Autonomy and Adaptive Behaviour
Towards a naturalized and biologically inspired definition of behavioural adaptive autonomy

Xabier Barandiaran
xabier @ barandiaran.net
http://barandiaran.net

Autonomía Situada
grey-walter @ sindominio.net
http://sindominio.net/autonomiasituada

08–07–03

Abstract

Within a naturalized dynamical approach to adaptive behaviour and cognition, *behavioural adaptive autonomy* is defined as: homeostatic maintenance of essential variables under viability constraints through self-modulating behavioural coupling with the environment, hierarchically decoupled from metabolic (constructive) processes. This definition allows for a naturalized notion of normative functionality, structurally and interactivelly emergent. We argue that artificial life techniques such as evolutionary simulation modelling provide a workable methodological framework for philosophical research on complex adaptive behaviour.

Keywords

Autonomy, adaptive behaviour, philosophy of biology, philosophy of cognitive science, cognition, dynamical systems, naturalization.

Copyleft 🄯

Autonomy and Adaptive Behaviour. *Towards a naturalized and biologically inspired definition of behavioural adaptive autonomy.* v.0.8

Versions

v.0.8	08–07–03	

Formats and Sources

html	http://barandiaran.net/textos/auto/auto.html
pdf	http://barandiaran.net/textos/auto/auto.pdf
ps	http://barandiaran.net/textos/auto/auto.ps
sources	http://barandiaran.net/textos/auto/

Cite

Xabier Barandiaran (2003) Autonomy and Adaptive Behaviour. Towards a naturalized and biologically inspired definition of behavioural adaptive autonomy. v.0.8. **url**:
http://barandiaran.net/textos/auto/auto.pdf

Contents

1 Introduction

In this paper we will try:

1. to explain the significance of autonomy (dynamically considered) for adaptive behaviour and cognitive science (section 2),

2. to clarify the different aspects of autonomy in relation to adaptive behaviour (section 3),

3. to specify a notion of *behavioural adaptive autonomy* and normative functionality in the framework of a dynamical approach to cognition and adaptive behaviour (section 4),

4. to justify (evolutionary) simulation modelling as a workable methodological framework for the study of behavioural adaptive autonomy (section 5) and

5. to extract some consequences of *behavioural adaptive autonomy* for a naturalized definition of cognition while illustrating some discusion with recent work on evolutionary simulation modelling (section 6).

The main thesis is:

- that *adaptive behavioural autonomy* shall be defined as:

  > homeostatic maintenance of essential variables under viability constraints through self-modulating behavioural coupling with the environment, hierarchically decoupled from metabolic (constructive) processes

- and that this definition of behavioural adaptive autonomy should serve as a lower boundary for a naturalist characterization of cognition as emergent from life but distinct from it.

In addition we provide a set of formal definitions at the end of the paper.

2 Why: Autonomy as a relevant concept in adaptive behaviour and cognitive science

The term *autonomy* and *autonomous* has been largely used in cognitive science and robotics (Maes, 1991) to describe an agent embodied and situated in the 'real world' and without external energy supply; we consider that a deeper sense of autonomy (as self-maintenance) allows for a richer characterization of cognition and adaptive behaviour. In this line the work of Varela (1979) provides a deeper sense of autonomy, recently developed by Ruiz-Mirazo and Moreno (2000) in the dimension of basic (autopoietic) autonomy and by Christensen and Hooker (2002) in the dimension of adaptation and cognition conceptualized through the notion of self-directedness; while Bickhard (2000) has analysed the consequences of autonomy for functional and representational normativity[1]. FALTA ENRIQUECER ESTE PÁRRAFO CON DIFERENTES AUTORES Y UNA BREVE HISTORIA DEL USO DEL CONCEPTO DE AUTONOMÍA EN LA SIMULACIÓN DE CONDUCTA ADAPTATIVA.

