Horizons for the Enactive Mind: Values, Social Interaction, and Play

Ezequiel A. Di Paolo, Marieke Rohde, Hanneke De Jaegher

CSRP 587

April 2007

ISSN 1350-3162

University of Sussex

Cognitive Science
Research Papers

Horizons for the Enactive Mind: Values, Social Interaction, and Play

Ezequiel Di Paolo, Marieke Rohde and Hanneke De Jaegher

Centre for Computational Neuroscience and Robotics (CCNR)
Centre for Research in Cognitive Science (COGS)
University of Sussex, Brighton BN1 9QH, U.K.

{ezequiel,m.rohde,h.de.jaegher}@sussex.ac.uk

To appear in *Enaction: Towards a New Paradigm for Cognitive Science*, J. Stewart, O. Gapenne, and E. A. Di Paolo (Eds), Cambridge, MA: MIT Press, forthcoming.

Abstract.

What is the enactive approach to cognition? Over the last 15 years this banner has grown to become a respectable alternative to traditional frameworks in cognitive science. It is at the same time a label with different interpretations and upon which different doubts have been cast. This paper elaborates on the core ideas that define the enactive approach and their implications: *autonomy*, *sense-making*, *emergence*, *embodiment*, and *experience*. These are coherent, radical and very powerful concepts that establish clear methodological guidelines for research. The paper also looks at the problems that arise from taking these ideas seriously. The enactive approach has plenty of room for elaboration in many different areas and many challenges to respond to. In particular, we concentrate on the problems surrounding several theories of value-appraisal and value-generation. The enactive view takes the task of understanding meaning and value very seriously and elaborates a proper scientific alternative to reductionist attempts to tackle these issues by functional localization. Another area where the enactive framework can make a significant contribution is social interaction and social undertanding. The legacy of computationalism and methodological individualism is very strong in this field. Enactivism allows us to see embodied social interaction and coordination at many different levels in an integrated manner, from the emergence of autonomous temporal structures that regulate interaction to the generation of socially mediated meaning. Finally, we also present some speculations about how the enactive approach may be the right tool to help us bridge knowledge of concrete embodied and situated practices and higher-level human cognition, thus becoming a serious contender to computationalism in all areas of cognitive science (and not just on low-level sensorimotor cognition). The language offered by the enactive perspective already proves very useful in formulating the problem of human cognition in a tractable manner. These speculations will centre on the role of *play* as an activity that allows the development of meaning-manipulation skills as well as a further level of autonomous cognitive self, one that is characteristic of human beings. For the enactive view play is seen as re-*creation* whereas for computationalism fun is a mystery.

These discussions will be supported and illustrated with examples from work in evolutionary robotics. The need for synthetic minimal models and their scientific role is a running theme of this paper.

1. Introduction

Fifteen years after the publication of *The Embodied Mind* (Varela, Thompson, & Rosch, 1991) – a book that advanced a new framework for understanding cognition, one that emphasizes the role of embodied experience, the autonomy of the cognizer and its relation of co-determination with its world – the term *enactive* has moved out of relative obscurity and has become a fashionable banner in many parts of cognitive science. It has found its way into the description of diverse areas, from education and human-computer interaction, to autonomous robotics and consciousness studies. On the surface, this acceptance measures the success of the ideas articulated by Varela and his colleagues. Their achievement was not only that of synthesizing a series of existing criticisms to a predominant computationalist paradigm, but also that of advancing a set of postulates to move the field forward. Indeed, the increasing use of enactive terminology could serve as an indication that the time is ripe for a new era in cognitive science. To a great extent we believe this to be so.

However, on closer inspection, a significant variety of meanings is revealed in the use of the word enactive (as it happens with closely associated terms such as *autonomous, embodied, situated*, and *dynamical*). Sometimes the label indicates only the partial adoption of enactive views, sometimes connections are vague, and in the worst cases we see the raising of implausible hybrids risking self-contradiction in their mixture of the old and the new. There is a lack of consensus about what constitutes enactivism and embodied cognitive science in general (Wilson, 2002). Enactive has sometimes been taken simply as synonymous of active, embodied as synonymous of physical, dynamical as synonymous of changing, and situated as synonymous of exchanging information with the environment, all properties that could be claimed by practically every cognitive theory, model and robot proposed since symbolic Artificial Intelligence (AI) first made its debut as the theoretical core of cognitive science 50 years ago. This situation can lead to confusion and eventually to the loss of meaning attached to these terms – indeed, a perceived ambiguity between revolution and reform has been noticed by early commentators such as (Dennett, 1993).

There are two reasons for this situation, both indicating pressing problems that must be addressed if enactive cognitive science is to get off the ground. The first one is a watering down of the original ideas of enactivism by their partial adoption or sublimation into other frameworks. The second, related reason is a genuine lack of enactive proposals to advance open questions in cognitive science that motivate more traditional frameworks, such as the problems of higher-level cognition. These reasons lead to the misappropriation of the above keywords in the acceptance of the lessons of enactivism but only for a restricted range of influence. In the opinion of many, the usefulness of enactive ideas is confined to the 'lower levels' of human cognition. This is the 'reform-not-revolution' interpretation. For instance, embodied and situated engagement with the environment may well be sufficient to describe insect navigation, but it will not tell us how we can plan a trip from Brighton to La Rochelle. Or enactive theories could well account for complex skills such as mastering sensorimotor contingencies in visual perception (O'Regan & Noë, 2001), or becoming an expert car driver (Dreyfus, 2002) but, important though these skills are, they remain cognitively marginal (Clark & Toribio, 1994) and fall short of explaining performances such as preparing for a maths final or designing a house. For some researchers, then, the usefulness of enactive ideas is confined to the 'lower levels' of human cognition. As soon as anything more complex is needed, we must somehow recover newly clothed versions of representationalism and computationalism (Clark & Toribio, 1994; Clark, 1997; Clark & Grush, 1999; Grush, 2004).

We would do wrong in ignoring such positions. They show us precisely what is at the core of the struggle between traditional and unorthodox temperaments in cognitive science today. Indeed, they indicate the dangerous fate that fresh and radical ideas may suffer: that of dilution into a background essentially indistinguishable from that which they initially intend to reject. We believe that it is mistaken to conclude that what enactivism cannot yet account for must necessarily be explained using an updated version of old ideas. But it will remain tempting to do so *as long as the principal tenets and implications of enactivism remain insufficiently clear*. It would also be wrong to ignore the importance of arguments that show the limitations of enactivism. They challenge enactivism as a theoretical framework and reveal how much is left to be done. *Enactivism is a framework that must be coherently developed and extended*.

In trying to answer the question "What is enactivism?" it is important not to straightjacket concepts that may still be partly in development. Some gaps may not yet be satisfactorily closed; some contradictions may or may not be only apparent. We should resist the temptation to decree solutions to these problems simply because we are dealing with definitional matters. The usefulness of a research programme also lies with its capability to grow and improve itself. It can only do so if problems and contradictions are brought to the centre and we let them do their work. For this, it is important to be engendering rather than conclusive, to indicate horizons rather than boundaries.

There are still many important areas in enactive cognitive science that demand serious development. These remain the stronghold of traditional conceptions. Most of the underdeveloped areas within the enactive approach involve higher-levels of cognitive performance: thinking, imagining, engaging in complex interactions with others, and so on. For as long as enactive ideas are taken as filling in details or as playing a contextual role in the explanation of such phenomena the situation will not change.

We dedicate this paper to clarifying the central tenets of enactivism and exploring some of its currently under-developed themes. In this exercise, following the logic of the central ideas of enactivism can sometimes lead to unexpected hypotheses and implications. We must not underestimate the value of a new framework in allowing us to *formulate the questions in a different vocabulary*, even if satisfactory answers are not yet forthcoming. Implicitly, the exploration of these questions and possible answers is at the same time a demonstration of the variety of methods available to enactivism, from phenomenology, to theory/experiment cycles, and to the synthesis of minimal models and validation by construction – an additional thread that runs through this paper and that we will pick up again in the discussion.

In particular, we focus first on value generation and question the coherence of the idea of a *value-system* in cognitive architectures (both computationalist and embodied) and similar modular structures whose function is to generate or judge the *meaning* of a situation. Many influential theories in cognitive science make use of the idea that value or meaning is some information appraised by an internal module within an agent's cognitive architecture. Whereas in an enactive perspective, meaning is inseparable from the whole of context-dependent, life-motivated, embodied activity, without being at all a hazy concept beyond the reach of scientific understanding. We also explore, furthering on the issue of the origins of meaning, the field of social cognition. Despite being the focus of many recent phenomenologically inspired criticisms (Thompson, 2001; Gallagher, 2001, 2005), we think that an enactive perspective on social understanding remains to be clearly formulated. Our exploration leads us towards a middle way between individualistic and holistic views of social interaction and to highlighting the central role played by temporality of social engagement in generating and transforming social understanding at different timescales through joint participation. In the final part, we take a speculative look at the embodied capability to manipulate the meaning of concrete situations by exploring the role of play in the development of human cognition. These explorations do not attempt to be complete, nor do they put the whole of human cognition within the reach of enactivism and forever banish representational/computational explanations. But they do extend the conceptual horizon and allow us to formulate the problem of higher cognitive performance in an alternative, enactive way.

2. The core of enactivism

It would be misleading to think of the enactive approach as a set of all radically novel ideas. It is much rather a synthesis of some new but also some old themes that mutually support each other. Overall, enactivism may be construed as a kind of non-reductive naturalism[1]. It sees the properties of living and cognitive systems as forming part of a continuum and consequently advocates a scientific program that explores several phases along this dimension.

We can find scientific predecessors to enactivism in, for example, Piaget's theory of cognitive development through sensorimotor equilibration (Piaget, 1936, 1967), in Poincaré's theory of the active role of movement in the construction of spatial perception (Poincaré, 1907), in Goldstein's theory of the self-actualizing organism (Goldstein, 1934), and in others. The very term 'enactive' has been similarly used before, for example by Bruner in the 60s, to describe knowledge that is acquired and manifested through action (Bruner, 1966). Equally, we find philosophical affinities with existential phenomenology (Heidegger, 1962; Merleau-Ponty, 1962), with Eastern mindfulness traditions, with Hans Jonas's biophilosophy (Jonas, 1966), and even with pragmatic thinkers such as Dewey (1929). Current compatibilities can be also found with many embodied and dynamical systems ideas in contemporary cognitive science (Beer, 2000; Chiel & Beer, 1997; Thelen & Smith, 1994; Hutchins, 1995a; Juarrero, 1999; Kelso, 1995), neuroscience (Bach–y–Rita et al., 1969; Damasio, 1994; Skarda & Freeman, 1987; Engel et al., 2001), evolutionary biology (Lewontin, 1983; Oyama, 2000) and AI/robotics (Beer, 2003; Brooks, 1991; Harvey et al., 1997; Nolfi & Floreano, 2000; Winograd & Flores, 1986). Some of these connections have been made explicit in *The Embodied Mind*, others have been elaborated later in the literature, and still others remain to be better established.

What is the core of the enactive approach? Views that take cognition as embodied and situated, or take experience seriously, or explore the purchase of dynamical systems ideas, will all share something with enactivism. But to call them enactive just because there is some conceptual overlap may only contribute to a meaningless proliferation of the term. This is unless we can show both that i) such views share or are developed from a basic core of enactive ideas, and ii) extensions to these ideas do not result in a contradiction of this basic core. We can identify five highly intertwined ideas that constitute the basic enactive approach (Varela et al., 1991; Thompson, 2005): *autonomy, sense-making, emergence, embodiment,* and *experience*. These ideas partially imply each other. We will not attempt to disentangle all of these connections in order to obtain a set of perfectly independent postulates. Indeed, these internal relations speak for the strength of their association under a single banner.

2.1 Autonomy

Living organisms are autonomous – they follow laws set up by their own activity. Fundamentally, they can only be autonomous by virtue of their self-generated identity as distinct entities. A system whose identity is fully specified by a designer and cannot, by means of its own actions, regenerate its own constitution, can only follow the laws contained in its design, no matter how plastic, adaptive, or life-like its performance. In order for a system to generate its own laws it must be able to build itself *at some level of identity*. If a system 'has no say' in defining its own organization, then it is condemned to follow an externally given design like a laid down rail track. It may be endowed with ways of changing its behavior depending on history, but at some level it will encounter an externally imposed functional (as opposed to physical) limitation to the extent to which it can change. This can only be avoided if the system's limitations result partly from its own dynamics.

The autonomy (or freedom) of a self-constituted system is by no means unconstrained (being able to influence one's own limitations does not imply being able to fully remove them; on the contrary it means being able to set up new ways of constraining one's own actions). Hans Jonas (1966) speaks of life as sustaining a relation of *needful freedom* with respect to its environment. Matter and energy are needed to fuel metabolism. In turn, metabolism sustains its form (its identity) by dynamically disassociating itself from specific material configurations.

It should be clear that by expressions like 'self-constitution' and 'generating its own laws' no mysterious vitalism is intended. However, the acceptance of an operational concept of emergence (discussed below) is implied. By saying that a system is self-constituted, we mean that its dynamics generate and sustain an identity. An *identity* is generated whenever a precarious network of dynamical processes becomes operationally closed. A system is operationally closed if for any given process that forms part of the system we can always find among its enabling conditions other processes that make up the system. This means that at some level of description, the conditions that sustain any given process in such a network always include those conditions provided by the operation of the other processes in the network, and that the result of their global activity is an identifiable unity in the same domain or level of description, (it does not, of course, mean that the system is isolated from interactions with the environment) Autonomy as operational closure is intended to describe self-generated identities at many possible levels (Varela, 1979, 1997).

Cognitive systems are also autonomous in an interactive sense in terms of their engagement with their environment as agents and not simply as systems coupled to other systems (Moreno & Etxeberria, 2005; Di Paolo, 2005). As such, they not only respond to external perturbations in the traditional sense of producing the appropriate action for a given situation, but do in fact *actively regulate* the conditions of their exchange with the environment, and in doing so, they enact a world or cognitive domain.

