nature which does not allow for a localizationist program to succeed: i.e. functional and structural decomposition of components and aggregative causal abstraction of mutual relations (Bechtel and Richardson, 1993). When localizationist strategies are thrown away the locus of the research enquiry regarding the nature and origin of cognition and adaptation is displaced towards: a) the specification of the dynamic structure of lower level mechanisms capable of implementing behavioral adaptive autonomy (i.e. capable of self-restructuring cohesive and recursive dynamics); and b) the search for the nature of intermediate explanatory patterns between the agent-environment structural coupling and the maintenance of essential variables under viability constraints: traditional explanatory concepts (such as information, representation, memory, processing, etc.) should be dynamically grounded if introduced at all in the proposed framework.

A-life and, more specifically, evolutionary simulation modelling[12] becomes a mayor research tool here through the synthesis and experimental manipulation and analysis of the behavior generated by embodied and situated DRNNs (dynamic recurrent neural networks). The simulation model acts as an artifactual blending between lower level neural mechanistic concepts and the global functional conceptualization of behavior (Barandiaran and Feltrero, 2003).

An interesting line of research has recently been proposed by Di Paolo (2003) in this direction. Di Paolo argues that behavior itself is underdetermined by survival conditions and proposes *habit formation* as the origin of intentionality. Habits are self sustaining dynamic structures of behavioral patterns, sensorimotor invariants homeostatically maintained by neural organization. Homeostatically controlled synaptic plasticity (Turrigliano, 1999) could be a relevant neural organization leading to such autonomy of behavioral patterns; as demonstrated by Di Paolo (2000). Although habit formation does not necesarily address the issue of the relation between metabolic and nervous autonomy Di Paolo points towards a fundamental step forward in current research trends: that structural coupling (and closed sensorimotor loops) is not all there is in a dynamical and situated approach and that a robust capacity of the agent to evaluate and restructure its coupling is the way to follow to achieve progresively higher levels of cognitive autonomy and intentionality. If Alife is to throw some light on the origin of cognition and adaptive behavior, far from equilibrium essential variables and value systems capable of specifying stability conditions for a given dynamical coupling with the

[12]Evolutionary robotics (Harvey et al., 1997; Nolfi and Floreano, 2000) and Randall Beer's minimally cognitive behavior program (Beer, 2004) being the major exponents here.

environment should be introduced in the simulation models. In particular essential variable based fitness functions in evolutionary simulation modelling are a particular instantiation of behavioral, internal and implicit fitness functions which (according to Floreano and Urlezai —2000) shall produce highlyl self-organized controll systems. This principles for evolutionary simulation modelling of autonomous agents were succesfully implemented in a foraging task with alternate profitability sources (Barandiaran, 2002).

In addition to this synthetic bottom-up methodology other analytic tools should be theoretically tunned. Complexity measures to understand functional integration in neural processes (Tononi et al., 1998) are producing interesting results, an could be used to better characterize the cohesive nature of the NS. An early exploratory example of such methodology is provided by Seth (2002), fusioning both evolutionary simulation modelling and complexity measures (using dynamical graph theory) of neural network dynamics to show that at least in one case, complex behavior requires complex mechanisms.

Conclusion

A wide use of the term autonomy is found in the Alife literature: from a set of engineering constraints in robotics to a fundamental organizational principle in biology. In relation to behavior and cognition it is not clear how to operationalize the term and what the relation is between behavioral autonomy and basic (autopoietic) autonomy. We have seen that the particular organization of the nervous system allows for a specific modelling domain of cognition and adaptive behavior: the domain of behavioral adaptive dynamics. Interlocked with basic/metabolic autonomy (through the requirement to actively maintain essential variables under viability constraints) *behavioral adaptive autonomy* becomes a process of cohesive maintenance of internal invariants through continuous interaction loops, which requires, in turn, a functionally integrated and plastic neural organization with a higher internal dynamic complexity than that established between the organism and its environment.

By providing an explicit definition of behavioral autonomy and functionality in dynamical terms we hope to have contributed something to the simulation modelling approach that traces the Alife route to AI; to the understanding of the transition that goes from basic forms of life to adaptive behavior and cognition.

References

Ashby, W. (1952). *Design for a Brain. The origin of adaptive behaviour*. Chapman and Hall, 1978 edition.

Barandiaran, X. (2002). Adaptive Behaviour, Autonomy and Value Systems. Master's thesis, COGS, University of Sussex, Brighton, UK. URL: http://citeseer.nj.nec.com/barandiaran02adaptive.html.

Barandiaran, X. and Feltrero, R. (2003). Conceptual and methodological blending in cognitive science. The role of simulated and robotic models in scientific explanation. In *Volume of abstracts of the 12th International Congress of Logic, Methodology and Philosophy of Science, Oviedo (Spain).*, page 171.

Bechtel, W. and Richardson, R. (1993). *Discovering Complexity. Decomposition and Localization as strategies in scientific research.* Princeton University Press.