But the relation between basic (autopoietic or self-maintaining) autonomy and behavioural or cognitive autonomy deserves some clarification in order to be introduced in dynamical system theory (as a conceptual framework) and evolutionary simulation modelling of adaptive behaviour (as a workable methodological framework). This is the main goal of this paper. At the same time by conceptualizing and modelling autonomy within the dynamical approach to cognitive science (van Gelder, 1998) and adaptive behaviour (Beer, 1997) a number of important goals could be achieved:

- To provide a autonomous normative criteria to interpret and evaluate adaptive and cognitive functionality, solving the frame of reference problem (Clancey, 1989) of computational functionalist approaches (Block, 1996).
- To naturalize such normative criteria on the dynamical organization of neural and interactive processes (and their relation with self-maintenance) giving rise to adaptive/cognitive behaviour as proposed by Bickhard, Christensen and Hooker (Cristensen and Hooker, 1999; Christensen and Bickhard, 2002; Christensen and Hooker, 2002); without recursion to evolutionary functionalism (Millikan, 1989a,b) or an absolute external observer in order to attribute structural/functional relations within the organism and between the organism and its environment.
- To integrate mechanistic, embodied and situated (interactive) explanations without recursion to pre-specified functional/behavioural primitives, thus integrating behavioural and structural complexity in a workable methodological framework. This will satisfy holistic, organismic or organizational criticisms (Gilbert and Sarkar, 2000) to traditional functionalist perspectives on cognition while providing synthetic and analytic criteria for advances in scientific research (thus avoiding the often questioned solipsist danger of such approaches).

[1] Contributions of authors to the development of the concept of autonomy (in relation to autopoiesis, cognition, functionality, normativity, etc.) did not happen in isolation, cross-referencing and collaboration has been a common practice so that specific contributions as outlined above shouldn't be taken too rigorously.

3 What: Autonomy, behaviour and adaptation

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[2]. 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.

3.1 Basic Autonomy, the root for normative functionality

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 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 main 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 systems and its environment for self maintenance (unlike dissipative structures which hold their organization only under a restricted set of external conditions that the system cannot modify). The membrane of a cell, controll of behaviour 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.

3.1.1 Functionality

It is the satisfaction of closure conditions that defines the function (Collier, 1999) of internal and interactive processes. Functionality is, thus, picked up at the level of their contribution to self-maintenance and not, as evolutionary functionalism proposes, at the level of selective history; nor, as computational functionalists defended, as externally (heteronomously) interpreted

[2] Although new physical laws will never be created by an organism, or any other system, it can always generate new constraints and internal control mechanisms

causal relations between computational states (and, when cognition is involver, their representational relation with external "states of affairs"). For, of course, contribution to self-maintenance is evolutionarily advantageous; but autonomy is to be seen not as a pure outcome of evolutionary processes but as the condition of possibility of such process.

3.1.2 Normativity

Functions become *normative* by means of the *dynamic presupossition* of that process in the overall organization of the system (Bickhard, 2000; Christensen and Bickhard, 2002). In other words because constructive and interactive functional processes are *the condition of possibility* of autonomous systems (as far from equilibrium systems) normativity emerges in nature. Normative asymmetry (adaptive/maladaptive, true/false, etc.) is transitive from the asymmetry between energy-well stability (rocks, atoms, etc.) and far-from-equilibrium stability. Functional normativity is thus naturalized: 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. Computational and evolutionary functionalism provide, both, heteronomous sources of normativity, unabling the very system for error detection behaviour or any other kind of normative re-organization.

3.2 The hierarchical decoupling of the nervous system

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 mouvements (motility) in big organisms (where the relative difference in velocity between metabolic reactions and body mouvement 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. behavioural control decoupled from metabolic (constructive) constraints. Following (Moreno and Lasa, 2003) in this argument, the relation between metabolic constructive processes (M) and the nervous system (NS) is characterized by:

1. **Hierarchical decoupling of the NS from M:** The NS is hierarchically decoupled from M by the:

   (a) **Bottom-up, local, constructive causation of NS by M:** constructive processes produce 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 (*decoupling*).

