

Mental Life

Conceptual models and synthetic methodologies for a post-cognitivist psychology

Xabier Barandiaran

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

Dept. of Logic and Philosophy of Science. University of the Basque Country, POBOX, 29900 / San Sebastian / Spain

to my grandfather,
whose
intellectual dedication
and lived subtlety
I will admire forever

“The question as to the nature of life, I believe, has been finally resolved, and is no longer a philosophical question. I hope something like this will happen to the so-called mind-body problem in the twenty-first century.”

JOHN SEARLE

“Now what makes the cell living? The soft organization of its inner events and occurrences. Thus, if we are looking for the fundamental laws, for the principle of life, we have to establish the connections of this soft organization”

TIBOR GÁNTI

1. New foundations for psychology?

Psychology is one of the most elusive fields of knowledge for current scientific standards. This is not a surprising fact given that the brain (which is at least partly responsible for psychological phenomena) is, as Isaac Asimov synthetically described it, “the most complex three pounds of matter in the universe”, with more possible combinations of

Copyright 2006 © Xabier Barandiaran. This is an open-access work distributed under the terms of the Creative Commons Attribution License 2.5, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. <http://creativecommons.org/licenses/by/2.5/>

Originally published as: Barandiaran, X. (2007) *Mental Life. Conceptual models and synthetic methodologies*. In B. Wallace, A. Ross, T. Anderson, J. Davies. (Eds.) *The World, the Mind and the Body: Psychology after cognitivism*. Imprint Academic.

You can freely download a copy of this work from:
http://barandiaran.net/textos/mental_life

neural states in the brain than there are atoms in the universe. In addition, brain activity is not determined in isolation (providing at least a workable experimental control condition). Nor is brain activity completely specified by anything like a genetic program subject to reverse engineering. On the contrary, a full range of bodily and environmental interactions shape neural activity, including the interactions with other (social) embodied bodies, historically shaped through multiple cultural and biographical contingencies, organized through a set of externalized technologies and languages. Besides its complexity preconceptions about the nature of the mind are everywhere active on the knowledge and power relationships that constitute our societies. However elusive, the question of what the mind is and how should it be studied, cannot be left aside with a “sorry, not yet amenable to science” sticker on it. The answer (however unsatisfactory, always provisional) is continuously operating on different scientific fields, public institutions, business operations and everyday life activity: from folk-psychology to psychiatric institutions, from education to neurobiological experiments, from inter-personal relationships to psycho-therapies. Attempts to create “definitive” scientific and rigorous foundations for psychology have been many. Each of them has equally enabled and limited our understanding of the mind with a great impact on society and other scientific disciplines. Some of those attempts include phenomenology, phrenology, behaviourism, gestalt psychology, psychoanalysis, computational functionalism, folk-psychology or eliminativist neuroscience.

The fact that we ask ourselves here about something like a post-cognitivist psychology uncovers two basic premises: (i) that something like a “cognitivist paradigm” has dominated mainstream psychological studies for a while, providing a fruitful foundational framework and (ii) that a possibility is envisioned to overcome this well established paradigm to make room for something like a post-cognitivist foundation for psychology. Both premises require that we make explicit what cognitivism might be in the first place. Following Wheelers analysis of the Cartesian inheritance in cognitive science (Wheeler, 2005), we

Table of Contents

1. New foundations for psychology?.....	1
2. Conceptual modelling. A MUNDane declaration of principles.....	2
3. Life: lessons from synthetic protocell biology.....	4
4. Simulating neurodynamic agents: an experimental framework for theoretical psychology.....	7
5. Mental Life.....	10
5.1. Modelling mental life.....	11
5.2. Five characteristic features of mental life.....	12
a) Emergent self.....	12
b) Situated openness.....	12
c) Normativity.....	13
d) Agency.....	13
e) Embodiment.....	13
5.3. The mind from the MUN.....	14
6. Conclusions: there is life beyond cognitivism.....	16
References.....	17

can depart by considering that *cognitivism* states that the foundation and demarcation of mental/psychological phenomena is given by: (i) a *subject-object* dichotomy (ii) in which the cognizer (the subject) manipulates *inner representational states* (of an immaterial nature: conscious-phenomenological, computational, or otherwise) (iii) according to the *rules of reason* (logical, linguistic, etc.); (iv) where representational content is acquired by *inferential procedures* and (v) used to process (deduce, transform) a plan in order to execute *actions* in the world (or, it could be equally appropriate, to satisfy the cognizer's mindfulness with the delight of truthful transformations). As such, cognitivism has established itself as a form of computational modernity with its faith on universal context-independent reason, its representational realism and its functionalist disembodiment, achieving a considerable success in the fields of Linguistics, Artificial Intelligence (especially within the subfields of expert systems and symbol manipulation based reasoning) and Philosophy of the Mind, among others.

The prefix in *post-cognitivism* does not imply a refusal of a cognitive subject, her reason, her reality or a linguistically structured mediation of behaviour. On the contrary, the "post" might be understood as the opportunity to experiment the limits of its predecessor in order to formulate questions beyond those limits and, eventually, to explore possible answers: what kind of processes, of a more fundamental character, make possible the appearance of the subject-object dichotomy that every cognitivist study presupposes? what is the origin of the imperative force of reason as a normative structure of mental processes? which are the physical and biological conditions that make its existence possible? what would happen if action is considered as the very condition for the production of perceived situations and not as the planned response to an objective state of affairs? This kind of questions demarcates the landscape where *post-cognitivism* might be able to make its more fruitful movements. In this sense it is perhaps the right time to review and push forward some of the *methodological* and *conceptual* innovations that are available to us and might permit to speak of a paradigmatic discontinuity that could properly be called "post-cognitivist". I shall attempt to tackle the methodological innovations by sketching some of the new insights that computer simulation models of neurodynamic embodied agents have permitted. On the conceptual side, and drawing some analogies from the emergent field of "synthetic protocell biology", I shall elaborate a conceptual model of *Mental Life*, merging together and pushing forward some of the conceptual achievements that nowadays populate the post-cognitivist landscape under the labels of dynamicism, embodiment and situatedness.

2. Conceptual modelling. A MUNDane declaration of principles

Before we attempt to define an alternative methodological and conceptual framework to that of cognitivism it is worth stating a set of epistemological principles that accurately define what this foundation should look like. This meta-theoretical exercise is somehow unavoidable in any attempt to approach foundational issues in psychology. A great deal of both theoretical and methodological debates in psychology do not directly deal with the content of psychological phenomena but with the definition of the very framework in which such a question should be made and answered. Explicitly stating a set of epistemological constraints should make clear how to evaluate the present approach and how it relates to scientific practice. I shall call these the MUN constraints, standing as a acronym for Minimalism, Universality and Naturalism. But first, let me write some preliminary words on models since these constraints are to be applied to the process of model building and interpretation.

Current philosophy of science has focused on models as the most important units of the scientific production and organization of knowledge (Cartwright, 1983; Giere, 1988; Morrison, 2000). Godfrey-Smith has recently summarized this model-based philosophy of science in the following way:

"A model-builder's usual goal is to construct and describe various hypothetical structures. These structures are used to help us understand some actual target system or systems. Generally, the understanding is supposed to be achieved via a resemblance relationship between the hypothetical and the real system. But both the degree and kind of resemblance that is sought are adjustable. (...) [T]he ability to describe and develop model systems in some detail, while remaining cautious or flexible about the particular respects in which the model might resemble the target system, is an essential tool. Modelling is specially useful when our knowledge of the target system is poor, and its workings are complex." (Godfrey-Smith, 2005:3)

We can see a promising avenue to settle some theoretical disputes on the foundations of psychology if we adopt a conceptual and simulation modelling paradigm. What this suggestion involves is that foundational concepts themselves be forged as models. Thus concepts should be constructed as characterizing a class of hypothetical systems or structures that, through a resemblance relationship with a class of target systems, should help us to understand some of their essential features. In particular, our goal is to *model Mental Life as a form of organization: i.e. as a class of hypothetical systems in which a set of component processes relate to each other in a specific (interdependent) manner giving rise to a set of characteristic features.*

Linguistically expressed conceptual models can be transformed and implemented into more tractable formal or

simulation models engaging with empirically testable research procedures. But a conceptual structure, mathematical construct or computer simulation on its own is not a model of anything, unless accompanied by an auxiliary framework made of the assumptions, generalizations and interpretative relationships that permits to relate the hypothetical structure to the target objects and evaluate its adequacy and epistemic scope. In this sense the MUN constraints shall make explicit not only how the conceptual model is to be built, but also how its auxiliary framework should be set up in order to transfer the model to the more empirically testable domain of scientific discourse. Let me start with Naturalism, then move to Universality and finally tackle the Minimalist constraint.

Naturalism is a widely spread philosophical position stating that *ad hoc* substances are not to be introduced in a model in order to explain the target phenomena. Thus, for instance, consciousness or information as theoretical primitives (Chalmers, 1995), representations or logical structures, should be avoided as foundationally privileged departure points. However, what constitutes an *ad hoc* substance or property is a difficult matter. The way out of this dilemma is to think on naturalism as a scientifically embodied philosophical practice both at its sensory and motor surfaces (so to speak): i.e. it should be grounded on available scientific knowledge and be able to feed-back to scientific practice (through its capacity to generate new hypotheses, to provide principles to reorganize knowledge, clarify concepts, uncover fallacies, etc.). In addition, one should wonder in which type of scientific field must one be embedded on. In this sense we shall expand our naturalist constraint to encompass an additional requirement: a *bottom-up* approach. By bottom-up we mean that concepts and components of our conceptual model should be built from the most simple and elementary (in relation to a given level of organization of empirical research) to the most complex and higher order ones. In particular we will defend a biological *grounding* by which components of our model should be derived or closely related to more fundamental biological processes¹. Thus, if we were to model meaning, for instance, it would be inappropriate to attribute semantic properties to component neural ensembles if meaning is defined strictly in higher level terms (for instance as linguistic performance or in reference to the use of dictionaries, with no reference to its biological and neural grounding). Higher level descriptions should be accompanied and grounded on bottom-up explanations of how those phenomena can be sustained and emerge from lower level organizational principles. A bottom-up approach includes a final naturalist constraint: that observer dependent properties (relational properties that are accessible to the external

observer's privileged position such as correlations between internal and environmental states, the designers intentions, etc.) should not be attributed to the model itself if no specific procedure is established to reconstruct them in a bottom-up observer-independent manner. This way we shall avoid the risk of projecting observer-dependent properties to the model and from the model to the explanation of the target system².

Thus our first MUN constraint mandates that we build our model from a naturalist perspective, which entails a bottom-up biological grounding of the concepts and components belonging to our model of *Mental Life*. Our second constraint is *Universalism*. Currently available biological systems amenable to experimentation and study are the result of a set of historical (evolutionary) contingencies. But knowledge has universalist aspirations. As Artificial Life founder Chris Langton (1989) claimed: it is not life-as-we-know-it but rather life-as-it-could-be that is of interest to the field. We could equally define our object of study the-mind-as-it-could-be rather than the-mind-as-we-know-it. This forces us to define universalizable patterns of life and mind rather than focusing on particular anatomical details of present mind-supporting brains and bodies. For instance if emotions are to be part of our final model it would be inappropriate to say that emotions are defined by the signals coming from a particular neural pathway, as if human brain anatomy was to determine what emotions are to be in the Universe. It might be the case that some psychological processes be unambiguously identified or correlated with certain brain areas, but this is not to say that what that process is be equivalent with a certain anatomical component or set of components that happen to be the locus of such emotions in planet-earth vertebrates.

