EXPLANATORY SCHEMES FOR SOCIAL COGNITION – A MINIMALIST INTERACTION-BASED APPROACH

Charles Lenay
Université de Technologie de Compiègne
charles.lenay@utc.fr

ABSTRACT: Should social interactions be explained by individual cognitive capacities? - or conversely, can these capacities be explained by social interactions? In order to answer this question we adopt a pragmatic perspective for which the cognitive activity cannot be understood as detached from the social milieu, which precedes the relations which it makes possible. In addition, if one takes the view that inter-individual interactions participate in the constitution of phenomena of social cognition, one must provide explanatory schemes making it possible to *account* for abilities such as the recognition of other subjects, the perception of their intentions or the imitation of their facial expressions, rather than presupposing them as given from the start. We contribute here to the search for such schemes by proposing a minimalist experimental paradigm in which one can observe the genesis or the functioning of such abilities. These observational situations, which are extremely simple, should then enable a precise discussion of the mechanisms at work by proposing a way of disentangling what explains what in the relation between individual abilities and social dynamics. In the framework of an enactive approach we will then be able to demonstrate the role of the body-object in the collective organisation of interactions and in the reconfiguration of individual structures.

1. Introduction

There is a general debate in current research in cognitive science between approaches which base their explanations solely on internal individual abilities (either by a system of symbolic representations or by neuronal structures and activities), and a multitude of approaches which, in a pragmatic perspective, seek to take into account the bodily, technical and social conditions of cognitive phenomena. The debate takes a particularly acute form when research bears on the nature and mechanisms of "social cognition", i.e. abilities such as the recognition of other subjects, understanding their intentions, the imitation of their behaviour and their facial expressions, joint attention, empathy, learning, language, and so on. Should one start from individual abilities to explain social phenomena, or should one start from social interactions in order to explain individual abilities?

On the one hand, in the form of classical methodological individualism in psychology, philosophy of mind or neuroscience, one considers that the cognition at the origin of social interactions is constituted solely by abilities possessed by the individual organism. Whether it be by inference based on internal representations (Theory Theory) or by the internal simulation of observed behaviours (Simulation Theory), it is a question of understanding how a given individual can guess the hidden intentions of another subject. Thus, for these approaches, the behaviour of intentional subjects is to be explained by internal mental states which are hidden to an external observer. Henceforth, the basis of social relations resides in the "mind-reading", i.e. the art of guessing the mental states of other subjects (Baron-Cohen 1997). It is only afterwards that joint attention and the construction of a social world can be achieved (Tomasello et al. 2005).

On the alternative view that we defend here, it is rather a question of taking into account social structures in the explanation of individual cognition. Thus, in this pragmatic framework, the so-called "4E" approaches posit that cognition is Embodied, Embedded, Extended and Enactive, i.e. that cognition cannot be explained without taking into account the concrete conditions of human action. The accent is then put on three related features:

i) There are always material, bodily and technical conditions for social relations. Interactions unfold in a context which is concretely structured and which defines the possibilities for the various actors to perceive, to act and to interact.

ii) Social relations are generally reciprocal, and engage the individuals concerned in a dynamic of interactions. The situation of an uncommitted external observer is secondary and must actually be artificially constructed.

iii) Cognitive or perceptual activities are accomplished by means of bodily engagement and sensori-motor coupling with the environment. They can thus potentially be perceived at the level of the activity itself (and do not have to be inferred indirectly via an internal reconstruction involving inference or simulation) (Gallagher 2008).

In this framework, it can be held that the social environment is not only a context for individual action, but that inter-individual interactions are a necessary condition for achieving social cognition or even, in a yet stronger version, that these interactions actually constitute such cognition (De Jaegher et al. 2010). A central role of this sort is attributed to inter-personal interactions in developmental science (Trevarthen 1979, 1993, Reddy 2008), in social neuroscience (Dumas 2011, 2014, Schilbach et al. 2013), in a dynamic systems approach to cognition (Thelen and Smith 1996), and certain enactivist approaches (De Jaegher et al. 2007, Froese 2016).

Already, in the pragmatic perspective developed by John Dewey cognition and perception should not be separated from the possibility of acting. What Dewey called the "reflex arc" is accomplished at different levels of "sensori-motor coordination". He writes for example : “we begin in perception not with a sensory stimulus, but with a sensorimotor coordination … it is the movement which is primary, and the sensation which is secondary, the movement of body, head and eye muscles determining the quality of what is experienced” (Dewey 1896, 358). Similarly in conception of active perception and in particular for enactive approaches, perception does not result from an internal treatment of sensory information, but is produced in the dynamics of the coupling between an autonomous agent and its environment (Varela 1991). By this activity, the agent brings forth, or “enacts”, the content of its experience which corresponds to the invariants of the sensorimotor dynamics (Kevin O’Reagan and A. Noe 2001). Ezequiel Di Paolo and Hanne De Jaegher (2016) therefore speak of “sense making” achieved in the regulation of the coupling between an autonomous adaptive system and its environment. The perception of an object corresponds to mastery of sensorimotor coupling with that object (and not in a mental representation distinct from the represented objects). The intentional content of a perception is constructed in this coupling. If we now consider a social perception as the recognition of another subject, of their intentions or expressions, such perception should be explained by a form of coupling with another subject, their intentions and expressions. Its intentional content should thus be constructed in a social coupling.

If one admits a sensorimotor approach to cognitive activity, we can understand that the encounter between the coupling activities of several organisms can form relational patterns whose behaviour and history surpass what each individual brings to it (Di Paolo and De Jaegher 2016). Shaun Gallagher proposes the term “Strong Interaction: a mutually engaged co-regulated coupling between at least two autonomous agents where the co-regulation and the coupling mutually affect each other, constituting a self-sustaining organization in the domain of relational dynamics." (Gallagher 2013, 59). This co-ordination leads to the emergence of a new domain of objects, actions and meanings for the subjects engaged in it. One can then define "participatory sense-making" as “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” (De Jaegher and Di Paolo 2007, 497). Social interactions play a constitutive role in the sense that the dynamics of interaction itself produces the performance of social cognition, as for example in the dance of a couple, a gesture of exchange, collaborative work or the attention of a new-born baby for her mother in the course of a proto-conversation.

