The enactive approach

Theoretical sketches from cell to society*

Tom Froese and Ezequiel A. Di Paolo
University of Tokyo / University of the Basque Country and
University of Sussex

There is a small but growing community of researchers spanning a spectrum of disciplines which are united in rejecting the still dominant computationalist paradigm in favor of the *enactive approach*. The framework of this approach is centered on a core set of ideas, such as autonomy, sense-making, emergence, embodiment, and experience. These concepts are finding novel applications in a diverse range of areas. One hot topic has been the establishment of an enactive approach to social interaction. The main purpose of this paper is to serve as an advanced entry point into these recent developments. It accomplishes this task in a twofold manner: (i) it provides a succinct synthesis of the most important core ideas and arguments in the theoretical framework of the enactive approach, and (ii) it uses this synthesis to refine the current enactive approach to social interaction. A new operational definition of social interaction is proposed which not only emphasizes the cognitive agency of the individuals and the irreducibility of the interaction process itself, but also the need for jointly co-regulated action. It is suggested that this revised conception of 'socio-cognitive interaction' may provide the necessary middle ground from which to understand the confluence of biological and cultural values in personal action.

Keywords: adaptivity, autonomy, cognition, enaction, sense-making, social interaction

1. Introduction

There is a small but growing community of researchers spanning a spectrum of academic disciplines which are united in rejecting the still dominant computationalist framework in favor of the late biologist's Francisco Varela's paradigm of *enaction* (e.g., Stewart et al. 2011; Torrance 2005, 2007). This enactive approach consists of a core set of ideas, namely autonomy, sense-making, emergence, embodiment, and experience, which find novel applications in a diverse range of

Pragmatics & Cognition 19:1 (2011), **1–36**. DOI 10.1075/pc.19.1.01fro
ISSN 0929–0907 / E-ISSN 1569–9943

disciplines such as biology, phenomenology, artificial life, social science, robotics, psychology, and neuroscience (Di Paolo et al. 2011). One specific area of research that is currently generating a lot of interest in the enactive approach is its approach to social interaction (Di Paolo 2009b). In contrast to the mainstream this account of sociality begins with an emphasis of biological autonomy and mutually coordinated interaction. It is recognized that the interaction process itself forms an irreducible domain of dynamics which can be constitutive of individual agency (De Jaegher and Froese 2009) and social cognition (De Jaegher et al. 2010). Moreover, it is possible to trace the influences of such irreducible interactions between autonomous systems all the way from cell to society and back again.

The main purpose of this paper is to serve as an advanced entry point into these recent developments. It accomplishes this task in a twofold manner: (i) it provides a succinct synthesis of the most important core ideas and developments in the theoretical framework of the enactive approach, and (ii) it uses this synthesis to refine the enactive approach to social interaction by specifying more precisely what characterizes the kinds of interaction found in three distinct types of inter-agent situations, namely within (organismic) multi-agent systems, (animal) social systems, and (human) socio-cultural systems. Of course, the enactive approach is still a very young research program, and certainly no claims of relative completeness can yet be made. In particular, this paper will say little about human language in itself although the topic is not beyond the scope of enactive thinking (see Bottineau 2011). Nevertheless, enough has been accomplished that it is possible to give some sketches of what such a complete picture might look like, and to provide a flavor of what may be needed to fill in the remaining gaps.

1.1 What is the enactive approach?

The enactive approach was initially conceived as an embodied and phenomenologically informed alternative to mainstream cognitive science (Varela et al. 1991).[1] Since then it has begun to establish itself as a wide-ranging research program with the potential to provide a new perspective on an extremely diverse variety of phenomena, reaching all the way from the single cell organism to human society (Thompson 2007). Moreover, the ongoing search for novel theoretical and methodological foundations has led to a series of systematic confrontations with some of the hardest questions known to philosophy and science: What is meaning and where does it come from? What defines cognition? What is the relationship between life and mind? What defines agency? What is special about social forms of interaction? What is the role of culture for human consciousness?

The way in which the enactive approach has approached these and other such questions has already stimulated productive debates within the specific domains of inquiry to which they are traditionally assigned. However, the goal of addressing these questions is not exhausted by such 'local' revisionism. The particular scientific discourse, though spread over such an apparently disparate set of research questions, is implicitly unified by the conceptual and methodological framework of the enactive approach. In other words, rather than being constrained by the traditional boundaries of any specific academic field, the research framework of this approach is inherently trans-disciplinary and driven by fundamental questions that are organized around the core ideas of autonomy, sense-making, emergence, embodiment, and experience (Di Paolo et al. 2011). The advantage of this conceptual coherence is a discourse that can integrate a diverse set of observations which are otherwise separated by disciplinary discontinuities.

This trans-disciplinary integration has to proceed along a delicate middle way: neither an eliminative reductionism nor a mysterious dualism will do. Observations drawn from distinct regions of phenomena must retain a relative independence with respect to each other. For example, even though the interactions between several embodied agents can enable the emergence of a domain of social phenomena, interaction in the social domain cannot be reduced to the behavior of the individuals (De Jaegher and Di Paolo 2007). Similarly, even though the interactions between an animal's neurons can enable the emergence of a domain of behavior and cognition, the latter domain cannot be reduced to the functioning of the nervous system (Barandiaran and Moreno 2006). And, even though the interaction between chemical constituents can enable the emergence of an autonomous individual, the existence of this individual cannot be reduced to the sum of the chemical components (Varela 1997). In other words, it is because the enactive approach starts with the concept of *autonomy* in *embodied* systems that it can speak about the non-mysterious *emergence* of non-reducible domains of activity, which are typically associated with qualitative shifts in *experience*. This re-enchantment of the concrete (Varela 1995) is the common denominator of the enactive approach, and it does not matter whether this approach is employed to investigate social, individual, or sub-individual phenomena.

One important consequence of such a unified discourse is that it becomes impossible to study any particular phenomenon, or even a whole domain of phenomena, in complete isolation. In addition to the demands of standard scientific practice, all systematic research within the enactive approach must face up to the immense challenges that are posed by non-linear interdependence of phenomena across all traditionally defined ontological regions (cf. Figure 1).

Even a discussion of the biological foundations of minimal agency cannot ignore how it is possible for metabolic values to give rise to detrimental but self-sustaining behavioral patterns (habits), or the way in which arbitrary socio-cultural norms can shape our metabolic constitution (Di Paolo 2009c). While such

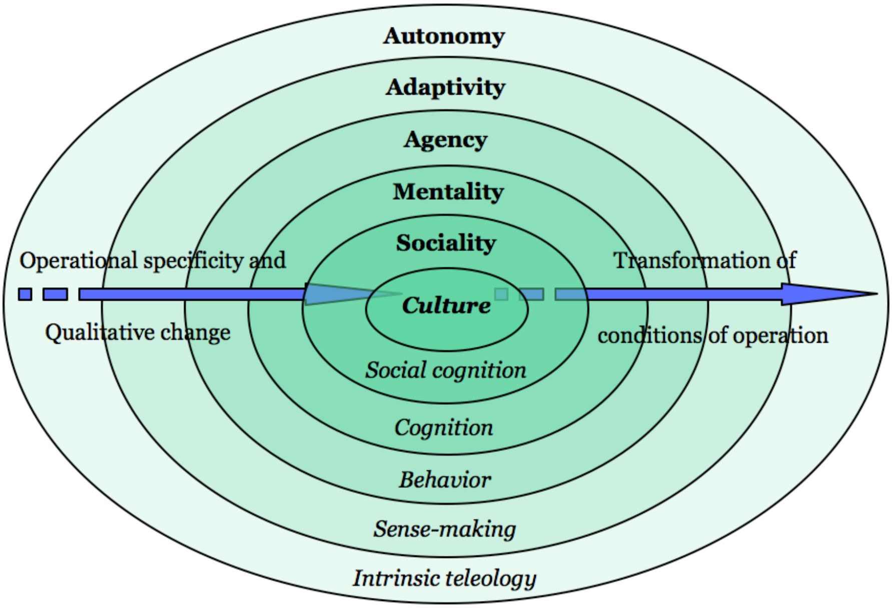

Figure 1. Schematic illustration of the core concepts of the enactive approach. Any inner layer necessarily depends on all of the outer layers, although complex relations can obtain between layers in both directions such that the emergence of new domains transforms the background conditions of operation. The necessary and sufficient conditions for each qualitative phenomenon specified at the bottom of a layer (e.g., 'sense-making') are specified by the operational requirements at the top of that layer, including those of all previous outer layers (e.g., 'autonomy' and 'adaptivity'). The term 'agency' refers to the ability of an autonomous system to achieve adaptation not only via internal re-organization, but also by adaptive regulation of its sensorimotor interactions. 'Mentality' denotes a form of agency whereby the norms of this regulatory activity are underdetermined by metabolic criteria alone (e.g., because of a nervous system), and 'sociality' additionally requires that the norms are partly determined by other-related concerns. The central layer, culture, is still in much need of further clarification by the enactive approach in both operational and phenomenological terms. As the operational specificity increases with each inner layer we can attribute an expansion of qualitative complexity to the perspective of the system. Note, again, that this is a two-way interaction since the effects of the emergent properties of the more specific layers can reenter into the constitutive background layers by transforming their conditions of operation.

considerations of interdependence can often be quite daunting in practice, they are in principle desirable because they help to prevent research from being diverted by false problems and solutions that owe their existence mainly to ill-considered abstractions. Moreover, a systematic recognition of interdependence entails that the investigation of any particular phenomenon can be used as an entry point from which to explore the entire range of the enactive approach. In this paper, for example, we will trace the insights of the enactive approach from organismic autonomy to human culture.

1.2 Overview of the paper

The aim of this paper is to critically review the enactive approach to social interaction. As noted above, in order to accomplish this goal a brief detour through other parts of the enactive approach will be necessary. In particular, we will introduce its conception of autonomous agency, as this will furnish us with the necessary conceptual framework, as well as provide us with a notion of individuality in relation to which a discussion of sociality makes sense. More specifically, we will make use of the concept of autonomy in order to illustrate the insights of the enactive approach from the most basic forms of system-environment interaction to the conditions of cultural interaction. The notion of agency is introduced as the most basic form of autonomous existence that can become part of a multi-agent system, namely a system in which the relational dynamics of inter-individual interactions can themselves take on an autonomous organization. This is followed by a consideration of the necessary conditions for social interaction, which requires a more specific form of agency (mentality) capable of constituting a cognitive domain. Finally, on this basis we discuss the role of the socio-cultural context.