2. **Downward causal dependency of M on NS:** Because the NS performs interactive functionality for the self-maintenance of the system, M depends of the proper functioning of NS.

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 NS's dynamics but the interactive closure: i.e. the input of matter and energy it gets from the environment.

We can now abstract a second domain in biological systems (hierarchically decoupled from basic autonomy): *the domain of the organism's behavioural adaptive dynamics*, specified by the dynamical coupling between the embodied nervous system and the environment and the metabolic meta-evaluation of that coupling.

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 behaviour. It is in this modelling that we will be able to define behavioural adaptive functionality and thus a new level of autonomy.

3.3 Autonomy and functionality in a dynamical approach to adaptive behaviour

Dynamically considered metabolism only acts as a set of control parameters for the nervous system; the behavioural domain is dynamically blind to metabolism's constructive functioning. 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 behaviour. The dynamical autonomy of the behavioural domain allows for a

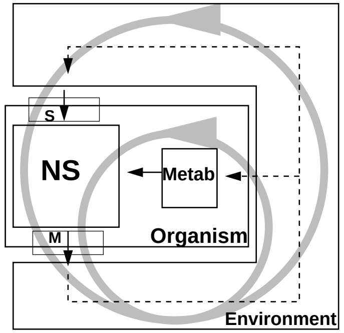

Figure 1: *Kinetic graph of a dynamical modelling of adaptive behaviour. Adapted from Ashby (1952): a closed sensorimotor loop (controlled by the nervous system) traverses the environment affecting metabolic processes, which, in turn feeds-back to the nervous system acting as a meta-regulatory mechanism. Embodiment is modelled by sensory (S) and motor (M) surfaces.*

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 behavioural domain (see figure 1).

3.3.1 Behavioural adaptive autonomy

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

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

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

This definition highlights three main aspects of behavioural adaptive autonomy:

Adaptivity: Homeostatic maintenance of essential variables under viability constraints assures a naturalized and autonomous criteria for (adaptive) functionality. Next section will further analyze the consecuences of functionality thus considered.

Agency: Self-modulation or self-restructuring of the interactive coupling provides a criteria for autonomous functionality (agency), excluding external contributions to adaptation such as parents' care. Because 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 behavioural autonomy, and a defining component of agency. The higher the agent's capacity for adaptively guided self re-structuring (plasticity) the higher it's behavioural adaptive autonomy and hence its agency.

Domain specificity: The hierachical decoupling of the nervous system from metabolic processes provides a naturalized criteria for the domain specificity of behavioural autonomy, distinct form 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 behavioural autonomy separated from local construtive aspects. Two kinds of autonomy are interlocked here: basic autonomy and behavioural autonomy. Both domains are mutually required, the behavioural domain satisfies interactive closure of basic autonomy and basic autonomy constructs the body and neural variables defining the behavioural domain while acting as a modulator of the structure of behavioural autonomy (see figure 2).

3.3.2 Behavioural functionality

Functionality, in the behavioural domain thus considered, can be defined as the mapping between agent-environment coupling and the essential variables. Normativity is transitive from basic autonomy to the behavioural domain through the maintenance of essential variables under viability constraints. Thus normative functionality (adaptation) is *a mapping between agent-environment coupling and the maintenance of essential variables under viability constraints*.

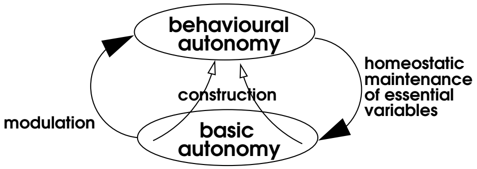

Figure 2: *Basic and behavioural autonomy interlocked.*

Because this definition of function does not compromise any structural decomposition in functional primitives (unlike traditional functionalism), a dynamical approach to behavioural functionality can hold two kinds of emergence[3]:

a) ***Structural emergence*****:** If the agent's 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 (localisation) is not possible. The functionality of the system *emerges* from local non-linear interactions between components. FALTA EXPANDIR UN POCO CON LA DEF DE EMERGENCIA DE BECHTEL

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.[4]

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. This way holism is preserved as a characterizing condition in complex systems but functionality can still be explicitly defined. Localisation is not a valid explanatory strategy but explanation is still possible (at least in principle). And this is so because functionality is the outcome of a *process* (not a structural relation between components): the dynamical coupling between agent and environment.