Minimalism is our third and final epistemological constraint. Minimalism might be seen as a direct consequence of our first naturalist bottom-up constraint but it is worth making it explicit as a specific requirement in itself. It states that our model must contain all but no more than those features necessary and sufficient to define the class of systems that it targets. So, rather than taking the higher level epistemic properties or language like sophisticated mental phenomena as a departure point, minimalism states that we should proceed making use of the simplest and more amenable components in order to build a model. This being said, and this is an important remark, the upper boundary for complexity increase must remain open. So, for instance, Neils Bohr, inspired by Rutherford, proposed the planetary model of the atom taking as a departure point "a simple system consisting of a positively charged nucleus of very

1 The bottom-up approach does not forbid to use top-down methods, in fact a top-down bottom-up circulation will be of significant importance. What the bottom-up approach emphasizes is that the role of top-down methodologies should be limited to a form of heuristics and guidance of the bottom-up grounding.

2 This form of projection of observer-dependent properties is a common mistake that appears on many human made and interpreted devices such as computer programs or robots. William Clancey (1989) has strongly argued on the danger of confusing three different frames of reference in robotic modelling: the robot designer's ontological preconceptions, the dynamics of a robot's interaction with its environment and an observer's descriptive theories of patterns in the robot's behaviour.

small dimensions and an electron describing closed orbits around it" (Bohr, 1913:3). Bohr's model, although focused on a simple system (the Hydrogen atom) to start with, was built with the rest of atomic forms in mind, so that components (electron orbits, nuclear forces and their relationships) could be aggregated to form more complex models once the minimal one was satisfactorily constructed and tested. Equally, we should tend to generalizable and expansible models, where a minimalist core stands as a foundational first step that permits to organize and discuss conceptual and empirical relationships. In the absence of a complete model, some properties might be studied on partial property-specific models. In this sense, formalization and computer simulation permit a common language to recombine and integrate achievements and components from different local or particular modelling experiments³.

But, unlike Bohr's case, we face a situation where there is no generally accepted and empirically available minimal target object to model. We are lacking the Hydrogen atom of the mind. What constitutes a genuine example of minimal cognition (not to speak of minimal mindfulness) remains an open issue which deserves much more attention than what it currently receives⁴. In such a situation, minimalism is a methodological remedy for the study of complex systems. So let us imagine that there was nothing like a one electron + one single proton atom left in the universe: only complex macromolecules to experiment with. In such a case we could proceed by creating something like an artificial atomic-physics that was to construct complex simulation models of non-existing atoms, out of which an artificial chemistry could be constructed which, finally, could be compared with experimentally available target macromolecules. The Hydrogen atom of the life and mind must be reconstructed from what we take to be coherent with our present knowledge of biological and neuro-psychological phenomena.

By generating such minimalist (but non directly empirically correlated) models we pay a considerable prize for the abstraction and idealization it necessarily involves. But we gain, on the other hand, an insight on the nature of complex systems that is not available otherwise. Elsewhere (Barandiaran & Moreno, 2006a) I have termed these models *conceptual simulation models* because they do not directly target any empirical object but remain, nevertheless, epistemically useful by providing the means for theoretical investigation, conceptual clarification and illustration, proofs of concept, knowledge reorganization and a set of other epistemic functions. This use of

conceptual simulation models has been called *opaque thought experiments*, in relation to the heuristic and conceptual role played by thought experiments in science while remaining analytically opaque due to the complexity of the simulation (Di Paolo *et al.*, 2000; Bedau, 1998)⁵. These models make computers virtual laboratories (Emmeche, 1994) where complex interactions among emergent dynamic structures can be extensively and intensively studied; improving our theoretical understanding of those natural phenomena which are much more complex than what the naked human mind or mathematical analysis alone are capable of exploring. In addition computer simulation models permit to stablish systematic experimental set-ups for those natural objects whose control conditions are difficult to fix. The embodied and situated brain/mind is one of such objects and artificial life robotic simulations (and realizations) stand for some of the most successful tools to model it at the conceptual level. These models remain far from the intricate complexities of natural brains-bodies and their subtle ecological environments, but stand, nevertheless, closer to a comprehensive conceptual understanding of the integrated and emergent patterns that might constitute the essence of *psyche*. I should come back to this topic throughout the rest of the chapter.

3. Life: lessons from synthetic protocell biology

If we are to develop a model of Mental Life it seems important to spend some time exploring the concept of life first (as it is understood and modelled in some current approaches). But the concept of life is not the only link with biology that we can benefit from. Since biology has suffered from most of the same conceptual and methodological problems of psychology, it should be equally expected for psychology to benefit from those conceptual and methodological remedies that biology is enjoying today. Both fortunes and misfortunes of biological sciences will contain important lessons to apply into psychology. In addition, biology has a much more detailed (and minimalist) understanding of living systems than the best available picture of brain activity or any other scientifically grounded psychological research fields. As a result, concepts and models developed in biology, with its fine grained molecular experimentalism and its computer modelling implementations, have acquired a high level of conceptual, methodological and empirical sophistication and accuracy. Importing some of this conceptual and modelling apparatus back to the realm of psychology stands as a promising avenue.

As Bechtel recently argued (2006), mechanistic explanations in biology have long underappreciated the importance of

³ Such is the case of some robotic systems that integrate partial neural models of functionally distinct anatomical parts (Brooks, 1997; Almásy *et al.*, 1998; Taylor & Taylor, 2000; to mention but a few).

⁴ There are a number of recent exceptions like Randall Beer's target article and its commentaries (Beer, 2003) on minimally cognitive robotic agents or van Duijn and colleagues' exploration into the principles of minimal cognition (Duijn *et al.*, 2006). Together with Alvaro Moreno I have also addressed this question somewhere else (Barandiaran & Moreno, 2006b).

⁵ Going even further Daniel Dennett (1994) has claimed that Artificial Life might be understood as a form of philosophy itself.

organization itself; i.e. how components get together in particular reactive arrangements creating the phenomenon under investigation. Far from the linear decomposition and isolated analysis of component's properties in which mainstream molecular biology has focused on, biological explanations will ultimately require models that include positive and negative feed-back loops, self-organized processes, coupled cycles and network properties that put on centre-stage the critical role of organization in living phenomena: "Only by keeping a keen eye on the organization at play in living systems is it possible to understand the mechanisms that figure in living organisms" (Bechtel, 2006).

Tibor Ganti's chemoton model (1971, 2003), Stuart Kauffman's autocatalytic network theory (1971), or Robert Rosen's M-R systems (1958). The original formulation of these models was done under a conceptual or linguistic form, accompanied by diagrammatic illustrations and by formalized descriptions. But since the early 70's, computer simulation models were used to illustrate the emergent order of the proposed organization, such is the case of Varela, Maturana and Uribe's pioneering work (1974). Also Kauffman explored the autocatalytic systems and other self-organizing biological processes making use of computers (1993) while Tibor Ganti's model's first simulation dates 1975 (Békés, 1975) followed by Csendes (1984).

Despite being often marginalized by mainstream biology these dynamic and organizational models of life have been further developed by a number of authors within the fields of artificial life, artificial chemistry, theoretical biology, complexity sciences, origins and synthesis of life, etc. Within the most recent approaches we shall focus on Mavelli and Ruiz-Mirazo's (2007) simulation model of a minimal self-reproducing cellular system that captures and integrates most of the essential features of the models mentioned above. Figure 1 graphically illustrates their model. At the nucleus of it we have an autocatalytic cycle: this is a network of chemical reactions that reproduces the components of the network itself through a cyclic loop of metabolites (A components). The first core idea is that of self-organization at the chemical level: a huge amount of microscopic elements adopt a global, macroscopic ordered pattern in the presence of a specific flow of matter and energy (represented by the continuous inflow of X precursors into the system with the outflow of W waste products, and the set of constraints on the equations that govern the reaction dynamics, expressed as kinetic constants k_n). Given the presence of precursor X the stochastic collision of A_1 molecules produces A_2 molecules that in turn produces A_3 molecules leading to A_4 which, closing the loop, generate A_1 molecules. The resulting pattern generates a form of *identity* in which, out of an undifferentiated chemical pool, a self-reinforcing order appears. The internal dynamic cohesion that constitutes this identity is not only a consequence of the material features of their components but also, and most importantly, of the achievement and maintenance of some type of circular dynamic causality. In other words, the very macroscopic pattern contributes to the maintenance of the dynamical cohesion at the microscopic level: the chemical cycle continuously regenerates its component processes. Thus, it is not only the local interactions that matter but the global patterns they generate: molecular properties are significant just in the context of massive stochastic collisions where the effect of a particular molecule will depend on the reaction rates of other components whose concentrations are continuously maintained in far-from-equilibrium stability conditions by the network of reaction cycles that constitutes the

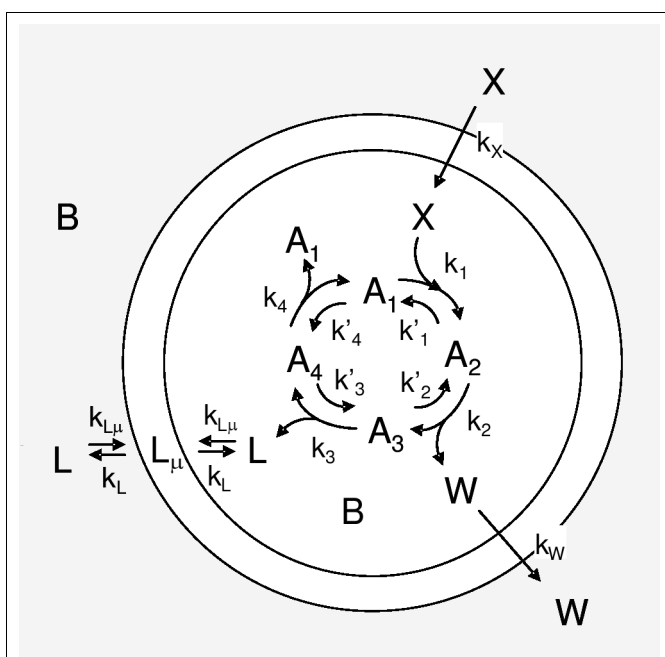


Figure 1: Graphical representation (by Mavelli & Ruiz-Mirazo 2007, with permission) of the simulation model of minimal protocell metabolism (what we here take as the basic organization or essence of life). A core autocatalytic network regenerates the components of the network (A components) and produces a membrane (L components) capable to manage the flow of matter through it (expressed through the precursor X and the waste product W). The value of the kinetic constants (k_n) together with the net flow of matter through the system keep it in far-from-equilibrium thermodynamic conditions: the coupled reactions are continuously sustaining the levels of concentration necessary to keep the system going.

Systems Biology (Ideker *et al.*, 2001; Kitano, 2002; O'Malley & Dupré, 2005) is the label under which current attempts to integrate data from molecular biology into organizational models (from developmental genetic regulatory networks to metabolic coupled cycles) are currently carried out. Of particular interest to us is the set of models of *minimal organization of life* that Systems Biology has started to develop, taking the cell as the basic unit and expression of life —what Solé and colleagues (2007) have labelled *Synthetic Protocell Biology*. Some of the early formulations of minimal models of life trace back to Maturana and Varela's autopoietic theory of life (1972, 1980),

system.