In this sort of approach, phenomena of social cognition can be explained by calling on the notion of “*intercorporeity*”, i.e. a form of concrete inter-individual

community in the situated processes of interaction, a community which precedes the constitution of individual intentions (Merleau-Ponty 1968, Gallagher 2013). For example, work in developmental science tends to demonstrate the existence, right from the earliest years of life, of a "Primary intersubjectivity", i.e. a pragmatic mutual understanding which comes about it the dynamics of the embodied emotional interactions of the proto-conversation between the infant and her caregiver (Trevarthen 1979, 1993, Trevarthen and Reddy 2007). The actions and perceptions of different organisms are harmonized in a "cross-modal sensorimotor system" (Gallagher and Meltzoff 1996); by mutual adjustment they give rise to a "common bodily intentionality that is shared across the perceiving subjects" (Gallagher 2013, 61). This inter-individual embodiment of primary intersubjectivity appears to be the substrate on which a "*secondary intersubjectivity*" can appear (from one year onwards), explaining for example the phenomena of joint attention and then *Communicative and narrative competencies* (from 2 to 4 years).

From the point of view of neurosciences, the discovery of the "Mirror Neuron System" (MNS) can be considered as providing a mechanism which explains the formation of this *intercorporeity*. Thus, these neuronal structures are activated *both* when the subject performs a certain action, *and* also when (s)he perceives the same action performed by another subject. They make it possible to understand the phenomena of imitation, in particular the unconscious imitation of movements, postures and expressions ("chameleon effect"). In the same vein, these neuronal structures seem to be involved in our capacity to share emotions by means of facial expressions and bodily postures (Gallese 2009). In this interaction-based framework, the Mirror Neuron System provides the sub-personal substrate of intercorporeity by explaining the engagement of organisms in behavioural adjustments before any conscious awareness or even distinction self and other (Gallese 2001). "It appears therefore that there is a we-centric dimension in the experience of a given emotional / affective state, and that it is underpinned by the activity of a common neural substrate." (Gallese 2009, 30). In this case, the neuronal structures do not pretend to underlie mental representations or even internal simulations; they serve rather to explain a set of practical interactions and inter-individual perceptions: "Mirror activation, on this interpretation, is not the initiation of simulation; it is part of an enactive intersubjective perception of what the other is doing" (Gallagher 2013, 64).

The work presented here is situated in this sort of interaction-based perspective, and aims at resolving a general difficulty that these approaches run up against. This difficulty concerns the initial capacities that one attributes to individuals in order to explain their engagement in meaningful social interactions. Are they capacities for social cognition that are necessary for the functioning of these interactions, or are they rather the products of such interactions? But if they are necessary, might they not also be sufficient? And if they are the product of such interactions, what is the mechanism that produces them and for what role?

In the interaction-based approaches that we have rapidly surveyed above, the fact is that one attributes substantial initial abilities to individual organisms in order to explain their participation in the interactions. In the case of "primary intersubjectivity", one admits from the outset that "innate or early developing sensory-motor capacities" include the capacities to recognize the face of another subject, to perceive their intentions and to participate in games of imitation and emotional regulation. All these capacities for what is called "social cognition" are considered as pre-given at the individual level, *before* the interactions that they allow actually occur.

But if these capacities dedicated to social cognition are necessary for setting up the interactions, may they not also be sufficient? If there is a need for specific internal structures in order to explain social cognition, is it not possible to interpret these structures in one form or another of internal individualism? These initial capacities would be sufficient to construct a system of personal individual representation or simulation, without

any need to confer on the actual occurrence inter-individual interactions any real explanatory role. This is indeed what on observes in certain interpretations of the MNS. Thus, these neuronal structures can very well be mobilized in an internalist framework to defend the "Simulation Theory": a form of mind-reading could start with a purely internal phase, independent of the environment, in which the intentions of another subject could be guessed at by imagining what one would experience if one performed oneself the behaviours that one perceives (Rizzolatti et al., 2002). This would amount to admitting that the social meaning of these capacities could be contained in purely *individual* internal structures.

If on the other hand one wishes to maintain an approach which is interaction-based, pragmatic or enactive, the social meaning of cognition should rather be deployed in the actual occurrence of inter-individual relations themselves. However, if the meaning of the individual structures which are mobilized in the interactions is only produced in these external relations, one runs the risk of a dangerous circularity: the capacities which participate in social interactions should be produced by the interactions themselves! If one wishes to avoid giving oneself in advance the individual capacities for social cognition, how can one interpret and explain the individual structures whose role is observed in inter-individual interactions?

This difficulty comes up repeatedly in the various approaches which refuse methodological individualism. One can understand that in "strong interactionnism" the capacities for the resolution of social tasks can be explained by the dynamics of the interaction present between the different subjects. But the individual capacities that are mobilized are presupposed as already established, since they are the condition for the interactions to occur. If intercorporeity is explained on the basis of pre-existing internal structures, we already have the entire explanation. The *explanans* is individual and the social is only the *explanandum*.

In the case of participatory sense making, this difficulty takes the form of a tension between the autonomy of the individual organisms and the autonomy of the collective (De Jaegher and Di Paolo 2007). One considers that the dynamics of the interaction can, as such, account for a situated social activity for each participant. But if one grants a form of autonomy to this collective dynamics, with its own properties and its own history, that means precisely that it evolves sufficiently independently of the individuals who participate in it. To the extent that the individual capacities are not affected, the emergent structures of the collective dynamics have no existence at the individual level. How then can one account for the existence of social meaning in personal sense-making activity? It seems that one is faced with an alternative: either the dynamics of the interaction modifies in a heteronomous fashion the sensorimotor activity of each organism; otherwise the dynamics of the interaction remains in a different domain from the activities that each individual pursues for himself.