2. Biological foundations

In this section we will briefly introduce the notion of biological autonomy, because this will provide us with the basic conceptual framework that we need in order to understand the biological foundations of the enactive approach. We then argue that adaptive autonomy is the minimal form of life, and that living is essentially a process of sense-making. On this basis we develop a definition of agency which will enable us to talk about multi-agent systems in the next section.

2.1 Biological autonomy: Identity, asymmetry, and normativity

Arguably the most foundational concept of the entire enactive approach is the notion of *autonomy*. This notion can be traced back to the seminal work of the Chilean biologists Humberto Maturana and Francisco Varela who proposed a description of the minimal organization of living systems, called *autopoiesis*, by reflecting on the metabolic self-production of single-cell organisms (Varela et al.

1974; Maturana and Varela 1987). The notion of autopoiesis continues to be influential in the enactive approach today, where it is typically closely associated with chemical forms of self-production. However, while such chemical autopoiesis is indeed a paradigmatic example of autonomy, it is not the only form that an autonomous system can take. Varela thus turned the lessons offered by the autonomy of minimal living systems into a more general operational characterization:

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

This definition of autonomy as *organizational closure* applies to living systems, such as single-cell and multi-cellular organisms, but moreover to a whole host of other systems such as the immune system, the nervous system, and even to social systems (Varela 1991). The self-reference inherent in the process of self-production, which forms the core of this definition of autonomy, has important implications: it allows us to talk about the interrelated notions of identity, precariousness, and the enaction of a meaningful world for the autonomous system. The problem of identity constitution has become especially pronounced in traditional robotics and AI because of the arbitrary choices that researchers are forced to make when distinguishing the system from its environment (Froese and Ziemke 2009). Without the autonomy afforded by organizational closure the system is incapable of defining its own identity as an individual; it remains an externally defined collection of components that we have merely chosen to designate as an 'agent' by convention. An autonomous system, on the other hand, is organized in such a way that its activity is both the 'cause and effect' of its own autonomous organization; in other words, its activity depends on organizational constraints, which are in turn regenerated by the activity itself. This gives it an essentially self-constituted identity because its own generative activity demarks what is to count as part of the system and what belongs to the environment.

Since the components that make up this autonomous system would normally disappear if it were not for this active realization of the self-producing organization, we should not think of the emergent identity as a static entity. It is more like an intrinsically open form of becoming whose continued existence is an ongoing achievement in the face of potential disintegration. We explicitly denote this situation by characterizing the identity of an autonomous system as being *precarious*. Disembodied systems such as computers, on the other hand, have a static identity that is imposed on a physical substrate from the outside and do not depend on any activity to persist. Generally, a component of such a system remains identical to itself even if isolated from the rest of the system. In sum, when we are referring to an autonomous system we are referring to a system composed of several processes that actively generate and sustain their systemic identity under precarious conditions.[2]

Since autonomous systems bring forth their own identity by actively demarcating the boundary between 'self' and 'other' during their ongoing self-production, it follows that they also actively and autonomously determine their domain of possible interactions, i.e., the potential manners in which the system can relate to its environment without ceasing to persist. Furthermore, what an autonomous system does, due to its precarious mode of existence, is to treat the perturbations it encounters during its ongoing activity from a perspective of significance which is not intrinsic to the perturbations themselves. In other words, the meaning of an encounter is not fully determined by that encounter itself. Instead, significance for the autonomous system is acquired in relation to the ongoing necessity for realizing its self-constituted identity, and thus constitutes a concern which is relative to the current situation of the system and its needs. This process of meaning generation in relation to the concerned perspective of the autonomous system is what is meant by the notion of *sense-making* (Weber and Varela 2002). It is important to note that the significance which is continuously brought forth by the endogenous activity of the autonomous system is what makes the lived world, as it appears from the perspective of that system, distinct from its physical environment, as it can be distinguished by an external observer (Varela 1997). In sum, sense-making is the enaction of a meaningful world by an autonomous system.[3]

The enactive approach to autonomy and sense-making entails that meaning is not to be found in the external environment or in the internal dynamics of the system. Instead, meaning is an aspect of the *relational* domain established between the two. It depends on the specific mode of co-determination that each autonomous system realizes with its environment, and accordingly different modes of structural coupling will give rise to different meanings. However, it is important to note that the claim that meaning is grounded in such relations does not entail that meaning can be reduced to those relational phenomena. There is an *asymmetry* underlying the relational domain of an autonomous system since the very existence of that relational domain is continuously enacted by the endogenous activity of that system (Barandiaran et al. 2009). In contrast to most embodied AI, where the relational domain exists no matter what the system is or does, the relational domain of a living system is not pre-given. It follows from this that any model of agency that only captures the relational dynamics on their own, as is the case with most work on sensorimotor situatedness, will only be able to capture the functional aspects of the behavior but not its intrinsic meaning. This is the root of the famous problem of meaning in the field of AI and robotics (see Froese and

Ziemke 2009). In order for these considerations to be of more specific use for the development of a more precise notion of agency, as required for our discussion of the dynamics of multi-agents systems, we first need to unpack the notion of sense-making in a bit more detail.

2.2 Adaptivity and sense-making

According to the enactive approach, the normativity inherent in sense-making implies that perturbations are somehow evaluated in relation to the autonomous system's viability. Varela (1997) has tried to situate the source of this sense-making in the occurrence of minor or major breakdowns of the autonomous system's active self-production (autopoiesis). However, the concept of autopoiesis (or constitutive autonomy more generally) by itself allows no gradation — either a system belongs to the class of such systems or it does not. The self-constitution of an identity can thus provide us only with the most basic kind of norm, namely that all events are good for that identity as long as they do not destroy it (and the latter events do not carry any significance because there will be no more identity to which they could even be related).

On this basis alone there is no room for accounting for the different shades of meaning which are constitutive of any organism's lived *Umwelt* (von Uexküll 1934/1957). Furthermore, the operational definitions of autopoiesis and autonomy neither require that the system can actively compensate for deleterious internal or external events, nor do they address the possibility that it can spontaneously improve its current situation. What is missing from these definitions? How can we extend the precarious perspective that is engendered by constitutive autonomy into a wider context of situated relevance such that we can talk about the enaction of the perceptual world?

Di Paolo (2005) has proposed a solution to this problem. He starts from the observation that even minimal autopoietic systems have a certain kind of tolerance or *robustness*. This means that they can sustain a certain range of perturbations as well as a certain range of internal structural changes before they lose their autopoiesis, where these ranges are defined by the organization and current state of the system. We can then define these ranges of non-fatal events as the system's *viability set*, which is assumed to be of finite measure, bounded, and possibly time-varying. However, in order for an autopoietic system to actively improve its current situation, it must (i) be capable of determining how the ongoing structural changes are shaping its trajectory within the viability set, and (ii) have the capacity to regulate the conditions of this trajectory appropriately. These two criteria are provided by the property of *adaptivity*. Similar to the case of robustness, the notion of adaptivity also implies that the autonomous system can tolerate a range of internal and external perturbations.[4] However, it entails a special kind of context-sensitive robustness which involves both actively monitoring perturbations and compensating for their deleterious tendencies. It is not necessary that adaptivity takes place by means of regulation of the system-environment coupling (but such sensorimotor adaptivity is required for agency, as will be clarified later). A more basic form of adaptivity involves internal regulation of metabolic pathways.

Note that an adaptive system's capacity to distinguish between positive and negative tendencies in relation to its current state does not contradict the organizational closure of the autonomous system: the system can measure the type and severity of a tendency according to the changes in the regulative resources required. Accordingly, if autopoiesis (or autonomy) suffices for generating a 'natural purpose' (Kant 1790), adaptivity reflects the organism's capability — necessary for sense-making — of evaluating the needs and expanding the means towards that purpose. While it is likely that some form of adaptivity was assumed to be implicit in the original definition of autopoiesis as constitutive of sense-making by Weber and Varela (2002), it is useful to turn this assumption into the explicit claim: adaptivity is necessary for sense-making.

Moreover, since sense-making depends on active regulation of the autonomous system's internal milieu, it is at this point that we can refer to the system's activity as a form of *living*. The adaptive regulation is an achievement of the autonomous system's internally generated activity rather than merely something that is simply undergone by it. It is therefore appropriate to consider adaptive autonomy as the most basic form of life, and sense-making as the most basic process of living (Thompson 2004). A living being does not only determine its own possible domain of interactions, as is the case for any kind of autonomous system, it also actualizes this domain of possibilities in a meaningful manner by means of adaptive behavior. Since these criteria are satisfied by all living beings, the question becomes how best to distinguish between different forms of life. For example, a plant does not have the same kind of relationship with its environment as does an animal which has to move and perceive in order to sustain itself. However, even a bacterium can do more than to just adaptively rearrange its internal metabolic pathways. It can actively improve its environmental conditions by seeking out areas with greater concentrations of nutrients both by random search and gradient following. In order to better capture these interactive forms of living we need a definition of agency.

2.3 Sensorimotor interaction and adaptive agency

Barandiaran, Di Paolo, and Rohde (2009) identify three conditions that a system must meet in order to be considered as a genuine agent: (i) a system must define its own individuality (identity), (ii) it must be the active source of activity in relation to its environment (interaction asymmetry), and (iii) it must regulate this activity in relation to certain norms (normativity). Accordingly, they put forward a definition of *agency* which holds that an agent is an autonomous system that adaptively regulates its interaction with its environment and thereby makes a necessary contribution to sustaining itself under precarious conditions. How does agency differ from adaptive autonomy?

As Barandiaran and Moreno (2008: 332) point out, an organism can realize the process of adaptive regulation in two distinct ways: (i) by means of the *internal* reorganization of constructive processes (metabolic adjustment), or (ii) the regulation of an *extended* interactive cycle (sensorimotor adjustment). In both cases there is some degree of decoupling from the basic constitutive processes since we are now talking about two dynamic 'levels' in the system: the constitutive level, which ensures ongoing self-construction, and the (now decoupled) interactive subsystem, which regulates boundary conditions of the former. It is only when the mechanisms of regulation operate by modulating structural coupling, such that adaptation is achieved through recursive interactions with the environment (interactive adaptivity), that we speak of *adaptive agency*. In contrast to internal compensation, this adaptive regulation of system-environment relations opens up a novel relational domain that can be traversed by means of behavior or *action* (i.e., regulated sensorimotor cycles).[5] This conception of adaptive agency and action is illustrated in Figure 2.

Of course, adaptive agency specifies only the most basic form of agency, as exemplified for example by a bacterium performing chemotaxis. But the notion still provides us with a useful inclusive category. In particular, it allows us to make some very general remarks about the dynamics of multi-agent systems that apply to the inter-individual interaction of a multitude of life forms, such as bacteria, invertebrates, animals, and humans.