This form of circular physico-chemical organization is a kind of dissipative structure (Nicolis & Prigogine, 1977), a far from thermodynamic equilibrium system that Schrödinger took to be at the core of living phenomena (Schrödinger, 1944). As such, the system, in order to maintain its constitutive order, needs a continuous flow of matter and energy. And, if it is to be robust against variations of this flow while maintaining its unity, a membrane is necessarily required both to retain or encapsulate the core metabolic organization and to “negotiate” its perturbations and needs (Ruiz-Mirazo & Moreno, 2004). This requirement is represented in the model by L molecules. The network produces L molecules that ensemble each other to produce L_{μ} molecules forming together a membrane that encapsulates the reaction-network. But the membrane is not just an envelop for the autocatalytic network, it selectively controls the diffusion of reactants between internal and external aqueous solutions. This is of fundamental importance since changes in the core autocatalytic network can modulate membrane properties to control the flow of matter and energy between the system and the environment. In turn, this leads to a qualitative difference in organization with regard to that of a single autocatalytic network (like Kauffman's —1971). This (minimal) proto-cell is capable to control its boundary conditions for self-maintenance: i.e., to regulate the input of matter and energy that ensures the ongoing regeneration of components while avoiding osmotic crisis and other organizational threads⁶.

From the simulation model just described and its interpretation as the basic organization of cells, a set of characteristic features or principles of minimal living systems can be extracted:

1. *Emergent Self*: Given a set of initial conditions (the presence of X precursor molecules above a certain threshold) a set of macroscopic correlations appear (an interdependent set of concentrations of A types of molecules) as a result of recurrent local interactions (stochastic collisions). In turn the occurrence of this local interactions recursively depends on the macroscopic correlation: the higher the concentration of A_1 molecules the higher the probability of a collision between A_1 molecules to produce A_2 molecules; and the

higher the number of collisions of A_1 molecules the higher will become its concentration (due to the circular set of reactions $A_1-A_2-A_3-A_4-A_1$) until this positive feedback loop reaches a steady state. The resulting macroscopic order is said to be *emergent* precisely because of the recursive micro-macro/local-global relationship and *self-sustaining* because of the circular causal loop that is established. Due to the chemical substrate of such an organization a physical boundary is required to retain the created emergent order. By producing a membrane as part of the reaction cycle the macroscopic order can be said to distinguish *itself* from its environment. Although expressed on much more sophisticated forms, all living systems ultimately respond to this logic of self-maintenance that determines their integrity as units of life. Now, such an organization cannot exist but as a far from thermodynamic equilibrium system which brings us to the second characteristic feature.

2. *Situated openness*: The system (the emergent self understood as a circular macroscopic correlation) can only exist as situated on a material and energetic environment in order to persist. But what this environment is (in relation to the system) is co-determined by its organization, i.e. by its form of self-maintenance. For instance, the way in which molecule X in the environment becomes relevant for living organization is not something determined exclusively on the basis of its objective molecular properties but in relation to the way in which X becomes a precursor of the nested set of reactions. In this sense, out of an in principle undifferentiated physical surrounding, living organization selectively cuts out an environment that becomes both a potential source of destructive perturbations and a necessary source of boundary conditions for self-maintenance. The system is thus constitutively open.
3. *Normative functionality*: As a consequence of the above features certain internal and interactive processes become normative. Independently of how the components of the system *are* interacting at a given time or the systems, as a whole, *is* functioning, there is something that it *ought* to do and a set of component interactions that *ought* to happen in a certain way. First, because of the circular interdependent organization, some internal processes *must* happen in a particular way for its continuing existence⁷; if collisions between

⁶ Such a kind of organization has been called *autonomy* (Varela, 1979) or, more concretely specified in the chemical and thermodynamic domain, *basic autonomy* (Ruiz-Mirazo & Moreno, 2004); naming the capacity of the system to create an identity, a self (*autos*) and to define its own rules or norms (*nomos*). It can be said that the system defines or created it norms in the sense that the global order is not determined by local properties in isolation but by the circular dynamics that govern *and* constitute the system as a unity; a unity that depends for its continuing existence on those higher level patterns of activity. A full sense of autonomy would require that the system performs some work, channeling the energy generated through its core metabolic cycle to produce an action on its environment that contributes to its self-maintenance through some control of its thermodynamic and physical boundary conditions (Kauffman, 2003; Ruiz-Mirazo & Moreno, 2004).

⁷ To use Christensen and Bickhard's terminology (2002) the rest of the component processes of the system *dynamically presupposes* a concrete way of functioning (from all the “physically” possible ones) for a given component, their stability and consequently their normative functioning, depends on it.

certain molecules do not produce the corresponding molecule at a certain rate, so that the reaction chain gets closed, the system collapses. Second, due to its far from thermodynamic equilibrium condition, some actions *must* be carried out by the very system in order to ensure its own existence (for instance to avoid an osmotic crisis). I shall call *normative* the stability dependencies that are created between the macroscopic variables of the system. Thus, a sense of good or bad, appropriate or inappropriate, adaptive or maladaptive, emerges from the very organization of the system and is not externally defined by a designer or observer that projects a desired functionality on it.

4. *Agency*: The emergence of a self-maintained and self-distinguished form of dissipative order, open to its environment will crucially depend (under internal and environmental variation) on its capacity to differentially negotiate the flow with the environment (or even to actively seek for the appropriate sources of matter and energy). Agency appears precisely when the system is capable to adaptively regulate its environmental conditions for self-maintenance⁸. The term adaptive regulation involves here a causal asymmetry on the determination of the system-environment relationship. A minimal example of agency is provided by the membrane's active ion-pumping that avoids an osmotic crisis: the system directs energy against the concentration gradient to control its boundary condition for self-maintenance.

What increasingly sophisticated and accurate simulations models add to the conceptual description of life is the possibility to systematically explore the emergent patterns and behaviours that such organizations are capable of achieving; thus providing a more precise and insightful understanding of its complexity. Within Mavelli and Ruiz-Mirazo's simulation, out of local stochastic reaction rules, the above characteristic features can be observed and measured along with cell division (driven by autocatalytic growth), buffering and other homeostatic properties, critical thresholds for self-organization and certain system behaviours (oscillations, instabilities, etc.). Computer implementations of the mathematical models make possible an automatized intensive and extensive exploration of the full range of organizational configurations. Conceptually, these models are of great importance since they permit to discuss in precise terms what set of components, processes, configurations, etc. are crucial to achieve increasingly more complex biological patterns of organization. These kind of

⁸ This feature of agency cannot yet be derived or interpreted within Mavelli and Ruiz-Mirazo's simulation model but it is part of the conceptual model that one of the authors has developed previously (Ruiz-Mirazo & Moreno, 2004).

simulation models have become an unavoidable tool to discuss and develop different theories of the origins of life and the necessary conditions for its appearance and synthesis. Ultimately, computer simulation models permit to generate detailed hypothesis to be tested in the "real" laboratory (Solé *et al.*, 2006).

This conception of life, modelled as a circular, emergent, self-sustaining and far from thermodynamic equilibrium chemical organization, comes to satisfy the MUN constraints stated above. It is a *naturalist* model since it stands grounded on the material and thermodynamic properties of the components and relationships that make up the system. No reference to *vitalist* forces is required to specify the essence of life. The model is *universal* for it can generalize the basic organization of life without reference to arbitrary or local contingent properties of life-as-we-know-it, while remaining coherent with its objective/material conditions of possibility. In addition, although the model focuses on a minimal cellular level, all its fundamental properties can be generalized to more complex living forms. Finally, it is out of question that it is a *minimal* model because it integrates all but only those component processes that are crucial to specify the most fundamental features of life⁹. Can we expect something like this minimal model of life to be possibly built in the realm of neural dynamics and situated embodied agents?

4. Simulating neurodynamic agents: an experimental framework for theoretical psychology

Evolutionary robotics (Cliff *et al.*, 1993; Harvey *et al.*, 1997; Nolfi & Floreano, 2000; Harvey *et al.*, 2005) together with computational neuroethology (Beer 1990, Cliff 1991), evolutionary autonomous agents (Rupin, 2002) or what Randall

⁹ I shall note that the above characterization of living organization fails to satisfy the requirement that the upper limit for complexity growth remain open within a minimalist model. The reason is that the complexity that the described chemical organization can achieve remains severely bounded if we don't integrate further, qualitatively distinct, components and processes (Ruiz-Mirazo *et al.* 2004). We are referring here to the genetic machinery. In fact, Tibor Ganti's original chemoton model already included such components (although in a very preliminary way): energetically stable, non-reactive and recombinable macromolecules. What such "template" molecules permit is the decoupling of a control regulatory subsystem within the whole organization. In the absence of such kind of components the adaptive capacities of the system are very limited. The introduction of informational talk is used in systems endowed with such components due to their recombinable capacity, energetically stable structure and the fact that they are "functionally interpreted" by the metabolic organization to produce specific molecules capable to create local constraints. In fact such molecules and processes would be crucial to achieve an open-ended increase of complexity ultimately leading to evolution as we know it and a proper account of life with all its evolutionary potential (Ruiz-Mirazo *et al.* 2004). For some of the authors supporting the picture of life presented here dynamical and self-organization models of biological processes (even autonomy) are not sufficient to account for life; informational or semiotic processes are required for such characterization. However they all agree on that autonomy is the most fundamental requisite and form of organization. Without it the very concepts of information, function or evolution could not be naturalized. For the purpose of this chapter I will take the form of organization described in this section (without recombinable molecular templates) to be an adequate model for the essence of life.

Beer has called a “minimally cognitive behaviour program” (Beer, 1996, 2003) is a modelling paradigm that permits to design embodied and situated agents capable to solve “minimally” cognitive tasks exploiting self-organized dynamics (both internal and interactive). In short, this modelling procedure (or one of its most successful variants) works as follows. A robotic body, an environment and a control architecture are simulated as dynamical systems in a computer. It is important to note that the dynamic controllers that are usually implemented in the robotic agent (Continuous Time Recurrent Neural Networks —CTRNN hereafter) are chosen so that, through variations of their parametric values and number of nodes, the controller can potentially approximate any possible dynamical system (Funahashi & Nakamura, 1993). Some parameters of the robotic architecture (specially those of the control system but, additionally, some body parameters too, specially those related to the sensors and motors) are left unspecified. Next, artificial evolution is used to optimize these parameter values. Those configurations that lead the robot (dynamically coupled to its environment) to perform a desired cognitive or behavioural task are selected. The best evolved agent is then tested through intensive experimentation and analysis to provide a dynamical causal explanation of how the task is successfully performed. What this kind of simulation permits (unlike other experimental approaches) is to integrate, in the same explanatory framework, detailed environmental, bodily and neural factors and their complex dynamic interplay. Neural patterns of activity and their stability or synchrony with environmental, body and sensor variability can be precisely determined and their connected causal dynamic organization made visible.