The aim of this article is to defend an interaction-based approach, in a truly strong sense, in which the social participates in the formation or the modification of individual cognitive activities. For that, it is necessary to propose explanatory schemes which make it possible to account for the way in which individual capacities could, at least in part, be genuinely *produced* by the social dynamics[1]. We must face up to a complex entanglement between the collective and the individual. There is a triple difficulty.

1. To be fully rigorous, it must be possible to define first of all the *non social* capacities which make it possible for a subject to engage in a coupling with the world (for example a perceptual coupling).

[1] One finds an aim of the same order in the "Interactive brain hypothesis » proposed by Di Paolo et De Jaegher (2012); but we shall see, our proposal is substantially different..

2. One must then explain how the engagement of these couplings in a dynamic of relatively autonomous inter-individual interactions allows for the appearance of collective properties that are genuinely social.

3. And finally, it must be shown how these collective properties can account for the formation, within each of the organisms in interaction, of what will then appear to be "individual" capacities for social cognition, i.e. capacities which make it possible for the individuals to engage in a coupling whose meaning is social, and which allow for inter-individual interactions which are richer than before.

To do this, we propose a minimalist experimental method which makes it possible to clarify as precisely as possible the articulations between individual and collective components of this interaction dynamics, their causal order and their logical relation as explanans or explandum.

These studies are based on the experimental paradigm of "perceptual crossing", and have already been used and taken up in the framework of various theoretical and philosophical discussions[2]. We will present them again here, in a systematic way, in order to bring out the general principles and the explanatory schemes that they make it possible to propose. We shall see in particular the importance of bodily engagement in the dynamics of perceptual interaction for the emergence of collective forms of social cognition; the meaning of a distinction between perceiving-body and body-object for understanding the processes of imitation and more generally the individual appropriation of social dynamics; and finally the role of a link between definite properties of the body-object and the social dynamics for the construction of individual associative structures.

In the following chapter, we shall describe the origin and the principles of the experimental method (ch.2). Then in a series of three chapter we shall present successively the experimental situations which make it possible to elaborate explanatory schemes for the recognition of other subjects (ch.3), the perception of his/her intentions (ch.4), and the imitation of her facial expression (ch.5). In the discussion (ch.6), we shall come back to the methodology employed, and we shall attempt to bring out some of the principles mobilized by these modes of explanation.

2. Minimalist experimental paradigm

The background to the minimalist method that we shall present is the general research programme centred on the thesis that "Technology is Anthropologically Constitutive" (the TAC thesis) (Stiegler 1998). In the framework of embodied, embedded and extended approaches, taking into account the technological dimension – tools, machines, environments – is a way of incorporating the role of culture and social history in the constitution of individual and collective cognitive activities. There again, there are weaker and stronger versions, going from a conception of the technical environment as a simple context or mode of deployment of a cognitive activity that remains essentially internal, up to the strong idea of a constitutive role for technology which creates the possibility of cognitive or perceptual operations which would simply not exist without it (Leroi-Gourhan 1964, Goody 1977, Clark 1998, 2004).

[2] The articles marked with an asterisk in the bibliography make explicit use of the results of this minimalsit approach, whether it be for spatial perception or social cognition.

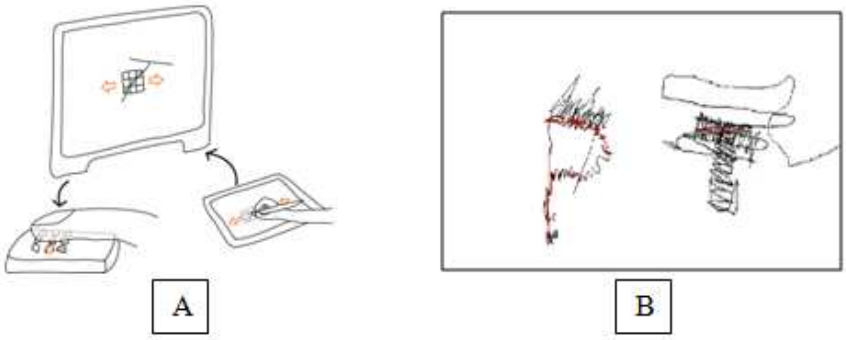

Fig. 1A. TACTOS system. The meeting between the matrix of 9 receptor fields and a form controls the activation of the 9 pins of two Braille cells. The forms recorded in the digital space of the screen are perceived by touch. **Fig. 1B**. Examples of perceptual trajectories. Here, we use just one receptor field, and a single all-or-nothing stimulation. We have coloured in red the points where the subject received a tactile stimulation. We see the perceptual trajectory when the subject recognizes a P or a T.

In order to systematically study the ways in which various technological mediations underlie specific domains of activity, one can start by examining the meaning of grasping a tool. An enactive approach is particularly useful here. Indeed, if cognition and perception arise in the coupling between an autonomous agent and its environment, the domain of possible operations will depend on the sensorimotor invariants that are available and hence on the possibilities of the agent for acting and sensing (Varela 1991). A tool that is taken up and grasped will modify the repertoire of these actions and sensations, and hence give rise to a specific field of possible operations. Here, the "technological constitution of cognitive operations" does not mean that these operations could be performed in a purely external manner on artificial machines, but rather that technological artifacts play a constitutive role by their participation in the coupling between organisms and their environment. The experimental study of this coupling consists of using the technological mediation in order to systematically control the sensory inputs and possible actions, and to observe the cognitive or perceptual operations which become possible.

We started by simplifying to the extreme the technological mediations in order to elucidate the minimal conditions necessary for the perception of spatial localization (Lenay et al. 1997, O'Regan and Noe 2001, Lenay and Steiner 2010) and for the recognition of shapes (Lenay et al. 2003). Here, it is only necessary to present briefly the technological device which was used for the recognition of shapes in a two-dimensional space, since this same device will serve as the basis for our experiments on social cognition.

The "Tactos system" we have developed makes it possible to "touch" digital shapes on the computer screen. The cursor, which the user controls with his mouse, is now replaced by a single receptor field. When this receptor field crosses the colored pixels of a shape, an all-or-nothing tactile stimulation is triggered under the index finger of the free hand. Even in this case, the users (whether blind persons or blindfolded adults) can learn to localize and recognize simple shapes. The perception is necessarily active because there is no intrinsic spatiality in the sensory input. It is thus realized essentially through a perceptual trajectory that can easily be recorded, analyzed, and modeled (Stewart and Gapenne 2004). We say that by its highly restrictive conditions, the device forces a spatial and temporal deployment of the perceptual activity.