Beer, R. (1995). A dynamical systems perspective on autonomous systems. *Artificial Intelligence*, (72):173–215.

Beer, R. D. (1997). The Dynamics of Adaptive Behavior: A research program. *Robotics and Autonomous Systems*, 20:257–289.

Beer, R. D. (2004). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behaviour*. in press.

Bickhard, M. H. (2000). Autonomy, Function, and Representation. In Etxeberria et al. (2000), pages 111–131.

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

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

Christensen, W. and Hooker, C. (2000). Autonomy and the emergence of intelligence: Organised interactive construction. In Etxeberria et al. (2000), pages 133–157.

Collier, J. (2002). What is autonomy? In Dubois, D., editor, *International Journal of Computing Anticipatory Systems. Partial proceedings of the Fifth International Conference CASYS'01 on Computing Anticipatory Systems, Lige, Belgium, August 13–18, 2001*.

Di Paolo, E. (2000). Homeostatic adaptation to inversion of the visual field and other sensorimotor disruptions. In Meyer, J.-A., Berthoz, A., Floreano, D., Roitblat, H., and Wilson, S., editors, *From Animals to Animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*, pages 440–449. Harvard, MA: MIT Press.

Di Paolo, E. (2003). Organismically inspired robotics. In Murase, K. and Asakura, T., editors, *Dynamical Systems Approach to Embodiment and Sociality*, pages 19–42. Advanced Knowledge International, Adelaide, Australia.

Etxeberria, A., Umerez, J., and Moreno, A., editors (2000). *Communication and Cognition – Artificial Intelligence. Special issue on "The contribution of artificial life and the sciences of complexity to the understanding of autonomous systems"*, volume 17 (3–4).

Floreano, D. and Urzelai, J. (2000). Evolutionary robots with online self-organization and behavioural fitness. *Robotics and Autonomous Systems*, 13:431–443.

Harvey, I., Husbands, P., Cliff, D., Thompson, A., and Jakobi, N. (1997). Evolutionary Robotics: the Sussex Approach. *Robotics and Autonomous Systems*, 20:205–224.

Hendriks-Jansen, H. (1996). In praise of interactive emergence, or why explanations don't have to wait for implementations. In Boden, M., editor, *The Philosophy of Artificial Life*, pages 282–299. Oxford University Press, Oxford.

Maes, P., editor (1991). *Designing Autonomous Agents*. MIT Press.

Maturana, H. and Varela, F. (1980). Autopoiesis. The realization of the living. In Maturana, H. and Varela, F., editors, *Autopoiesis and Cognition. The realization of the living*, pages 73–138. D. Reidel Publishing Company, Dordrecht, Holland.

Moreno, A. and Lasa, A. (2003). From Basic Adaptivity to Early Mind. *Evolution and Cognition*, 9(1):12–24.

Nolfi, S. and Floreano, D. (2000). *Evolutionary Robotics: The Biology, Intelligence and Technology of Self-Organizing Machines*. MIT Press.

Prem, E. (1997). Epistemic autonomy in models of living systems. In *Proceedings of the Fourth European Conference on Artificial Life.* MIT Press, Bradford Books.

Ruiz-Mirazo, K. and Moreno, A. (2000). Searching for the Roots of Autonomy: the natural and artificial paradigms revisited. In Etxeberria et al. (2000), pages 209–228.

Seth, A. K. (2002). Using dynamical graph theory to relate behavioral and mechanistic complexity in evolved neural networks. Unpublished. Url: http://www.nsi.edu/users/seth/Papers/nips2002.pdf.

Smithers, T. (1997). Autonomy in Robots and Other Agents. *Brain and Cognition*, (34):88–106.

Steels, L. (1991). Towards a Theory of Emergent Functionality. In Meyer, J. and Wilson, R., editors, *Simulation of Adaptive Behaviour*, pages 451–461. MIT Press.

Tononi, G., Edelman, G., and Sporns, O. (1998). Complexity and coherency: integrating information in the brain. *Behavioural and Brain Sciences*, 2(12):474–484.

Turrigliano, G. (1999). Homeostatic plasticity in neuronal networks: The more things change, the more they stay the same. *Trends in Neuroscience*, 22:221–227.

Varela, F. (1979). *Principles of Biologicall Autonomy*. North-Holland, New York.

Varela, F. (1992). Autopoiesis and a biology of intentionality. In McMullin, B., editor, *Proceedings of a workshop on Autopoiesis and Percetion*, pages 4–14.

Varela, F. and Bourgine, P. (1992). Towards a Practice of Autonomous Systems. In Varela, F., editor, *Towards a Practice of Autonomous Systems. Proceedings of the First European Conference on Artificial Life*, pages xi–xvi.

Acknowledgments

I want to acknowledge the Basque Government for financial support through the doctoral fellowship BFI03371-AE and the University of the Basque Country for the research project grant 9/UPV 00003.230-13707/2001. I want to thank Alvaro Moreno and Ezequiel Di Paolo for usefull comments and discussion of the topics presented here.