Using this artificial experimental framework, models of shape recognition (Cliff *et al.*, 1993), learning (Tuci *et al.*, 2002), communication (Quinn, 2001) and a wide range of cognitive phenomena (Slocum *et al.*, 2000) can be built where no theoretical assumptions are previously introduced in the model. This is so because the agent is designed in abstract dynamical terms, with no pre-specified anatomical/structural components and predefined functions, and artificial evolution “blindly” generates behaviourally efficient agents. Selection operates at the level of the brain-body-environment continuum, thus no *a priori* task decomposition or functional presuppositions need to be made in relation to how the agent “should” solve the task. This leaves room for self-organizing dynamics to emerge in the simulation, for different and previously difficult to imagine dynamic modes of behavioural organization to appear. What we get is a kind of *emergent dynamic functionalist* approach in which a potentially universal dynamic controller is constrained to achieve an online embodied functionality out of a previously non-specified sensorimotor architecture.

The problem of these models, as some reader will by now be

willing to point out, is that they involve no reference to real existing biological cognitive agents, they have no direct empirical target to correlate with. This fact makes them paradigmatic examples of the conceptual models described before. It is precisely in this sense that evolutionary robotics might turn out of much help as a methodology to elaborate conceptual foundations for psychology and cognitive science. In fact, evolutionary robotic models have already been used to rise some interesting foundational discussions and achievements (Harvey, 2000; Di Paolo *et al.*, 2003; Beer, 2003; Wheeler, 1996, 2005; Clark, 1997; Chemero, 2000).

Some of the theoretical achievements (that most evolutionary roboticists would be willing to accept as post-cognitivist) include precise, even formalized, accounts of the critical role played by a number of key principles in cognitive systems. *Situatedness* defines the sensorimotor context of an agent as dependent on its controllable relative motion. As a result an agent can exploit environmental cues and sensorimotor correlations to solve cognitive problems that would otherwise require a high cognitive load (or even an exponential growth of context-independent inferences necessary to solve the task). Early in evolutionary robotics (Cliff *et al.*, 1993) the relevance of situatedness was made an explicit and tractable feature of intelligent behaviour by exploring the emergent dynamics of shape recognition; where cognitive behaviour was the result of distributed system-environment dynamic loops. *Sensorimotor embodiment* is another feature that these models have helped precisely defining. Embodiment is shown as a function of bodily properties in relation to the situated sensorimotor coupling of the agent with its environment, where architectural and mechanical constraints have being shown to be necessary (an sometimes sufficient) to achieve a number of cognitive behaviours that where previously thought to require explicit and sophisticated symbolic procedures. On the one hand the motor embodiment defines a limited and biased interface with the world where constrained degrees of freedom (joint angles, elasticity, shape, etc.) facilitate some characteristic interactions (grasping, walking, etc.). On the other hand embodied sensors are not continuous full-range measuring devices but, on the contrary, appear limited and specialized on specific ranges, transformations and filtering of sensory perturbations. For instance the coclea, the retina, or an artificially evolved sensor become embodiments of sensory surfaces that exploit physical and relational features to transform environmental perturbations into pre-organized signals. In addition, sensory and motor surfaces that evolved together appear coupled through recurrent sensorimotor and somatosensory interactions increasing the effect of embodiment. *Dynamicism* is another aspect of adaptive behaviour that evolutionary robotic experiments have shown to be irreducible to representational-computationalist positions.

Dynamical concepts and tools have been proved to be the most adequate framework to account for the complex adaptive behaviour that natural and artificial embodied and situated agents can display (Beer, 2003).

Despite the important conceptual achievements that it has enabled, evolutionary roboticist Ezequiel Di Paolo (2003) recently raised a concern upon the limits of the current situated, embodied and dynamicist robotic paradigm (the criticism is even more valid for traditional AI). His main claim can be summarized as the observation that “a robot failing in its performance does not show any signs of preoccupation”. Something is missing in current approaches to model minds with robots: a dynamical, embodied and situated sensorimotor loop is not enough to account for mental properties (in particular for intentionality). This shows up on that, unlike human made devices, animals *do* have concerns about the performance of their actions; something vital is going on through their actions. Following previous work by Hans Jonas (1966), Maturana and Varela (1980), von Uexkull (1940: 1982) and others, Di Paolo explores the hypothesis that metabolic living organization is the genuine source of value and intentionality that animals benefit from. Due to the circular and far from equilibrium condition of their metabolic body, in constant precarious existence, organisms are capable to intrinsically evaluate and “suffer” the consequences of their behavioural performance (their motivation for action is inscribed on the metabolic constitution of their flesh). It is the intrinsic normative functionality of their interactions as living systems that makes them genuine agents; while robotic machines are externally designed to perform a task that is intrinsically irrelevant to their continuing existence as mechanical systems. This leads to a difficult situation for roboticists since, if survival of metabolic organization is the source of all value and intentionality, there would be no choice other than creating self-producing chemical systems in order to achieve the goal of creating and synthetically exploring genuine intelligent phenomena. The alternative seems to abandon robotics altogether. A way out of this dilemma, Di Paolo follows, might be to create, within the domain of behavioural dynamics, self-sustaining patterns that could be considered to have equivalent properties to those argued to be on the basis of genuine intentionality in living organisms.

Far from a mere declaration of principles and intentions Di Paolo built a robotic simulation model to proof the concept¹⁰.

¹⁰ The experiment was originally inspired in a well documented psychological phenomena that could not be properly explained by available cognitivist approaches. During the 60s Kohler (1962) systematically studied re-adaptation to visual inversion. After a period of two weeks of severe difficulties to coordinate behaviour, experimental subjects wearing inverted goggles started to behave coherently, they reported that the whole perceptive up-down regularities started to emerge again in their perceptual experience of the world re-inverting the visual effect of the goggles. After goggles were removed subjects reported that their visual field appeared

Based on the of previous experiments of visual inversion in humans (Kohler, 1962) and the additional neuroscientific evidence that synaptic plasticity is homeostatically regulated (Turrigiano, 1999), Ezequiel Di Paolo (2000b) evolved a robotic simulation model where robotic agents (controlled by dynamic recurrent neural networks with homeostatic Hebbian plasticity) were capable to readapt to sensory inversion (without the agents being selected for that task during evolution). After artificially evolving the agents to perform phototaxis (with the additional requirement to maintain internal synaptic stability), Di Paolo's agents were tested for visual inversion. At the beginning of the trial agents performed phototactic behavior (the agent was able to navigate a 2 dimensional space approaching light). Later on the trial, right and left light sensors were inverted, subsequently disrupting phototactic behavior. Agents were not evolved to adapt to sensory inversion... how could, then, phototactic behavior be recovered? The hypothesis is that, by evolving the agents for phototaxis while selecting for internal synaptic stability, both synaptic stability and behavioral stability got evolutionarily coupled. Thus, “normal” phototactic behavior was sustained by a stabilized set of synaptic parameters in the agent's control architecture. When visual inversion was introduced into the agents their internal synaptic dynamics entered an unstable region. The instability of synaptic parameters, in turn, produced behavioural instabilities (the agents performed “random” movements). As a result, the synaptic parameter space was explored until phototactic behavior was recovered which, in turn, stabilized the values of synaptic parameters. In short, loss of phototactic behavior induced synaptic plastic change until phototactic sensorimotor correlations were recovered, since the stability condition for synaptic plasticity got evolutionarily coupled with phototactic behaviour.

This mechanism provides a model for neurodynamic behavioural self-maintenance where neurodynamic structures emerge that recursively depend on the behaviour they sustain and, inversely, behavioural stability and coherency depends on the stability of neurodynamic structures. A stabilized set of synaptic parameters produces phototactic behavior but, if phototactic behavior is disrupted synaptic stability is lost; and, inversely, if synaptic stability is lost coherent behavior disappears until both coupled dynamics are stabilized again. To this elementary mode of neurodynamic self-maintenance Di Paolo called *habits*. And habits might provide an interesting inside into value and intentional phenomena, non reducible to biological adaptive constraints: “Habits, as self-sustaining dynamic structures, underlay the generation of behaviour and so it is *them* that are challenged when behaviour is perturbed. An interesting hypothesis is that often when adaptation occurs in

upside-down and only recovered “normal” vision until a new process of re-adaptation occurred.

the animal world this is not because organismic survival is challenged directly but because the circular process generating a habit is.” The way out of the roboticist dilemma can now be envisioned, the metabolic living organization needs not be modelled in order to grasp intentionality: “The interaction and commerce between these structures of behaviour, and not this or that particular performance, would become the object of robotic design, and the *conservation of an organised meshwork of habits*, the basis on which to ground artificial intentionality” (Di Paolo, 2003; italics added to the original).

5. Mental Life

Di Paolo's conceptual simulation experiments discover two fundamental issues: (i) that life-like self-sustaining emergent patterns can be found within the behavioural domain and, I shall now argue in more detail, (ii) that metabolically driven agency might be insufficient, even unnecessary, for mindfulness; a specific form of life might be required to achieve psychological properties: Mental Life.

If we recover the notion of agency at the metabolic level (discussed above) it becomes evident that there are many opportunities for sustainability based on fast and flexible motility. The problem a minimal model of life is that the unicellular form of organization is severely limited to occupy such mode of existence¹¹. This had to wait until the appearance of specialised cells, within multicellular systems, capable to channel electro-chemical action potentials connecting sensory and motor surfaces in a fast, integrated and selective manner: the nervous system (NS hereafter)¹². What we get with the appearance of the NS is that on top of the basic metabolic organization a new dynamical system emerges (controlling the sensorimotor coupling with the environment) and whose dynamics are somehow locally decoupled from the underlying metabolic processes.

Thus, unlike plants and unicellular systems, organisms endowed with neural tissues can control their behaviour independently of the continuous processes of metabolism, cell replication and growth. Within multicellular life cycles, a new

dynamic domain appears (that of electrochemical action potentials) free from having to satisfy more immediate metabolic functions. As a result behavioural interactions can be quickly and efficiently achieved¹³. But this freedom (that permits a form of decoupling between interactive and constructive living processes) is accompanied by a set of global constraints that ensure the functional integration of the NS within the organism. In particular, if we take a careful overview on the set of constraints that operate creating functional order in neural and behavioural dynamics, we can abstract three general types:

1. *Architectural constraints*: composed of genetic and developmental constraints specifying some innate conditions of the architecture and “parameters” of the NS (number and types of neurons, type of connectivity, neuromodulator pathways, etc.) and the organisms embodiment (body mechanical properties and sensory modalities).
2. *Biological adaptive signals*: Internal signals from other body organs generally causally correlated with metabolic and sexual needs (pain, pleasure, etc.) with a high modulatory capacity over neural activity.
3. *Self-generated constraints*: those that the very activity of the NS generates through environmental interactions.

Type 1 and type 2 constraints subordinate the activity of the NS to satisfy biologically adaptive needs and from the point of view of neural sensorimotor dynamics these constraints appear as given, as “externally” fixed¹⁴. In such a case, behaviour, so to speak, is the “slave” of metabolism—as well as other, larger time-scale (philogenetic) self-maintaining needs (including sexual mating, kin care, etc.). In some animals neural activity is mostly pre-specified by type 1 and type 2 constraints. In fact, *C. elegans* has been shown to have highly stereotyped behaviours and a very homogeneous neural circuit among different individuals of the same species (to the extent that the number and function of neurons is identical among them—White et al. 1986, Hobert 2005). If evolution (acting on self-organized developmental processes) fixes a set of constraints that almost fully specifies the behaviour of a *C. elegans*, there is not much of a difference between such an organism and a simple Braitenberg vehicle¹⁵. Whether the sensorimotor system is

11 The kind of internal organization that is found at the cellular scale does not permit to increase the sensorimotor complexity for two reasons: (i) as more complex forms of sensorimotor interaction start to form the more likely it is that catastrophic interferences between the core metabolic network and sensorimotor mediation occur (they both share the same biochemical medium) and (ii) higher complexity requires higher size and a costly trade-off emerges between increase in size and the capacity of the unicellular organism to efficiently connect sensory and motor surfaces while moving as unity in space. As we know this problem had to wait quite a long evolutionary time to be solved.