We can now apply this minimalist method to the study of social cognition. Indeed, when the Tactos device made it possible to "touch" digital forms on the computer screen, it became evident that this space could be shared between several users. However, in order for a perceptual interaction to be possible it is necessary to add something essential: the "perceiving-body" in the

space of interaction, which corresponds here to the receptor field, must be linked to a "body-object", i.e. a shape which can be perceived by the other participants. We can then construct situations of minimalist inter-individual interactions, simplifying as much as possible the complexity of the shared context of interaction and the repertoires of actions and sensations of the participants. We can then observe the functioning of the collective coupling and the capacities for social cognition which it allows. We have applied this method to study the phenomena of the recognition of another subject, the perception of her intentions, or the imitation of facial expressions. Our hope is that the mechanisms elucidated in these artificially simplified conditions will subsequently serve as explanatory schemes accounting for the far more complex cases of ordinary daily social interactions.

We may note that studying social cognition by way of technological mediation makes it possible to propose a simple definition of what we mean by "social cognition". Since the participants are distant from each other, the conditions of their interaction must be completely specified, without ambiguity, by the information transmitted for their actions and perceptions. Thus, they share a common space of interaction if but only if the actions of one participant animate in real-time a body-object that can be perceived in the space of action of the other participant, and *vice versa*. For such a definition, simply technological, there can be forms of inter-individual interaction even if the subjects are not engaged in an activity which is explicitly co-ordinated. This makes it possible to avoid the difficulties and ambiguities that lie in store if one is looking for cognitivist definitions such as the one proposed by Schoenherr and Westra (2017, 7) : "When two or more conscious human beings mutually and knowingly affect one another's actions, they are engaged in a social interaction". Introducing notions of "consciousness" and "knowledge", right from the very definition of social interaction, makes it impossible in advance to understand how the interactions can *account* for this social consciousness or knowledge.

3. Perceptual crossing and the Recognition of Others

This study on the recognition of others will be presented relatively briefly since it has already been the object of many discussions (Auvray and Rohde 2012) and has led to complementary experiments and modelling.

Classically, in the framework of the philosophy of mind and the representationalist paradigm in cognitive science, one considers that the problem of the recognition of another subject comes down to the question of the adoption of an "intentional stance". The question is to determine the criteria and mechanisms used by the subjects in deciding to treat the perceived objects, either as simple "things" which obey a mechanical causality, or else as "intentional agents" who act on the basis of internal representations and goals. In a coherent manner, experimental studies carried out in this framework establish a strict separation between the observing subject and the scene that is observed. As behind a one-way mirror, the observer is in the very artificial position of a voyeur observing the potential other subject.

If instead we adopt an interactionist perspective, it is clear that living organisms are always already engaged in relations with others. It follows that the question of the recognition of others must fundamentally be mutual. The recognition of an intentional subject ought to take place during an interaction where the perceived subject can reciprocally recognize the observer himself as an intentional subject. We therefore wished to study these situations where two perceptual activities meet, as for example in mutual touching, looks where both subjects "catch each other's eyes", or proto-conversation between mother and infant. We designate them by the term "perceptual crossing" (Lenay et al. 2006, Auvray et al. 2009). Our hypothesis was that perceptual crossing allows the mutual recognition of the two subjects, not as the cognitive attribution of intentionality but as a direct experience, the experience of "seeing the other seeing me" (Gallagher 2014, 3).

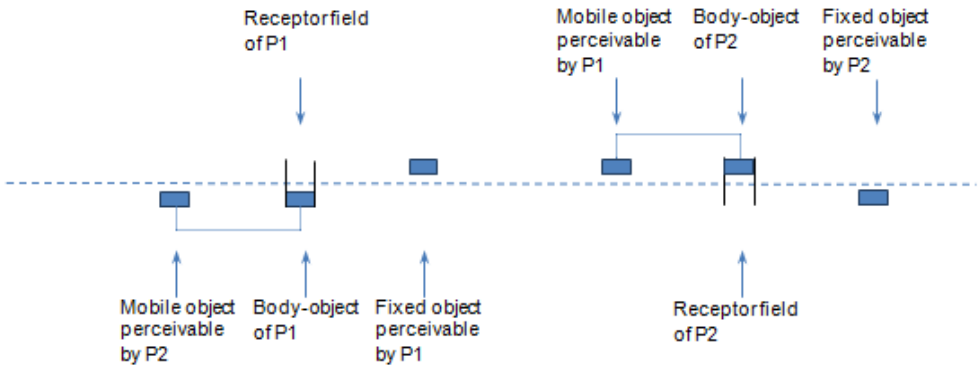

Fig. 2. The participants P1and P2 share a one-dimensional space without edges (the two ends join up to form a torus). They move their receptor field (identical with their object-body) from left to right by the computer mouse. When the receptor field of one participant encounters a black pixel, she receives a tactile stimulus on the finger of her free hand. The participants are instructed to click when they consider that they encounter their partner.

3.1. Experimental setup for perceptual crossing

In order to study this situation we have used the minimalist experimental paradigm that we have just described so as to set up an elementary form of perceptual crossing. If the perceiving-body and the body-object coincide, when the receptor field of one participant encounters the body-object of another participant, the receptor field of this second participant will necessarily encounter the body-object of the first participant. There will be perceptual crossing.