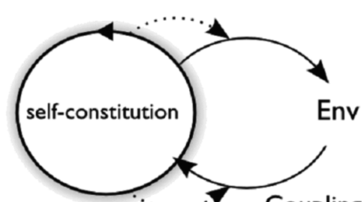

Figure 2. The relationship between constitutive autonomy and adaptive agency: the *autonomous system* self-constitutes an identity which is conserved during structural coupling with its environment (full arrows); *adaptive agency* requires additional regulation by the system which is aimed at adjusting this coupling relationship appropriately (dotted arrows). (Copyright 2010 E. Di Paolo. Licensed under Creative Commons Attribution 3.0 Unported [http://creativecommons.org/licences/by/3.0])

3. Interactive foundations

We have introduced the concept of autonomy in order to develop the enactive account of adaptive agency. In this section we will once again make use of this concept when describing the self-sustaining structures that can emerge on the basis of an interaction process between two or more adaptive agents, i.e., we are interested in the autonomy of multi-agent systems as such. This will provide us with the conceptual foundations for the next section, in which we will analyze what is special about social interactions.

3.1 Multi-agent system: The autonomy of the interaction process

As we argued in the previous section, adaptive autonomy is a sufficient condition for living (sense-making) as such, but it requires an interactive realization of adaptivity, one which depends on the regulation of sensorimotor interaction cycles, for the presence of adaptive agency. This adaptive agency is only the most basic form of agency, but it is sufficient to allow us to consider a simple extension to the basic scenario that was shown in Figure 2. We can imagine two adaptive agents encountering each other in a shared environment, as depicted in Figure 3.

In the case of a solitary embodied agent the sensory stimulation of the agent is largely determined by its own structure and movements, thus giving rise to a closed sensorimotor loop. This closed loop makes it possible for the agent to engage in sensorimotor coordination so as to structure its own perceptual space (see Pfeifer and Scheier 1999: 377–434). However, in the case where two adaptive agents share an environment, one agent's movements can affect that environment in such a way that this results in changes of sensory stimulation for the other agent, and *vice versa*. Moreover, when these changes in stimulation for one agent in turn

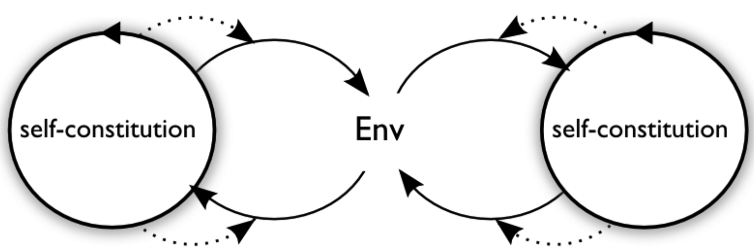

Figure 3. The relationship between two adaptive agents sharing the same environment: the manner in which one agent's movements affect the environment can result in changes to sensory stimulation for the other agent, and vice versa, creating the basis for a multi-agent recursive interaction. (Copyright 2010 E. Di Paolo. Licensed under Creative Commons Attribution 3.0 Unported [http://creativecommons.org/licences/by/3.0])

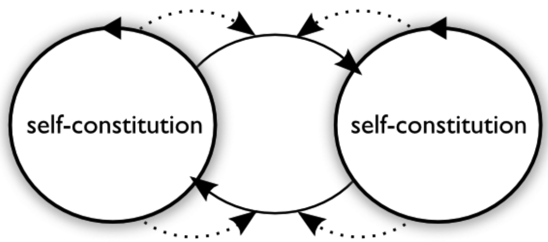

Figure 4. Schematic of a multi-agent system. It is possible that when two adaptive agents who share an environment begin to engage in mutual sensorimotor coupling, that their activities become entwined in such a manner that their mutual interaction results in an interaction process that is itself characterized by an autonomous organization, i.e. an emergent structure in its own right. (Copyright 2010 E. Di Paolo. Licensed under Creative Commons Attribution 3.0 Unported [http://creativecommons.org/licences/by/3.0])

lead to changes in its movement that change the stimulation for the other agent, and so forth in a way that recursively sustains this mutual interaction, the emergent result is a special configuration of coordinated behavior. More precisely, the inter-individual interaction process itself can now be characterized as being an autonomous structure in the relational domain that is constituted by the interacting agents (De Jaegher and Di Paolo 2007). Accordingly, we can simply modify the schematic of Figure 3 by emphasizing the autonomous organization of the interaction process, as shown in Figure 4.

We are now in the position to define the concept of a *multi-agent system* as referring to an interaction process that is constituted by the mutually coordinated behavior of two or more adaptive agents whereby that interaction process is itself characterized by an autonomous organization. More precisely defined:

> Multi-agent interaction is the regulated coupling between at least two adaptive agents, where the regulation is aimed at aspects of the coupling itself so that it constitutes an emergent autonomous organization in the domain of relational dynamics, without destroying in the process the adaptive agency of at least two individuals involved (though their scope can be augmented or reduced).

This definition of interaction in a multi-agent system is based on a related one proposed by De Jaegher and Di Paolo,[6] but it puts more specific requirements on the necessary form of agency (adaptive), and refers to this type of interaction as 'multi-agent' rather than 'social'. The requirement of adaptive agency is merely making explicit the minimal necessary conditions for a system to become a component member of an inter-agential interaction process. The motivation for the avoidance of the term 'social' for this general type of multi-agent interaction is that this gives us a more fine-grained conceptual handle on the variety of phenomena that involve more than one agent, including a more specific definition of the social which we will develop in the next section.

3.2 Examples of multi-agent systems

In order to better illustrate the concept of multi-agent system we will consider some cases that are useful in delineating the concept's and scope of applicability: (i) aggregations of single-cell organisms, and (ii) inter-individual interaction between humans.

Can aggregates of single-cell organisms (adaptive agents) regulate their interactions so as to form multi-agent systems? Famous examples of such systems are the so-called slime molds (*mycetozoans*). In the case of the slime mold *Dycostelium discoideum* the spores begin life as unicellular amoebae and multiply by mitosis while feeding on bacteria. Once the food supply runs out the amoeboid individuals are capable of merging into a multi-cellular 'slug' which allows them to migrate to a more favorable environment. Eventually this 'slug' settles down and transforms into a fruiting body for the distribution of new spores. In the case of this particular species these transformations happen without cellular fusion and with a clear diversity of cellular types. In other words, the mutual interaction between several autonomous individual cells is organizationally closed so as to constitute an autonomous multi-cellular organization in its own right.

But is this final 'second-order' autonomous system strictly speaking a multi-agent system as we have defined it? The answer to this question is not immediately clear because it depends on whether the individual members of the collective 'slug' still retain agency in their own right during this stage of their lifecycle. The question therefore becomes an empirical one: are the collective system's cells achieving their adaptivity through the regulation of interactive cycles? If yes, then the 'slug' is not only an adaptive agent, but also a multi-agent system, which would make it an example of a 'second-order' adaptive agent. In practice, however, the boundaries between these different stages may be fuzzier than our operational definitions suggest, and future work should look more closely at the transitions between them. Slime molds are convenient targets for this endeavor because we can trace transformations from (i) behavior of individual free-floating amoebae (i.e., 'lower level' adaptive agents) to (ii) mutual interaction between several free floating amoebae (i.e., a multi-agent system) to (iii) the formation of one differentiated physical aggregate (i.e., a second-order autonomous system) and, possibly, to (iv) the dissolution of individual agents into one collective organism (i.e., a 'higher level' adaptive agent). Depending on the operational conditions obtaining at these different transformations, we may be in the presence of a multi-agent system only at certain stages of the process or not at all.

Of course, since we have tried to define the most general form of multi-agent system, the applicability of this notion is not limited to the domain of single-cell organisms. In fact, since the theoretical framework of the enactive approach is an extension of general systems theory, its insights are not limited to the concrete domain from which they were originally derived. Even some properties of the social interaction between human beings can be accounted for in terms of a multi-agent system. It is helpful to briefly illustrate this possibility by means of a simple concrete case study.

A recent psychological study by Auvray, Lenay, and Stewart (2009) has investigated the dynamics of human interaction under minimal conditions. Two participants were asked to locate each other in a simple 1-D virtual environment using only left-right movement and an all-or-nothing tactile feedback mechanism, which indicated whether their virtual 'avatar' was overlapping any objects within the virtual space. They could encounter three types of objects: (i) a static object, (ii) the avatar of the other participant, and (iii) a 'shadow' copy of the other participant's avatar that exactly mirrored the other's movement at a displaced location. Since all virtual objects were of the same size and only generated an all-or-nothing tactile response, the only way to differentiate between them was through the interaction dynamics that they afforded. And, indeed, participants did manage to locate each other successfully because ongoing mutual interaction afforded the most stable situation under these circumstances. Thus, even though the participants 'failed' to achieve the task individually, i.e., there was no significant difference between the probability of a clicking response to the other's avatar and the other's shadow object (Auvray et al. 2009: 39), they still managed to solve the task collectively because of the self-sustaining dynamics of the interaction process (see De Jaegher et al. 2010). Even though it is impossible to distinguish the active partner from her irresponsive copy on an individual basis, it turns out that most clicks are made correctly because a mutual interaction is more likely to persist and participants are therefore more prone to face each other.

In this case the organization of the whole facilitated the behavior of the individuals, but this is not a necessity for the formation of multi-agent systems. Indeed, De Jaegher and Di Paolo (2007) argue that it might actually be more revealing to investigate situations in which the individual interactors are attempting to stop interacting, but where the interaction process self-sustains even in spite of this intention. That can easily occur, for instance, when two people attempt to walk past each other in a corridor, but happen to move in mirroring directions at the same time. They thereby co-create a symmetrical coordinated relation, which is likely to result in them moving in mirroring directions again, thus leading to further interaction. In this case the individual intention of terminating the interaction process is actually prevented from being realized due to the emerging coordination patterns at the inter-individual level. In other words, in these kinds of cases the overall organization of the interaction subsumes the individual actions of the interactors in such a way that the autonomous identity of the interactive situation is retained, at least temporarily, despite their own best efforts to the contrary. For better or worse, whole and parts mutually enable and constrain each other.[7]

3.3 Overcoming the 'cognitive gap' of the enactive approach

The discussion of the slime mold and some forms of human interaction is illustrative of the strengths and challenges of the enactive approach. It highlights one of the most outstanding problems faced by this research program, namely that there still is a legitimate concern as to what extent the insights gained on the cellular level scale about to multi-cellular organisms, including humans. We can refer to this problem as the 'cognitive gap' of the enactive approach (see Froese and Di Paolo 2009; De Jaegher and Froese 2009). Since the next sections will introduce concepts that are specific to animals and humans, it is pertinent to sketch the outlines of a response to this worry here. How do we get the enactive approach from cells to society? A partial response to this challenge has already been indicated by the use of the same systemic concepts in both examples. This demonstrates that the concepts which are applicable to slime molds can also tell us something, though surely not everything, about the interaction dynamics between humans. In the introduction to this article we noted this conceptual continuity of the enactive approach and it will again become evident when we come to consider social cognition in the next section.