12 A more universalist formulation of this innovation requires to abstract the kind of organizational difference that the NS introduces on living organization. Some of my colleagues and latter on myself (Moreno & Laso 2004, Barandiaran 2004, Etxeberria & Moreno 2005, Barandiaran & Moreno 2006a, 2006c) have termed this transition *hierarchical decoupling* or *informational decoupling* (Moreno et. al. 1997) of the sensorimotor control from metabolism (the underlying chemical self-constructing and self-repair machinery).

13 In addition, the evolution of the NS is accompanied by a set of changes in bodyplan that enables the channelling of metabolically recruited chemical energy into mechanical work through a musculoskeletal system. What we get as a result of this hierarchical decoupling of sensorimotor interactions (embedded on a living body) is *adaptive behaviour*.

14 Natural selection operating on internal and interactive self-organizing processes, the inner (non agency-dependent) structure of the organism, the agents organic and material constituency etc., can be viewed as good candidates for this fixation.

15 The name Braitenberg vehicle is used to refer to a class of robotic systems first proposed by Valentino Braitenberg (1986). One of the most simple forms of his vehicles consists of a robot with two wheels and two light sensors. The light sensor that is situated on the left side of the robots front

materially constructed by “the rest” of the organism and functionally integrated in its metabolic self-maintenance (as in C-elegans) or externally built and functionally decoupled from its underlying structure (as in a Braitenberg vehicle) is completely extraneous to the dynamic causal organization of behaviour. Take the example of a mutated C-elegans with disrupted sensors producing anti-adaptive behaviour (as there are many and, in fact, currently used for comparative studies) and its interactions with the environment (from the point of view of its neurodynamic phenomenology) will be as “significant” or “insignificant” as those of a non-mutant (and metabolically sustainable) one.¹⁶

But, as the size and connectivity of neural ensembles increases in encephalized animals, adaptive signals and architectural constraints are not enough to instruct the dynamics of the NS so as to produce adaptive behaviour (even if local and interactive self-organizing patterns are evolutionarily exploited). Adaptive signals can be correlated with metabolic needs and evaluate the effect of behavioral interactions on the body dynamics but can not specify *how* to achieve adaptive behavioral success. In relation to the architectural constraints, as the size of the NS increases the number of innate constraints play a smaller role on the specification of neural architectures, leaving it open to the recursive activity of the network and its history of interactions with the environment¹⁷. A space of freedom is thus created when neurodynamic mediation of adaptive behavior overcomes the regulatory capacity of adaptive signals and architectural constraints. Thus the NS needs to generate its own regulatory constraints in continuous interaction with the environment and adaptive body signals, triggering a process of dynamic self-determination that transcends metabolic values. At this point a new form of life appears, embedded on biological life but capable to generate its own normativity and value, its own distinctive identity and

is directly connected to the right motor and the light sensor on the front-right connected to the left motor. Activity on the light sensor induces activity on the motor so that if the robots receives a higher light intensity on its left side it will turn to that side (due to the differential activation of the motors). As a result the vehicle approaches a light source.

16 Paraphrasing Hans Jonas describing the lack of genuine intentionality of machines, the mutant C-elegans “may just as well be said, instead of being distressed, to abandon itself with relish to its wild oscillations, and instead of suffering the frustration of failure, to enjoy the unchecked fulfilment of its impulses. ‘Just as well’ amounts of course to ‘neither’” (Jonas 2001: 112). Surely, the C-elegans’ metabolism will suffer the effects of its sensorimotor failure and might even get stresses (forcing its metabolic dynamics to compensate the effects of behavioural failure). But this “metabolic stress” might equally be blind to its causal correlation with a particular behaviour; suffering its maladaptive condition as externally given. If natural selection fixes the correlation between specific behavioural performances and their functional contribution to metabolic self-maintenance then failure on behavioural performance does not necessarily imply that metabolic closure be affected other than externally and inaccessibly to its capacity to detect and compensate that failure as properly behavioural, intentional, failure.

17 The reason is that a bottleneck exist on how much of the brains circuitry can be genetically specified. As Elman et al. (1996) have noted in human beings only global architectural and chronotopic constraints participate on the development of the NS. *Chronotopic constraints* affect the timing of certain developmental processes and *global architectural constraints* specify global neural pathways, kinds of connectivity between neurons, etc. But none of these constraints can specify the dynamic structures that produce behaviour in adult brains.

world, a new mode of agency: that resulting from the preservation of an internal coherency of experience, the coherency of the developmental organization of neurodynamic patterns. *Mental Life* appears.

5.1. Modelling mental life

Recovering Di Paolo's notion of habits as dynamic structures, we are now in place to build a conceptual model of Mental Life in five conceptually distinguishable steps than synthesize its form of organization:

1. Neurodynamic structures are created that sustain different sensorimotor couplings with the environment (the formation of this structures might originally be due to the fixation of self-organized patterns by body adaptive signals and supported by early architectural constraints).
2. Interactive stability dependencies are created at least between some neurodynamic structures and the behaviour they sustain.
3. Internal stability dependencies are created *between* neurodynamic structures.
4. A nested web of neurodynamic structures appears when dynamic structures become progressively more independent from biological adaptive signals and innate architectural constraints and more dependent on: a) higher order stability dependencies between them and, b) the interactions that they altogether sustain with the environment.
5. The adaptive regulation of behaviour to preserve the stability dependencies of the web of neurodynamic structures becomes the main organizational principle of brain activity and behaviour.

Mental Life appears when the adaptive conservation of the internal organization of neural dynamics becomes the main principle of sensorimotor regulation. From a set of initial conditions of huge developmental plasticity, triggered by biological adaptive signals and channelled by architectural constraints, the NS generates more and more internal constraints and interdependencies between behaviourally emergent self-organized patterns, until the preservation of the internal coherency of these nested structures takes over the regulation of embodied brain dynamics.

This way “a form of life” appears in the realm of sensorimotor dynamics. The minimal model of cellular organization sketched above represented the essential causal

structure of life as a circular network of self-sustaining chemical reactions. It was shown how this network distinguished itself from its environment through a selective membrane that actively regulates the thermodynamic and material flow required for its continuing existence. We can now envision an analogous process in the domain of neurodynamic organization: a web of interdependent dynamic structures is progressively created through a continuous sensorimotor flow, regulated by the behavioural activity of the system. Emergent patterns of brain activity are analogous to chemical reactions, the sensorimotor flow analogous to the thermodynamic and material flow, the selective action of the membrane analogous to the behavioural control of the sensory flow and the behavioural and neurodynamic tendency towards the preservation of an internal (experiential) coherence might be equivalent to the metabolic self-sustaining organization of life¹⁸.

5.2. Five characteristic features of mental life

It is now time to revisit the 4 model features of living organization extracted previously (including an additional one not present in metabolic life: its living and lived embodiment). Both isomorphisms and dissonances between mental and biological life as well as their mutual relationships will help on further elaborating and evaluating the model while completing it with some post-cognitivist research trends.

a) Emergent self

Like minimal cellular chemical dynamics, brain dynamics also involves an irreducible emergent organization where the role of neurons and neural ensembles is contextually defined within a circular causal structure where micro and macro levels of correlation mutually constraint each other. Brain activity cannot be appropriately studied by locally decomposing units and putting them back together through simple computational relationships. Rather than a point to point information transfer between functionally specific modules, large scale brain activity responds to the transient correlation of distributed neural ensembles, as a result of multiple feedback loops between different brain regions. Thus, despite the anatomical and functional modularity that might be found among certain brain

regions, mental properties such as meaning and intentionality (Freeman, 1997, 2000) or conscious awareness (Edelman & Tononi, 2000) have been argued to be the result of emergent and circular dynamics. There is an increasing number of experiments and simulation models supporting this form of conceptualizing brain activity from a dynamical and holistic perspective: chaotic approaches to large scale brain activity (Freeman, 2000; Tsuda, 2001), the dynamic core hypothesis (Varela, 1995; Varela *et al.*, 2001), timing nets (Cariani, 2001), adaptive resonance theory (Carpenter & Grossberg, 2003), analysis in terms of transient correlations (Friston, 2001), to mention but a few. Not surprisingly, some analytic and modelling techniques that are used in synthetic protocell biology are found at the level of neurodynamic research: dynamic and stochastic models, network analysis, chaos, complexity measurements, criticality, power law distributions, etc.

Finally, a crucial feature of our model of biological life was that of self-maintenance. Similarly the activity of the NS can be seen as continuously regenerating itself, through multiple reverberating circuits, self-generated or spontaneous activity, etc. (Cariani, 1999). A significant aspect of this self-maintenance, as we have seen before, is that it is closed through sensorimotor interactions. Which leads us to the second feature of our model.

b) Situated openness

The necessity of all living systems to maintain an open thermodynamic and material flow with their environments in order to sustain their dissipative organization might be seen as somehow isomorphic with the necessity for psychological identity to be situated in a sensorimotor world extracting, through it, a set of coherent correlations that are necessary to maintain its organization (Di Paolo's model of habits provides a minimalist instance of self-maintaining situated openness that certainly inspires this analogy). We are continuously dealing with our biological, emotional, social and cognitive world out of which our psychological identity (our personality) is created and maintained. Consequently, the activity of the NS should not be seen as stimulus driven but as continuously engaged with the world on the maintenance and regulation of its dissipative organization; an organization that can only be complete *through the environment*. This feature of mental life puts action in the centre. Action does not appear as a final step of a representational planing process, a receipt to be executed in the world. On the contrary, the inter-active flow that sustains neurodynamic organization is constitutive of Mental Life. As a result, what-the-system-is is intertwined with what-the-system-does: neurodynamic organization is cause and effect of the interactions it sustains. Thus, it is not only that cognitive processes are situated and context-dependent (which is one of

¹⁸ A similar analogy between a model of metabolic circular causality and the situated activity of the nervous system was first formulated by Maturana & Varela (1972, 1980) the model of such type of organization was conceptually formulated as "operationally closed". In Varela's words: "The operational closure of the nervous system then brings forth a specific mode of coherence, which is embedded in the organism. This coherence is a cognitive self : a unit of perception/motion in space, sensory-motor invariances mediated through the interneuron network. The passage to cognition happens at the level of a behavioral entity, and not, as in the basic cellular self, as a spatially bounded entity. The key in this cognitive process is the nervous system through its neuro-logic. In other words the cognitive self is the manner in which the organism, through its own self-produced activity, becomes a distinct entity in space, but always coupled to its corresponding environment from which it remains nevertheless distinct. A distinct coherent self which, by the very same process of constituting itself, configures an external world of perception and action." (Varela, 1992).

the most celebrated theoretical achievements of post-cognitivist approaches —Wheeler, 2005; Clark, 1997) but, going even further, that Mental Life exploits this situatedness to generate and regulate its internal organization (and not just the structure of behaviour). It follows that isolation from the environment will destroy Mental Life (as a prolonged disruption of the thermodynamic and material flow will destroy a living system). This is, in fact, the case if we are to follow some studies regarding the psychological effects of sensory deprivation and solitary confinement (Haney, 2003), showing how severe personality disorders follow from long periods of isolation.

c) Normativity

The evidence for an specifically mental normativity, distinct from evolutionary or metabolic one, shows up on the fact that failure of behavioural performance does not necessarily imply failure for biological adaptation. Inversely success on cognitive performance does not necessarily involve biological success. In terms the model of Mental Life, an internal neural process or a behavioural interaction becomes functional if it contributes to the self-maintenance of neurodynamic organization. Genuine mindful normativity appears when the adaptive values are lost as initial conditions of the development of the NS and progressively substituted by the preservation of the web of internal and interactive stability dependencies. As a consequence the model of Mental Life encompasses a wider range of normative dimensions that the purely evolutionary adaptive or the epistemic and referential.