Viewing cognitive systems as autonomous is to reject the traditional poles of seeing cognition as responding to an environmental stimulus on the one hand, and as satisfying internal demands on the other – both of which subordinate the agent to a role of obedience. It is also to recognize the 'ongoingness' of sensorimotor couplings that lead to patterns of perception and action twinned to the point that the distinction is often dissolved[2]. Autonomous agency goes even further than the recognition of ongoing sensorimotor couplings as dynamical and emphasizes the role of the agent in constructing, organizing, maintaining, and regulating those closed sensorimotor loops. In doing so, the cognizer plays a role in determining what are the laws that it will follow, what is the 'game' that is being played.

2.2 Sense-making

Already implied in the notion of interactive autonomy is the realization that organisms cast a web of significance on their world. Regulation of structural coupling with the environment implies that there is a direction that this process is aiming at: that of the continuity of the self-generated identity or identities that initiate the regulation. This establishes a *perspective on the world* with its own normativity, which is the counterpart of the agent being a centre of *activity in the world* (Varela, 1997; Weber & Varela, 2002; Di Paolo, 2005; Thompson, 2007). Exchanges with the world are inherently significant for the cognizer and this is the definitional property of a cognitive system: the creation and appreciation of meaning or *sense-making* in short.

It will be important to notice already, and this issue is treated more extensively in the following section, that the concept of sense-making in this view is an inherently active concept. Organisms do not passively receive information from their environments, which they then translate into internal representations. Natural cognitive systems are simply not in the business of accessing their world in order to build accurate pictures of it. They directly participate in the generation of meaning by their action; *they enact a world*. In this point enactivism differs from other non-representational and dynamical views such as Gibsonian ecological psychology (Varela et al., 1991, pp. 203-4). For the enactivist, sense is not an invariance present in the environment that must be retrieved by direct (or indirect) means. Invariances are instead the outcome of the dialogue between the active principle of organisms in action and the structure of the environment. The 'finding' of meaning must be enacted; it is always a *formative* activity, never about the extraction of information as if this was already present. This is another idea that sets the enactive framework apart from more traditional views in cognitive science: a dynamical, biologically grounded, theory of sense-making. Like few ideas in the past, this concept strikes at the heart of what is to be cognitive. We will elaborate this point in the next section and show how elusive this way of thinking can be even among researchers who have taken embodiment and situatedness very seriously.

2.3 Emergence

The overarching question in cognitive science is: How does it work? For the enactive approach the connected concepts of autonomy and sense-making already invoke some notion of emergence in addressing this question. Autonomy is not a property of a collection of components, but the consequence of a new identity that arises out of dynamical processes in operational closure. Meaning is not to be found in elements belonging to the environment or in the internal dynamics of the agent, but belongs to the relational domain established between the two.

The idea of emergence has been much debated in various domains from metaphysics to epistemology and it has had a furious revival over the last three decades with the advent of the sciences of complexity. Beyond the debates about the possibility of ontological emergence (Kim, 1999; Silberstein & McGeever, 1999), there is a pragmatic application of the term that stems from the well-understood phenomenon of self-organization, which has served to remove the air of mystery around emergence in order to bring it back in line with a naturalistic project. There is also a demand for emergentist explanations in biology where hierarchical organization is all too evident (e.g., genetic regulation, cells, extra-cellular matrix, tissues, organs, organism, dyads, groups, institutions, societies).

Emergence is used to describe the formation of a novel property or process out of the interaction of different existing processes or events (Thompson & Varela, 2001). But in order to distinguish an emergent process from simply an aggregate of dynamical elements, two things must hold: 1) the emergent process must have its own autonomous identity and 2) the sustaining of this identity and the interaction between the emergent process and its context must lead to constraints and modulation to the operation of the underlying levels[3]. The first property indicates the identifiability of the emergent process whose characteristics are enabled but not fully determined by the properties of the component processes. The second property refers to the mutual constraining between emerging and enabling levels (sometimes described as circular or downward causation).

We find the clearest example of emergence in life itself. The property of continuous self-production, renewal and regeneration of a physically bounded network of molecular transformations (autopoiesis) is not to be found at any level below that of the living cell itself. Being a self-sustaining bounded network of chemical transformations is not (it cannot be) the property or the responsibility of single components in this network. The new level is not only autonomous in terms of exhibiting its own identity and laws of transformation, it also introduces, through interaction with its co-defined context, modulations to the boundary conditions of the lower level processes that give rise to it.

This phenomenon is to be found at various different levels in multicellular organisms and in particular animals and humans. Variations on this theme have been used to describe the emergence of the self/non-self distinction in immune networks (Stewart & Coutinho, 2004), the generation, maintenance, and eventual dissolution of coherent modes of synchronous activity in the brain (Engel et al., 2001; Varela et al., 2001), and also between these coherent modes and action/perception cycles (Rodriguez et al., 2001; Le Van Quyen & Petitmengin, 2002). Emergent phenomena, as indicated in the last examples, can be fleeting. Single acts can bear a relation of emergence with respect to their sensorimotor component phases.

Taking emergence seriously makes the enactive approach very skeptical about the localization of function of one level in specific components at a lower level (homuncularity) and consequently it leads to the rejection of 'boxology' as a valid method to address the 'how does it work' question. Any labeling of sub-systemic components and variables with names belonging naturally to properties of emergent levels (e.g., value-systems, cognitive maps, emotional modules, mirror neurons) should be treated with extreme caution.

Having said all this, emergence remains problematic due often to its opaqueness and the ease with which the term can be misused. The weight of explaining how a given phenomenon constitutes a proper case of emergence remains with the supporters of this view. The very blurring of distinctions between levels that the enactive approach criticizes of cognitivism was what allowed the latter paradigm to connect personal and sub-personal levels. The properties of higher levels are thus explained in terms of lower ones because they are already somehow present there. For the emergentist, the connection and the interaction between levels becomes a problem to be addressed case by case often by recourse to complex concepts and tools derived from dynamical systems theory. It is clear that much work is still needed for clarifying and operationalizing the concept of emergence. Synthetic models can prove very valuable as tools for grasping emergent phenomena.

2.4 Embodiment

For enactivism, cognition is embodied action. Originally, this assertion is meant to emphasize bodily-mediated cognition as opposed to the performance of internal computations in order to find out about the world and *then* act on it. Embodied action is temporally and spatially embedded.

In a concrete and practical sense, a cognitive system is embodied to the extent to which its activity depends non-trivially on the body. However, the widespread use of the term has led in some cases to the loss of the original contrast with computationalism and even to the serious consideration of trivial senses of embodiment as mere physical presence – in this view a word-processor running on a computer would be embodied, (cf., Chrisley, 2003). It is easy to miss a fundamental motivation behind embodiment. It is not a question of moving the mind from a highly sheltered realm of computational modules in the head into wet and messy bodily structures. Such an idea remains Cartesian in its separation between the mind on the one hand and the body on the other. By contrast, embodiment means that mind is inherent in the active, worldful body, that the body is not a puppet controlled by the brain but a whole animate system with many autonomous layers of self-coordination and self-organization and various degrees of openness to the world that create its sense-making activity.

Indeed, to say that cognition is embodied is to express a tautology – it simply cannot *but* be embodied. But pointing to this has been (and still is) necessary in the computational/representational climate that gave rise to the embodied turn in cognitive science. Unfortunately, this means that as long as we must continue to emphasize mind as embodied, the main point of the criticism has not yet been understood (Sheets-Johnstone, 1999). For this reason, it is important to do much more than just *saying* that cognition is embodied. The debate must be moved to the concrete realm of seeing exactly how the animate body in its world *is* a mind. Any discussion of embodiment *in abstracto* will be highly impoverished.

Fortunately, concrete explorations on embodiment abound. The clearest are the simplest. Consider for instance the work of Charles Lenay and colleagues on perceptual augmentation (Lenay, 2003). In an experiment where the sensor channel is minimized to a single on/off tactile signal on the skin, blindfolded subjects must point a photoreceptor attached to their forefinger in order to locate the direction of a source of light. Every time the receptor is active the tactile signal is provided. A subject whose wrist is restricted in movement is only able to locate the general direction of the light source, but when the wrist is released, she can learn to perceive not just the direction but the distance as well. A freeing up of a degree of freedom results in a qualitatively different percept via the coordination of the wrist and other joints. Perception depends on the possibilities and organizational skills of the body and the kinesthetic experience that arises in coupling with the world. Similar exploitations of body movement and proprioception to generate stable perception are commonplace in autonomous robotics (Pfeifer & Scheier, 1999; Salomon, 1998). These minimalist examples should not be taken to imply that the relevance of the body is only restricted to ongoing coping with concrete sensorimotor activities. There is much evidence that higher-level cognitive skills, such as reasoning and problem solving, mental image manipulation, and semantic forms depend crucially on bodily structures (Wilson, 2002; Lakoff, 1987).

There is a further twist to the role played by the body in the case of human cognition – one that could explain the resilience of Cartesian modes of thinking. Even though our bodies are not puppets, to say that we control our bodies is, in a sense, not entirely wrong. We certainly do. But we do so in subtle ways that relate to the emergence of forms of reflexive autonomy, this time of a socio-linguistic nature. Like an alien presence, I set new aims for my body (I decide to embrace the pain of a yoga class, I decide to go on a diet). Being able to support and transform new identities is one way in which the body creates the experience of a self not quite the same as the metabolic self. This is an experience that nurtures Cartesianism. In fact, the body, by further manipulating its sense-making activity and transforming its value-making, is capable of putting itself in a novel situation which is partly its own creation. In doing so, it is playing a highly skillful dual role. This is afforded by the high flexibility and plasticity of the human body, but would not be possible without instauration within a symbolic order and the social mediation that makes our bodies fit to a scheme of control and observation of behavioral and cultural norms thus giving rise to socio-linguistic and narrative selves.

2.5 Experience

For enactivism, experience is central both methodologically and thematically. Far from being an epiphenomenon or a puzzle as it is for cognitivism, experience in the enactive approach is intertwined with being alive and enacting a world of significance. As part of the enactive method, experience goes beyond being data to be explained. It becomes a guiding force in a dialogue between phenomenology and science, resulting in an ongoing pragmatic circulation and mutual illumination between the two (Gallagher, 1997; van Gelder, 1999; Varela, 1996, 1999).

Many modern accounts of cognitive activity already take experience seriously. For instance, Dreyfus's defense of non-representational skill acquisition (Dreyfus, 2002) is based on paying careful attention to the experience of undergoing a process of task improvement. As we make the journey from beginners to experts through practice, not only is skilful performance improved but experience is also transformed. This is to be expected if embodiment is taken seriously. If experience and the body in action were to relate to each other as two mutually external systems, we would expect either an unchangeable or a fleeting relation between our bodies and our experience. Instead we find a lawful relation of bodily and experience transformations. Becoming a wine connoisseur is certainly an achievable goal but expertise in this field (as in any other) is not obtained through gaining the right kind of *information* but through the right kind of *transformation* – one that can only be brought about by appropriate time-extended training (experimenting, making mistakes). Experience is altered in a lawful manner through the process. It is itself a skillful aspect of embodied activity.

An embodied perspective results in serious attention being paid to isomorphisms between mechanisms and experience. Varela (1999) and van Gelder (1999) provide different, but related, dynamical systems accounts of mechanisms that might underlie the protentive and retentive structure of time consciousness as described by Husserl. Kelly (2000) considers neural models of pointing and grasping that run parallel to Merleau-Ponty's concepts of the intentional arc and maximal grip. Wheeler (2005) explores isomorphic relationships between embodied-embedded accounts of situated action and Heideggerian categories such as the ready-to-hand, breakdowns, and present-at-hand. What is interesting in many of these accounts is that the process of circulation is not one of assimilating scientific hypotheses into phenomenology, but may itself inform phenomenology. This is as it should be in a proper dialogue and such is the methodology advocated by first-person methods in the joint study of experience and brain-body activity (Varela, 1996; Lutz, 2002).

Experience may also serve the role of clarifying our commitments. Hans Jonas (1966) looks into the world of living beings and sees that life is a process with interiority. Metabolism has all the existential credentials of concernful being. It is precarious, it separates itself from non-being, it struggles to keep itself going and preserve its identity, and it relates to the world in value-laden terms. However, the inward aspect of life cannot be demonstrated using our current scientific tools. This does not make it any less factual for Jonas. He knows that all life is connected along an evolutionary continuum, and he knows that we ourselves are embodied living creatures with an inner life. This is how we can then know that living beings are forms of existence and that they also have an inner life.

This example is telling because it already contains a difficult to swallow aspect of using experience in a dialogue with science, which is, at the same time, perhaps its most revolutionary implication. Phenomenologically informed science goes beyond black marks on paper or experimental procedures for measuring data, and dives straight into the realm of personal experience. No amount of rational argument will convince a reader of Jonas's claim that, as an embodied organism, he is concerned with his own existence if the reader cannot see this for himself. Jonas appeals to the performance of a gesture that goes beyond comprehending a text. The implication is that in order to work as a source of knowledge, enactivism will contain an element of personal practice. It is necessary to come back to the phenomenology and check that our theories make sense, but this means we must become skillful in our phenomenology as well – personally so.

3. Values and the limits of evolutionary explanations

The previous section shows that there are certain ideas in cognitive science that the enactive approach clearly rejects, e.g., homuncularity, boxology, separability between action and perception, and representationalism. In this section we will revisit some of these themes in a more focused manner.

In everyday life we experience the world in value-loaded terms. This fact is hard to avoid and has been the subject of much philosophical debate throughout the ages. For enactivism, value is simply an aspect of all sense-making, since sense-making is, at its root, the evaluation of the consequences of interaction for the conservation of an identity. Perhaps as a reaction to the subjective overtones of this issue, traditional cognitive science has not dwelled much on the explicit mechanisms involved in value judgement as an inherent aspect of cognitive activity. In general, questions about value have or natural purposes have been dealt with separately, preferably with reference to evolutionary history (Millikan, 1984): everything living beings do is ultimately reduced to survival strategies in situations their ancestors encountered, or to the urge to spread their genes as widely as possible. In a more traditional modeling framework this idea translates to values being 'built-in' by evolution; phylogenetically invariant yardsticks against which actual lifetime encounters are measured and structured, and from which cognitive mechanisms that are themselves independent of these values deduce the meaning of situations, actions and perceptions.