But we may still ask: does cellular biology really have something to say about the sciences of the human mind? What about all the phenomena that have only appeared after billions of years of evolution? At first sight the task of establishing life-mind continuity starting on the basis of insights gained from minimal, single-cell forms of life appears to equate the cognitive gap with the whole history of life on earth. Surely it would be better to start with a paradigmatic model of agency that is at least of medium complexity, like a simple animal? This stance is typically advocated by embodied approaches to cognitive science and AI (e.g., Brooks 1991). But notice that whereas Brooks' insect-like robots still face an immense *phylogenetic* gap (and hence the provocative title of Kirsh's (1991) paper "Today the earwig tomorrow man?"), the single-cell models often favored by the enactive approach can be viewed as confronting us with an *ontogenetic* gap instead. After all, we all start as single-cell organisms. With this shift in perspective the gap has been narrowed from the entire extent of evolutionary history on earth, to the developmental lifespan of a single human individual. It is therefore essential that the enactive approach pays closer attention to research in developmental systems.

The incorporation of a developmental perspective would go some way to addressing the cognitive gap, but by itself it will not take us all the way from cell to society. Froese and Di Paolo (2009) have argued that a crucial aspect of overcoming the cognitive gap is to realize that multi-agent systems can be a potent source of interactive scaffolding. The core idea of this proposal, namely that many defining aspects of human cognition are realized by our situatedness in a socio-cultural context, has already been advanced in the cognitive sciences from the perspective of anthropology (e.g., Hutchins 1995). What the enactive approach contributes to this perspective is a generalization of this insight to a much wider range of inter-individual interactions. A simple multi-agent system might not provide as much scaffolding as a well developed social interaction, yet the effects of either kind of interaction process are similarly irreducible to individual capacities alone, and either can significantly shape an individual's behavioral domain (De Jaegher and Froese 2009). Indeed, the introduction of the multi-agent system as an intermediate level of analysis between the individual agent and the properly social domain works in favor of overcoming the cognitive gap: it underlines the transformative potential of basic multi-agent interactions *even without* the presence of interactions in which the other agent is explicitly involved *as* another agent.

4. Social interaction

The operational definition of multi-agent system has provided us with a general systemic way of characterizing interactions between adaptive agents that result in the emergence of autonomous structures in their own right. Moreover, a multi-agent system can radically alter the behavioral domains of the interactors in terms of its own normativity, either in accordance with or despite of the goals of those individuals. However, in many contexts as it stands the notion of a multi-agent interaction is too broad to capture what is specific about *social* interactions. To specify what is essential about sociality we must first introduce some additional qualifications. We will focus on two especially important ones, namely the generation of non-metabolic values and the appreciation of other 'selves'.

4.1 Mentality: From adaptive behavior to cognition

The meaning of sense-making and adaptive behavior is strictly related to the viability range of the autonomous identity by which they are enacted. This limits the adaptive organism's normativity to *self*-related values that are based on the individual's metabolic requirements alone. However, in order to make sense of another agent *as* another agent it is a necessary for there to be a capacity of sense-making based on non-metabolic *other*-related values: the presence of the other agent must be perceivable as a foreign locus of goal-directed behavior, i.e., as another self with its own self-related values.[8] The necessary conditions for adaptive agency are by themselves not sufficient to accomplish such a decentralization of significance.

As a first step to explain this insufficiency we can note that there is a mismatch of values: failure to regulate a social interaction does not necessarily imply a direct failure of self-maintenance and metabolic self-production. The values governing the unfolding of social interactions preserve a relative independence with regard to the norms of physical realization and regeneration. However, for an adaptive agent the constitution of relatively independent norms for social purposes is impossible because its capacity for regulating its interactions is, while partially decoupled from constructive processes, still too closely tied to its own metabolic existence. To be sure, the realization of the norms that are constitutive of its regulatory activity can be constrained by the autonomous dynamics of a multi-agent system, but they cannot be simply transformed into specifically social norms because their success is largely determined by basic energetic and material needs.

What is needed is the addition of a new domain of operations in which the behavior is guided by a normativity that is highly underdetermined by metabolic values. In fact, in our bodies there are several such partially decoupled systems, the most famous being the immune system and the nervous system. Both of them are involved in making self-other distinctions in their own way (Varela 1991). But it is the nervous system which is of special interest to us here, because it governs the sensorimotor interactions which are essential for social interaction. Moreover, the nervous system also enables the emergence of autonomous dynamics that are relatively decoupled from metabolic processes such that the regulation of sensorimotor behavior is freed from the strict confines of self-related normativity and can instead be about something other. We argue that this kind of other-related 'aboutness' or *mentality* is a prerequisite to sociality: only a cognitive agent can be a social agent.

But what precisely is *cognition* according to the enactive approach? The question of cognition is obviously one of the most foundational questions faced by the traditional cognitive sciences, and it similarly poses considerable challenges to the enactive approach. However, the challenges it faces are of a fundamentally different kind than those faced by computationalism. In the past it has followed the autopoietic tradition in biology by simply equating cognition with the process of living as autopoiesis (e.g., Stewart 1992). Then, under the influence of the biophilosophy of Kant (1790) and Jonas (1966), this formula has been updated such that cognition becomes equivalent to sense-making (Thompson 2004). However, this position is still not fully satisfactory because adaptive behavior could be restricted to the realization of direct coping only, while cognition can also involve concerns that are not immediately related to ongoing physiological or environmental events. Ultimately, the process of cognition must be flexible enough so that it can be shaped into abstract thought, the phenomenon which has been the target of investigation by the mainstream cognitive sciences.

As a first step toward this goal we can draw on the work of Barandiaran and Moreno (2006, 2008) who have been refining the biological foundations of the enactive approach so as to better account for what is unique about cognition. Effectively, they have focused on the relative independence of the operation of the nervous system with regard to the rest of the living body as the basis for the emergence of a novel domain of autonomous structures. They argue that cognition consists in the adaptive preservation of a dynamical network of autonomous sensorimotor structures sustained by continuous interactions with the environment and the body. More precisely:

> The hierarchical decoupling achieved through the electrochemical functioning of neural interactions and their capacity to establish a highly connected and non-linear network of interactions provides a dynamic domain with open-ended potentialities, not limited by the possibility of interference with basic metabolic processes (unlike diffusion processes in unicellular systems and plants). It is precisely the open-ended capacity of this high-dimensional domain that opens the door to spatial and temporal self-organization in neural dynamics and generates an extremely rich dynamic domain mediating the interactive cycle, overcoming some limitations of previous sensorimotor control systems (Barandiaran and Moreno 2008: 338).

A paradigmatic example of such autonomous structures are habits, which encompass partial aspects of the nervous system, physiological and structural systems of the body, and patterns of behavior and processes in the environment (Di Paolo 2003). Due to the relative independence of the nervous system from metabolic-constructive processes, i.e., the hierarchical decoupling of its electro-chemical activity, the normative regulation of sensorimotor interaction is underdetermined by basic material and energetic needs. The essential upshot of this relative independence is that the stability of an autonomous cognitive structure largely depends on the electro-chemical activity of the nervous system as well as the way that this structure is coupled to sensorimotor cycles.[9] Only an agent that is capable of regulating its sensorimotor cycles in this non-metabolic manner can be characterized by a form of *cognitive agency*. In sum, following on from Barandiaran and Moreno, we can define cognitive interaction as follows:

> Cognition is the regulated sensorimotor coupling between a cognitive agent and its environment, where the regulation is aimed at aspects of the coupling itself so that it constitutes an emergent autonomous organization in the domains of internal and relational dynamics, without destroying in the process the agency of that agent (though the latter's scope can be augmented or reduced).

To some extent, the additional requirement of non-metabolic regulation of sensorimotor interaction cycles is already a restricted possibility even for adaptive agents, because the very mechanisms of adaptive regulation are at least partially decoupled from the metabolic-constructive processes (Barandiaran and Moreno 2008). This is not a problem, however, as we would expect the process of cognition to be somehow prefigured in the process of adaptive behavior. Nevertheless, the behavioral domain of adaptive agents is severely limited because the regulatory goals are largely determined by metabolic needs, rather than by the activity that is generated via sensorimotor interaction and within the adaptive mechanism itself. Cognition, on the other hand, is based on an almost open-ended domain of potential behavior. It only becomes possible when the bulk of adaptive mechanisms are hierarchically decoupled from the rest of the living body in such a way that novel autonomous structures can arise via recurrent dynamics (cf. Barandiaran and Moreno 2006: 180). Once the requirements for *cognitive agency* are in place it is possible that the continuation of certain patterns of sensorimotor interaction become goals in themselves, for example due to the autonomous dynamic structures which they induce in neural activity. Moreover, these patterns can involve coordination with another agent in multi-agent system. Thus, only a cognitive agent can give rise to a social domain that is defined by its own specific normativity.

4.2 Sociality: From participatory sense-making to social cognition

Now that we have outlined the enactive account of cognition, can we say something more specific about *social cognition*? It should be clear by now that we need to specify the conditions of emergence for social norms in a cognitive domain, and that these norms must be related to the other agent as a foreign locus of goal-directed behavior. But what is the precise role of the other agent during a social interaction? De Jaegher and Di Paolo (2007: 492) rightly insist that

> if the autonomy of one of the interactors were destroyed, the process would reduce to the cognitive engagement of the remaining agent with his non-social world. The 'other' would simply become a tool, an object, or a problem for his individual cognition (such a situation would epitomise what we have diagnosed traditional perspectives on social cognition as suffering from: namely, the lack of a properly social level).