Our model of mental life entails a significant shift on cognitivist assumptions in relation to the normative (truth) status of mental states: from a representational approach where normativity is defined by the correlation between internal states and external states of affairs to a view where normativity is defined in terms of an interactively maintained internal consistency and coherency of experience. This view does not rule out the notion of semantics or even that of adequacy, it just does not reduce it to a causal correlation between internal states and “states of affairs” in the environment (Dretske, 1988) or to an evolutionary selective history that ensures a correspondence relationship (Millikan, 1984); making semantics, in both cases, external to the internal causal organization of the system (Bickhard, 2000). Within the model of Mental Life just sketched, intentional semantics might be best viewed as affordances or canalizations of possibilities for action which might, or might not, turn out to be adequate to what the consequent engagement with the world permits. The relationships between different dynamic structures within neural organization will be accordingly regulated to preserve an internal coherency regarding future interactive expectations and sensorimotor correlations.

d) Agency

The situated openness of life introduces a problem of demarcation between system and environment. At the level of minimal metabolic life the problem is solved by the encapsulation of the core chemical organization within a self-generated membrane while the selective action of the membrane on the system-environment diffusion processes demarcates a control asymmetry giving rise to agency. Models of cognitive processes that put the emphasis on extended, distributed and situated dynamics have to face the problem of how to define the identity of the subject as distinct from the environmental distributed features that are functionally integrated on the production of behaviour. If cognitive behaviour is the result of a non-trivial causal spread, as Wheeler has called it (2005), ... can we really speak of agency? In order to answer this question we need to take into account how the situatedness and agency of metabolic life is different to that of Mental Life.

Metabolism needs to be situated in a material and thermodynamic flow and it needs to regulate the inflow and the outflow. But this flow does not constitute a cycle: i.e. the material and thermodynamic outflow (heat and waste products) does not *recursively* feed-back into the inflow generating a *closed loop*. For instance, the effect of a molecule or a change of temperature in a cellular system is directly specified by the relational properties of the molecule (as a physico-chemical entity) and the metabolic organization of the cell. The very appearance of motility (providing the domain in which Mental Life should latter appear) produces a completely different mode of situatedness and, consequently, of agency. As a result, the way in which objects and processes in the world become significant or functional for Mental Life becomes different from its metabolic counterpart. Mental Life's mode of situatedness is circular, transformations in motor surfaces have a direct effect on sensory surfaces and neurodynamic activity is continuously engaged on that circularity. For Mental Life environmental objects and processes have no direct effects but only through the way in which they are engaged within the sensorimotor cycle. There are two complementary aspects in which mental agency shows up: (i) the causal asymmetry provided by the circular and self-sustaining organization of internal dynamics controlling its sensorimotor flow and (ii) the selective engagement with the environment Mental Life's regulatory dynamics creates, cutting out a world of interactions that is continuously shaped by the goals and intentions of the subject.

e) Embodiment

Mental Life has an additional characteristic feature not present at the metabolic level: its living and lived embodiment. While metabolism is embodied on non-living components (molecules), Mental Life (as-we-know-it) is embedded on a living body. This

includes not only the mechanical musculoskeletal system and sensory organs (as current views on embodied cognition have repeatedly addressed) but also, and more fundamentally, an organismic living body with its metabolic regulatory needs. In fact, as Moreno and Lasa (2003) have pointed out, the appearance of mind is strongly linked to a set of bodyplan transitions where encephalized brains require and enable a neurally regulated bodyplan (i.e. not only the brain is embodied but the body is also embrained). As a consequence, part of the nervous system is not dedicated to deal with the external world but with an internal metabolic environment, that is in charge of regulating (from digestion to blood circulation). As Damasio (1994, 1999) has repeatedly argued, the feedback relationships between body-regulation dynamics and the sensorimotor dynamics constitute an emotional world that becomes constitutive of mental processes. In fact, the neural activity in charge of body regulation drives much of the early developmental process of Mental Life as a sophisticated form of body signalling. This is a necessary requirement for those neural systems not fully determined by innate architectural constraints since behavioural adaptivity must be continuously adjusted and evaluated on the basis of the effect (on body homeostasis) of the interactions that the NS maintains. Latter in the developmental process the modulatory capacity of the NS of the interior, as Edelman has called it (1989), is recruited by the sensorimotor nervous system to regulate its increasingly complex organization. Thus, the embrained living body takes part on both the formation and the maintenance of Mental Life.

5.3. The mind from the MUN

It is now time to evaluate if the conceptual model of Mental Life sketched here satisfies the MUN constraints stated at the beginning of this chapter; making it explicit how, and to which extent, it might be integrated into scientific research. The model is *Minimalist* on that it contains all but no more than the necessary and sufficient conditions to specify the domain of the mental as a specific domain in itself with its own level of normativity and agency. The mind is not just a sophisticated form of biological adaptation, behaviour, chaotic dynamical system, developmental process or situated and embodied neural activity. The hypothesis proposed here is that the mind is defined by a specific form of organization: Mental Life. For some, the model might be too demanding on that simple forms of sensorimotor behaviour that are often taken to be minimal cases of cognition or mindfulness (such as chemotactic behaviour in bacteria) will be left out as non-mental. For others (specially for those that take human intelligence as the paradigmatic reference) the model might be considered too minimal and below the level of complexity that is necessary to characterize genuine psychological phenomena. But what the conceptual model of Mental Life just sketched somehow implies is that there

exists a gradient towards the mental, defined by *increasingly* interdependent number of behaviourally generated neurodynamic structures and, particularly, by the progressive appearance of a regulatory principle of conservation of the resulting organization. Thus, rather than a lower level boundary, this model of Mental Life works as a limit concept that specifies a gradient of neurodynamic autonomy (a behavioural self-determination that might never be complete). As such, the model itself does not permit to establish a clear cut distinction specifying when does Mental Life exactly start (in evolution or ontogenetic development). However, if the model is fully and adequately naturalized, it should be possible to make measurements and comparisons between different systems regarding their degree of mindfulness. And it might turn out that natural systems are non-homogeneously distributed on the mindfulness axis and that there exists a non-linear transition from mentally inanimate to mentally alive forms of behaviour (probably due to some complex evolutionary feedback between brain, body and social environment).

The abstract formulation of this hypothetical organization satisfies the constraint of *Universalism*. No reference to specific anatomical or functional structures is required to define psychological phenomena and, on the other hand, it is in relation to this essential organization that learning, emotions, intentions, etc. can be defined. Not only the organization but the domain in which it appears, was also formulated in universalist terms. What the NS is was not defined by any live-on-earth particularity of neural cell types, but by the abstract properties of neurons as components capable to create relatively unconstrained sensorimotor dynamics¹⁹. The satisfaction of a universalist formulation should be able to identify mental-life-as-it-could-be and becomes, thus, of fundamental importance to solve the problem of the possibility of artificial minds. In this sense it is evident that standard robots isolated or subject to input-output deprivation are equally stable than if situated in an environment: what they *are* is independent from what they *do*. Current robots do not suffer from the thread of mental death nor do they benefit from Mental Life. But the question is whether it is *in principle* possible for robots to have minds. Unlike the cognitivist hypothesis about the nature of cognition, in which computer implementations of symbolic computations are but instances of the nature of cognition, computer simulation models of Mental Life are not realizations. A numerical simulation of the states of the variables of a dynamical system cannot be ontologically equivalent with a real dynamical system (Pattee, 1995). But this still leaves open the question of whether Mental Life can be realized by artificial systems. A more precise universalist formulation of the model, together with a clear definition of the term “artificial”, should be able to provide a

¹⁹ This abstract property is formulated as the decoupling of the NS from metabolic organization (see Barandiaran & Moreno, 2006c, for a more detailed explanation).

definitive answer. In this sense it is crucial to elucidate the nature of the *stability dependencies* between dynamic structures and the sense in which the sensorimotor organization might be required to be dissipative or *far from equilibrium*. A strong interpretation of these terms would require that dynamical changes in the system be irreversible which, in turn, will imply that the model would not possibly be realized as a mechanical system (in terms of Newtonian mechanics). Thus the most genuine “artificial” realization of Mental Life we can nowadays try to achieve might require robotic embodiments of cultured cells (living components that are truly far from equilibrium dissipative structures capable to support altogether far from equilibrium neurodynamic structures)²⁰. But another crucial aspect is the role played by the embrained living body on the formation and regulation of neurodynamic organization in mental-life-as-we-know-it. Does the model of Mental Life necessarily imply that it be embodied on a living systems? If Mental Life is about a specific form of neurodynamic situated organization and not something directly defined by metabolism... Would an artificial internal environment (that feeds-back to the sensorimotor system in the same manner as the living body does) suffice to create genuine Mental Life in a robotically embodied set of cultured neurons?

I do not mean to provide a definitive answer to these questions here but to show how the proposed conceptual model permits to approach these issues and to highlight which are the relevant questions that need to be further developed. In this sense what is importantly required is to achieve an accurate simulation model of minimal Mental Life that could be tested and complemented with empirically grounded theories of large scale and situated brain dynamics. We are here entering the requirements of the *Naturalist* constraint. As it stands now, the model does not break any naturalist constraint: no *ad hoc* substances needed to be included, nor observer dependent properties were taken to be causally relevant components of the model and the five characteristic features of mind were grounded or inferred from the model and its application to known psychological phenomena and neuroscientific studies. Somewhere else (Barandiaran & Moreno 2006c) I have, together with Alvaro Moreno, traced in more detail the bottom-up transitions that lead from the origin of life to the domain of adaptive behaviour; providing a proper biological grounding of what was here taken as a bottom-line theoretically primitive causal domain for Mental Life: that of embodied and situated neural dynamics. And here stands, precisely, one of the epistemological strengths of the present model: on that, unlike some cognitivist formulation of mindfulness, it takes as theoretically primitive (i.e., as the basis on which the rest of the theoretical foundations are to be built) a mathematically

formalizable domain that is, in addition, directly measurable (not without difficulties) in terms of physical sensory properties, mechanical behaviour and neural activity (electrical and biochemical). In this sense the model integrates some of the tools and theoretical achievements that are more prominently post-cognitivist: the dynamical system approach (Beer 1995, van Gelder & Port 1995). In addition, the model entails that the mind cannot be sufficiently characterized merely as a causal domain (be it dynamical or computational, disembodied or situated) but as a specific form of organization within that domain (a further development of this topic might be found at Barandiaran & Moreno, 2006b).