Explanations of this kind are in tension with the principles of enactivism, in particular with the concept of sense-making. In this section, we juxtapose such traditional views, where ultimate ends come in evolutionarily sealed boxes, with an alternative, more enactive view that explains values and meanings as consequences of the kind of dynamical system a living organism is. We discuss an enactive theory of value in its rudimentary form, which is based on the theory of autopoiesis. A number of open questions, such as the explanation of non-metabolic values or transitions in value-generating mechanisms are raised and implications for computational models of cognition are discussed.

3.1 Values: built-in or constructed?

Weber and Varela (2002) have been the first to derive intrinsic teleology, natural purposes and the capacity of sense-making from autopoiesis, drawing on Kant's *Critique of Judgement* and Hans Jonas's philosophy of biology (Jonas, 1966), and the position argued for here commits to this general idea. In this kind of reasoning, the struggle for continuing autopoiesis – in other words survival – is at the core of intrinsic teleology and the capacity of sense-making. Even though survival plays a central role in both autopoietic and evolutionary explanations of value (one must first survive in order to reproduce), there are essential differences between the claim that, what affects an organism's autopoietic organization is of value, and the claim that values are built-in because they benefit survival and hence have been selected for.

If values are built-in, they need to have some form of priority over the living acting creature, either temporally or logically. Typically, claims about biological traits being built-in are about them being part of the genetic package. 'Values' is a term that describes the meaning of organismic behavior, not one of its physiological or mechanistic properties, like, for instance, the blood type. Therefore, the idea of built-in values relies on some kind of *a priori semantics*: parts of the genetic code are thought to execute according to pre-programmed rules and, thereby, generate values. This automated 'sense look-up' is not the same as sense-making, which we identified as one of the central concepts of the enactive approach. Similarly, we are dealing with *pre-factum* evolutionary teleonomy, not with autonomy. Instead of emergence, we find a direct reduction of evaluative function to physical structures. Instead of embodiment, we find abstract principles that are presumed independent of embodied interaction. Finally, lived experience is subdued as secondary to historical selection pressure – whether value is manifested experientially seems irrelevant. This is where the idea of built-in values and the enactive approach diverge.

This may sound like a very black-and-white picture. Maybe not all that living organisms do can be explained through built-in values, but there are surely some basic properties and behaviors that will always benefit survival, e.g., oxygen, food, water, and light will always be good for most animals, so what is wrong with claiming that there are some built-in basic values like 'water is good', 'light is good', 'this food is good'? The point is not to argue that such norms do not exist across individuals of a species, but rather that they should be searched for on the emergent level of autonomous interaction, not on the level of mechanism. If we imagine that a mechanistic structure inside a living organism was solely responsible for the generation of values, does that mean that the remainder of the organism is value-agnostic, that the values generated by this mechanism are arbitrary? Would that not imply that a mutation of the genetic code that tells the organism that 'food is good' could result in the generation of the value 'poison is good'? For the mutant system, poison would then be a positive value, just as food was for its ancestor, even if this mutation would eventually kill it, which seems a strange idea. The facts that food and water and light are good and that poison is bad are a result of the kind of system that an organism is and that they are of consequence for its conservation. In this sense, no mutation can create the value 'poison is good' without changing the organization of the system so that it thrives on 'poison'. The value for this organism would again be 'food is good', not 'poison is good'. The organism is an ontological centre that imbues interactions with the environment with significance they do not have in its absence; and this significance is not arbitrary. It is dynamically constructed, and that is the essence of the idea of sense-making.

The thrust behind the idea of pre-coded values is the assumption of a kind of isomorphism between what is genuinely good or bad for the organism and what the executed genetic value programs say is good. They are thought to predict the effect of lifetime encounters for metabolism, on the basis of phylogenetic experience. Therefore, they have to rely on phylogenetic constancies. It is cases where we can observe a change of relation between a value and an organism that demonstrate the ontological priority of biological autonomy. The most striking examples of such value changes, which can shatter the functionality of established relations, are illness, perceptual supplementation, and other perturbations to the body (distortion or impairment). Bach-y-Rita and colleagues (1969) have demonstrated the amazing human capacity to perceive visually, despite a loss of sight, by relaying pixeled images, recorded with a head-mounted camera, to arrays of tactile stimulators. What kind of pre-existent, built-in value mechanism could be made responsible for assigning the meaning of light patterns to tickling stimuli on the skin?

Or consider a patient who, during the course of a disease, is subjected to increasing dosages of a pharmaceutical agent, with the result that he not only survives dosages of the drug that would be fatal to the average human being, but also that his metabolism relies on the medicine in a way that deprivation would cause his death. The value of this substance for the metabolism is inverted as a consequence of the changes undergone by the organism. But the transformation is not arbitrary. On the contrary, the kind of system that the organism becomes will determine the drug’s altered value, and this determination cannot be attributed to a local module, evolutionarily dedicated to the task of assigning meaning, but to the system as a whole. If constancies break down, we observe that local mechanisms gradually undergo a change in how their function relates to meaning such that local processes are not anymore about the same thing they were about once they were selected for. We call this phenomenon *semantic drift*; it comes up again in section 3.3.

Even if it is true that specific internal structures play a fundamental role in the value-appraisal process, reducing the latter to the former seems a category mistake; it confounds the domains of mechanism and of behavior. To localize the correlated function in these structures is like saying the speed of a car is in the gas pedal.

3.2 Kinds of values

We propose to define value as *the extent to which a situation affects the viability of a self-sustaining and precarious process that generates an identity*. The most widely discussed and most intensely analyzed such process is autopoiesis, the continuous material regeneration of a self-bounded, self-constructing network of molecular transformations in a far-from-equilibrium situation. Encounters will be good or bad depending on their effect on autopoiesis. Up to now, our discussion has exclusively argued the case of this basic 'metabolic value', as it seems the least controversial. It now remains to be established what kinds of other processes might be self-sustaining, precarious and generate an identity, this is to say: what other processes might generate values.

Logically, there are two possibilities for value-generation by processes other than metabolism itself: value generation alongside autopoiesis and value generation independent of autopoiesis. Both scenarios immediately lead to further questions. If there are self-sustaining precarious processes that generate an identity, but are fully independent of living organisms, where does teleology come from? Can we really say that such processes generate value, and if yes, value for whom or for what? By contrast, if such processes 'parasite' on the process of living how do the values they generate relate to the basic metabolic value? What happens in case of a conflict? The enactive paradigm leaves space for a multitude of possible positions on these matters; these questions are far from settled and this section cannot but present a few existing positions and our own thoughts in progress.

Varela's own perspective on the organism as a 'meshwork of selfless selves' (Varela, 1991, 1997) can be seen as an exploration into value-generating mechanisms, mainly of the first kind, i.e., based on autopoiesis as the most basic form of autonomy and identity generation. Identity generation, for him, entails that an invariant quality is maintained coherently by an operationally closed process whose primary effect is its own sustained production. Varela studied three mechanisms to bring about such processes: autopoiesis (cellular identity), the immune system (multicellular identity) and the nervous system (neurocognitive identity). He acknowledges the existence or possibility of other levels of identity, reaching from pre-cellular identity (e.g., identity of self-replicating molecules) to socio-linguistic identity and super-organismic identity. Similar ideas are elaborated by Jonas (1966).

Figure 1 about here

We want to touch on some examples from a non-exhaustive listing of transitions in value-generating mechanisms (figure 1) that we consider particularly important or interesting, drawing on some of Varela's and Jonas's ideas. The first three stages of this scale are frequently not treated as distinct. However, it has recently been argued (Di Paolo, 2005) that mere autopoiesis, according to the original definition, even though it is sufficient to generate natural teleology and metabolic value, does not entail active appraisal of metabolic value: an autopoietic entity can be robust to perturbations without the logical necessity to actively monitor its own state and act to improve the conditions for continued autopoiesis. Only *adaptive* autopoietic entities that improve the conditions for continued autopoiesis, by actively monitoring their own state, identifying at least some tendencies that bring them closer to the boundary of viability and acting to counter these tendencies can be actual ‘sense-makers’, not just robust to perturbations. A similarly subtle distinction is the one between adaptive organisms and interactive regulators (Moreno & Etxeberria, 2005): whilst the former act to counter hostile tendencies by changing their internal organization, the latter act on the environment and thereby exhibit the most fundamental form of agency. An example of a just-adaptive organism is the sulphur bacterium that survives anaerobically in marine sediments whereas bacteria swimming up a sugar gradient would, by virtue of their motion, qualify for minimal agency.

The further stages on the scale are largely adopted from Jonas’s work. Animals, through their motility, exhibit the capacity to act and perceive as well as desire or fear something distal. And humans, through capacities such as image-making and ultimately of constructing a self-image, gain the ability to regard situations objectively and define themselves as subjects, to distinguish truth from falsehood, and to experience happiness and frustration (Jonas, 1966; Di Paolo, 2005). This ladder follows the ‘gradient of mediacy’. It connects increasing degrees of mediation between an urge and its satisfaction to higher degrees of precariousness, and to the consequent liberation of ways to generate values. For instance, only a sense-making organism is capable of deception by virtue of the mediacy of urge and satisfaction. A bacterium that swims up the ‘saccharine’ gradient, as it would in a sugar gradient, can be properly said to have assigned significance to a sign that is not immediately related to its metabolism, even though it is still bound to generate meanings solely based on the consequences for its metabolism. With increasing mediacy, the possibilities to create meaning for signs become less and less constrained by the instantaneous metabolic needs of the organism. Such hierarchies of processes bringing about different kinds of identities and values relate to the study of the major transitions in evolution like the evolution of the eukaryote cell, of sex or of multicellularity, as described by Maynard Smith and Szathmáry (1995). However, even though different organizations of living creatures enable new and more complex kinds of value-generating processes, transitions in structure cannot immediately be equated with transitions in value, the evolution of value-generating processes proceeds in a more gradual and continuous fashion. The exact relation between complication of material organization through processes of reproduction and selection and the evolution of values is largely unexplored territory that certainly deserves future attention.

One of the big unknown variables in this equation is how different kinds of values are tied together to form a unitary self. By calling the organism a ‘meshwork of selfless selves’ Varela (1991) avoids the answer to this question:

> “Organism as self, then, cannot be broached as a single process. We are forced to discover ‘regions’ that interweave in complex manners, and, in the

> case of humans, that extend beyond the strict confines of the body into the socio-linguistic register" (Varela, 1991, p. 102).

It is certainly true that levels of value generation can be in conflict: how can it be that your body will fight for its life despite the deliberate attempt to end autopoiesis through an overdose of sleeping pills? Or, the other way around, how can the body attack itself in an autoimmune disease, to the dismay of the layer of self that is able to express itself linguistically? Here, we disagree with Weber (2003), who seems to imply that value is always primordial and one-dimensional, i.e., that everything that is of value to an organism can be ultimately derived from metabolic value (he calls it 'existential value'). Such reductions may provide adequate description for forms of life that do not involve high degrees of mediacy, but not if several levels of value generation are in conflict. For a smoker, the mechanisms of addiction may be explained with reference to metabolism but it does not follow that smoking is in any way *about* survival in the way that breathing is.

Figure 2 about here

How do different, sometimes competing self-sustaining and precarious processes, spanning various levels of identity generation, sometimes exceeding the boundaries of the autopoietic individual, relate to the cognizing subject? Could there be genuine values without autopoiesis? These are the big mysteries that remain to be solved. But it seems clear that drawing a box labeled 'value' is an unsatisfactory answer to these complex questions.

3.3 Modeling values

In this section, we want to discuss how to model values following the enactive approach. We see a large potential to advance the enactive approach through the adequate use of synthetic models. However, it is very difficult to avoid remnants of Cartesian ways of thinking that are concealed in apparently innocent modeling assumptions. Partially replicating previous arguments by Rutkowska (1997), we want to uncover such 'lurking homunculi' in 'value system architectures', a class of architectures that feature a local mechanism to assign values.

The term 'value systems' is taken from the Theory of Neuronal Group Selection (TNGS), mainly forwarded by Edelman et al. (e.g., Edelman, 1989), who define value systems as neural modules that are "already specified during embryogenesis as the result of evolutionary selection upon the phenotype" (Sporns &Edelman, 1993, p. 968) and that internally generate reinforcement signals to direct future ontogenetic adaptation. For instance, a value system for reaching would become active if the hand comes close to the target. A functional and structural division between behavior-generating mechanisms and mechanisms of value-based adaptation is at the core of this type of architecture.

In order to point out the difficulties that result from such a separation of value judgement (built in) from value execution (ongoing), we now present two examples of our own research in computational modeling. The deliberately simple first set of simulation experiments is described in more detail in (Rohde & Di Paolo, 2006) and illustrates the difficulties of embedding functional modules into another wise dynamic and embodied system. A mobile, two-wheeled agent is controlled by a neural network, which is generated automatically, using an evolutionary algorithm, such that the agent's behavior optimizes a formal performance measure. This 'evolutionary robotics' technique mimics the principles of Darwinian natural selection in a simplified manner and is pleasant to the enactivist for several reasons. Since the performance criterion rates the *behavior* of an agent in a given environment, not its input-output mappings, it provides a natural account of the situatedness, embodiment and dynamics of behavior. Also, whilst the experimenter determines function by specifying the performance criterion, she underspecifies the mechanism that brings it about – it is left to be shaped by automated search. Thereby, prior assumptions about the relation between function and mechanism are minimized, which can lead to behavior emerging from mechanism in ways that the experimenter could not have come up with, be it due to implicit prior assumptions, or due to cognitive limitations in dealing with complex dynamical systems (Harvey et al., 2005). Despite these principal advantages, the usefulness of these models in any particular case will depend on many more factors and design decisions.

Figure 3 about here

Value system architectures are inspired by findings on neural assemblies whose activity corresponds to salient events in the agent-environment interaction that are interpreted as internally generated reinforcement signals. In order to explore just how such a 'value signal' could be generated, without caring yet about its function, an agent moving on a plane is evolved to perform light seeking behavior for a set of light sources presented sequentially and, at the same time, to generate a signal that corresponds to how well its approach to the light is being performed. Therefore, this value signal should go up only when the agent is progressing in its task.