[3] We are here talking of weak emergence in the sense of an holistic, recursive and distributed causal structure.

[4] 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)

herre

4 How: Methodological perspectives on the study of adaptive behavioural autonomy

Now, the problem with behavioural 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 localisationist program to succeed: i.e. functional and structural decomposition and aggregative causal abstraction of mutual relations (Bechtel and Richardson, 1993).

When localisation is thrown away the locus of philosophical enquiry regarding the nature and origin of cognition (if a naturalist and biologically inspired philosophical approach is to be adopted) is displaced towards;

- the specification of the dynamic organization of lower level mechanisms capable of implementing behavioural adaptive autonomy and FALTA EXTENDER

- 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 concepts (information, representation, memory, processing, etc.) should be dynamically grounded.

This task is genuinely philosophical and distinct from specific modelling of biologicall targets.

A-life (Langton, 1996; Dennet, 1995; Moreno, 2000) and, more specifically, evolutionary simulation modelling[5] becomes a mayor philosophical tool here, not for a mere synthesising of behavioural autonomy but for philosophical research through opaque though experiments (Di Paolo et al., 2000) with conceptually (dynamically) complex systems, produced, implemented and manipulated in a computer. The simulation acts as an artefactual blending (Fauconnier and Turner, 1998) between lower level neural mechanistic concepts and global functional conceptualization of behaviour (Barandiaran and Feltrero, 2003).

Evolutionary simulation modelling works by:

1. Definition of a set of body, environment and neural structures (unspecified on their parameter values). Neural structures are abstractions

[5] Evolutionary robotics (Harvey et al., 1997; Nolfi and Floreano, 2000) and Randall Beer's minimally cognitive behaviour program (Beer, 1996; Slocum et al., 2000; Beer, 2001) being the major exponents here.

of a set of lower level neural mechanisms from neuroscientific models (functionally unspecific), and body structures are a set of robotic idealizations.

2. Artificial evolution of parameters according to a given fitness function.
3. Reproduction/simulation of system behaviour with numerical methods allowing for qualitative analysis of complex dynamical systems.

Highly connected CTRNNs (continuous time recurrent neural networks) are used in evolutionary robotics to model de agent's control architecture. The dynamics of such networks are highly complex, capable (in principle) to emulate any other dynamical system with a finite number of variables (Funakashi and Nakamura, 1993).

Because the lower level mechanisms are functionally unspecific and artificial evolution is used to achieve emergent functionality, evolutionary simulation modelling has long being used as tool to produce proofs of concept regarding the relation between lower level mechanisms and global behaviour. Examples of such proofs of concept include the production of minimally cognitive behaviour without explicit internal representations (Beer, 2001), autonomous learning in neural networks without synaptic plasticity (Tuci et al., 2002) or the achievement of functional readaptation to sensorimotor disruption (through homeostatic synaptic plasticity) without disruption ever being present on the evolutionary history of the simulated agent (Di Paolo, 2000). This simulation models do not pretend to model any specific biological target, but are rather used as philosophical or intratheoretical experiments and their consequences for the philosophy of biology and philosophy of the mind are significant on that they keep testing theoretical assumptions and illustrating conceptual re-organization.