It is certainly the case that the other agent must remain autonomous for an interaction to be characterized as social. The question that remains, however, is whether a cognitive interaction between two or more individuals in a multi-agent system is also a sufficient criterion. What is needed is a notion of sociality that not only excludes interactions that *destroy* the autonomy of the other, but also exclude those situations in which the other is simply *encountered* as a mere tool, object, or problem to be solved by an individual's cognitive ability (if the other appears as something to be encountered at all). In order to ground this second distinction, i.e., the exclusion of situations in which a cognitive agent makes sense of the other agent as a non-agential being, the notion of a multi-agent system of cognitive agents is not specific enough. There are situations in which cognitive agents can interact (such that all of De Jaegher and Di Paolo's requirements are fulfilled), but in which the other agent is simply treated as part of the physical environment. A well known example would be the cognitive domain of a severely autistic person who is embedded within the social world of others, but who does not perceive others as such. In such cases there are certainly mutual interactions between cognitive agents, and these interactions can give rise to autonomous structures that enable and constrain individual behavior (multi-agent systems), but there is no sociality in the joint sense-making. Such *participatory sense-making* can be achieved by adaptive agents in a multi-agent system, and makes no special use of specifically other-related normativity made available by a cognitive domain.

A nice illustration of participatory sense-making without social cognition is provided by the psychological experiment by Auvray et al. (2009), which we described in the previous section. In this case the human participants constitute an autonomous interaction process, but without actually being able to meaningfully differentiate between the socially contingent and non-contingent situations. What this example demonstrates is that it is not sufficient for two cognitive agents to give rise to an autonomous interaction process if they are to break out of their individual-centered cognitive domains. While the behavior of the participants is, unbeknownst to them, guided by the global dynamics of the interaction process to an appropriate solution to the given task, their individual sense-making remains qualitatively unaffected with respect to its solitary point of reference. It is practically impossible for individuals to distinguish between the movements of the other participant and her copy, even though they are 'collectively' solving the task due to the autonomous dynamics of the multi-agent system. Accordingly, this experiment demonstrates that mutual interaction between cognitive agents in a multi-agent system is a necessary but not sufficient condition for the constitution of social significance.

Since we have argued that it is regulation of structural coupling which is constitutive of the qualitative aspect of sense-making activity (i.e., adaptive agency), we need to take a closer look at this regulative aspect. What kind of regulation could be characteristic of a social interaction such that it attains meaning as a social event for the agent? When De Jaegher and Di Paolo (2007: 497) first introduced the notion of participatory sense-making, they provided the following description:

> If regulation of social coupling takes place through coordination of movements, and if movements — including utterances — are the tools of sense-making, then our proposal is: social agents can coordinate their sense-making in social encounters. [...] This is what we call *participatory sense-making*: the coordination of intentional activity in interaction, whereby individual sense-making processes are affected and new domains of social sense-making can be generated that were not available to each individual on her own.

This emphasis on the coordination of behaviors in social interaction is a good starting point. However, as Gallagher (2009) has pointed out, we must be careful to differentiate between two distinct kinds of inter-individual situations which are being conflated in this first attempt: participatory sense-making, which is a more general term, and social cognition, which is a specific form. For the first case, there are interactions in a multi-agent system, whereby the actions of individuals can mutually enable and constrain each directly, as well as indirectly (because of the autonomous interaction process). Nevertheless, even though new and otherwise unattainable domains of sense-making can be opened up in this mutually interactive manner, thereby establishing participatory forms of sense-making, they do not necessarily involve any sense of the other agent *as such*. This is the case, for instance, in bacterial colonies, ecosystems, and even in much of our globalized culture. We can only buy a book online because we are embedded in an extensive multi-agent system, but all the underlying coordination and interaction is actually anonymous and hidden from view. An individual's interaction with a shopping website is not a social experience. The enaction of social quality in relation to others requires a special form of participatory sense-making, namely social cognition: regulated sensorimotor coordination whereby the other is recognized as such.

4.3 Socio-cognitive interaction: Interacting with others who are recognized as such

Now we can put all the previous operational definitions together in order to formulate a precise question about the origins of social cognition: How can we explain the emergence of social cognition in terms of participatory sense-making taking place in a multi-agent system of cognitive agents? Many researchers have noted that human newborns appear to exhibit some form of primary intersubjectivity already, such as in cases of neonate imitation (e.g., Trevarthen and Reddy 2007; Gallagher 2005). To be sure, it could be argued that this kind of interaction is more closely related to participatory sense-making in general than to full-blown social cognition, especially since the former can already involve complex forms of coordination without presupposing any additional requirements. The crucial difference between these two forms of mutual interaction, however, is that only the latter is related to the presence of the other in relation to the unfolding of the coordinated interaction.

How can the other's presence in our perceptual world come about? It must be based on a certain type of sensitivity to social contingency, and the enactive approach proposes that this is an interactive property (e.g., De Jaegher et al. 2010; Gallagher 2008; De Jaegher 2009). However, a cognitive agent's sensitivity to social contingency is only a necessary but not sufficient condition for attributing social cognition. This is because it is possible to demonstrate that movement which appears to an external observer as being regulated in relation to social sensitivity can instead be an emergent outcome of the autonomous interaction process between two or more coupled dynamical systems (e.g., Froese and Di Paolo 2008). We have seen an example of this in our discussion of the experiments in perceptual crossing, where stability of the interaction process was the deciding factor outside of the awareness of the individuals (Auvray et al. 2009). What is additionally required is a corresponding regulation of the cognitive agent's actions in relation to an other-directed normativity which is specifically related to that sensitivity.

At this point we meet with the infamous 'problem of other minds': in what way does the cognitive agent have to regulate its actions in relation to its sensitivity to the contingency of its interaction such that it makes sense of the other's contingent responses as belonging to another agent as such? This remains one of the outstanding problems of the enactive account of social cognition and we don't pretend to fully resolve it here. Nevertheless, we can still sharpen our intuitions by considering a case study. For instance, De Jaegher and Di Paolo (2008), drawing on Fogel (1993), provide an insightful description of what we could consider as a paradigmatic social action: the act of giving. Fogel describes a filmed session between a 1-year-old baby and his mother, in which the infant extends his arms with an object, and keeps them relatively stationary, only to gently release the object as the mother's hand takes hold of it. From this description it is already evident that the act of giving has an essentially different goal structure from individual-centered cognitive engagements. In essence, in order for the social action to be completed successfully, *it requires acceptance from the other agent*. In a more recent paper Di Paolo (2009c: 59–60; emphasis added) comments:

> Assuming for a moment that the infant is the initiator of the act, we realise that *he must create an opening by his action that may only be completed by the action of the mother*. The giving involves more than orientation of the mother's sense-making; it involves a request for her not only to orient towards the new situation, but also to create an activity that will bring the act to completion. In other words: to take up the invitation for an intention to be shared. […] an invitation to participate is experienced as a request to create an appropriate closure of a sense-making

> activity that was not originally hers. To accept this request is to produce the 'other half of the act' bringing it to a successful completion.

The regulation involved in social interaction between cognitive agents is indeed of a special kind: one cognitive agent's regulation of interaction creates an opening for an action that can only be realized through the complementary regulation of interaction by another. In other words, social interaction between cognitive agents is realized by the coordination of regulation of mutual interaction whereby the success of regulation essentially depends on appropriate coordination. In order to distinguish this particular kind of interaction from the notion of 'social interaction' more broadly conceived (e.g., any multi-agent system), as well as to distinguish it from the traditional conception of 'social cognition' (which takes as paradigmatic the case where one agent perceives another agent in a unidirectional manner), we propose to introduce the concept of *socio-cognitive interaction*. More precisely:

> Socio-cognitive interaction is the co-regulated sensorimotor coupling between at least two cognitive agents, whereby the regulation of each agent is aimed at aspects of the mutual coupling itself such that:
> 1. A new autonomous organization emerges from the interaction process spanning at least two internal and a shared relational domain of dynamics, and
> 2. The cognitive agency of at least two of the individuals is not destroyed in the process (though their scope can be augmented or reduced), and
> 3. A cognitive agent's regulation of sensorimotor coupling is complemented by the coordinated regulation of at least one other cognitive agent.

This operational definition of socio-cognitive interaction builds on all of the concepts that we have introduced so far. Criterion (1) largely remains the same, except that it has been adjusted so that the nervous-system-based regulation of sensorimotor interaction cycles now involves at least two cognitive agents. Criterion (2) might appear superfluous because of criterion (3), but it allows us to exclude marginal cases (e.g., the final act of submission of a gazelle which has fallen prey to a lion). It is criterion (3) that does most of the work: for an action to be social it has to be a joint action. Of course, this does not say anything about the way in which this joint effort is actually accomplished. It could involve roles that are relatively synchronous in their realization (e.g., some forms of dancing), or they could be complementary in an asymmetrical manner (e.g., the act of giving, which must involve offering and accepting). The essential factor is that the unfolding of the sensorimotor interaction is *co-regulated*, because it is this interactively coordinated regulation of interaction that imbues the situation with a social quality (Froese 2009: 69–70; De Jaegher et al. 2010).

We can draw experimental predictions from this definition. We know that when cognitive agents mutually enable and constrain their sense-making activities in a multi-agent system, they can open up behavioral domains that would have otherwise remained inaccessible to the individual agents. But without co-dependent regulation we do not expect there to be any social phenomenology. Indeed, this hypothesis is nicely supported by the psychological experiment conducted by Auvray et al. (2009): the relative stability and instability of the mutual interaction process causes the participants to collectively succeed at a task which they are individually incapable of solving, but the participants do not report any social qualities (and individually fail to recognize the other above chance level). According to our proposed definition of social cognition this is to be expected because the participants are essentially given an individual-based task (i.e., click when *you* find the other). For a qualitative change to happen, on the other hand, the task should be changed such that an intended activity of one participant can only become realized by the coordinated activity of the other. For example, the experimental task can be modified so that the participants are required to interact so as to agree on a common direction of movement (left or right), and then continue interacting while trying to cover as much distance as possible. It has been demonstrated with modeling experiments that this task modification can lead to novel behaviors (e.g., Froese and Di Paolo 2011), but it is still in need of phenomenological verification in actual psychological experiments. Since the agreement on a common trajectory requires co-regulated interaction, we predict that successful coordination will result in an experienced sense of sociality.

Finally, we can speculate that a solution to the problem of other minds may be found by considering the way in which an autonomous interaction process can entangle cognitive agents with each other. This is because we have argued that the social quality of an action is essentially dependent on the manner in which it is completed. Accordingly, the action does not necessarily have to be intended as a social gesture; it suffices if it happens to be completed as one. De Jaegher and Di Paolo (2008), for example, have suggested that when we remove the assumption that the infant in the case study intentionally originated the act of giving, we open up new interpretive possibilities. In that case "a certain movement extending the object in the direction of the mother, without yet intending to give it, may now be opportunistically invested with a novel meaning through joint sense-making. Latent intentions become crystallised through the joint activity so that not only the completion of the act is achieved together, *but also its initiation*" (Di Paolo 2009c: 60 emphasis added). If we accept the idea that it is possible to retroactively reinvest meaning into a previously lived situation in relation to its outcome, then we can also imagine the emergence of a social event out of two non-social acts which just happened to complement each other in the right kind of way (e.g., two kids happen to stumble into each other by accident and start to play fight).