The network controller evolved to control the two-wheeled simulated agent is extremely simple, but amazingly good at estimating how close the agent is to alight source, despite the poor sensory endowment (two light sensors generating on-off signals) and the consequent massive ambiguity in the sensory space. The encircled group of three neurons is the part of the structure that generates the value signal (figure 3(A)). When investigating what this 'value system' does, we find that it responds positively to activity on the left light sensor, but negatively to activity on the right light sensor, which, intuitively, does not make a lot of sense. The successful judgement can only be understood by taking the sensorimotor context into consideration, i.e., the agent's light seeking strategy (figure 3(B)). If the agent does not see the light, it turns to the right, until it senses the light with both sensors. It then approaches the light from the right, constantly bringing the light source in and out of range of the right sensor. In the end, the agent cycles around the light source in small circles, perceiving the light with the left sensor only. Knowing this, it is much easier to understand how the 'value system' achieves a correct estimation of the distance. The approach behavior only starts when the light is in range of the left light sensor, and this sensor remains activated from then on, which explains the positive response to left sensor activation. The right light sensor, however, is only activated during the approach trajectory, and for increasingly short intervals, but not once the light source has been reached, and therefore is negatively correlated to progress in performance.

This simple example demonstrates an important theoretical possibility: a value signal that correlates to behavioral success, even if it is generated by a neural structure that is disconnected from the motor system, can exploit and rely on an existing sensorimotor context. Why is this possibility important? Because it undermines the very idea of top–down behavioral adaptation on the basis of value system judgment: By identifying a correlation between activity in a separate cell assembly and behavioral success, we infer that this module is a value system (figure 2(A)) that informs the organism if a performed behavior is successful. But what if this module relies in itself, in a circular fashion, on behavioral invariants in order to perform its judgement?

In order to explore this question, in a second experiment, we allow the synaptic weights between sensors and motors (behavior generating sub-system) to change in order to maximize the output of the value system from the previous experiment. Such 'neural Darwinism' is proposed as the source of adaptation in TNGS. In fact figure 3 (C) illustrates how, in contrast, with the embodied value system described above, this type of modulation quickly results in a *deterioration* of performance. In a system that exploits sensorimotor couplings to generate a value signal, if these couplings are modified, their semantic contribution to the generation of meaningful judgment is gradually withdrawn, and we observe a semantic drift of the value signal: Activity in the value system causes a change in behavior, which in turn causes a change of 'meaning' of the activity of the value system, which causes a change in behavior, and so on. The system described above, in isolation, rewards activity of the left sensor and punishes activity of the right. So if the semantic contribution of the sensorimotor couplings is gradually modified, the agent ends up avoiding the light source in a large circle, because this is the behavior that optimizes value system output, but not phototaxis.

This deterioration of performance is hardly surprising, given the structure of the value system and the way it works. But it demonstrates that value system architectures as outlined are not guaranteed to work without taking on board further premises. It has to be ensured that a value system estimates performance independent of the presence of reciprocal causal links, feedback loops and semantic drift of local structures. If a value system is implemented in a rigid context, as it has been done in some robots with a limited behavioral domain (Verschure et al., 1995), the meaning of the signal can be preserved independent of the modulation of behavior, such that the proposed circuits of adaptation do indeed work. However, in order to be convincing as a biological theory, it is necessary to specify how such a rigid wiring and disembodiment of value systems is realized in a living organism that is in constant material flux. This is exactly the kind of problem that classical computationalist approaches have failed to answer satisfactorily. Indeed, we see value systems, because of their disembodied nature and top–down supervision of adaptation, as leftovers from a Cartesian mode of thinking. Such leftovers are not surprising; decades of exercising a computationalist methodology persist in the very language used to formulate questions.

An enactive approach, however, is based on the idea that functional invariants, such as values, self-organize and emerge from a constantly varying material substrate. They are not reduced to local physical structures, such as a value system, and therefore there are no problems of explaining the semantic rigidity of material subunits.

Figure 4 about here

We now discuss an evolutionary robotics experiment that we conceive of as a first step towards a model of sense-making (Di Paolo, 2000b). The task and agent are similar to the experiment described earlier, i.e., seeking a sequence of different light sources (see figure 4(A) for a sketch of the agent). The controller consists of a network of homeostatic units, i.e., neurons that regulate their connections to other neurons so that their own activity is maintained within a target range. This regulation is achieved by inducing local changes in the weighted connections, a design that is inspired by Ashby’s homeostat (Ashby, 1960). These networks were set-up to achieve both phototaxis and internal homeostasis by artificial evolution.

Every displacement of the light source (peaks in distance) is followed by a quick approaching behavior (figure 4(B)). The interesting fact about this agent is that it adapts against left-right swapping of its sensors (figure 4 (C)): even though initially, the agent moves away from the light source–as we would expect if the visual field is inverted – over time it changes its behavior back to approaching the light, i.e., the agent *reinterprets* its sensory channels according to the alterations of sensorimotor coupling it experiences, even though it had never been subjected to such alterations during evolution.

To what extent can these experiments be seen as more enactive than value system architectures? First we ask, why does adaptation to visual inversion occur at all? Internal homeostasis acts as a dynamical organization trying to conserve itself, a minimal case of a self-sustained identity. The changes thus introduced can be said to conserve the *autonomy* of the neural process. A conservation that, through evolution, has been intrinsically linked to behavioral performance, i.e., phototaxis. Hence light is of positive value for this agent. When the body is disrupted performance is disrupted as well which can only be 'interpreted' by the autonomous dynamics as a challenge to its conservation. The recovery of homeostasis results also in the reinterpretation of the sensorimotor coupling (and eventually in the regaining of phototaxis). However, the positive value of light demonstrated by the adaptive process cannot be reduced to the local plastic dynamics, it *emerges* through the ongoing internal and interactive dynamics of the agent in its environment. The meaning of light sensor activity and its functional role for phototaxis is dynamically constructed during the interaction. This minimal dynamic *sense-making* is very different from the *a priori* semantics of value systems, which have to be protected from semantic drift. We find stability of neural dynamics, even if the system is not explicitly designed to serve as adaptation mechanism for a particular class of predicted problems, and this emergent meaningful adaptation can be explained through the study of mechanism and the parallel study of the behavior it brings about.

This example also demonstrates the usefulness of simulation modeling as a method for the enactive framework by showing how problems of functional reduction can be avoided, and even some degree of dynamical autonomy can be achieved that brings about adaptation through emergent value-generation.

4. Enacting social meaning: rhythm and coordination

In this section, we explore what an enactive approach to social understanding would look like. Initial steps in the direction of such an approach have been taken, but they generally have a long way to go still. Some authors have suggested ways of conceptualizing social understanding that touch upon some enactive ideas as outlined in section 2, though most of these proposals are not fully developed yet. In Gallagher's (2001) proposal, for instance, the basis of social understanding lies in the abilities of primary intersubjectivity (Trevarthen, 1979). These include intentionality detection, the detection of eye-direction, imitation, the perception of emotion and meaning in postures and movements. Thompson (2001) has suggested that we understand each other as part of an ongoing 'self-other co-determination' that takes place when we are in interaction.

And Ami Klin and his colleagues (2003) have also produced a so-called enactive approach to the social domain; however, it remains chiefly focused on perception.

4.1 Towards enactive social understanding

Before laying out our proposal for an enactive approach to social understanding, let us have a look at the gaps in traditional takes on social cognition. The underlying assumption of central paradigms such as Theory of Mind theory (ToM) and simulation theory is that minds are enclosed and opaque, and hence others are puzzles for us to solve. The proposal of ToM accounts as regards social understanding is that we cognitively figure out others: we understand others by applying a capacity to draw logical inferences to sets of knowledge and perceptions. This often includes rules (knowledge) about how the social world works. Simulation theory was proposed in reaction to the 'cold reasoning' stance of ToM. In Gordon's radical simulation approach, one uses "one's own motivational and emotional resources and one's own capacity for practical reasoning" (Gordon, 1996, p. 15). This proposal is supposedly *hot* because it involves a subject's own resources. We find out about what another is thinking or doing through an internal simulation of their behavior. Simulation comes in roughly two guises. There is Gordon's radical simulationism in which we act out the other's stance, that is, we 'become the other' for a short while in order to understand her. On the less radical version of simulation we imagine ourselves in the shoes of the person we are trying to understand. All these approaches, the different versions of ToM and of simulation theory, presuppose a thorough disconnection between subjects. In a social situation, we are confronted with an impenetrable other and so we find ourselves again and again thrown upon our own resources of reasoning and/or imagination.

Apart from internal contradictions (De Jaegher, 2006a), this kind of approach is obviously not enactive. The body plays no essential role. Issues of autonomy, emergence and self-organization are not implicated. As regards sense-making: meaning is derived from good old fashioned information processing, on the basis of which we explain and predict other people's behavior. Experience could be said to come into simulation approaches, but we would have to wax very lyrical about it – too much so – for the kind of experience implicated here to be anything like what it is understood to mean in an enactive approach.

Alternatives to both mindreading and simulation approaches have been suggested, many of which have some root in phenomenology and/or in sociology. Gallagher's embodied approach, for instance, is predominantly based on phenomenology and empirical work in various aspects of cognitive science. Gallagher has criticized both the mindreading and simulation approaches because of their assumption that minds are private. Instead of this, he suggests that what we think, intend, desire, and so on is practiced, and as such expressed and recognized, in our body. That is, you express yourself in your bodily comportment, and this can be picked up by other persons, because of their own lived bodies. Gallagher's claim is that "in most intersubjective situations we have a direct understanding of another person's intentions because their intentions are explicitly expressed in their embodied actions, and mirrored in our own capabilities for action" (Gallagher, 2005, p. 224). Gallagher calls his approach the 'embodied practice of mind'. We have an embodied, lived and immediate experience of the other in social situations and this is so because we are naturally and in an embodied way coupled with others from infancy. This coupling is possible because of the inter-modal link between proprioception and visual perception that exists from birth and connects the body schema (a set of subpersonal sensorimotor processes that are dynamically involved in movements and posture maintenance) with the perception of others. According to Gallagher, "the conception of an innate, intermodal visual-proprioceptive/sensory-motor linkage suggests that the link to the other person is immediate; experientially, and not just objectively, we are born into a world of others" (2005, p. 82). Basically, we know others because of our own embodied experience, not because their bodies look like ours, but because we experience them through our own bodies. We are not confronted with an object to dismantle in encountering another, but with someone that we already relate to at a very basic, bodily level.

A drawback of Gallagher's proposal is that it presupposes coupling between persons. How people interact therefore does not become a topic for investigation[4]. Even though Gallagher makes the point that the third-person approach to understanding others is wrong, and suggests instead that second-person interaction is the primary way of understanding others, he does not thematize the interaction itself, does not put it up for investigation. Individuals in interaction are, according to him "*always already* 'coupled' " (Gallagher, 2005, p. 81, original emphasis).

If we are to take social understanding seriously and investigate it in the manner of enactivism, however, we need to pay special attention to the process of social interaction itself (De Jaegher, 2006b). We suggest that, in order to understand social cognition, the embodied aspects investigated by Gallagher and others need to be supplemented by an approach to social interaction, in analogy with the interaction between agent and world as described in section 2. In order to fully understand how meaning comes about in social understanding, we suggest, we need to not only focus on the embodiment of interactors, but also on the interaction process that takes place between them.

There have been suggestions along those lines. Hobson (2002), for example, discusses ‘interpersonal engagement’. This is the intersubjective sharing of experiences, which infants are already good at and which forms the fertile ground for the development of our capacity for thinking. There is also a large amount of research on dialogue and interaction, where – sometimes – the actual interaction as such is studied: in conversation analysis and context analysis (Kendon, 1990; Schiffrin, 1994). This work has generated interesting findings, but the research in these fields has often been geared towards notes on empirical findings, more than towards theoretical principles of communication or interaction. In order to get at the latter, we need to look more concretely at the mechanism of social interaction as such, which is also what Thompson seems to suggests when he puts forward the notion of self-other co-determination (Thompson, 2001).

4.2 Interaction and coordination

In order to combine the findings and ideas of the above researchers, and make the combination amenable to an approach that will eventually integrate the core ideas of enactivism, we introduce a framework for studying interaction and coordination. This framework, initially based on ideas of embodiment and self-organization, forms the fertile ground for incorporating other aspects of enactive cognition.

Interaction is here understood as the coupling between an agent and a specific aspect of its world: another agent. Interaction is the mutual interdependence (or bi-directional coupling) of the behaviors of two social agents. Precisely which behaviors of the agents are implicated in this process will depend on the specific interaction and the situation in which it takes place. What is of most interest right now, however, is precisely what kinds of interdependence can exist.

Systems can be *correlated*, i.e., we may find similarities or coherences of behavior over and above what would be expected from what is known about their normal functioning. Of all the correlated behaviors, some are *accidentally correlated*, and some are *non-accidentally correlated*. We are most interested here in the latter form of interdependence and we will call it coordination. In social situations, coordination thus refers to the non-accidental correlation of behaviors of two or more social agents. It is brought about by one or more common and/or connecting factors.

Imagine two people walking down the street. Suddenly, both of them turn their heads. Suppose we notice that their head-turning behavior has been prompted by someone screaming behind them. Their behaviors are thus *externally coordinated* because there is a common triggering factor. In the absence of such a factor, their behaviors might have been a fortuitous correlation or the result of *pre-coordination*. When two people turn their heads at the same time because they are both, say for some strange neurological reason, set to turn their heads every hour on the hour, the observed coherence is brought about by a pre-established coordination: their shared predisposition for hourly head-turning. Again, there is a common factor: an internal 'head-turning clock'. Common factors in pre-coordination can be of diverse origin, but often it is a similar mechanism or shared aspects of history in the individual systems.