In addition to this synthetic bottom-up methodology other analytic tools should be philosophically tuned. Complexity measures to understand functional integration in neural processes (Tononi et al., 1998) are producing interesting results. An early exploratory example of such methodology is provided by Seth (2002), fusioning both evolutionary simulation modelling and complexity measures of neural network dynamics. Complexity measures showed that when the evolved networks dynamics are analysed by random activation of neurons results are significantly different than analysed when coupled with the environments for which they where evolved. The experiment aiming to demonstrait that, at least in one case, complex behaviour requires complex mechanisms, ends up showing that structural analysis of network connectivity is never enough and that coupling (situatedness) becomes, once again, a fundamental characteristic of behaviour. FALTA ELABORAR UN POCO MÁS EL TRATAMIENTO DE LA INTEGRACIÓN EN EDELMAN.

5 Conclusion

The strength of the proposed dynamical perspective on behavioural adaptive autonomy (as a lower boundary condition for a naturalist characterization of cognition) is given by the shift from:

- viewing cognition as computations between 'representational' automaton states, whose representational normativity is fixed by an heteronomously interpreted functional equivalence with states of affairs in the world

to:

- and interactive dynamical process whose normativity is given by its satisfaction of closure criteria and functionality is grounded on the embodied and situated nature of behavioural dynamics (structurally and interactively emergent and capable of self-restructuring according to the metabolic evaluation of the interactive coupling).

But further discussion and clarification is required in this direction. Although behavioural adaptive autonomy satisfies the goals mentioned in section 2, further developments in, at least, two directions seem plausible.

5.1 Further characterization towards higher level cognition

Adaptive behavioural autonomy underdetermines cognitive behaviour. Homeostatic maintenance of essential variables and self re-structuring capacity is necessary but not sufficient for a characterization of a gradual notion of cognition. In this sense we believe that the work of Christensen and Hooker (2002) on self-diretedness is a natural step forward. FALTA EXTENDER ESTO UN POCO Y VOLVER A HABLAR DE INFORMACIÓN BREVEMENTE, INCLUYENDO UNA MENCIÓN A "INFORMATION AND AUTONOMY" CON ALVARO.

5.2 Other sources of normativity should be considered

Could a characterization of cognition come from other sources rather than maintenance of essential variables within viability constraints? Could a sort of sensorimotor coherence become an alternative source of normativity? possibly a kind of minimal structural metaestability condition for any kind of coherent behaviour (whether this behaviour is adaptive or not)?

If something different to behavioural adaptivity was necessary for complex cognitive behaviour to happen (e.g. the metaestability above mentioned), that condition would enable adaptive behaviour, but not everything enabled by that condition would be adaptive. Because that condition would be necessary for cognition it could be considered normative (as condition of possibility) and thus a new normative domain would appear decoupled from interactive closure criteria. This will admit non adaptive behaviour to still be cognitive. An interesting line of research has recently been proposed by Di Paolo (2003) in this direction. Di Paolo argues that behaviour itself is underdetermined by survival conditions and proposes *habit formation* as the origine of intentionality. Habits are self sustaining dynamic structures of behavioural 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 behavioural patterns; as demonstrated by Di Paolo (2000).

Rather than providing conclusive results, what such research on synthetic bottom-up simulation modelling is showing (within a dynamical approach to cognition) is that the time is ready to address important philosophical issues in a workable methodological and conceptual framework for the study of behavioural autonomy. By providing an explicit definition of behavioural adaptive autonomy in this framework we hope to have contributed something in this direction.

References

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

Barandiaran, X. and Feltrero, R. (2003). Conceptual and methodological blending in cognitive science. The role of simulated and robotic models in scientific explanation. Accepted paper for the 12th International Congress of Logic, Methodology and Philosophy of Science, Oviedo (Spain), August 7–13, 2003.

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

Beer, R. D. (1996). Toward the evolution of dynamical neural networks for minimally cognitive behaviour. In Maes, P., Mataric, M., Meyer, J. A., Pollack, J., and Wilson, S., editors, *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behaviour*, pages 421–429. Harvard, MA: MIT Press.

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

Beer, R. D. (2001). The dynamics of active categorical perception in an evolved model agent. *submitted to Behavioral and Brain Sciences*. Downloaded on 13/3/02 from http://vorlon.cwru.edu/˜beer/.