The question is whether the proposed form of organization can be properly naturalized in the sense of being formalized and introduced in empirical and synthetic research. In that sense, the way to proceed requires to find mathematical formulations for the components and relationships of the model in terms of dynamical system theory. A task that should not, in principle, show major difficulties: the notion of dynamic structure can be understood as a local attractor (possibly requiring a more sophisticated formulation such as that of a chaotic attractor, a neuronal transients, etc.), stability dependencies might be formulated as global interdependencies between the structure of different local attractors, while transitions between dynamic structures and their mutual transformations might be studied through the concept of chaotic itinerancy (Tsuda, 2001). Finally the notion of adaptive regulation of the web of dynamic structures (or attractor landscape) implies the major theoretical and mathematical challenge. The difficulty resides on that there is a form of self-reference involved: it is the web itself that regulates its stability dependencies (not an external source of control that can measure and operate separately upon the dynamics of the network) so that an explicit distinction between control parameter, controlled variables and controller subsystem might not be made. Ashby's notion of ultrastability has been proposed to approach such form of organization (Di Paolo, 2003) while there are also other, more radical formulations, that deny the possibility of a dynamical formalization of the kind of closure involved in mental and other forms of life (Rosen, 1991; Kampis 1991).

In any case, it is clear that simulation models might be able to implement maximal approximations of the conceptual model of Mental Life presented here. A conceptual modelling approach that remains minimalist but universally generalizable (and close to what can be implemented in computer simulations) is necessary if we are to understand, in its complexity, the kind of interactive and neural organization that supports our mental lives. I have sketched here a conceptual model of Mental Life that meets these demands. However, the construction of this model needs to be tested and feed with a continuous circulation

²⁰ For an overview robotic research with cultured neural cells see Bakkum *et al.* (2004).

from neuroscience (specially from large-scale models of brain activity), while simulation models of embodied neurodynamic agents might be able to integrate the results of this circulation and display the complex and emergent properties that we take to be essential to mind. The final picture is still far from definitive but the tools and concepts are progressively closer to it.

6. There is life beyond cognitivism

Something analogous to the formation of a circular emergent organization at the origins of life could be in place underlying the elusive foundations of psychological phenomena. The idea that an analogous form of organization might be generating living and mental phenomena has a long standing tradition (from Aristotle to Maturana & Varela). What current models of synthetic protocell biology permit is to understand the fundamental and minimal organization of life (its essence) in a much more detailed account than what we had previously available (assembling through computer simulation models the huge amount of data that molecular biology has produced during the last decades). The accuracy that present models of minimal life have gained might also help on the task of building analogue conceptual (simulation) models in the realm of neural embodied dynamics. I have outlined how current modelling paradigms in cognitive science (in particular the field of evolutionary robotics) permits to construct minimal simulation models of embodied and situated neurodynamic agents. Unlike past attempts to define the mind in terms of systemic and holistic properties, the binding that nowadays is being carried out between simulation techniques and large scale brain dynamic studies permits to develop a much more detailed and scientifically fruitful theoretical foundation for psychology than what we could have ever done before. I have proposed here a conceptual model of Mental Life as a hypothetical organization made of sensorimotor (neuro)dynamic structures which are nested through internal stability dependencies and altogether dependent on the behavioural interactions they sustain, where the preservation of such stability dependencies becomes the main organizational principle. From this model a set of characteristic features were inferred: (i) the formation of an emergent *self*, (ii) the *openness* of mental life as a constitutively situated process, (iii) the *normative* character that certain internal and interactive processes acquire as functionally integrated in the regenerative and self-maintaining character of neurodynamic organization, (iv) how mental life provides the means for the constitution of an *agent* that creates its own world throughout its selective coupling with the environment and (vi) the living and lived *embodiment* in which the mind-as-we-know-it is embedded on, providing a sort of internal environment that becomes constitutive of the process that bring forth and sustain Mental Life.

We keep making sense of our lives. And science provides extremely accurate and powerful models and metaphors to do so, cognitive technologies we live by. As cognitive agents we cannot escape the urge to conceptualize, model, and inhabit our situated and precarious existence. Theoretical foundations for psychology permeate our lives in multiple dimensions: through the institutional policies that they justify, the therapies they design, the technologies that accompany them, the metaphors they inspire... Cognitivism has long dominated our understanding of the mind, conceptualizing it as a computational and representational machine. But however difficult it might turn out to avoid linguistic, propositional and representational descriptions of some of the scenes of our everyday mental lives, reducing them to computational processes of that sort will amount to reduce living phenomenology to a differential reproduction of a set of genetic permutations bearing representational relationships with phenotypic states of affairs. But this reduction is not necessarily the best available and scientifically guaranteed model today. There is life beyond cognitivism. Other metaphors, models and technologies populate our scientific ecosystems. The conceptual model I sketched here, synthesizing existing trends and certainly inspired by them, condenses some of the new opportunities that are opened when models of biological organization, together with simulation techniques and dynamicist neuroscience, make it possible to reconceptualize the foundations of psychology.

The systemic and integrative (holistic) view of the activity of the NS is not new but the conceptual model of Mental Life proposed here, as a central notion for a post-cognitivist psychology, might be able to integrate some current post-cognitivist trends in psychology and cognitive science (dynamicism, embodiment, situatedness). In particular, the model captures a number of phenomena that remain alien to traditional cognitivist computationalism but that constitute, nevertheless, core phenomenological aspects of our mental lives. In contrast with cognitivism, the basic, fundamental organization of Mental Life is not that of a syntactical representation of the objective world whose correlation is measured by an external observer or natural selector. This is not to say that, like in biological life (where molecular templates permit to build increasingly complex molecular constraints for self-organized biochemical and biophysical processes) recombinable or compositional structures might not become powerful “technologies” in the domain of the mental. Symbolic and computational structures might be integrated or emergent from the fundamental form of organization that constitutes life. Higher level regularities (such as those found in linguistic structures) might be seen as internally structuring (constraining and enabling) brain dynamics or scaffolding the situated and

distributed dynamic environment that the brain is embedded on (like instrumental technologies, cognitively structured environments or socially constrained protocols and institutions—Clark, 1997). But mental concerns, meanings, intentions, values, habits, pauses, trauma, desires ... would never be understood without taking into account how the underlying brain-body-environment dynamics make them be there, as patterns of the sub-symbolic neurodynamic organization that constitutes our mental lives; continuously engaged in a world that is both the result and the condition of possibility of its permanent re-creation.

References

- Almásson, N., Edelman, G.M. & Sporns, O. (1998), 'Behavioral constraints in the development of neuronal properties: A cortical model embedded in a real-world device', *Cerebral Cortex* **8**:346—361.
- Bakkum, D. J., Shkolnik, A. C., Ben-Ary, G., Gamblen, P., DeMarse, T. B. & Potter, S. M. (2004) Removing some 'A' from AI: Embodied Cultured Networks, In Iida, F., Pfeifer, R., Steels, L. and Kuniyoshi, Y. (Eds.) *Embodied Artificial Intelligence*, Springer Verlag, pp. 130—145.
- Barandiaran, X. (2004) Behavioral Adaptive Autonomy. A milestone in the Alife route to AI? In Pollack, J., Bedau, M.A., Husbands, P., Ikegami, T., Watson, R.A. (Eds.) *Proceedings of the 9th International Conference on the Simulation and Synthesis of Living Systems (ALife9)*, MIT Press, pp. 514—521.
- Barandiaran, X. & Moreno, A. (2006a) ALife models as epistemic artefacts. In Rocha, L.M., Jaeger, L.S., Bedau, M.A., Floreano, D., Goldstone, R.L. & Vespignani, A. (Eds.) *Proceedings of the 10th International Conference on the Simulation and Synthesis of Living Systems (ALifeX)*. MIT Press, pp. 513—519.
- Barandiaran, X. & Moreno, A. (2006b) On what makes certain dynamical systems cognitive. *Adaptive Behavior* **14**(2): 171—185.
- Barandiaran, X. & Moreno, A. (2006c) Adaptivity: from metabolism to behaviour. Submitted
- Bechtel, W. (2006) Biological mechanisms: Organized to maintain autonomy. In F. Boogerd, et al., (Eds.) *Systems Biology; Philosophical Foundations*. New York: Elsevier, in press.
- Bedau, M. A. (1998) Philosophical content and method of artificial life. In Bynum, T.W. and Moor, J.H. editors. *The digital phoenix: How computers are changing philosophy*. Oxford, UK: Basil Blackwell. pp. 135—152.
- Beer, R. (1990) *Intelligence as Adaptive Behaviour: An Experiment in Computational Neuroethology*. Academic 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. (1995) A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence* **72**:73—215.
- Beer, R.D. (2003) The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, **11**(4): 209—243.
- Békés, F. (1975) Simulation of kinetics of proliferating chemical systems. *Biosystems* **7**: 189—195.
- Bickhard, M.H. (2000) Information and Representation in Autonomous Agents. *Journal of Cognitive Systems Research*, **1**(2):65—75.
- Bohr, N. (1913) On the Constitution of Atoms and Molecules (Parts 1, 2 and 3). *Philosophical Magazine* **26**: 1-25, .476-502, 857-875.
- Braitenberg, V. (1986) *Vehicles: Experiments in Synthetic Psychology* (MIT Press)
- Brooks, R.A. (1997) From Earwigs to Humans. *Robotics and Autonomous Systems*, **20**(2-4):291—304.
- Cariani, P. (2001) Neural timing nets. *Neural networks* **14**(6-7): 737—753.
- Carpenter, G.A. & Grossberg, S. (2003) Adaptive resonance theory. In M.A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks, Second Edition*, Cambridge, MA: MIT Press, 87—90.
- Cartwright, N. (1983) *How the Laws of Physics Lie*. Oxford: Oxford University Press.
- Chalmers, D. (1995) Facing up the problem of consciousness. *Journal of Consciousness Studies* **2**(3):200—219.
- Chemero, A. (2000) Anti-representationalism and the Dynamical Stance. *Philosophy of Science* **67**: 625—647.
- Christensen, W.D. & Bickhard, M.H. (2002) The Process Dynamics of Normative Function. *Monist* **85** (1): 3—28.
- Clancey, W. (1989) The frame of reference problem in cognitive modelling. *Proceedings of the 11th Annual Conference of the Cognitive Science Society*, Ann Arbor. Lawrence Erlbaum Associates, pp. 107—114.
- Clark, A. (1997) *Being there. Putting Brain, Body, and World Together Again*. MIT Press, Cambridge, MA.
- Cliff, D. T. (1991). Computational Neuroethology: A Provisional Manifesto. In Meyer, J. A. and Wilson, S.W., (Eds.) *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior (SAB'91)*. MIT Press, Cambridge, MA.
- Cliff, D., Harvey, I., and Husbands, P. (1993). Explorations in evolutionary robotics. *Adaptive Behavior*, **2**(1):71—104.
- Csendes, T. (1984) A simulation study on the chemoton. *Kybernetes*, **13** (2): 79.
- Damasio, A.R. (1994) *Descartes' Error. Emotion, Reason and the Human Brain*. G.P. Putnam's Sons, NY.
- Damasio, A.R. (1999) *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. Heinemann, London.
- Dawkins, R. (1976) *The selfish gene*. Oxford University Press.
- Dennett, D.C. (1994) Artificial Life as Philosophy. *Artificial Life* **1**: 291—292.
- 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., (Eds.), *Dynamical Systems Approach to Embodiment and Sociality*, pages 19—42. Advanced Knowledge International, Adelaide, Australia.
- Di Paolo, E., Noble, J. & Bullock, S. (2000). Simulation models as opaque thought experiments. In Bedau, et al. (eds.), *Proc. Artificial Life VII*. Cambridge MA: MIT Press, pp. 497—506.
- Dretske, F. (1988) *Explaining Behavior*. MIT Press.
- Duijn, M., Keijzer, F. & Franken, D. (2006) Principles of Minimal Cognition: Casting Cognition as Sensorimotor Coordination *Adaptive Behavior*, **14**(2):157—170.
- Edelman, G.M. (1989). *The remembered present*. Basic Books.
- Edelman, G.M. (1992) *Brilliant Air, Brilliant Fire: On the Matter of Mind*. Basic Books, New York.
- Edelman, G.M. & Tononi, G. (2000) *Consciousness: How Matter Becomes Imagination*. Penguin Books, London.
- Elman, J.L., Bates, E., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: a connectionist perspective on development*. Cambridge, MA: MIT Press.