Hardly any social encounter, even if it is characterized by some pre-coordination, can unfold on the basis of pre-coordination alone (conversely there is some pre-coordination present in almost all social encounters, even if only by a common cultural background). In the work of Gallagher, for example, behaviors of interacting individuals seem to be based on a pre-coordination that is itself grounded in their innate body schema. But only when interactors actually get together, and the interaction process unfolds, can social understanding take place. More on-the-spot coordination than mere pre-coordination is needed, and it will be argued here that most of it is *interactional coordination*. This refers to role of the interaction process itself in generating or facilitating coordination. Imagine two people trying to pass each other in a narrow corridor, whereby each repeatedly steps out of the way, but to the same side. Coordination here is achieved by the joint action of the individuals involved as interaction unfolds. This may even consist of one person not doing anything, standing still, and letting the other make the decisive move – a kind of solution that is negotiated contingently on the interaction process.

On the other hand, coordination can also make interaction more likely to happen and continue. An example of this is making an appointment in order to meet. Coordination thus can have an interactional function. This we call *functional coordination*. A beautiful example of this is the case of wolf circling (Moran et al., 1981). Sometimes, as a wolf walks past another one that is seated, the second one gets up and starts to move in the opposite direction. However, rather than pass each other and walk away, they start to move in a circle together, head to tail. This behavior makes it possible for the wolves to size each other up as it were, and to decide upon fighting or not, which can be said to be the function of this bodily coordination. Such coordination often serves an interactional function, namely that of facilitating or continuing the interaction, whatever it may lead to or change into. Interactional coordination and functional coordination are not easy to separate; they are two sides of the same coin and describe the reciprocal influence between coordinated behavior and interaction as a process. As an extreme case of coordination through interaction we find the phenomenon of *one-sided coordination*. This happens when an individual coordinates *to* rather than *with* another. This distinction is illustrated in the models described below.

In the following section, we will discuss some examples of investigations of these aspects of social interaction, two robotics models, one of which is based on an empirical study of 'perceptual crossing'. Following from this, we will make the link to meaning generation in social interaction via the introduction of the notions of *interaction rhythm* and *participatory sense-making*.

4.3 Modeling embodied coordination

One approach to the question of how coordination between social interactors may be established is illustrated by some evolutionary robotics work on social interaction. Already more than half a century ago, simple forms of social robot coordination were explored by W. Grey Walter and his tortoises (Walter, 1950). Such experiments demonstrated how a couple of very simple individual behaviors (such as wandering around and approaching a source of light) could result in complex, dance-like, coordination when two such robots were put in mutual interaction. Recent studies using evolutionary methods also demonstrate this. For instance, a simple model of simulated robots that must interact through an acoustic medium is presented in (Di Paolo, 2000a). This work shows how different kinds of coordination are a direct result of the embodied interaction between agents over time.

Figure 5 about here

The model is deliberately simple. Two mobile agents are placed in an unbounded two-dimensional arena. Their bodies are circular and can move by differential steering of two opposing wheels, which are controlled by a small continuous-time neural network. These agents are also provided with a loudspeaker that they use to regulate continuously the volume of sound they emit. They also have two microphones located symmetrically in their bodies, which are used to pick up any sound in the environment, including the sound they themselves produce. There are no other kinds of sensors – agents can only interact through the sound produced by their loudspeakers. So, there is an inherent problem of distinguishing a signal produced by an external source and by the agent itself since all sound signals are added up. A sound signal that travels through the body of an agent decays in intensity so that there is a significant difference to the sensor activity if the sound impinges directly on it or must go through the listener's body first; this self-shadowing property is indeed used by many mammals to detect sound source location.

With this setup, the task set for the agents is to locate and remain close to each other. There are no other restrictions to the agent's activity: they are allowed to evolve any kind of continuous sound signal or move in any way. The problem is nontrivial because of the lack of other sensors and the single sound channel. Shouting at the top of their voices will not work because the self-produced signal will overwhelm the sensors, but remaining quiet will not give any clue as to the agent's position that can be used to achieve the task. Consequently, sound must be used strategically. Because of their random initial positions coordination between the agents must be achieved in order to facilitate a continuing interaction.

Successful agent pairs acquire a coordinated pattern of signaling in which individuals take turns in emitting sound so that each may hear the other's production. They solve the 'self/non-self ' distinction problem by making use of the self-shadowing property. If an agent constantly rotates, an external source of sound produces a regularly rhythmic pattern in the agent's sensors, while the sensing of its own signal is unaffected. A simple embodied strategy simplifies what would otherwise be a complex pattern recognition problem. This regular pattern affects their own sound production so that they also signal rhythmically, and finally through a process of mutual modulation the production of sound is coordinated in an anti-phase entrainment of signals. Further coordination is observed during interaction in proximity when patterns of regularly alternate movement are produced that resemble a dance (figure 5). Both the sound and movement coordination patterns are achieved through a process of co-adaptation – tests on individual agents show that they are not capable of producing any of these behaviors in the presence of a non-contingent recording of a partner from a previous interaction, i.e., they are capable of interactional coordination but not of one-sided coordination.

This and similar models demonstrate that the achievement of coordination *through* the interaction process is indeed something that we can expect to happen (as opposed to something that demands purpose-built mechanisms) in a broad range of dynamical systems in interaction. The agents in this example use their bodies and the time-structure of their own movement to generate coordination. The generated patterns themselves help maintain the continuous coordination and periods of breakdown followed by recovery of coordination are observed.

Of course, social coordination even in simple systems may take more complex forms apart from entrainment. Other models have explored how coordination can be used to generate a division of labor and breaking of the symmetry of an interactive situation, (Quinn, 2001). Quinn shows that if two identical agents are required to move together in a straight line, they must first 'decide' who is going to lead and who is going to follow (a situation analogous to the narrow corridor scene described before where no pre-coordination exists). The symmetry is broken using stereotypical movements that result in the orientation of one agent by the other in an offer to become the leader – a proper minimal act of communication.

Experiments like these are sometimes disregarded because they seem so simple and 'low level' that it seems hard to see how they relate to human cognition. An alternative challenge for enactivist synthetic modeling is to try and account for empirical research conducted on human subjects that is driven by a similar aspiration for minimalism; for instance, the perceptual supplementation study by Lenay (2003) cited earlier.

Auvray, Lenay and Stewart (2006) have investigated the social phenomenon of 'perceptual crossing' in a similarly minimalistic manner. Blindfolded human subjects interacting in a shared minimal virtual environment are asked to recognise the presence of each other. The only possibility to act is to move the cursor left and right along a virtual 'tape' that wraps around. Subjects sense the presence of an object or the other player only through a touch sensor whenever their own cursor 'steps' on them. To make the task non-trivial, there is also a static object of the same size as the other subject on the tape (fixed lure), as well as a mobile object that shadows the motion of the other subject at a constant distance (attached lure). The problem is therefore not only distinguishing moving from non-moving entities along the tape using the touch feedback, but distinguishing between two entities that move exactly the same, only one of which represents the 'sensing' position of the other subject. The momentary sensory patterns therefore do not suffice to distinguish the three entities that may be encountered. Even so, recognition still results from the mutual search for each other.

Successful recognition relies on sensorimotor coordination, rather than on an individual's capacity to express a confident judgement on whether a stimulus is contingently caused by the partner or not. When subjects encounter a stimulus they tend to oscillate around it and these scanning movements only remain stable in the case that both players are in contact with each other. A subject could be fooled by the other player's attached lure, but only to the point that the other player is oscillating on the spot (one-way interaction). This situation is unstable, as the other player will eventually move away to continue the search. Only when the two-way interaction condition is established does the situation remain globally stable. Hence the solution is interactional because it is established by both partners searching for each other, but does not rely on individuals performing the right kind of perceptual recognition between responsive and non-responsive objects. The distinction in this case is made by the other partner moving away in the situation of a one-way interaction.

Figure 6 about here

We have applied the technique of evolutionary robotics to gain further insight into this task (Di Paolo et al., 2007). The virtual environment and task are the same and the agents are controlled by a neural network. The resulting global strategy is similar[5] but it raises an interesting further possibility regarding the role of interactional coordination. The empirical study shows that human subjects do not confuse a static lure with another subject. At first sight, it seems obvious that telling a mobile stimulus from a static one is the easiest task to solve in this experiment. Humans could, for instance, rapidly learn to discount changes to stimuli generated by their own movement using proprioception. The agent evolved in our model has another solution to the problem. If we take a closer look, we find a striking similarity between sensorimotor patterns for perceptual crossing involving the other player and for scanning a fixed lure (figure 6(A) and (B)). Encountering any stimulus makes the agent revert its direction of movement, which leads to another encounter followed by another inversion of velocity, and so forth. When we inspect the duration of the stimulus upon crossing a fixed object, we realize that it lasts longer than when crossing a moving partner. This is because the fixed object does not move itself. Therefore the *perceived size* differs for the two cases: longer in the case of a fixed object and shorter in the case of a moving object. The agent seems simply to rely on integrating sensory stimulation over time to make the distinction. This can be confirmed from the fact that the agent is quite easily tricked into making the wrong decision if the size of the fixed lure is varied.

What is interesting is that the smaller perceived size in the case of perceptual crossing depends on encounters remaining in an anti-phase pattern (figure 6 (A)). In other words, it depends on interactional coordination. Hence a systematic distinction in individually perceived size (between objects having the same objective size) is *co-constructed* during coordinated interaction, and in turn, individuals respond to the apparently smaller object by remaining in coordinated interaction. Here we see the importance of simple models as generators of ideas. Even though the general solution to the task is similar to the human scenario, on this last point the agents behave differently with respect to fixed objects. But their strategy highlights how interactional coordination can be exploited.

These examples demonstrate the potential of an enactive modeling approach for the study of social interaction: instead of limiting the view to what happens inside one individual, the interaction process is taken seriously, and, thereby, these models have the possibility to capture the rich dynamics of reciprocity that are left outside of traditional individualistic approaches. The models demonstrate the importance of *timing* in interaction and suggest how it can affect sensorimotor processes at the individual level to the point that recognizing an interaction partner is possible thanks to the interplay and mutual modulation between the interaction and individual cognitive properties.

4.4 Rhythm and participatory sense-making

How do we get from here to meaning generation in social encounters between humans? How do interactors understand each other? We believe that meaning generation and transformation take place in the processes of interaction and coordination, and are more in particular dependent on their timing, as has also been suggested by the experiments discussed. Interactional coordination and functional coordination can be seen as the processes by which social encounters self-organize. In social situations in the human world, meaning is generated ongoingly in the interaction out of this self-organization, in combination with the histories, backgrounds, expectations, thoughts and moods of the interactors.

How? It may be that enacting the social world happens in the precise timing of the functional and interactional coordination processes taking place in social situations. We call this timing *interaction rhythm*. Interaction rhythm refers to the diverse aspects of the temporality of the interaction – a necessary, though not sufficient, aspect of establishing, maintaining and closing social interactions. Timing coordination in interaction is done at many different levels of movement, including utterances (which can be conceived of as a kind of movement, see e.g. Gallagher's exposition on 'expressive movement' (Gallagher, 2005)), posture maintenance and so on. Rhythm as a term is preferred over the more general 'temporality' because it captures the *active* role that these elements play in the generation and organization of social interactions. As used here, the term 'rhythm' does not refer to a continual strict temporal regularity (which is one of the more generally received connotations of the word), but rather to the possible and actual temporal variability of timing in interaction, including, at times and at certain levels of behavior, regular timing.

Interaction rhythm refers to the self-organization in time of several elements and processes that *span* the individuals, i.e. the temporal organization of elements across and *between* individuals. This process can take on a strong role and momentum of its own, i.e., it can itself become an autonomous phenomenon. If the interaction process is like this, then it can alter and have an effect on the behaviors of the individuals involved in the interaction, e.g., the perception of object size in the perceptual crossing experiment. In human interactions, the individuals involved are autonomous themselves, and this makes for the complexity of social interaction. If we are to understand meaning generation or sense-making in social interaction, we need to grasp what goes on in this interplay between the different states of the interaction process itself and those of the individuals engaged in it. Using the notion of interaction rhythm in the endeavor of grasping how social cognition works enables us to conceptualize social understanding as something that takes place, is enacted, in the interaction. Our aim in this domain too, is to gain an insight into an aspect of cognition, not by positing a specialized module in the brain, but by the actual dynamics of interaction between the cognizer and the environment.

A factor we think is responsible for our social aptitude is what we call a 'rhythmic capacity', which is not a capacity strictly of an individual, but one that comes about in interaction and is changed by both the interactional process and the individuals involved. We define this central capacity of social cognition as: flexibly temporally coordinating through the interaction with another person. Through such flexible coordination, the rhythm of an interaction can be adapted to changing circumstances, changes of goals, moods, etc. This capacity is crucially dependent both on the individual interactors and at the same time on the process of engagement that ensues between them in every interaction. A corollary of this is that the rhythmic capacity of a person is never complete outside of an ongoing interaction. The rhythmic capacity may well play a role in explaining why we interact very differently with some people than with others. In some encounters we may feel shy, while in others we can bring out our flamboyant side, even without an obvious cause such as a big difference in age or social status. Of course, how we interact with which particular person is not set in stone. More factors than who we may be interacting with are at play, such as the situation in which the interaction occurs, our history of encounters with that person, our respective backgrounds and how much of them we share, our mood and so on. We cannot say who is in charge of the process of the interaction, it is not either one of the interactors alone, nor can a choice be made between whether it is the interactors or the interaction process that is responsible for any social understanding that takes place. Here again, therefore, interactions as wholes are important to study, plus their histories. Social meaning generation relies on the mutual attunement of individual sense-making, achieved in the interaction rhythm.

To conclude, we propose the notion of *participatory sense-making* for social understanding in an enactive framework. Participatory sense-making is the extension of the enactive notion of sense-making into the realm of social cognition.