Bickhard, M. H. (2000). Autonomy, Function, and Representation. *Communication and Cognition – Artificial Intelligence*, 17(3–4):111–131. Special Issue on: The contribution of artificial life and the sciences of complexity to the understanding of autonomous systems. Guest Editors: Arantza Exteberria, Alvaro Moreno, Jon Umerez.

Block, N. (1996). What is Functionalism. Online revised entry on functionalism in the . In Borchert, D., editor, *The Encyclopedia of Philosophy Supplement*. MacMillan. URL:
http://www.nyu.edu/gsas/dept/philo/faculty/block.

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

Christensen, W. and Hooker, C. (2002). Self-directed agents. *Contemporary Naturalist Theories of Evolution and Intentionality, Canadian Journal of Philosophy*, 31 (special issue).

Clancey, W. (1989). The Frame of Reference Problem in Cognitive Modeling. In Arbor, A., editor, *Proceedings of the 11th Annual Conference of The Cognitve Science Society*, pages 107–114. Lawrence Erlbaum Associates.

Collier, J. (1999). Autonomy and Process Closure as the Basis for Functionality. In Chandler, J.L.R./van de Vijver, G., editor, *Closure: Emergent Organizations and their Dynamics. Volume 901 of the New York Academy of Sciences.*

Cristensen, W. and Hooker, C. (1999). An Interactivist-Constructivist Approach to Naturalism, Intentionality and Mind. In *Presented to Naturalism, Evolution, and Mind. The 1999 Royal Institute of Philosophy Conference*. University of Edimburg.

Dennet, D. (1995). Artificial Life as Philosophy. In Langton, C., editor, *Artificial Life. An overview*, pages 291–2. MIT, Cambridge, MA.

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: homeostatic adaptation and teleology beyond the closed sensorimotor loop. Working paper.

Di Paolo, E., Noble, J., and Bullock, S. (2000). Simulation Models as Opaque Thought Experiments. In Bedau, M., McCaskill, J., Packard, N., and Rasmussen, S., editors, *Artificial Life VII: The 7th International Conference on the Simulaiton and Synthesis of of Living Systems*. Reed College, Oregon, USA.

Fauconnier, G. and Turner, M. (1998). Conceptual Integration Networks. *Cognitive Science*, 22(2):133–187.

Funakashi, K. and Nakamura, Y. (1993). Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks*, 6:1–64.

Gilbert, S. F. and Sarkar, S. (2000). Embracing Complexity: Organicism for the 21st Century. *Developmental dynamics*, 219:1–9.

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.

Langton, C. (1996). Artificial Life. In Boden, M., editor, *The Philosophy of Artificial Life*, pages 39–94. 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.

Millikan, R. G. (1989a). Biosemantics. *Journal of Philosophy*, 86 Issue 6 (June):281–297.

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

Moreno, A. (2000). Artificial Life as a bridge between Science and Philosophy. In Bedau, M.A., M. J. P. N. and Rasmussen, S., editors, *Artificial Life VII: The 7th International Conference on the Simulaiton and Synthesis of of Living Systems*. MIT Press.

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

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

Ruiz-Mirazo, K. and Moreno, A. (2000). Searching for the Roots of Autonomy: the natural and artificial paradigms revisited. *Artificial Intelligence*, 17 (3–4) Special issue: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.

Slocum, A. C., Downey, D. C., and Beer, R. D. (2000). Further experiments in the evolution of minimally cognitive behavior: From perceiving affordances to selective attention. In Meyer, J., Berthoz, A., Floreano, D., Roitblat, H., and Wilson, S., editors, *From Animals to Animats 6: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, pages 430–439. Harvard, MA: MIT Press.

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.

Tuci, E., Harvey, I., and Quinn, M. (2002). Evolving integrated controllers for autonomous learning robots using dynamic neural networks. In *Proceedings of The Seventh International Conference on the Simulation of Adaptive Behaviour (SAB'02).*

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

van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioural and Brain Sciences*, 21:615–665.