Moreover, these kinds of encounters might happen more often than expected because an autonomous interaction process can sustain itself even despite the intentions of the interacting individuals. Social cognition could then be an outcome of the desire to disentangle oneself from one foreign locus of influence (the interaction process) by joint coordination with another foreign locus (the other agent). This is nicely illustrated by De Jaegher and Di Paolo's corridor example, where the interactive situation happens to constrain the individuals' movements so that they continue to mirror each other's actions and thereby repeatedly block the way. A resolution of the conflict requires the individuals to become more aware of each other's actions and thus to take control of the situation together. The constraining impact of the autonomous interaction process is thereby finally overcome by being transformed into a social interaction during which it can be jointly resolved. On this view, the self-other distinction, which is so fundamental to all social cognition, might develop as a process of individuation within an integrated multi-agent system. The methodological individualism of the mainstream is thus turned on its head: becoming an independent individual is essentially a socio-cultural achievement.[10]

5. The role of culture

In the previous section we have suggested that what used to be the foundational problem of social cognition, i.e., the so-called problem of other minds, can be dissolved once we realize that the 'self-other' distinction can crystallize out of the mutual interactions in a multi-agent system. In other words, it turns out that individuation and socialization are essentially two complementary sides of the same developmental coin. One crucial aspect of this proposal, which we have neglected so far, is the constitutive role of culture. There is in fact a growing interest in culture within the enactive approach (e.g., Thompson 2007, 2001; Steiner and Stewart 2009; Di Paolo 2009c), but clearly much more remains to be done. The aim of this final section is to very briefly sketch the outlines of what an enactive account of enculturated cognition could consist in, while pointing out some of the main challenges that must still be resolved.

5.1 Enculturation: Incorporating cultural heteronomy

The act of giving, as a paradigmatic social act, is widespread throughout the animal kingdom. It is most often found in the context of parenting (e.g., giving food) or courtship (e.g., making more or less arbitrary offerings). As such, it is one of the most fundamental social acts on the basis of which other forms of sociality can develop. The act itself does not presuppose much and, following De Jaegher and Di Paolo's (2008) interpretation of the infant giving an object to its mother, it is possible that none of the interactors intentionally originated the act. An arbitrary exchange can be subsequently invested with social significance when its joint completion changes the very meaning of the relationship to that of 'giver' and 'receiver'.

However, do the abstract categories of 'giver' and 'receiver' actually have any meaning in the animal kingdom apart from their use by human beings? Typically, we would expect that the roles are much more concretely situated in non-human cases of social interaction, e.g., as 'feeder' and 'fed' or 'courter' and 'courted'. The example of the object exchange between the infant and its mother thus points to the need for some additional clarification. Where do the norms which guide the mother's response to the infant's behavior come from? And how do they provide a measure for the successful completion of the act as a whole?

It is here that the socio-cultural background, in which the interactors and the unfolding interaction process are embedded, comes into play. Indeed, the mother might be moved to accept the object because that is 'what one does' when offered something by another. From her perspective, treating the gesture as the infant's attempt to 'give' the object is a 'natural' way of making sense of the situation, and this sense-making is implicitly achieved in terms of a pre-established socio-cultural practice. Moreover, this meaning, once it has been actualized in the situation, is not lost on the infant, either, who has now discovered a novel way of interacting with his mother. In other words, to characterize this example as a socio-cognitive interaction alone misses the fact that we are dealing with a process of *enculturation*. The case study of the infant-mother interaction demonstrates that human interactions can go beyond the strict confines of our definition of socio-cognitive interaction to include historical values derived from a pre-established, traditional heritage.

The appeal to a pre-existing order of shared practices indicates that our treatment of socio-cognitive interaction, which has only focused on the momentary constitution of norms during the interaction, is not sufficient to capture the whole of sociality. In particular, it is missing what is specific about human kinds of socio-cognitive interactions, namely that they always unfold within a cultural context. As Steiner and Stewart (2009) have emphasized, the latter kinds of socio-cognitive interactions can also include a form of *heteronomy*, i.e., the abiding by a heritage of pre-established social structures. Indeed, the claim that there are cultural values that guide our behavior and understanding points to a more general phenomenon, since the process of enculturation has similarly profound effects on our solitary behavior. A castaway like Robinson Crusoe does not immediately cease to behave like an Englishman when he finds himself socially isolated on a tropical island. Enculturation thus involves at least some form of *incorporation* of heteronomy (Vygotsky 1978).

Steiner and Stewart argue that only enculturated forms of interaction deserve to be called social interactions, in order to distance them from the kind of 'social' interactions that are paradigmatic of De Jaegher and Di Paolo's original approach. However, while we agree that the latter approach was too inclusive, which is why we have re-conceptualized it in terms of multi-agent interaction and turned it into a necessary but not sufficient condition for socio-cognitive interaction, Steiner and Stewart's approach is overly exclusive. They make sociality a specifically human phenomenon, thereby excluding everything from the so-called social insects to our closest primate relatives. In contrast to both of these approaches, the definition of socio-cognitive interaction that we have provided in the previous section takes up a middle ground. On the one hand, it excludes cognitive interactions that merely contingently happen to involve another agent (i.e., within a multi-agent system but without other-related values), but on the other hand it includes co-regulated interactions that are not already guided by pre-established cultural norms. Of course, this is not to deny that Steiner and Stewart are correct in insisting that there is something special about many *human* forms of sociality, including their heteronomous character, but this specificity is perhaps better captured by the notion of *culture* rather than by sociality as such.

An important problem that still remains for the enactive approach is to explain how an agent capable of socio-cognitive interaction is turned into one capable of socio-cultural interaction by being shaped by 'external' cultural values. How can we account for the incorporation of heteronomous norms? How does common sense arise out of participatory sense-making? The details of this developmental process still need to be worked out, perhaps by striking up new interdisciplinary collaborations. But the key concepts of the enactive approach already offer us a clue. After all, the autonomy of the interaction process, when viewed from the perspective of the interacting agents, also is a form of heteronomy that has its own intrinsic teleology (Torrance and Froese 2011). Of course, future work will need to determine more precisely what is special about the heteronomy of human culture. In particular, how is it possible that the behavior of an isolated individual automatically adheres to cultural norms even when others are not immediately present? But even here we should be able to approach this problem from the perspective of socio-cognitive interaction, especially social learning. If we want to know how culture can continue to shape our behavior even outside of an immediate social context, then we first need to better understand how an agent involved in a socio-cognitive interaction, faced with the heteronomy of another agent and the heteronomy of the interaction process itself, can undergo a change in behavior that we would call learning. There is also the question of pedagogy which needs to be addressed. One case of socio-cultural interaction that especially deserves further consideration in this regard is the acquisition of language. Some of the recent work on the origins of 'distributed language' could be informative in this respect (e.g., Cowley 2006), and it would be of mutual interest if future research compares the central ideas of that work with the conceptual framework of the enactive approach.

5.2 Life and mind: Biology or culture?

A final question to consider is whether the constitutive impact of cultural values is not a problem for the enactive approach. Do we not have to provide a biological foundation for these values? Yes and no. Yes, in the sense that these values can only exist for certain kinds of sense-making agents, and these agents are biological in that they are alive (autonomous and adaptive). No, in the sense that this is not a reduction of cultural values to their biological conditions of possibility; the socio-cultural domain retains its own relatively independent autonomy. As such, the emergence of the heteronomy of culture is the appearance of another discontinuity in the system of discontinuities which constitutes life, mind, and sociality. More specifically, a coherence of discourse is preserved because the heteronomy of culture turns out to be mutually interdependent with the heteronomy of sociality, and the same conceptual framework of autonomy that forms the foundation of the enactive approach is applicable to both.

It is already clear that, like the previous transitions along the 'life-mind continuity', a cognitive agent's entrance into a cultural domain is both enabling and constraining. It is constraining because taking part in shared practices requires the alignment of an individual's autonomy with a pre-established normativity. But despite this constraining, or rather because of it, there is also an expansion of possibilities. A good example of this is play, the freedom of which lies in a players' capability to create new meaningful constraints by which it can steer its sense-making activity and set new laws for itself and others to follow (Di Paolo et al. 2011). Moreover, by inaugurating a historical trace of shared individual and social practices that can go beyond an individual's lifetime, cultural interaction provides the foundation for cumulatively building on previous more or less viable ways of living. This is important because every increase of autonomy also has the effect of an increase in arbitrariness, which tradition helps us to fill in a meaningful way.

Finally, it should be emphasized again that these considerations of socio-cultural cognition are nothing but preliminary remarks to stimulate further debate. Much more needs to be said about the emergence of language, writing, and other modern technology (see Stewart 2011). But it is also important that the enactive approach does not fall into the trap of reinventing the wheel. For instance, the cognitive impact of socio-cultural practices and technological objects is already being systematically investigated from a perspective of anthropology in a way that closely matches the interests of the enactive approach to social cognition (e.g., Hutchins 1995). It is essential to further strengthen these novel inter-disciplinary links. To be sure, cognitive anthropology is sometimes mentioned as one of the core disciplines of the cognitive sciences, but its actual role has usually been marginalized (i.e., it has turned into the 'missing discipline', see Boden 2006: 515–543). This is surely not surprising considering the cognitivist methodological individualism that still prevails in the mainstream, and which attempts to reduce all socio-cultural factors into internal mental representations.[11] In contrast, the enactive approach assigns a certain amount of autonomy to socio-cultural processes while firmly linking them to biological individuals, and it combines structural and phenomenological research into one method. It is therefore likely that the enactive approach will incorporate anthropology as another one of its core disciplines.

6. Conclusion

We have begun this paper with a consideration of the scope of the enactive approach as represented in Figure 1. We have argued that the more specialized phenomena (inner layers) depend necessarily (and not just historically, i.e., evolutionarily and developmentally) on the existence of all of the former, more inclusive phenomena (outer layers). At the same time we have made sure to emphasize that even though every new domain emerges on the basis of activity in the preceding domains, it cannot be reduced to that enabling activity and, moreover, it can alter the conditions of realization of pre-existing domains. This operational asymmetry between domains is what the recurring concept of autonomy provides. This concept is also what guarantees that we are actually dealing with a non-reductive life-mind continuity, rather than a progression of heuristics that could be collapsed into a purely physical level on the basis of a more advanced science. In this manner we have traced the current state of the theoretical framework of the enactive approach from cell to society, from cellular biology to cultural anthropology.