The question how we understand each other can be answered in an enactive framework as follows. In sense-making, active coupling with the world brings forth a realm of significance. In a social situation, the active coupling is with another social agent. Social agents can be engaged in individual sense-making, but when they start interacting, their sense-making is modified in accordance with the specific aspect of the world they are now interacting with – another social agent – in accordance with the specifications laid out above. Social meaning generation relies on the mutual attunement of individual sense-making, achieved in the interaction rhythm and by the rhythmic capacity. Not only this, it also opens up domains of sense-making that are not open to the individual alone. Participatory sense-making constitutes a continuum from highly participatory to less participatory sense-making. At the latter end of the spectrum, we find for instance *orientation*, in which individual A orients B to aspects of B's cognitive domain. This is not very participatory, because there is not much mutuality to the sense-making. As we move away from this end of the range, the sense-making activities of the individuals involved are increasingly mutually changed by their coordinated sense-making, and also change it. At the extremely participatory end of the spectrum, individuals truly connect their sense-making activities, with consequences for each in the process, in the form of the interactional generation of new meanings and the transformation of existing meanings. Academic collaborations are a good example of this. Sometimes, when the partnership is especially fruitful, a completely new vantage point on a problem arises, or a fresh interpretation of a result, which were not there before. Sometimes it is quite impossible to attribute this development to one of the participants only.

5. Play: enactive re-creation

We come back to some of the problems raised in the introduction. This section will draw on what we have learned so far about the horizons of enactivism to approach the general question of human cognition (the umbrella term under which cognitive scientists gather conceptual thinking, planning, language, social competences, etc.).

We have already mentioned that the impact of the enactive approach in cognitive science, and that of embodied and dynamical views in general, has been acknowledged by many sectors, but not yet as a proper replacement for representationalism in what concerns higher level cognition. Some arguments have been advanced regarding the very possibility of a non-representational framework for this task. Clark and Toribio (1994) question how the very situatedness of action-oriented and richly dynamical couplings between agent and environment is not at the same time responsible for 'tying down' cognition to the present situation. Internal representations, the argument concludes, will have to re-enter the picture to account for activities that seem decoupled from the current situation, such as picturing the house of your childhood.

The argument is right in that indeed, from an enactive perspective, such high level skills are still unexplained. But the argument simply assumes that they are also unexplainable in enactive terms. Importantly, the argument relies on a misunderstanding of what situatedness is about. To say that we are present in a situation with our bodies does not mean that the situation boils down to the physical couplings that we encounter, i.e., that we are shackled to our present circumstances. This is why the concept of sense-making is so interesting. It is all too easy to interpret this idea in a one-sided manner – events in the world are given meaning by the agent – and ignore the crucial possibility that the cognitive agent may also be an *active creator of meaning* and that such creation can be subject to change and eventual control by emergent levels of cognitive identity.

Could this point be a way of making progress in an enactive account of human cognition? Let us try to formulate the essence of the problem first. What is essential to human cognition as opposed to other forms of animal cognition? Margaret Donaldson (1992) formulates the issue in a very useful way. She puzzles about the amazing human capability of constantly inventing new goals so that we invest them with value and submit passionately to them (sports, hobbies, record-breaking). An explanation of human intelligence should perhaps not concentrate so much on issues such as, say, how do we manage to do maths. It should bring to the centre the question of *why* we do those things at all? When did they become valuable for us?

Donaldson describes different ways to be a human mind. As a developmental psychologist, she concentrates on how transitions between these different modes occur throughout a lifetime. The question parallels how Jonas and others have treated the history of mind as transitions in scales of mediacy. Donaldson distinguishes four modes in which we function as minds depending on the focus of concern. This is amenable to the whole of our previous discussion. To have different foci of concern is no more or less than to have different modes of value-generation. The *point mode* deals with here-and-now coping (most animal activity, skilful practices in humans). The *line mode* expands the focus of concern to the immediate past and the possible future as well as to other spatial localities (understanding of immediate causes and consequences of events). The *construct mode* produces a de-centering of cognitive activity; concern focuses on events that have happened or may happen at some point in time or somewhere, and not necessarily involving the cognizer (induction, generalization). Finally, the *transcendent mode* has no locus; it deals with nowhere, no-time (abstract thought, metaphysics).

These modes are manifested to different degrees in different circumstances and with respect to different mental 'components' such as perception, action, emotion, and thought[6]. The modes are transversed developmentally, building upon previous stages. The generation of different kinds of intention and the manipulation of our own consciousness are the central factors in this development. This backdrop can help us describe our problem as that of formulating an enactive account of how to move beyond the point mode and into the line and construct modes.

This transition indicates the development of a capacity to 'unstick' meanings from a given situation and 'stick' novel ones onto it, or generally the capacity to influence meaning generation. This has confusingly been described as offline intelligence (Clark, 1997; Wheeler, 2005), whereas 'de-centering' or 'meaning manipulation' may be better labels. Such a capability is indeed a challenge for dynamical accounts of cognition that emphasize coupling with the environment. It would seem that cognitive activity is 'glued' to the here-and-now in such accounts, i.e., always in the point mode. By contrast, cognitivism sees no challenge in this. Manipulation of representations to deal with the here-and-now is not fundamentally different from manipulation of representations to deal with the there-and-then, or with no-where, no-time. This is hardly surprising. Cognitivism starts at the high end of the spectrum. It is based on non-temporal, non-spatial, un-situated mechanisms.

If we look historically or developmentally for an activity that could play a part in this transition we must conclude that a) it should be an embodied activity, accountable for by means of the many skills that we can already explain in enactive terms, and b) it should allow for ambiguity of meaning as well as the generation of novel kinds of value. The worst possible candidates are concrete goal-directed activities where meaning is well defined by situational constraints. The best candidates are those goal-generating activities where meaning is fluid. Jonas points to image-making which is indeed an excellent example. But it is already too sophisticated and immediately invites representational thinking. More parsimonious possibilities include dance, music, ritual, and play. Here we briefly explore the latter.

Can we sketch an enactive account of play? There is a significant literature on play in animals as well as different forms of play in human children and how play relates to socialization, self-regulation, attachment, use of language, and the development of cognitive capabilities[7]. The interesting fact for our present discussion is that elements of the meaning manipulation that this activity can afford are already present in all forms of play. We have already mentioned the possibility of sense-making leading to increasingly removed manipulation of meaning. Might not the presumed bacterium swimming up a saccharine (not sugar) gradient and the young baboon accepting to be chased around by the smaller playful infant share something in common? Are not both deceived to different degrees in their sense-making activities, the one unknowingly, the other willingly?

The first thing to note about play is that it is hard to define and easy to recognize. Miller (1973) lists some properties of play such as the repetition of motor patterns, lack of economy, exaggeration, lack of a direct practical end, production of novel sequences of behavior, combinatorial flexibility, egalitarianism, etc. Play occurs only in the absence of more urgent motivations related to survival; hence it is the privilege of species where individuals have enough spare energy, time and protection. Not all animals do it, and in those species that play, mainly infants and juveniles do it – exceptions are humans and species that are given safety through their adulthood such as cats, dogs, and domesticated monkeys and apes. Evolutionary explanations of play abound. They typically refer to beneficial by-products such as training of motor skills. The merits of such explanations must be assessed in each individual case, but in general terms understanding even quite 'unsophisticated' bodily play (rough and tumble, simulated pursuit-evasion, etc.) cannot be fully achieved without an experiential approach. Much is missed if we cannot understand why animals are *interested* in play. Maxine Sheets-Johnstone (2003) answers this question by indicating the dimension of kinesthetic feeling that animals explore in play: the dimension of corporeal powers, the *I-can* and *I-can-not*.

The experiential dimension of value explored in this way is opened up by the element of social interaction and the forms of participatory sense-making that it affords. It is here that kinaesthetic pleasure turns into make-believe. Running may be fun, running from or after someone even more. The excitement of aggressive or sexual encounters can be safely explored if distinguished from real ones by appropriate signals and conventions. It is this novel way of socially exploring the meaning of fake situations using real and concrete interactions that is taken to its pinnacle by humans in the form of pretend play. Here we are already at the other side of the transition because if the arrangement of wooden cubes can be a house and the pen a spaceship, the root capacity of meaning creation and manipulation is already going strong.

Cognitivist accounts of pretence in play, such as (Leslie, 1987), go very much in line with similar accounts of social understanding already examined, and their criticisms, e.g., (Hobson, 1990) complete the parallel. Piaget's views on pretend play are closer to the enactive approach (Piaget, 1951). For him the beginnings of play are rooted in the assimilative function whereby new situations are coped with using existing sensorimotor schemas. A 15-month old infant deals with a pillow using certain actions (touching, laying her head on it, going to sleep). As soon as another object (a blanket) is assimilated into the same structures, it becomes a make-believe pillow. The infant finds pleasure in the assimilative function and smiles. Donaldson (1992) criticizes this view (see also Sutton-Smith, 1966). If only assimilation were taking place, the blanket and the pillow would be indistinguishable. There would be no reason to smile unless there was a simultaneous awareness of the difference between the two cases and the sense of 'getting away with something'. Make-believe relies crucially on the *combined similarity and difference* between two situations, one concrete, tied to physical events, the other in terms of manipulated meaning, (the tension of this combination reappears in other creative activities such as making images).

The view of play as predominant assimilation misses out the active element of construction of new environmentally- and bodily-mediated meaning. Play breaks from the constraints of self-equilibrating cognition. It does not have the structure of a cognitive confrontation with an environment that places demands on the agent. Play is precisely *not* a problem requiring a solution. In fact, play is the breaking of this pattern; or rather it's re-deployment into an active construction of meaningful action where no such sense-making is directly demanded from the environment or from definite internal needs. The urge to play (at least during the creative phases of play) is indeed present but remains undefined until the activity of play itself helps the child make this urge clearer.

How is this possible within an accommodation/assimilation/equilibration dynamics? It seems not possible if we resist the active participation of the child in transforming her world. Vygotsky (1966) gives us a glimpse of how such manipulation of sense-making could happen. In play, the child begins to detach meaning from a situation and to regulate such meanings first with respect to objects and later to her own actions. This is motivated by the inability to satisfy immediate needs. Play becomes a way of substitution for real satisfaction and a way of dealing with an insurmountable mediacy. Soon the value-generating properties of play become evident and the activity is done for its own sake. 'Detachment' is an embodied activity. It begins by relying on concrete similarities – a doll resembles a person – but soon these similarities are mostly given by the child's own use of gestural schemas and not the objects themselves (Watson & Jackowitz, 1984). If something is treated as a horse, if it is made to move and sound like a horse, then the child accepts it as a horse (without forgetting it is not one). This is the ambiguity that, according to Donaldson, can produce laughter, the bringing into presence of what is not there, a cheating of 'reality'.

Once objects in the environment are imbued with meaning by actions that in turn demand from the child an (adaptive) interpretation, these objects become toys, would-be cars, houses and creatures. The child is now acting at the pinnacle of her capabilities because she is bringing forth an alienated meaning through her gestural schemas and then – and here is the equally radical trick – submitting to the reality thus created through adaptive equilibration (the absence of which would make play unchallenging and 'un-real').

The combination of a concrete embodied situation with alienated meaning is the freedom-engendering paradox of play. But it would not be a paradox if all there was to pretence was the manipulation of internal representations. This would result in no sense of ambiguity. Cognitivism cannot explain fun. When the child becomes the regulator of play, the activity takes off as a proper form of life. The child explores the new freedom by following pleasurable activities, but at the same time she learns to generate new rules, new constraints that structure and re-evaluate reality and that must be followed strictly (otherwise play becomes random and boring). The child is unhappy if she cannot bounce the ball more than the nine times she has managed so far. The norm is arbitrary, invented by the child, but in allowing her body to submit to it, it becomes as serious as any biological norm. The player is the lawgiver and the rule-follower, the question-maker and the responder. Play is thereby autonomous in the strict sense advocated by enactivism.

Pretending is only possible if a novel way of generating norms and values co-arises with exploratory play. The best players are those that create new rules in a contextual manner so that they can continue to play and fun does not run down by exhausting the possibilities of the game. Rules are made-up in play; they are solidified versions of norms. Fun is the exploration of the limits thus imposed on bodily activity and social interaction. But when the possibilities are extinguished the game becomes boring. Fun is also the change and revision of norms that reopen play. Over time, play is a self-structuring process governed by the dialectics of expansion and contraction of possibilities. Its freedom lies in the capability that players acquire of creating new meaningful (not arbitrary) constraints. The playful body is a new form of autonomous being, a novel mode of the cognitive self. It can now steer its sense-making activity and set new laws for itself and others to follow. This might help to answer the question we raised at the end of our discussion about embodiment in section 2.4.

We find that play is an area particularly rich for the exploration of enactive themes from emergence of identities and levels of social coordination, to manipulation of sense-making through experientially-guided bodily action. Perhaps no other framework is better placed to explain play and its paradoxes and this may be why there is such a paucity of references to play in cognitive science. When a child skillfully supplements the perceptual lack of similarity between a spoon and a car by making the spoon move and sound like a car he has grasped in an embodied manner the extent to which perception can be action-mediated. With his body he can now alter his sense-making activity, both on external objects, as well as his own actions and those of others. He has become a practitioner of enactive *re-creation*.

6. Conclusion

A proper extension to the enactive approach into a solid and mainstream framework for understanding cognition in all its manifestations will be a job of many and lasting for many years. This paper has only attempted to point to specific directions and show that enactivism can be made into a coherent set of ideas, distinguishable from other alternatives and that it can provide the language to formulate problems and the tools to advance on issues that are sometimes out of the focal range of traditional perspectives. The strength of any scientific proposal will eventually be in how it advances our understanding, be that in the form of predictability and control, or in the form of synthetic constructions, models, and technologies for coping and interacting with complex systems such as education policies, methods for diagnosis, novel therapies, etc. For this, it is crucial for ideas to be intelligible and promising.

In this respect, we would like to draw attention to the valuable role played by minimal models and experiments. Their function goes beyond the study of a given phenomenon. Minimal modeling provides crucial conceptual training that would be hard to obtain otherwise (Beer, 2003; Harvey et al., 2005). Analytical thinking is at home with linear causality, well-defined and unchanging systems, and reduction. The alternatives of emergent, many-layered, causally spread, non-linear systems in constant constitutive and interactive flux are very hard to manage conceptually. This is an important focus of resistance to many enactive ideas. It is here that synthetic modeling techniques may have their major impact: in producing novel ways of thinking and generating proofs of concept to show that some proposals may not be as coherent as they sound (as in our critical study on value system architectures) or to demonstrate that apparently hazy concepts find clear instantiations even in simple systems (as in the case of emergent coordination through social interaction processes). Methodological minimalism is, therefore, a key element contributing to the acceptability of enactive ideas.