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

Definitions

Environment (1): The environment take in isolation will be defined by $\dot{\mathbf{x}}_{\mathcal{E}} = \mathcal{E}(\mathbf{x}_{\mathcal{E}})$; where $\mathbf{x}_{\mathcal{E}}$ is the state vector of the environment and $\mathcal{E}$ the function gouverning its change. If the agent can induce changes in the environment we can open the model and in the following way: $\dot{\mathbf{x}}_{\mathcal{E}} = \mathcal{E}(\mathbf{x}_{\mathcal{E}}; \mathbf{m}(t))$; where $\mathbf{m}(t)$ is the agent's motor activity vector acting on the environment.

Agent (1): We take the agent to be the mechanism specifying agency in the adaptive system (i.e. not considering metabolic processes, constructive autonomy). Thus: $\dot{\mathbf{x}}_{\mathcal{A}} = \mathcal{A}(\mathbf{x}_{\mathcal{A}}; \mathbf{s}(t))$; where $\mathbf{x}_{\mathcal{A}}$ is the state vector of the agent, $\mathcal{A}$ the lows governing the agents variables and $\mathbf{s}$ the sensory input vector to the agent.

Essential variables: Our modelling only captures relevant operational structures *for* adaptive behaviour. In this context we model essential variables as enclosing mainly thermodynamic conditions in relation to the agent as a *far from equilibrium* system. Essential variables must be under viability constraints for the system to maintain its biological organization. Essential variables are non controled variables; i.e. the agent can only modulate those variables through the environment. Thus we define the essential variable vector as $\mathbf{x}_{\mathcal{T}} = \mathcal{T}(\mathbf{i}(t))$ where the input to essential variables $\mathbf{i}(t) = \mathbf{I}(\mathbf{x}_{\mathcal{E}})$. Where $\mathbf{I}$ is a function of the environment determining the input to essential variables. We shall, thus, define the rate of change of essential variables as $\mathbf{x}_{\mathcal{T}}(t)$, where $\dot{\mathbf{x}}_{\mathcal{T}} = \mathcal{T}(\mathbf{x}_{\mathcal{T}}; \mathbf{i}(t))$.

Agent-Environment coupling: Because the sensory input is a function of the environment (in relation to the agent) we can specify that $\mathbf{m}(t) = \mathbf{M}(\mathbf{x}_{\mathcal{E}})$. And converselly: $\mathbf{s}(t) = \mathbf{S}(\mathbf{x}_{\mathcal{A}})$.

Environment (2): Thus we can re-define the environment (introducing the agent's effect) as $\dot{\mathbf{x}}_{\mathcal{E}} = \mathcal{E}(\mathbf{x}_{\mathcal{E}}; \mathbf{M}(\mathbf{x}_{\mathcal{A}}))$.

Agent (2): If we consider that the agent might get some signal vector $\mathbf{v}$, so that $\mathbf{v} = \mathbf{V}(\mathbf{x}_{\mathcal{E}})$ from its essential variables and considering the agent-environment coupling we get $\dot{\mathbf{x}}_{\mathcal{A}} = \mathcal{A}(\mathbf{x}_{\mathcal{A}}; \mathbf{S}(\mathbf{x}_{\mathcal{E}}); \mathbf{V}(\mathbf{x}_{\mathcal{T}}))$.

Adaptive System: We can define the adaptive system as the coupling between $\mathbf{x}_{\mathcal{A}}$ and $\mathbf{x}_{\mathcal{T}}$.