To repeat, we should not misunderstand the operational asymmetry between domains as prescribing a one-sided interaction only. On the contrary, once the different domains of activity have been established for an agent, their relationship is not one of hierarchical dependence, but rather of multiple interdependences. For any agent it is possible (and likely) that its activities in the different domains all mutually constrain and enable each other in various non-trivial ways. Thus, even cultural norms can be re-inscribed back into the normativity operative on the metabolic level (Di Paolo 2009c). Accordingly, we can identify multiple yet integrated, interdependent, mutually enabling and constraining autonomous systems within and across different domains. Working out how these multiple interdependencies precisely operate, and how they combine to bring forth coherent forms of human agency, including an individual perspective, is one of the most important research problems for enactive cognitive science.

In conclusion, this paper has demonstrated that the enactive approach has the potential to constitute one systematic theoretical framework that retains its conceptual continuity from life to mind and from cell to society. This framework is not complete by any means, although it is beginning to form a coherent research program. To facilitate this process we have offered explicit definitions of the key concepts so that they can be debated and improved. In so doing this paper has also provided an advanced entry point into the ongoing debates of the enactive approach, but without presupposing detailed knowledge of the primary literature (which admittedly can sometimes be quite inaccessible to the general reader). To stimulate further research in this area we have tried to show how the concepts of the enactive approach can make difficult areas of traditional scientific terrain more fruitful, and at the same time we have pointed out some important omissions within the enactive approach which present exciting opportunities for further developments.

Acknowledgements

* We would like to thank the participants of the Life and Mind group for their many helpful discussions, and especially Stephen Cowley for his extensive comments on an earlier version of this article. We would also like to acknowledge the encouraging comments made by Marcelo Dascal, Natika Newton, and one anonymous reviewer. Froese was supported by a research grant of the Japanese Society for the Promotion of Science.

Notes

1. In order to avoid confusion it is important to emphasize right from the start that there are other approaches in the cognitive sciences that have begun to use the label 'enaction' after Francisco Varela introduced the term following Jerome Bruner. For example, there is Hutto's (2005) 'radical enactivism', Noë's (2004) enactive approach to perception, as well as the enactive account of mental representations by Ellis and Newton (Ellis and Newton 2010; see also Ellis 2006 and Newton 2004). Most criticism of 'enactivism' is actually targeted specifically at Noë's version, which has some differences to the enactive framework pursued here (see Thompson 2005). More recently, Noë (2009) seems to have aligned his position more closely to the approach discussed here. Nevertheless, further research is still needed in order to determine what (if any) the essential differences between these different 'enactive' approaches are (Kiverstein and Clark 2009).

2. It is important not to be tempted to reify this notion of identity, especially because an autonomous system's identity cannot be localized as a particular entity. Nevertheless, despite this 'emptiness' it still has operational efficacy as an emergent existence, which is not nothing, but which shapes and is shaped by the system's operations. Accordingly, the notion of a precarious identity is better thought of as denoting an existence between the extremes of self and non-self, permanence and impermanence. From this perspective we can understand why the enactive paradigm has always been interested in Buddhist epistemology (e.g., Varela et al. 1991), since some strands of Buddhism also include profound insights into the possibility of a 'middle way' between the extreme poles of different dualities (Lopez Jr. 2004: 350–361).

3. Note that the notion of sense-making could serve to formulate a partial response to the 'hard problem' of consciousness because it is supposed to account for the lived quality of being-there, i.e., that there is 'something it is like to be' that system. Of course, a full response would need more unpacking, including a deeper appreciation of the first-person perspective (e.g., Hanna and Thompson 2003). Still, what should be clear already is that in this respect the enactive approach differs significantly from a mere sensorimotor approach: the former begins with an account of meaningful situatedness in terms of the enacted world as a totality, whereas the latter is only concerned with establishing why there is a differentiation in perceptual quality according to sensorimotor contingencies. It is doubtful, however, whether the concept of worldhood can be recovered from this latter position because a mere summation of distinct qualities does not by itself constitute a meaningful totality. Of course, the enactive approach must still explain how such a totality, once brought into existence, could become differentiated. One possibility of achieving this is with the concept of adaptivity, which we will introduce next in the next subsection.

4. Note that this concept of adaptivity, as a type of regulatory mechanism, must be clearly distinguished from the more general biological notion of 'adaptedness'. This latter notion is typically used to indicate all viable behavior that has evolutionary origins and contributes to reproductive success. It is therefore an observer-relative notion that has no operational counterpart within the observed organism. Adaptivity, on the other hand, refers to the activity taking place within the organism as it compensates perturbations.

5. Here we have another crucial difference between the enactive approach and the sensorimotor approach: the former attempts to provide operational criteria to distinguish between mere physical change (e.g., your hair moving in the wind), living (e.g., your body regulating internal temperature), and behavior or action (e.g., walking home). Moreover, both living and action are forms of sense-making, so they are inherently meaningful, with their lived quality depending on the particular form of regulation. The sensorimotor approach, on the other hand, lacks a proper definition of action, even despite its insistence on the role of 'action in perception' (e.g., Noë 2004). This is a significant shortcoming because this insistence on the role of embodied action is what essentially distinguishes it from Gibson's ecological sensorimotor approach to perception (Mossio and Taraborelli 2008).

6. De Jaegher and Di Paolo's definition reads: "Social interaction is the regulated coupling between at least two autonomous agents, where the regulation is aimed at aspects of the coupling itself so that it constitutes an emergent autonomous organization in the domain of relational dynamics, without destroying in the process the autonomy of the agents involved (though the latter's scope can be augmented or reduced)" (2007: 493). We will consider this definition more fully later on.

7. The 'downward' effect of the emergent interaction process is no mere theoretical speculation. It is possible, for example, to modify Auvray et al.'s (2009) experimental setup such that the 'global' and 'local' goal structures are in tension with each other, and to investigate the ensuing dynamics in a detailed manner by means of evolutionary robotics modeling experiments (Froese and Di Paolo 2010).

8. The case of predator–prey interaction deserves some further thought. If the predator perceives the prey as another agent then there is an other-related value, but it is directly connected to self-related (metabolic) values (i.e., the other as food). If the predator is successful the autonomy of the prey is lost ('other' becomes 'self'), but it is interesting to consider what happens before that, during the pursuit. Some aspects of the dynamical coupling between predator and prey, at least at some points in time, may fulfill the requirements we have set for bona fide interactions. The question is empirical.

9. The *form* of these autonomous structures is relatively independent from their neuro-behavioral realization in a particular concrete situation. To be sure, their actual expression and meaning will be largely dependent on the cognitive agent's individual circumstances, but their specific form, as a general systemic property, can be instantiated by other cognitive agents as well. Moreover, there is no need for any reflective consciousness in order for their autonomous dynamics to have a meaningful effect. Accordingly, it appears that there is an opportunity for incorporating a Jungian psychology of the unconscious into the enactive approach to the cognitive sciences, especially for his concept of the 'archetype' (see Jung 1972).

10. The idea that the presence of a relatively independent 'self' is not a pre-given starting point, as assumed by methodological individualism, but rather the result of a dialectic between an individual and its social milieu, is certainly not new. This creates new openings for collaboration between disparate disciplines. For instance, there is a strong potential for incorporating some of the insights of Vygotskyan psychology (see Vygotsky 1978) into the enactive approach to the cognitive sciences. Similarly, there is an opportunity to establish better links with the tradition of phenomenological psychology, which has already explored the topic of social individuation (e.g., Merleau-Ponty 1964).

11. In fact, the opposite approach is likely to be more productive, namely to account for the idea of 'internal mental representation' in terms of socio-cultural factors.

References

Auvray, M., Lenay, C., and Stewart, J. 2009. "Perceptual interactions in a minimalist virtual environment". *New Ideas in Psychology* 27(1): 32–47.

Barandiaran, X., Di Paolo, E.A., and Rohde, M. 2009. "Defining agency: Individuality, normativity, asymmetry, and spatio-temporality in action". *Adaptive Behavior* 17(5): 367–386.

Barandiaran, X. and Moreno, A. 2006. "On what makes certain dynamical systems cognitive: A minimally cognitive organization program". *Adaptive Behavior* 14(2): 171–185.

Barandiaran, X. and Moreno, A. 2008. "Adaptivity: From metabolism to behavior". *Adaptive Behavior* 16(5): 325–344.

Boden, M.A. 2006. *Mind as Machine: A History of Cognitive Science*, vol. 1. Oxford: Oxford University Press.

Bottineau, D. 2011. "Language and enaction". In J. Stewart, O. Gapenne, and E.A. Di Paolo (eds), *Enaction: Towards a New Paradigm for Cognitive Science*. Cambridge, MA: The MIT Press, 267–306.

Brooks, R.A.1991. "Intelligence without representation". *Artificial Intelligence* 47(1–3): 139–160.

Cowley, S.J. 2006. "Distributed language: Biomechanics, functions and the origins of talk". In C. Lyon, C. Nehaniv, and A. Cangelosi (eds), *The Emergence and Evolution of Linguistic Communication*. London: Springer, 105–129.

Di Paolo, E.A. 2003. "Organismically-inspired robotics: homeostatic adaptation and teleology beyond the closed sensorimotor loop". In K. Murase and T. Asakura (eds.), *Dynamical Systems Approach to Embodiment and Sociality*. Adelaide, Australia: Advanced Knowledge International, 19–42.

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

Di Paolo, E.A. 2009. "Extended Life". *Topoi* 28(1): 9–21.

Di Paolo, E.A. 2009b. "Editorial: The social and enactive mind". *Phenomenology and the Cognitive Sciences* 8(4): 409–415.

Di Paolo, E.A. 2009c. "Overcoming autopoiesis: An enactive detour on the way from life to society". In R. Magalhães and R. Sanchez (eds), *Autopoiesis in Organizations and Information Systems*. New York: Elsevier Science, 43–68.

Di Paolo, E.A., Rohde, M., and De Jaegher, H. 2011. "Horizons for the enactive mind: Values, social interaction, and play". In J. Stewart, O. Gapenne, and E.A. Di Paolo (eds), *Enaction: Toward a New Paradigm for Cognitive Science*. Cambridge, MA: The MIT Press, 33–87.