In order to facilitate precise analysis of the dynamics of perception and interaction, the space of action of the participants is reduced to a shared one-dimensional space (i.e. a straight line; left-right movements of their receptor field are commanded by the computer mouse); and the repertoire of sensory feedback is reduced to a single all-or-nothing stimulus (a tactile stimulus is delivered to a finger of the free hand) when the receptor field passes over a black pixel. The participants are blind-folded and can only interact via the experimental device. For each subject, three objects are present:

- The body-object of the other participant (which corresponds exactly to her receptor field), 4 pixels wide. When the two participants are in the same position, each receives an all-or-nothing tactile stimulation. This is the situation of "perceptual crossing".
- A fixed object (4 pixels wide). The fixed objet for the participant 1 is invisible for participant 2, and is placed in a different position.
- A moving object (4 pixels wide) that we call the "mobile lure". In order to ensure that the mobile lure would have the same richness of movement as the body-object of the other participant, but without being responsive to perceptual crossings, we attached it by a rigid virtual link to the receptor field of the partner. The mobile lure thus follows exactly, but at a constant distance, all the movements performed by the partner.

The participants are informed that there are these three objects, but they do not know that the mobile lure is attached to the receptor field of their partner. They are instructed to click when they consider that they have encountered their partner.

This experimental configuration makes it possible to test our theoretical hypothesis: even though the mobile lure and the body-object of the partner have objectively exactly the same movements, will the participants be able to distinguish them on the sole basis that the receptor field of the partner is sensitive and animated by a perceptual activity turned towards their own perceptual movements?

The results are rich and apparently contradictory (Auvray et al. 2009). On one hand, it is a success : the results for all the participants and all the sessions showed that the majority of clicks (63%) occurred when the two partners were indeed in front of each other, i.e. in a situation of perceptual crossing (cf. Fig. 3). These results seem to show that the subjects succeed in distinguishing the receptor field of their partner, which is surprising since by construction the mobile lure has exactly the same objective movement as this receptor field. Here again, the reduction of the sensory input forces a spatial and temporal deployment of the perceptual activities, which can be easily recorded for analysis. The interest of the minimalist method is that we can know everything that happened to the subjects during their interaction, that is to say all their actions and the sensory feedback they received.

Now closer analysis of the stimulations received by each partner reveals that behind this collective success there is hidden a failure at the individual level. 52% of the stimulations received come from a perceptual crossing, 33% come from encounters with the fixed object and only 15% from encounters with the mobile lure. If we now calculate the *ratio* between sensory stimuli and clicks, we find 0.33 for the fixed object, 1.26 for the perceptual crossing and 1.51 for the mobile lure (Table 1).

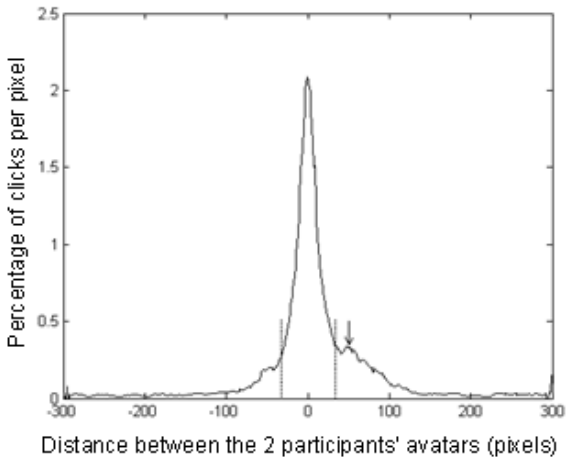

Fig. 3. If we look at the distribution of frequencies of clicks as a function of the distance between the receptor fields of the two participants we observe a very sharp peak for zero, when the participants are actually face to face, and a small peak at 50 pixels corresponding to clicks made on the mobile lure.

The relatively low ratio of 0.33 shows that that the fixed object was usually (although not always) recognized as such because the subjects almost did not often click although they encountered it frequently. However, the ratio between clicks and stimuli shows that overall the participants do not seem to distinguish between stimuli due to perceptual crossing and stimuli due to the mobile lure (1.26 vs. 1.51). The difference in clicks on the mobile lure and on the receptor field of their partner (23% vs. 66%) can be explained entirely by strategies of movement which are such that encounters with the mobile lure are much less frequent than encounters due to perceptual crossing (15% vs. 52%).

In this first experiment where the aim is to discriminate the presence of another subject, the individuals fail whereas the collective action succeeds. Individuals fail to recognize another subject in terms of a sensory input, which shows that the collective dynamics is established independently of any such recognition. The collective success is not explained by an individual capacity to recognize another subject by means of a particular sensation. The collective success of the participants is explained entirely by their capacity to find each other face-to-face, and not because they were able to recognize in the pattern of stimuli a clue allowing them to distinguish between the presence of the partner and that of the mobile lure. The only difference resides in the interaction itself.

	Receptor field		Mobile lure		Fixed object	
Percentage of clicks	**65.9 %**	± 3.9	**23.0 %**	± 10.4	**11.0 %**	± 8.9
Percentage of stimuli	52.2 %	± 15.2	15.2 %	± 6.2	32.7 %	± 11.8
Ratio clicks / stimuli	1.26		1.51		0.33	

Table 1. Ratio between clicks and stimuli, for each sort of stimulus.

It is easy to understand how perceptual trajectories cling to one another during perceptual crossings (Lenay et al. 2011), and this has also been confirmed by various forms of robotic modelling (Di Paolo et al. 2008, Iizuka and Di Paolo 2007). We may note to start with that there is a very general perceptual strategy employed by each agent: when a participant encounters a stimulation, and in particular when she meets her partner, she will invert her movement while the latter will do the same. The two receptor fields will thus enter into a sort of dance. The only difference between the receptor field of the other participant and the mobile lure attached to it is that only the former is sensitive to my presence; and this sensitivity is linked to a perceptual activity which constantly aims at remaining in the vicinity of a singularity. This is precisely a sufficient condition for the formation of an attractor in the joint dynamics which tends to augment the probability that the partner will be present.

Here the social interaction is constitutive of the recognition of another subject, to the extent that the corresponding task (clicking on the partner) is collectively solved, independently of individual judgements. We can also say that the dynamics of the interaction facilitates the individual task of discriminating the other subject, by bringing the partners face to face (Di Paolo 2016). However, this collective solution is not accessible as such to the individual participants, since they did not succeed in associating the presence of the other subject with this or that sensory stimulation (Froese and Di Paolo 2011, Michael and Overgaard 2012)[3]. This collective capacity does not therefore allow each individual to engage in more complex activities or relations which would be based on full, specific recognition of the other subject.