Agent-Environment coupling (2): The agent-environment coupled system will be:

$$\dot{\mathbf{x}}_{\mathcal{C}} = \begin{bmatrix} \dot{\mathbf{x}}_{\mathcal{A}} \\ \dot{\mathbf{x}}_{\mathcal{E}} \end{bmatrix} = \mathcal{C}(\mathbf{x}_{\mathcal{C}}) = \begin{bmatrix} \mathcal{A}(\mathbf{x}_{\mathcal{A}}; \mathbf{S}(\mathbf{x}_{\mathcal{E}}); \mathbf{V}(\mathbf{x}_{\mathcal{T}})) \\ \mathcal{E}(\mathbf{x}_{\mathcal{E}}; \mathbf{M}(\mathbf{x}_{\mathcal{A}})) \end{bmatrix}.$$

Universe:

$$\dot{\mathbf{x}}_{\mathcal{U}} = \begin{bmatrix} \dot{\mathbf{x}}_{\mathcal{A}} \\ \dot{\mathbf{x}}_{\mathcal{E}} \\ \dot{\mathbf{x}}_{\mathcal{T}} \end{bmatrix} = \mathcal{U}(\mathbf{x}_{\mathcal{U}}) = \begin{bmatrix} \mathcal{A}(\mathbf{x}_{\mathcal{A}}; \mathbf{S}(\mathbf{x}_{\mathcal{E}}); \mathbf{V}(\mathbf{x}_{\mathcal{T}})) \\ \mathcal{E}(\mathbf{x}_{\mathcal{E}}; \mathbf{M}(\mathbf{x}_{\mathcal{A}})) \\ \mathcal{T}(\mathbf{x}_{\mathcal{T}}; \mathbf{I}(\mathbf{x}_{\mathcal{E}})) \end{bmatrix}.$$

Sensorimotor transformation: Sensorimotor transformations are the trajectories of $\mathbf{x}_{\mathcal{A}}$

Behaviour: Metalevel description of agent-environment interaction. Behaviours must be defined in the $\mathcal{C}$ domain, adaptive behaviour in the $\mathcal{U}$ domain.

Organization: The set of metainvariant relations between the variables defining a system: the functions governing variables.

Structure: The set of invariant relations between the variables defining a system: it is composed of functions *and* parameters. When some variables of a system remain stable for long periods of time (comparing to the rate of change of other variables) they can be considered parameters; lets call them *weak parameters*.

Plasticity: Capacity of a given organization to induce structural changes, i.e. weak parameter changes.

Adaptive (normative) Function: Because de Agent is a far from equilibrium system it must keep some of its essential variables under viability constraints activelly (since they will tend to decay according to the second law of thermodynamics). Some essential variables represent (in the model) this thermodynamic condition, from which the agent gets indications trough $\mathbf{V}$. Other essential variables (like body integrity) must be keept within viability constraints during the interaction processes. Thus adaptive functionality $\mathcal{F}$ can be defined as a mapping from the coupled agent-environment system to the essential variables, so that the essential variables are keep between viability constraints; i.e. $\mathcal{F} : \mathbf{x}_{\mathcal{C}} \rightarrow \mathbf{x}_{\mathcal{T}} \in \mathcal{V}_{\mathcal{T}}$; where $\mathcal{V}_{\mathcal{T}}$ is the viability subspace of the essential variables. FALTA: SOME ESSENTIAL VARIABLES DON'T DECAY (BODY INTEGRITY). TAMBIÉN HAY QUE DISTINGUIR ENTRE FUNCIÓN DE UN PROCESO Y FUNCIÓN PROPIA DE UNA ESTRUCTURA

Behavioural Adaptive Autonomy: An adaptive system is autonomous if it is capable of homeostatically maintaining $\mathbf{x}_{\mathcal{T}}$ under viability constraints throught self-modulating behavioural coupling with the environment. Adaptive functionality is always the outcome of the agent-environment coupling, thus adaptive autonomy is a continous measure of the agent's active contribution to the satisfaction of adaptive functionality.

Mechanistic Explanation: The answer to the question "which subset of $\mathcal{C}$ is causally relevant to the performance of a function?". FALTA INTRODUCE HYPERDESCRIPTIONS HERE (READ CHRIS).

Localization: Functional components' mapping into structural components. Localization requires structure and functionality to be decomposed, which might not be always possible if struture is integrated (i.e. if interaction between components is higher than within components).