- Emmeche, C. (1994) *The Garden in the Machine*. Princeton University Press
- Freeman, W.J. (1997) Nonlinear neurodynamics of intentionality. *Journal of Mind and Behavior* **18**(2-3): 291–304.
- Freeman, W.J. (2000) *How Brains Make Up Their Minds*. New York: Columbia University Press.
- Friston, K.J. (2000) The labile brain (I, II & III). *Phil. Trans. R. Soc. Lond. B* **355**: 215–265.
- Funahashi, K. and Nakamura, Y. (1993). Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks*, **6**:1–64.
- Gánti, T. (1971) *The Principle of Life*. 1st edition. (In Hungarian) Gondolat, Budapest.
- Giere, R. N. (1988) *Explaining Science: A cognitive approach*. University of Chicago Press.
- Godfrey-Smith, P. (2005) Folk Psychology as a Model. *Philosophers' Imprint* **5**(6). <www.philosophersimprint.org/005006/>
- Haney, C. (2003) Mental Health Issues in Long-Term Solitary and "Supermax" Confinement *Crime & Delinquency*, Vol. 49 No. 1, January 2003 124-156.
- Harvey, I. (2000) Robotics: Philosophy of Mind using a Screwdriver. In Gomi, T. (Ed.) *Evolutionary Robotics: From Intelligent Robots to Artificial Life*, Vol. III, AAI Books, Ontario, Canada, pp. 207–230.
- Harvey, I., Di Paolo, E., Wood, R., Quinn, M., & Tuci, E.A. (2005) Evolutionary Robotics: A new scientific tool for studying cognition. *Artificial Life*, **11**(1-2): 79–98.
- Harvey, I., Husbands, P., Cliff, D., Thompson, A., and Jakobi, N. (1997). Evolutionary Robotics: the Sussex Approach. *Robotics and Autonomous Systems*, **20**:205–224.
- Hobert, O. (2005) Specification of the nervous system of C-elegans. The C. elegans Research Community (Ed.) *WormBook*, doi/10.1895/wormbook.1.12.1, <http://www.wormbook.org>.
- Husbands, P., Harvey, I., Cliff, D., and Miller, G. (1997). Artificial Evolution: A New Path for Artificial Intelligence? *Brain and Cognition*, **34**:130–159.
- Husbands, P., Smith, T., Jakobi, N., and O'Shea, M. (1998). Better living through chemistry: Evolving GasNets for robot control. *Connection Science*, **10**(3-4):185–210.
- Ideker, T., Galitski, T. & Hood, L. 2001, A new approach to decoding life: Systems Biology, *Annu. Rev. Genomics Hum. Genet.* **2**, 343–372.
- Jonas, H. (1966) *The phenomenon of life: towards a philosophical biology*. Northwestern University Press.
- Kampis, G. (1991) *Self-modifying systems in biology and cognitive science*. Pergamon Press.
- Kauffman, S.A. (1971) Cellular homeostasis, epigenesis and replication in randomly aggregated macromolecular systems. *Journal of Cybernetics* **1**: 71–96.
- Kauffman, S.A. (1993) *The Origins of Order: self-organisation and selection in evolution*. Oxford University Press.
- Kauffman, S.A. (2003) Molecular autonomous agents. *Phil. Trans. R. Soc. Lond. A* **361**:1089–1099.
- Keller, E.F. (2005). The Century Beyond the Gene. *J. Bioscience* **30**(1): 3–10.
- Kitano, H. (2002) Systems Biology: a brief overview. *Science* **295**:1662–1664.
- Kohler, I. (1962) Experiments with goggles. *Scientific American* **206** (5): 62–72.
- Langton, C.G. (1989) Artificial Life. In Langton, C. (Ed.) *Artificial Life*, SFI Studies in the Sciences of Complexity, Proc. Vol. VI. Redwood City, CA: Addison-Wesley.
- Maturana, H. & Varela, F. (1972) *De Máquinas y Seres Vivos: Una teoría sobre la organización biológica*, Santiago de Chile: Editorial Universitaria.
- Maturana, H. & Varela, F. (1980) *Autopoiesis and Cognition: The Realization of the Living*. Dordrecht: Reidel.
- Mavelli, F. & Ruiz-Mirazo, K. (2007) Stochastic simulations of minimal self-reproducing cellular systems. *Philosophical Transactions of the Royal Society of London B*, in press.
- Millikan, R.G. (1984) *Language, Thought and Other Biological Categories*. MIT press, Cambridge MA.
- Moreno, A. & Etxeberria, A. (2005) Agency in natural and artificial systems. *Artificial Life* **11**(1):161–176.
- Moreno, A. & Lasa, A. (2003) From Basic Adaptivity to Early Mind. *Evolution and Cognition*, **9** (1):12–30.
- Moreno, A., Umerez, J. & Ibañez, J. (1997) Cognition and Life. The Autonomy of Cognition. *Brain and Cognition* **34**(1): 107–129.
- Morrison, M. (2000) *Unifying Scientific Theories*. Cambridge: Cambridge University Press.
- Nicolis, G. & Prigogine, Y. (1977) *Self-organization in Non-equilibrium Systems*. Wiley, New York.
- Nolfi, S. & Floreano, D. (2000). *Evolutionary Robotics: The Biology, Intelligence and Technology of Self-Organizing Machines*. MIT Press.
- O'Malley, M.A. & Dupré, J. (2005) Fundamental issues in Systems Biology. *BioEssays* **27**: 1270–1276.
- Pattee H.H. 1995. Artificial Life Needs a Real Epistemology. In Moran F., et al. (Eds.) *Advances in Artificial Life*. Berlin: Springer, pp. 23–38.
- Port, R. & van Gelder, T. (1995) *Mind as motion: Explorations in the dynamics of cognition*. MIT Press.
- Quine, W. (1969) Epistemology Naturalized. In *Ontological Relativity and Other Essays*. Columbia University Press.
- Quinn, M. (2001) Evolving communication without dedicated communication channels. In Kelemen, J. and Sosik, P., editors, *Proceedings of ECAL01*, pages 357–366. Springer Verlag.
- Rosen, R. (1958). A relational theory of biological systems. *Bulletin of Mathematical Biophysics*. **20**: 245–341.
- Rosen, R. (1991) *Life Itself - A Comprehensive Enquiry into the Nature, Origin and Fabrication of Life*. Columbia University Press, New York.
- Rosen, R. (1991). *Life itself. A comprehensive enquiry into the nature, origin and fabrication of life*. New York: Columbia University Press.
- 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.
- Rupin, E. (2002) Evolutionary Autonomous Agents. A neuroscience perspective. *Nature Reviews Neuroscience* **3** (February): 132–141.
- Schördinger, E. (1944) *What is Life - the Physical Aspect of the Living Cell*. Cambridge University Press.
- Slocum, A.C., Downey, D.C. & Beer, R.D. (2000). Further Experiments in the Evolution of Minimally Cognitive Behavior: From Perceiving Affordances to Selective Attention. In J. Meyer, A. Berthoz, D. Floreano, H. Roitblat and S. Wilson (Eds.), *From Animals to Animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*. Cambridge, MA: MIT Press.
- Solé, R.V., Munteanu, A., Rodriguez-Caso, C. & Macía, J. (2007) Synthetic protocell biology *Philosophical Transactions of the Royal Society of London B*, in 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.
- Taylor, N.R. & Taylor, J.G. (2000) *Hard-Wired Models of Working*

- Memory and Temporal Sequence Storage and Generation. *Neural Networks* **13**:201–224.
- Tsuda, I. (2001) Toward an interpretation of dynamic neural activity in terms of chaotic dynamical systems, *Behavioral and Brain Sciences* **24**(5):793–847.
- 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)*.
- Turrigiano, G.G. (1999) Homeostatic plasticity in neuronal networks: the more things change, the more they stay the same. *Trends in Neuroscience* **22**:221–228.
- van Duijn, M., Keijzer, F., & Franken, D. (2006) Principles of Minimal Cognition: Casting Cognition as Sensorimotor Coordination. *Adaptive Behavior*, **14**(2): 157–170.
- Varela, F. (1979) *Principles of biological autonomy*. Elsevier, New York
- Varela, F. (1992) Autopoiesis and a Biology of intentionality. In McMullin, B. & Murphy, N. (eds.) *Autopoiesis & Perception*. pp. 1–14. Proceedings of a Workshop held in Dublin City University, August 25th and 26th 1992. School of Electronic Engineering Technical Report, Dublin, 1994.
- Varela, F. (1995) Resonant Cell Assemblies: A new approach to cognitive functions and neuronal synchrony. *Biological Research* **28**:81–95.
- Varela, F., Lachaux, J.P., Rodriguez E. & Martinerie, J. (2001) The brainweb: phase synchronization and large-scale integration. *Nature Reviews Neuroscience* **3**:229–239.
- Varela, F., Maturana, H., & Uribe, R. (1974). Autopoiesis: the organization of living systems, its characterization and a model. *Biosystems* **5** 187–196.A
- Varela, F., Thompson, E., & Rosch, E. (1991) *The embodied mind. Cognitive Science and Human Experience*. Cambridge MA: MIT Press
- von Uexküll, J. (1940). *Bedeutungslehre*. (Bios 10. Johann Ambrosius Barth, Leipzig), [translated by Thure von Uexküll, 1982: The theory of meaning. *Semiotica* 42(1): 25-82].
- Wheeler, M. (1996) From Robots to Rothko: the bringing forth of world. In Boden, M. (Ed.) *The Philosophy of Artificial Life*, Oxford University Press, pp.209–236.
- Wheeler, M. (2005) *Reconstructing the cognitive world*. MIT Press.
- White, J.G., Southgate, E., Thomson, J.N. & Brenner, S. (1986) The structure of the nervous system of the nematode *Caenorhabditis elegans*, *Phil. Trans. Roy. Soc. London Ser. B* 314: 1–340.

Acknowledgements

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 the author has the support of the doctoral fellowship BFI03371-AE from the Basque Government. I would like to express my intellectual debt with the San Sebastian group of Philosophy of Biology (IAS-research). All the ideas presented here are, one way or another, a continuation of the work that has inspired me within that group for several years now. Thanks then to Arantxa, Jon, Kepa and specially to Alvaro Moreno for creating such an inspiring intellectual environment. Although the present formulation is mine, most of the ideas presented here have been developed in continuous discussion and joint intellectual work with Alvaro. I have also benefit from Kepa's careful revision of the section on cellular living organization. Finally I would like to thank Laura for her continuous support and love, as well as multiple revisions and discussion on early and final drafts of this work.