If we now wish to account for the formation of an individual recognition of the presence of the partner, we must envisage a second step. But we may note that we have a good starting position. If one wishes to propose an explanatory scheme for the constitution of individual social cognition, it is necessary to give oneself an initial situation in which it is sufficiently absent, while social (i.e. collective) properties are present. Our hypothesis is that it is only subsequently, by the individual appropriation of the emergent properties of the collective dynamics that one can account for the constitution of individual social cognition. In order to understand this, two new experiments were carried out.

3.2. A new experiment, a collaborative task

Tom Froese has taken up the perceptual crossing experiment, bring to it just two subtle adjustments (Froese et al. 2014): he placed the pairs of subjects in the spirit of a collaborative task (they consider themselves as forming a team which must gather more correct clicks than the other teams); and above all he asked each subject to click only once for each one-

[3] "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" (Froese and Di Paolo 2011, 20).

minute session, which reinforces the collaborative attention. These apparently minor changes in the instructions had a striking effect: there were now 88,5% correct clicks on the partner, 8,6% on the mobile lure and 2,9% on the fixed object. With 73% of stimulations due to the receptor field of the partner, 13% due to the mobile lure and 14% to the fixed object, these results show that, even at the individual level, the probability of clicking is greater for a stimulation due to the receptor field of the partner than for the mobile lure.

Examination of the trajectories for perceptual crossing shows that the subjects succeeded in the task because they set up a richer dynamics of interaction. The great difference with the first experiment is that now the shared aim with the partner is to avoid “false-friends” (i.e. periods of encounter with the mobile lure that can lead to error because unluckily the lure appears to be attached to a subject). It is only after a sufficiently long period of interaction that the subjects risk making a click. For that, they have to determine the position of a sensory singularity having the maximum likelihood of corresponding to the partner.

It is most interesting to note that the trajectories show the emergence of a quite general organisation involving a strategy of "turn taking": one of the two participants remains immobile for a few moments, offering herself as a localizable object for the perception of her partner and receiving stimuli passively; then resuming her active perceptual movements while the partner remains immobile to offer herself as an object in turn. In this way there comes into play a sort of dialogue or competition between two subjects each trying to objectify the other, and only succeeding by alternating the roles of subject and object (Sartre 1943). We will come back to this explanatory scheme in the discussion.

Before that, we will now see how one can obtain the same result with another experiment where there is a clear distinction between the determinate component of the other subject, and the dynamic component of the interaction.

3.3. Experiment with auditory discrimination

In this variation, we again use the initial experimental protocol of perceptual crossing, but this time there is no tactile stimulation: the three objects (fixed object, mobile lure, receptor field of the partner) are associated with three different sounds that are easily distinguished (Lenay and Stewart 2012). The precise association between the three sounds and the three objects is changed in random fashion for each experimental session. The task is to determine, for each session, which sound is associated with the receptor field of the partner. In these conditions, the subjects are highly successful: 85% correct identification of the partner from the 4th session of two minutes onwards.

Examination of the perceptual trajectories allows us to explain this result: as in the first experiment, there is an attractor of perceptual activities of the two partners. The difference is that now the sensory returns can be differentiated; it is then relatively easy for the subjects to recognize when they are indeed in contact with their partner. Thus, the perceptual crossing is characterized for the subject by frequent sensory stimulation combined with an indeterminate position, contrary to the fixed object which can give frequent sensory stimulation but with a well-defined position, and contrary to the moving object which gives no frequent sensory stimulation.

We will return later to this explanatory scheme. Before that, let us see how the same method makes it possible to address other aspects of social cognition.

4. Perception of intentions of others

Many studies in neuroscience and experimental psychology aim at demonstrating and explaining the perception of the intentions of others (Meltzoff 2005, Pacherie 2005). The discovery of neuronal structures such as the mirror neuron system (MNS), which are activated in specific fashion both on the occasion of perception of an intentional action by another subject

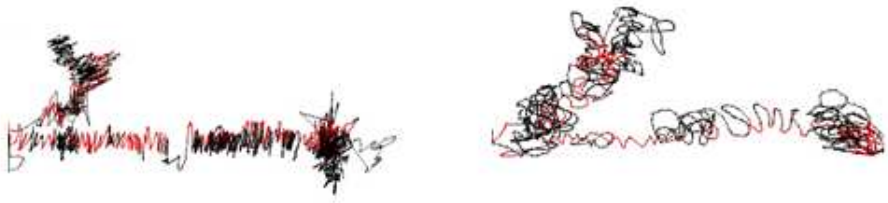

Fig. 4. The trajectories of the two participants (which were superposed) are shown here separately. In this experiment the "guide" (trajectory on the right) received the paradoxical instruction to follow the "follower" (trajectory on the left). Through the dynamics of interaction there is the emergence of a shared intention. The participants invent together the orientations of a path, here to go left and then go back up diagonally to the right.

and when the subject herself executes the same action, has been invoked as an explanation: the intention behind a perceived movement is interpreted by the intention that the observer would have if she performed the same movement (Gallese et al. 2004). Thus these structures appear to give each individual the capacity for a direct understanding, without any explicit reasoning, of the intentions of others. These structures clearly have a social meaning; but the question remains as to whether this meaning is in the last resort an internal construction, or is it the product of social interactions? In the former case one could come back to an internalist conception of cognition, and concrete social interactions would only be the secondary consequence of these individual internal capacities. In the latter case, it would be the social interactions which explain the genesis of these internal structures, and not the other way round. In order to find the means to envisage this second possibility, we can propose a new minimalist experiment.