Models that attempt to illuminate the enactive framework will have to take into account the core ideas of enactivism. A serious take on embodiment will depend on the extent to which a system's behavior relies non-trivially on its body and its sensorimotor coupling with the environment as opposed to input-output information processing. Emergent properties and functionality will contrast with misplaced localization in sub-agential modules. Autonomy, to the extent that it can be captured in simulation or robotic models, will depend on how the model instantiates the dynamics of self-constituted precarious processes that generate an identity and how such processes create a normativity at the interactive level that leads to sense-making. Enactive modeling must also relate to experience. As a scientific tool it belongs to the realm of third person methods and so the relation will have to find its place in the process of mutual constraining that has been proposed for the empirical sciences and first person methods already mentioned above.

Alongside the explorations presented in this work and the horizons of questions, methods, and explanations that they open, there will be many other areas where enactive views could make a contribution. We repeat that we have not aspired to be exhaustive neither in breadth nor in depth. But we do think we have moved in the direction where enactivism could grow the strongest: the direction towards higher forms of cognition. Some of the ideas we explored raise more questions than definitive answers. And this is as it should be in the current context. Focusing on the core concepts of enactivism has been a way of changing perspectives on well-known problems. This will inevitably lead to novel questions, which we have raised throughout the paper. How do different modes of value-generation co-exist in a human subject? How does sense-making get socially coordinated through different kinds of participation? How is the creation of novel meaning achieved in transitional activities such as play? Each of these areas indicates a direction where much further work is needed and that might possibly lead to newer horizons.

References

Ashby, W. R. (1960). *Design for a brain: The origin of adaptive behaviour* (Second edition). London: Chapman and Hall.

Auvray, M., Lenay, C., & Stewart, J. (2006). The attribution of intentionality in a simulated environment: the case of minimalist devices. In *Tenth Meeting of the Association for the Scientific Study of Consciousnes, Oxford, UK, 23–26 June, 2006.*

Bach–y–Rita, P., Collins, C. C., Sauders, F., White, B., & Scadden, L. (1969). Vision substitution by tactile image projection. *Nature*, 221, 963–964.

Beer, R. (2003). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, 11, 209–243.

Beer, R. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, 4, 91–99.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47, 139–159.

Bruner, J. (1966). *Toward a theory of instruction.* Cambridge MA: Harvard University Press.

Chiel, H. J., & Beer, R. (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in Neuroscience*, 20, 553–557.

Chrisley, R. (2003). Embodied artificial intelligence. *Artificial Intelligence*, 149, 131–150.

Clark, A., & Toribio, J. (1994). Doing without representing? *Synthese*, 101, 401–431.

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

Clark, A., & Grush, R. (1999). Towards a cognitive robotics. *Adaptive Behavior*, 7, 5–16.

Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.

De Jaegher, H. (2006a). Are people with autism the mindreaders here? In

European Society for Philosophy and Psychology Annual Meeting, Belfast, 24–27 August.

De Jaegher, H. (2006b). *Social interaction rhythm and participatory sense-making. An embodied, interactional approach to social understanding, with implications for autism*. Unpublished DPhil Thesis, University of Sussex.

Dennett, D. C. (1993). Review of F. Varela, E. Thompson and E. Rosch, The Embodied Mind. *American Journal of Psychology*, 106, 121–126.

Dewey, J. (1929). *Experience and Nature* (Second edition). New York: Dover. (1958).

Di Paolo, E., Rohde, M., & Iizuka, H. (2007). Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. *New Ideas in Psychology*. Forthcoming.

Di Paolo, E. A. (2000a). Behavioral coordination, structural congruence and entrainment in acoustically coupled agents. *Adaptive Behavior*, 8, 27 – 47.

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

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

Donaldson, M. (1992). *Human minds: An exploration*. London: Penguin Books.

Dreyfus, H. L. (2002). Intelligence without representation - Merleau-Ponty's critique of mental representation. The relevance of phenomenology to scientific explanation. *Phenomenology and the Cognitive Sciences*, 1, 367–383.

Edelman, G. M. (1989). *The remembered present. A biological theory of consciousness*. Oxford: Oxford University Press.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, 2, 704–716.

Fagen, R. (1981). *Animal play beha*vior. Oxford: Oxford University Press.

Fink, E. (1968). The Oasis of Happiness: Toward an Ontology of Play. *Yale French*

Studies, 41, 19–30.

Gallagher, S. (1997). Mutual enlightenment: Recent phenomenology in cognitive science. *Journal of Consciousness Studies*, 4, 195 – 215.

Gallagher, S. (2001). The practice of mind: Theory, simulation or interaction?. In Thompson, E. (Ed.), *Between ourselves: second-person issues in the study of consciousness*, pp. 83–107. Exeter: UK: Imprint Academic.

Gallagher, S. (2005). *How the body shapes the mind*. New York: Basic Books.

Gallese, V. (2001). The 'Shared Manifold' hypothesis: From mirror neurons to empathy. In Thompson, E. (Ed.), *Between ourselves: second-person issues in the study of consciousness*, pp. 33–50. Exeter: UK: Imprint Academic.

Goffman, I. (1961). Encounters. Indianopolis: Bobbs-Merrill.

Goldstein, K. (1995/1934). *The organism*. New York: Zone Books.

Gordon, R. M. (1996). 'Radical' simulationism. In Carruthers, P., & Smith, P. K. (Eds.), *Theories of Theories of Mind*, pp. 11–21. Cambridge University Press, Cambridge.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery and perception. *Behavioral and Brain Sciences*, 27 (3), 377–396.

Harvey, I., Di Paolo, E. A., Wood, R., Quinn, M., & Tuci, E. (2005). Evolutionary robotics: A new scientific tool for studying cognition. *Artificial Life*, 11, 79–98.

Harvey, I., P., H., Cliff, D., Thompson, A., & Jakobi, N. (1997). Evolutionary robotics: the Sussex approach. *Robotics and Autonomous Systems*, 20, 207 – 224.

Heidegger, M. (1962). *Being and Time*. Oxford: Blackwell. Trans. J. Macquarrie and E. Robinson.

Hobson, P. (1990). On acquiring knowledge about people and the capacity to pretend: Response to Leslie (1987). *Psychological Review*, 97, 114–121.

Hobson, P. (2002). *Cradle of thought*. London: Macmillan.

Huizinga, J. (1949). *Homo Ludens*. London: Routledge.

Hutchins, E. (1995a). *Cognition in the wild*. Cambridge MA: MIT Press.

Izquierdo Torres, E., & Di Paolo, E. A. (2005). Is an embodied system ever purely reactive?. In Capcarrere, M., Freitas, A., Bentley, P., Johnson, C., & Timmis, J. (Eds.), *Advances in Artificial Life: Eighth European Conference on Artificial Life*, pp. 252–261. Springer Verlag, Berlin.

Jonas, H. (1966). The phenomenon of life: Towards a philosophical biology. Evanston, IL:
Northwestern University Press.

Juarrero, A. (1999). *Dynamics in action: Intentional behavior as a complex system*. Cambridge MA: MIT Press.

Kelly, S. (2000). Grasping at straws: Motor intentionality and the cognitive science of skillful action. In Wrathall, M., & Malpas, J. (Eds.), *Heidegger, coping, and cognitive science: Essays in honor of Hubert L. Dreyfus, vol. II.* Cambridge MA: MIT Press.

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge MA: MIT Press.

Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters*. Cambridge: Cambridge University Press.

Kim, J. (1999). Making sense of emergence. *Philosophical Studies*, 95, 3–36.

Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: Lessons from autism. *Philosophical Transactions of the Royal Society London B Biological Sciences*, 358, 345–360.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago Press, Chicago.

Le Van Quyen, M., & Petitmengin, C. (2002). Neuronal dynamics and conscious experience: an example of reciprocal causation before epileptic seizures. *Phenomenology and the Cognitive Sciences*, 1, 169–180.

Lenay, C. (2003). Ignorance et suppléance : la question de l'espace. *Hdr 2002, Université de Technologie de Compiègne.*

Leslie, A. (1987). Pretence and representations: The origins of 'Theory of Mind'. *Psychological Review*, 94, 412–426.

Lewontin, R. C. (1983). The organism as the subject and object of evolution. *Scientia*, 118, 63 – 82.

Lutz, A. (2002). Toward a neurophenomenology as an account of generative passages: A first empirical case study. *Phenomenology and the Cognitive Sciences*, 1, 133–167.

Maynard Smith, J., & Szathmáry, E. (1995). *The major transitions in evolution*. Oxford: W. H. Freeman.

McGee, K. (2005). Enactive cognitive science. Part 1: Background and research themes. *Constructivist Foundations*, 1 (1), 19–34.

Merleau-Ponty, M. (1962). *Phenomenology of perception*. London: Routledge.

Miller, S. (1973). Ends, means and Galumphing: Some leitmotifs of play. *American Anthropologist*, 75, 87–98.

Millikan, R. G. (1984). *Language, thought and other biological categories: New foundations for realism*. Cambridge: MIT Press.

Moran, G., Fentress, J. C., & Golani, I. (1981). A description of relational patterns during "ritualized fighting" in wolves. *Animal Behaviour*, 29, 1146–1165.

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

Nolfi, S., & Floreano, D. (2000). *Evolutionary robotics. The biology, intelligence, and technology of self–organizing machines*. Cambridge MA: MIT Press.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24 (5), 883–917.

Oyama, S. (2000). *The ontogeny of information: Developmental systems and evolution* (Second edition). Duke University Press.

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

Piaget, J. (1936). *La naissance de l'intelligence chez l'enfant.* Delachaux et Niestlé, Neuchátel-Paris.

Piaget, J. (1951). *Play, dreams and imitation in childhood*. London: Routledge.

Piaget, J. (1967). *Biologie et connaissance: Essai sur les relations entre les régulations organiques et les processus cognitifs.* Gallimard.

Poincaré, H. (1907). *La science et l'hypothèse*. Paris: Flammarion.

Quinn, M. (2001). Evolving communication without dedicated communication channels. In Kelemen, J., & Sosik, P. (Eds.), *Advances in Artificial Life: Sixth European Conference on Artificial Life*, pp. 357–366. Berlin: Springer Verlag.

Rodriguez, E., George, N., Lachaux, J.-P., Matinerie, J. Reanault, B., & Varela, F. J. (2001). Perception's shadow: long-distance synchronization of human brain activity. *Nature*, 397, 430 – 433.

Rohde, M., & Di Paolo, E. A. (2006). 'Value signals': An exploration in evolutionary robotics. *Cognitive science research paper 584, COGS, University of Sussex*.

Rutkowska, J. C. (1997). What's value worth? Constraints on unsupervised behaviour acquisition. In Husbands, P., & Harvey, I. (Eds.), *Proceedings of the Fourth European Conference on Artificial Life*, pp. 290 – 298, Cambridge, MA: MIT Press.

Salomon, R. (1998). Achieving robust behavior by using proprioceptive activity patterns. *BioSystems*, 47, 193–206.

Schiffrin, D. (1994). *Approaches to Discourse*. Oxford: Blackwell.

Schwartzman, H. B. (1978). *Transformations: The anthropology of children's of play*. New York: Plenum.

Sheets-Johnstone, M. (1999). Emotion and movement: A beginning empirical-phenomenological analysis of their relationship. *Journal of Consciousness Studies*, 6, 11–12.

Sheets-Johnstone, M. (2003). Child's play: A multidisciplinary perspective. *Human Studies*, 26, 409–430.

Silberstein, M., & McGeever, J. (1999). The search for ontological emergence. The *Philosophical Quarterly*, 49, 182–200.

Skarda, C. A., & Freeman, W. J. (1987). How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences*, 10, 161–195.

Sporns, O., & Edelman, G. M. (1993). Solving Bernstein's problem: A proposal for the development of coordinated movement by selection. *Child Development*, 64, 960 – 981.

Stewart, J., & Coutinho, A. (2004). The affirmation of self: A new perspective on the immune system. *Artificial Life*, 10, 261–267.

Sutton-Smith, B. (1966). Piaget on play: a critique. *Psychological Review*, 73, 104–110.

Sutton-Smith, B. (1997). *The ambiguity of play*. Cambridge, MA: Harvard University Press.

Thelen, E., & Smith, L. B. (1994). A dynamic systems approach to the development of cognition and action. Cambridge, MA: MIT Press.

Thompson, E. (2001). Empathy and consciousness. In Thompson, E. (Ed.), *Between ourselves: second-person issues in the study of consciousness*, pp. 1–32. Exeter: UK: Imprint Academic.

Thompson, E. (2005). Sensorimotor subjectivity and the enactive approach to experience. *Phenomenology and the Cognitive Sciences*, 4, 407–427.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge, MA: Harvard University Press.

Thompson, E., & Varela, F. (2001). Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences*, 5, 418–425.

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In Bullowa, M. (Ed.), *Before speech*, pp. 39 – 52. Cambridge: Cambridge University Press.

van Gelder, T. (1999). Wooden iron? Husserlian phenomenology meets cognitive science. In Petitot, J., Varela, F. J., Pachoud, B., & Roy, J.-M. (Eds.), *Naturalizing phenomenology*, pp. 245–265. Stanford, CA: Stanford University Press.

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

Varela, F. J. (1991). Organism: A meshwork of selfless selves. In Tauber, A. I. (Ed.), *Organism and the origin of the self*, pp. 79 – 107. Netherlands: Kluwer Academic.

Varela, F. J. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies*, 3, 330 – 350.

Varela, F. J. (1997). Patterns of life: Intertwining identity and cognition. *Brain and Cognition*, 34, 72 – 87.

Varela, F. J. (1999). The specious present: A neurophenomenology of time consciousness. In Petitot, J., Varela, F. J., Pachoud, B., & Roy, J.-M. (Eds.), *Naturalizing phenomenology*, pp. 266 – 314. Stanford, CA: Stanford University Press.

Varela, F. J., Lachaux, J.-P., Rodriguez, E., & Matinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2, 229 – 230.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.

Verschure, P. J., Wray, J., Sporns, O., Tononi, G., & Edelman, G. M. (1995). Multilevel analysis of classical conditioning in a behaving real world artifact. *Robotics and Autonomous Systems*, 16, 247–265.