Ellis, R.D. 2006. "Phenomenology-friendly neuroscience: The return to Merleau-Ponty as psychologist". *Human Studies* 29(1): 33–35.

Ellis, R.D. and Newton, N. 2010. *How the Mind Uses the Brain (to move the body and image the universe)*. Chicago, IL: Carus Publishing Company

Fogel, A. 1993. *Developing Through Relationships: Origins of Communication, Self and Culture*. London: Harvester Wheatsheaf.

Froese, T. and Di Paolo, E.A. 2008. "Stability of coordination requires mutuality of interaction in a model of embodied agents". In M. Asada, J.C.T. Hallam, J.-A. Meyer, and J. Tani (eds), *From Animals to Animats 10: Proceedings of the 10th International. Conference on Simulation of Adaptive Behavior*. Berlin: Springer, 52–61.

Froese, T. and Di Paolo, E.A. 2009. "Sociality and the life-mind continuity thesis". *Phenomenology and the Cognitive Sciences* 8(4): 439–463.

Froese, T. and Di Paolo, E.A. 2010. "Modeling social interaction as perceptual crossing: An investigation into the dynamics of the interaction process". *Connection Science* 22(1): 43–68.

Froese, T. and Di Paolo, E.A. 2011. "Toward minimally social behavior: Social psychology meets evolutionary robotics". In G. Kampis, I. Karsai, and E. Szathmary (eds). Berlin: Springer, 420–427.

Froese, T. and Ziemke, T. 2009. "Enactive artificial intelligence: Investigating the systemic organization of life and mind". *Artificial Intelligence* 173(3–4): 366–500.

Gallagher, S. 2008. "Direct perception in the intersubjective context". *Consciousness and Cognition* 17(2): 535–543.

Gallagher, S. 2009. "Two problems of intersubjectivity". *Journal of Consciousness Studies* 16(6–8): 289–308.

Hanna, R. and Thompson, E. 2003. "The mind-body-body problem". *Theoria et Historia Scientarum* 7(1): 24–44.

Hutchins, E. 1995. *Cognition in the Wild*. Cambridge, MA: The MIT Press.

Hutto, D. 2005. "Knowing what? Radical versus conservative enactivism". *Phenomenology and the Cognitive Sciences* 4(4): 389–405.

De Jaegher, H. 2009. "Social understanding though direct perception? Yes, by interacting". *Consciousness and Cognition* 18(2): 535–542.

De Jaegher, H. and Di Paolo, E.A. 2007. "Participatory sense-making: An enactive approach to social cognition". *Phenomenology and the Cognitive Sciences* 6(4): 485–507.

De Jaegher, H. and Di Paolo, E.A. 2008. "Making sense in participation: An enactive approach to social cognition". In F. Morganti, A. Carassa, and G. Riva (eds), *Enacting Intersubjectivity: A Cognitive and Social Perspective on the Study of Interactions*. Amsterdam: IOS Press, 33–47.

De Jaegher, H., Di Paolo, E.A., and Gallagher, S. 2010. "Can social interaction constitute social cognition?". *Trends in Cognitive Sciences* 14(10): 441–447.

De Jaegher, H. and Froese, T. 2009. "On the role of social interaction in individual agency". *Adaptive Behavior* 17(5): 444–460.

Jonas, H. 1966 [2001]. *The Phenomenon of Life: Toward a Philosophical Biology*. Evanston, IL: Northwestern University Press.

Jung, C.G. 1972. *Four Archetypes: Mother, Rebirth, Spirit, Trickster*. Translated by R.F.C. Hull. London: Routledge & Kegan Paul.

Kant, I. 1790 [1987]. *Critique of Judgment*. Translated by W.S. Pluhar. Indianapolis, IN: Hacket Publishing Company.

Kirsh, D. 1991. "Today the earwig, tomorrow man?". *Artificial Intelligence* 47(1–3): 161–184.

Kiverstein, J. and Clark, A. 2009. "Introduction: Mind embodied, embedded, enacted: One church or many?". *Topoi* 28(1): 1–7.

Lopez, Jr., D.S. 2004. *Buddhist Scriptures*. London: Penguin Books.

Maturana, H.R. and Varela, F.J. 1987. *The Tree of Knowledge: The Biological Roots of Human Understanding*. Boston: Shambhala Publications.

Merleau-Ponty, M. 1964. "The child's relations with others". In W. Cobb (trans), J.M. Edie (ed), M. Merleau-Ponty *The Primacy of Perception And Other Essays on Phenomenological Psychology, The Philosophy of Art, History and* Politics. Evanston, IL: Northwestern University Press, 96–155.

Mossio, M. and Taraborelli, D. 2008. "Action-dependent perceptual invariants: From ecological to sensorimotor approaches". *Consciousness and Cognition* 17(4): 1324–1340.

Newton, N. 2004. "The art of representation: Support for an enactive approach". *Behavioral and Brain Sciences* 27(3): 411.

Noë, A. 2004. *Action in Perception*. Cambridge, MA: The MIT Press.

Noë, A. 2009. *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*. New York, NY: Hill and Wang.

Pfeifer, R. and Scheier, C. 1999. *Understanding Intelligence*. Cambridge, MA: The MIT Press.

Steiner, P. and Stewart, J. 2009. "From autonomy to heteronomy (and back): The enaction of social life". *Phenomenology and the Cognitive Sciences* 8(4): 527–550.

Stewart, J. 1992. "Life = cognition: The epistemological and ontological significance of artificial life". In F.J. Varela and P. Bourgine (eds), *Towards a Practice of Autonomous Systems: Proceedings of the 1st European Conference on Artificial Life*. Cambridge, MA: The MIT Press, 475–483.

Stewart, J. 2011. "Foundational issues in enaction as a paradigm for cognitive science: From the origin of life to consciousness and writing". In J. Stewart, O. Gapenne, and E.A. Di Paolo (eds), *Enaction: Towards a New Paradigm for Cognitive Science*. Cambridge, MA: The MIT Press, 1–32.

Stewart, J., Gapenne, O., and Di Paolo, E.A. (eds). 2011. *Enaction: Toward a New Paradigm for Cognitive Science*, Cambridge, MA: The MIT Press.

Thompson, E. 2001. "Empathy and consciousness". *Journal of Consciousness Studies* 8(5–7): 1–32.

Thompson, E. 2004. "Life and mind: From autopoiesis to neurophenomenology. A tribute to Francisco Varela". *Phenomenology and the Cognitive Sciences* 3(4): 381–398.

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

Thompson, E. 2007. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Harvard University Press.

Thompson, E. and Varela, F.J. 2001. "Radical embodiment: Neural dynamics and consciousness". *Trends in Cognitive Sciences* 5(10): 418–425.

Torrance, S. 2005. "In search of the enactive: Introduction to special issue on enactive experience". *Phenomenology and the Cognitive Sciences* 4(4): 357–368.

Torrance, S. 2007. "Introduction to the second special issue on enactive experience". *Phenomenology and the Cognitive Sciences* 6(4): 425.

Torrance, S. and Froese, T. 2011. "An inter-enactive approach to agency: Phenomenology, dynamics, and sociality". *Humana.Mente* 15: 21–53.

Trevarthen, C. and Reddy, V. 2007. "Consciousness in infants". In M. Velmans and S. Schneider (eds), *The Blackwell Companion to Consciousness*. Oxford: Blackwell, 41–57.

von Uexküll, J. 1934 [1957]. "A stroll through the worlds of animals and men: A picture book of invisible worlds". In C.H. Schiller (trans, ed), *Instinctive Behavior: The Development of a Modern Concept*. New York: International Universities Press, 5–80.

Varela, F.J. 1979. *Principles of Biological Autonomy*. New York: Elsevier / North Holland.

Varela, F.J. 1991. "Organism: A meshwork of selfless selves". In A.I. Tauber (ed), *Organisms and the Origins of Self*. Dordrecht, Netherlands: Kluwer Academic Publishers, 79–107.

Varela, F.J. 1995. "The re-enchantment of the concrete: Some biological ingredients for a nouvelle cognitive science". In L. Steels and R. Brooks (eds), *The Artificial Life Route to Artificial Intelligence*. Hove, UK: Lawrence Erlbaum Associates, 11–22.

Varela, F.J. 1997. "Patterns of life: Intertwining identity and cognition". *Brain and Cognition* 34(1): 72–87.

Varela, F.J., Maturana, H.R., and Uribe, R. 1974. "Autopoiesis: The organization of living systems, its characterization and a model". *BioSystems* 5: 187–196.

Varela, F.J, Thompson, E., and Rosch, E. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: The MIT Press.

Vygotsky, L.S. 1934 [1978]. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Weber, A. and Varela, F.J. 2002. "Life after Kant: Natural purposes and the autopoietic foundations of biological individuality". *Phenomenology and the Cognitive Sciences* 1(2): 97–125.

Authors' addresses

Tom Froese
Ikegami Laboratory
The University of Tokyo
3-8-1 Komaba, Meguro-ku
Tokyo 153-8902
Japan

t.froese@gmail.com

Ezequiel A. Di Paolo
Ikerbasque, Basque Science Foundation
Department of Logic and Philosophy of Science
University of the Basque Country (UPV/EHU)
Av de Tolosa, 70
20080 San Sebastian
Spain

ezequiel@sussex.ac.uk

About the authors

Tom Froese received his D.Phil. in Cognitive Science from the University of Sussex in 2009. The focus of his doctoral research was the relationship between life, mind, and sociality, which he explored using a variety of methods, including evolutionary robotics and philosophical phenomenology. More recently he has been interested in the cross-section of phenomenological and systemic approaches in the study of consciousness. In 2010 he was a post-doctoral researcher at the Sackler Centre for Consciousness Science in Brighton, and now he is a JSPS post-doctoral researcher in the Ikegami Laboratory of the University of Tokyo.

As of January 2010, **Ezequiel Di Paolo** has been appointed Ikerbasque Research Professor at the Department of Logic and Philosophy of Science of the University of the Basque Country (UPV/EHU) in San Sebastian. Before that appointment, he was a Reader in Evolutionary and Adaptive Systems at the University of Sussex and co-director of the Evolutionary and Adaptive Systems (EASy) MSc programme. He continues to be a Visiting Researcher of the Centre for Computational Neuroscience and Robotics (CCNR) and the Centre for Research in Cognitive Science (COGS) at the University of Sussex. He is Editor-in-Chief of the journal *Adaptive Behavior*, and is the author of over 100 peer-reviewed publications. His interests include: adaptive behavior in natural and artificial systems, biological modeling, evolutionary robotics, embodied cognition, philosophy of mind, and philosophy of biology.