In this experiment, participants share a bi-dimensional space. Each participant has only one receptor field coupled to a single tactile stimulator. The body-image is a square of the same surface as the receptor field. As with touch, the perception is exactly reciprocal: I cannot touch another person without being touched by him. There is neither a fixed object, nor a mobile lure. One of the two participants is given a very simple mission that the other must guess: "go to the top right", "go left", and so on. The follower tries to find his guide and to keep contact as much as possible. The guide should be collaborative and try as much as possible to help his partner guess his intention. They have no other means of interaction than the system itself. Most of participants succeed in this task, which is not easy since there is no spatial information in the sensory input: the receptor field is only a point (it is not a retina which could give a local indication of a movement). At each moment the follower has no means of guessing the direction of the movement of the guide other than allowing a loss of contact (for example by a small sweeping movement around the guide), and then engaging in a certain direction in the hope of managing to find the guide again. Progressively, with care and attention, the follower manages to anticipate the direction of the movement indicated by the guide. The extreme technical restrictions of this interactional situation produce an equivalent between perceiving the gesture of the other subject and accomplishing the gesture oneself. There is no perception of the movements of the partner prior to imitating or simulating it, but rather hitching the perceptual activity of the follower to the movements of the guide.

We may add that the minimalist conditions of the interaction are such that the perception of the movement of the guide can only be the perception of the *intention* behind this movement, since at each moment the follower must make a bet on the intention. There is an understanding of the intentions of others by sharing his intentions. This can clearly be seen when one introduces a new and tricky mission to the guide which is… "follow your partner". The dynamics of the interaction in this situation, where the participants follow each other mutually, leads to the emergence of collective dimensions which seem to be intentional even though no intentional direction pre-exists.

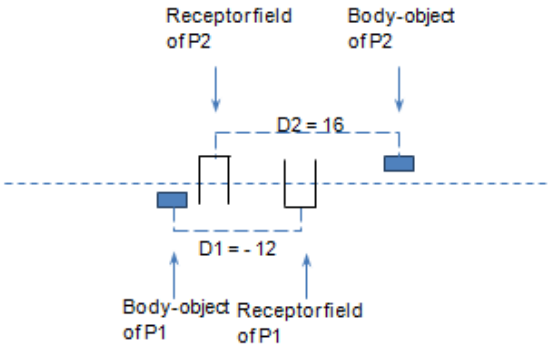

Fig. 5. The participants share a one-dimensional space (a torus) in which they move their receptor field together with their object-body (which is linked at a distance D). When the receptor field of one participant encounters the object-body of the other participant, she receives a tactile stimulus under the finger of the free hand. Here D2 > D1; if the participants seek a perceptual crossing, they will be carried in a collective drift towards the right.

In this situation, there is only understanding of the intention of the other subject by *sharing* the emergent intentions (Merleau-Ponty 1945, Wilkerson 1999). We will examine later in the general discussion whether this extreme situation can shed light on the question of the status of capacities for the perception of intentions. Before that, we will present a third experiment about the imitation of facial expressions (Lenay and Stewart 2012, Froese et al. 2012).

5. Imitation of facial expressions

How is it possible to explain the initial capacity for the imitation of gestures and expressions by the new-born infant? How can we understand that in the first days after birth, the infant can establish a relation of equivalence between the movements she observes and the motor and proprioceptive data concerning her own actions? In order to solve this "correspondence problem" (Brass and Heyes, 2005), it would seem difficult to invoke a learning process, especially when it is a question of *opaque actions* such as her own facial movements that she cannot see herself. Meltzoff and Moore (1977) therefore invoke the existence of innate individual capacities, an "Active Intermodal Matching" system (AIM) which is supposed to perform a supra-modal representation of bodily actions which are seen or performed. In the same way, one can also postulate the existence of an "innate mirror-neuron system" which participates in a neuronal cabling between perceived facial expressions and the expressions that are produced (Rizzolatti et al. 2002).

We have seen that in an interaction-based framework, this neonatal imitation is considered to be an initial capacity allowing the appearance of a "primary intersubjectivity", i.e. a very early process that pulls the infant into a dyadic interaction with the other (Trevarthen 1979). The cross-modal integration of vision and proprioception allows the infant to make some kind of pragmatic sense of the other's expression, in a way that calls forth a response (Gallagher and Meltzoff 1996). This makes it possible to understand the advent of a dynamics of reciprocal interaction between the new-born infant and her caregiver, a dynamics which produces various forms of synchronization in sequences which escape in part from the control of one or other participant.

However, here again, if one were to accept an innate "Active Intermodal Matching" system or an "innate mirror-neuron system", the interaction-based perspective would be weakened. In this case, the social cognition is not explained primarily by the inter-individual interactions themselves, but rather by the functioning of internal structures. The question remains to understand the origin or the genesis of these structures.

The path we propose to explore here consists of seeking the conditions for the appearance of "mimetic phenomena" in the very dynamics of the perceptual interactions, i.e. to create a situation where it is not the imitation that explains the interaction but the dynamics of interaction which produces the imitation. We have thus taken up our experiment of minimalist perceptual crossing; but this time, the participants can modify what is presented to their partner. In accordance with our approach, we have chosen as a minimal modification of the body: the relative distance between the body-object and the receptor field. The receptor field is no longer directly perceivable by the partner. All that is perceivable is the body-object that is attached to the receptor field. The distance between his receptor field and his own body-object can be actively modified by the participant. By clicking on the right or left button of the mouse, he can displace his body-object to the right or to the left relatively to his receptor field, 2 pixels at a time for each click. With their computer mouse the participants move their receptor field in a shared one-dimensional space, and seek to mutually perceive each other. However, they do not know the position of their body-object, relatively to their own receptor field.

The objective external description of "imitation" will be a similarity in the relative distances between receptor field and body-objects of the two subjects. We call D1 the position of the body-object of participant 1 with respect to his receptor field, and D2 for participant 2. When (D1 + D2) = 0, the receptor field of the first participant is exactly in front of the body-object of the second participant and the receptor field of this participant is also directly in front of the body-object of the first. In this configuration, there should be no problem for achieving perceptual crossing since the two partners perceive each other mutually at the same time. However, if (D1 + D2) < 0, the perceptual crossing is unbalanced, each participant moving to the left to find his partner. Similarly, if (D1+D2) > 0, the perceptual crossing should drift to the right. In the course of their interaction, the participants will likely feel this drift, and use the clicks of the computer mouse to alter D1 and D2 such as to stabilize their perceptual crossing.