Vygotsky, L. S. (1966). Play and its role in the mental development of the child. *Soviet Psychology*, 12, 62–76.

Walter, W. G. (1950). An imitation of life. *Scientific American*, 182(5), 42 – 45.

Watson, M. W., & Jackowitz, E. R. (1984). Agents and recipient objects in the development of early symbolic play. *Child Development*, 55, 1091–1097.

Weber, A. (2003). *Natur als Bedeutung. Versuch einer Semiotischen Theorie des Lebendigen*. Königshausen & Neumann, Würzburg.

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

Wheeler, M. (2005). *Reconstructing the cognitive world: The next step*. Cambridge MA: MIT Press.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, 9, 625–636.

Winnicott, D. (1971). *Playing and reality*. London: Routledge.

Winograd, T., & Flores, F. (1986). *Understanding computers and cognition*. Norwood NJ: Ablex.

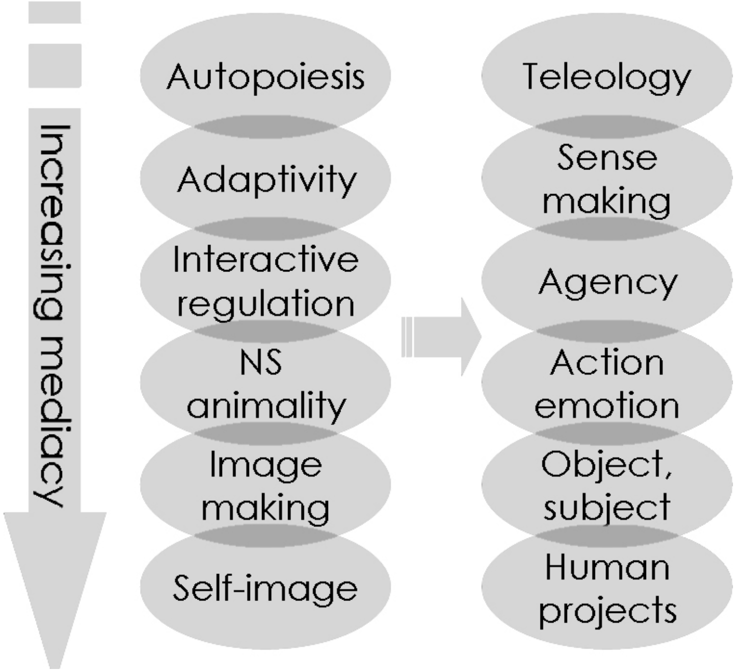

Figure 1: Life-mind continuity and the scale of increasing mediacy, (see text).

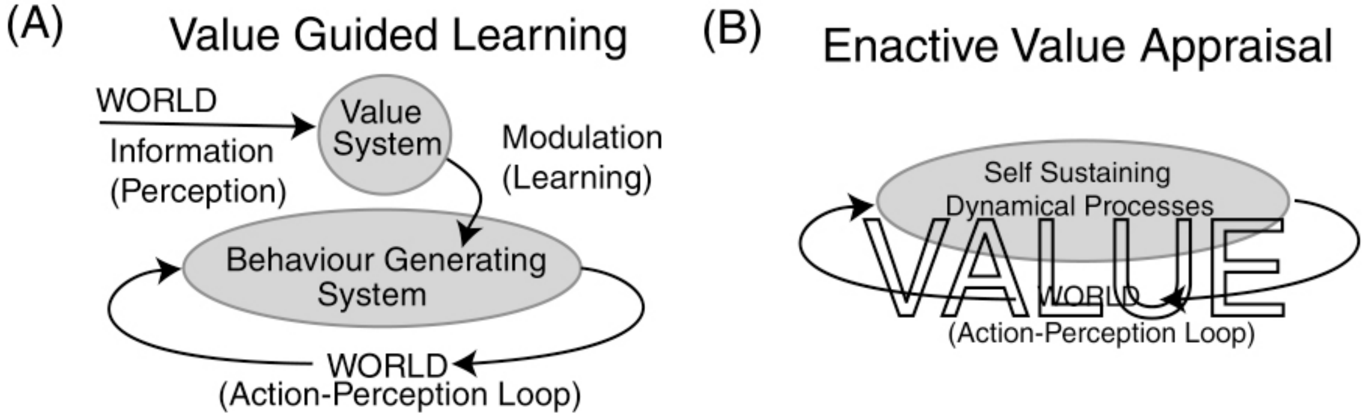

Figure 2: An illustration of the value systems (A) and the enactive approach (B) to conceptualizing values.

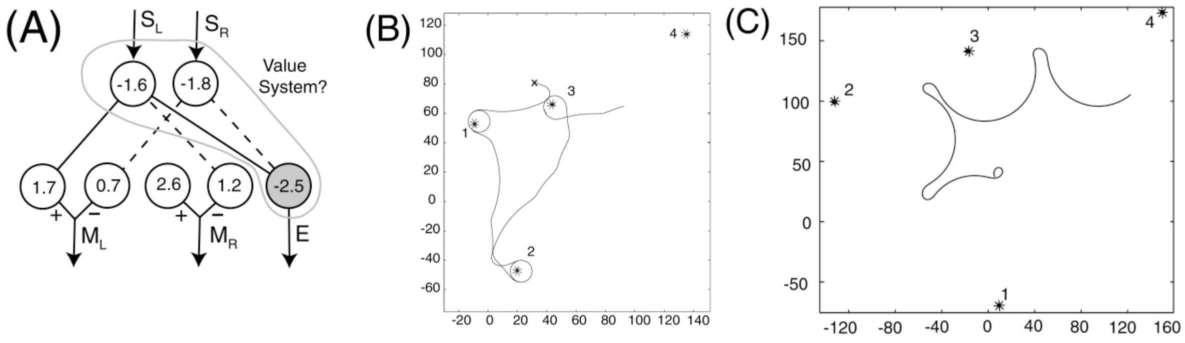

Figure 3: (A) The value judging and light seeking agent controller. (B) The successful light seeking behavior (C) The deterioration of light seeking through applications of the principles of neural Darwinism.

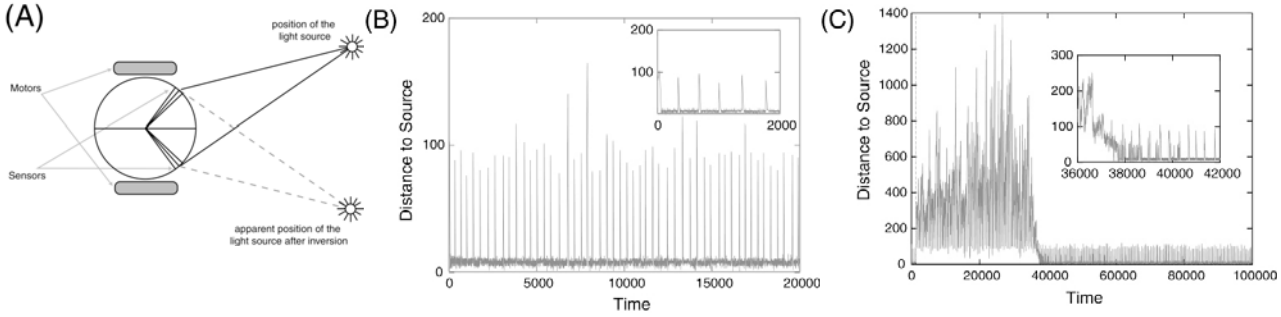

Figure 4: Experiments in homeostatic adaptation using a two-wheeled light seeking agent (A). The agent's distance to a long series of light sources is plotted as a function of time both for the case of normal (B) and inverted (C) visual fields. In (B) the agent approaches each new source of light that replaces the old one; in (C) immediately after sensors are inverted, the agent moves away from new light sources in its vicinity until adaptation ensues and light seeking behavior recovers. See text for details.

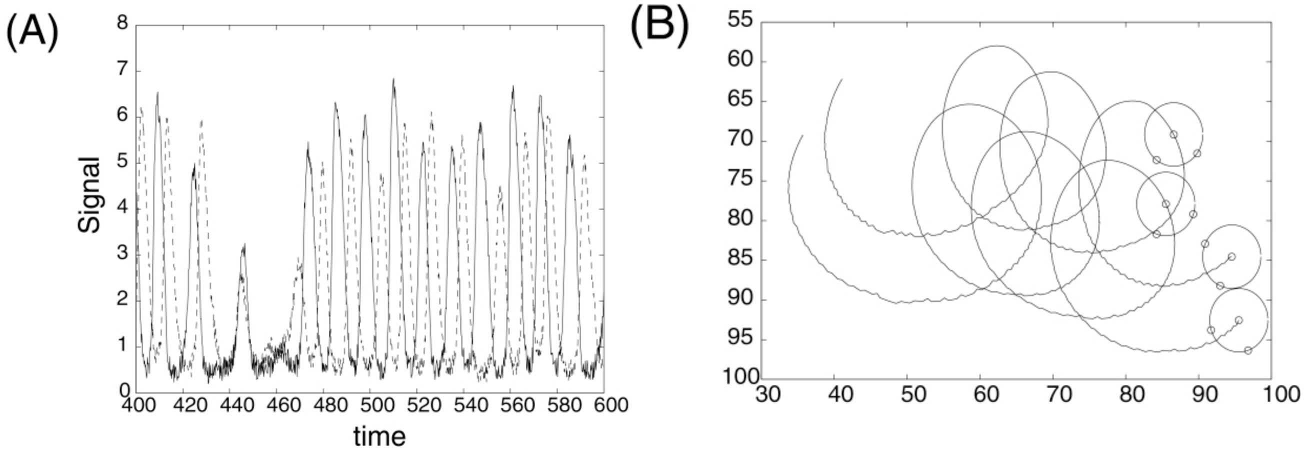

Figure 5: Sound patterns of agents in coordination (A) showing turn-taking activity. (B): trajectories of agents in coordination.

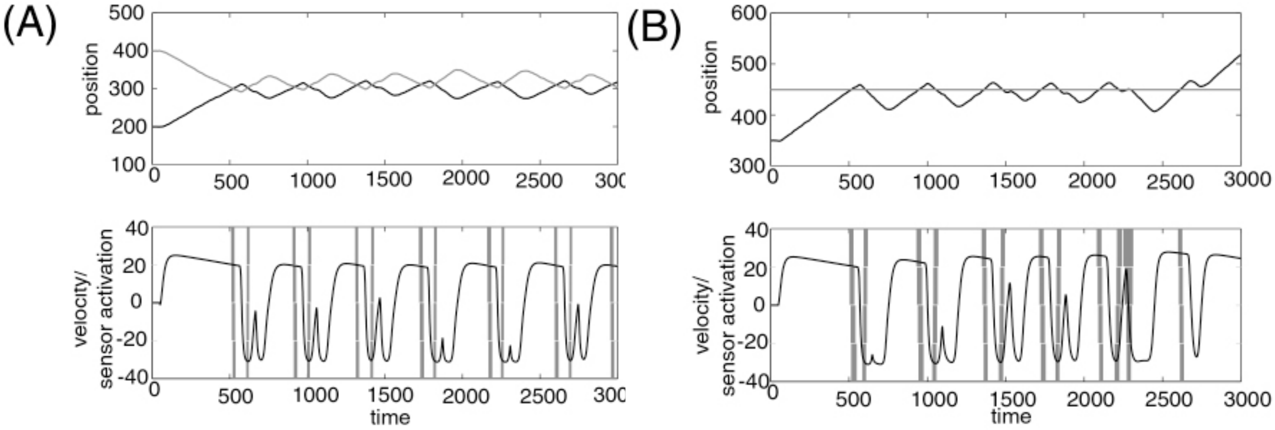

Figure 6: Perceptual crossing model. Top plots show the trajectories of agents over time; plots at the bottom show the motor commands (dark line) and sensor input (gray line). (A): Stabilized social perceptual crossing, (B): scanning of a fixed object.

Notes

[1] We are not concentrating here on the ontological implications of the enactive approach, which have often been interpreted in constructivist terms (McGee, 2005). Without retreating from this issue we want to focus on how enactive ideas can work and generate novel understanding. The emphasis on the ontological implications of enactivism has, in our opinion, often produced the negative effect of sympathetic thinkers disassociating themselves from the use of the term e.g., (Clark, 1997). We want to pragmatically focus on the enactive approach as a scientific framework.

[2] This is shown by work on evolutionary robotics. Beer (2003) has explored how the process of 'making a decision' between two actions in a simple agent is in fact extended over time and does not happen in an instant. Izquierdo-Torres and Di Paolo (2005) have demonstrated the role of time-extended action to disambiguate perceptual tasks in similar agents.

[3] Emergence in this view is close to the notion proposed in (Thompson & Varela, 2001; Thompson, 2007) with the exception that our second requirement is there presented only as a possibility. We favor a stronger definition because we want to emphasize the role of mutual causation in order to introduce a sharper contrast between enactivism and reductionism.

[4] A problem shared by other sensorimotor theories of social cognition such as those built upon the role of 'mirror neurons' (Gallese, 2001) – additionally, such neural correlations themselves should be treated as suspect of the meaning reduction criticized in section 3.

[5] Interestingly, the agent's behavior resembles the human subject's behavior only if we include a delay between an agent's encounter of an object and input to the neurocontroller. If such a delay is not present, the agent's position eventually converges to a fixed point and stands still. This result raises an interesting question: why do subjects keep oscillating around each other, rather than to just 'stand on top of each other' after recognition? Our model predicts that sensory delays play a role in this phenomenon and that the amplitude of the scanning oscillations around a target is positively correlated with the amount of delay.

[6] One of Donaldson main points is how, since the Enlightenment, Western civilization has emphasized the development of the higher modes mainly for thought but not for emotion. We do not have university degrees in being a good happy person, for instance.

[7] Although there is a paucity of research on play strictly from within cognitive science, important relevant works on the subject can be found in the fields of cultural anthropology (Schwartzman, 1978), developmental psychology (Sutton-Smith, 1997), phenomenology (Fink, 1968), animal behavior (Fagen, 1981), psychoanalysis (Winnicott, 1971), and social science (Goffman, 1961; Huizinga, 1949).