Overall, there was clearly a convergence towards the situation where (D1 + D2) = 0, that we may identify as a situation of imitation. Even though the participants do not know the position of their body-object, either at the beginning or at the end, their joint search for a situation of balanced perceptual crossing rapidly leads to a similarity in these positions. In 3 minutes, the disequilibria in (D1+D2) are reduced to less than 30% of their initial values. The clicks of both participants act on a common spatial variable, the relative distance (D1+D2), which determines the balance of the perceptual interaction. By bringing this common variable to zero, they produce a stabilisation of the perceptual crossing which, from the point of view of an external observer, corresponds to a mirror-resemblance of the images that are presented to the partner. We may note that there are an infinite number of possible solutions to this "resemblance" between the participants, i.e. pairs of values such that (D1 + D2) = 0, and indeed the variance of the solutions increases over the course of interaction sequences (Lenay and Stewart 2012).

Analysis of the experimental results indicated that participants succeed in matching their bodies by responding to the relative stability of the interaction process, because mismatches in relative bodily configuration introduce systematic sideward drifts into the flow of the interaction. This drift cannot be reduced to actions of one of the participants; on the contrary, it emerges out of their interaction. Mimicry was therefore enabled by a collective property of the interaction process as a whole. It is explained because the tuning of the synchronization of the perceptual crossing is realized by the spatial variables of the form of the bodies of the participants.

In this experiment, "imitation" is the result and not the cause. Mimicry cannot be achieved by comparing external perception of the other's body configuration with external self-monitoring, or by comparing external perception of the other's body configuration with

internal self-observation. However, in a second phase, this *de facto* imitation should allow for learning of this sort. We will come back to this point.

6. Discussion

6.1. On methodology: Schematism and minimalism

The radicality of the minimalism of the experiments we have just presented forces an externalist approach to perception. The perceptual activities were externalised in the form of perceptual trajectories which can be easily recorded, permitting a complete analysis of the sensorimotor relations. This is what we called a spatial and temporal deployment of perceptual activity. Conversely, it is also possible to study the effect of an increase in the number of receptor fields. If, rather than a single receptor field, the user controls the displacements of a retina composed of a matrix of unit receptor fields (for example 4 x 4 receptor fields), there is a parallelism in the simultaneous sensory inputs corresponding to the surface covered by the retina. We have studied the way in which the perceptual trajectories are modified when this parallelism increases. One then observes an economy of movements, and hence a sort of internalization of the perceptual or cognitive processes (Sribunruangrit 2004). Parallelism in the sensory input leads to a relative separation between perception and action, even if active movements of the retina remain necessary.

In the case of social interactions, it is particularly revealing to employ a situation where the externalization is maximal, by attributing only a single bit of information at each moment and restricting the space of action to one or two translational dimensions. The resulting dynamics, unveiled before our eyes, seem to realize the empirical conditions for a theoretical discussion by reducing conceptual ambiguities to a minimum. This should allow for a precise discussion of the individual or social components of human cognition.

There is of course a huge gap between this kind of artificial experimental conditions and natural realization of human activities. The spaces of action and sensation are reduced to a minimum, there is no pre-existing general social structure and, worst of all, the interactions only concern a single pair of two subjects. But I would like to make two remarks.

i) The first is that my aim is only to propose methods for understanding phenomena, ways of explaining, what I call explanatory schemes. It is not a question of totally resolving all the issues of human social cognition in all its complexity, but only to propose schemes which might serve as models for explanations adapted to real situations. A structure of the dynamics of interaction highlighted in minimalist conditions to explain the constitution of a capacity for social cognition can perhaps be applied in the complex conditions of "natural" situations to explain a similar ability. It is then a purely empirical question to verify if this schema still works and if it proposes a sufficiently parsimonious understanding of the phenomena. Thus, we hope that the explanatory scheme established under minimalist conditions is a tool that can be used to capture phenomena, a tool that works or does not work depending on the situation.

ii) The second point is that such an explanatory approach is not reductionist. It is not a matter of bringing an observed phenomenon back to simpler and different causes that would only explain its appearance, as when we pretend to reduce a cognitive phenomenon to a neuronal dynamics. Rather, it is a question of building a world that works, even if it is a very simple world. Success is measured by the viability of an enaction, and not by the adequacy of a representation. It is a technological mode of validation, in the fashion of what Francisco Varela called “validation by construction” in the field of artificial life (Varela 1997). Here, even if the modes of interaction and the structures of the environment are as simple as possible, they still allow us to understand the enaction of a world shared between the subjects in interaction. We may add that this type of

experimental situation is sufficiently simple that examination of the coupled activities of the various subjects can be the basis for an abduction leading to the construction of minimal hypotheses concerning the underlying strategies of action.

6.2. Social approach to social cognition

We remarked in the introduction that in order to overcome the difficulties involved in "mind reading", one should start by recognizing that the majority of intentional activities can be perceived directly at the level of behaviours, without any need for inference concerning mental representations. However, that is not sufficient.

If the mechanisms of this perception of behaviours were themselves essential internal, we would come back to the problem of hidden intentions. This is the great risk involved in supposing from the start that there are individual capacities for recognizing faces, recognizing intentions or imitating expressions (Gallagher 2013, 61). The intentions of others with whom one considers entering in relation in a pre-reflexive coupling concern precisely in the first place the orientation of their perceptual activity. It is therefore necessary to add that the perceptual activity itself must be sufficiently embodied and expressed in the behaviour. This is what is accomplished with the minimalist conditions we have adopted.

But even that is not yet sufficient. If one wishes that in social cognition the sense-making of individual perceptual activity bears on the inter-individual interactions, it is necessary that this perceptual activity itself can be configured by means of the dynamics of interaction. At any rate, that is one of the specificities of the explanatory schemes proposed here.

Those three experiments that we have presented correspond to three explanatory schemes for the recognition of others, for the perception of their intentions, and for the imitation of facial expressions (we have also conducted the same sort of experiment for the study of joint attention (Deschamps 2016)). The interest of these explanatory schemes is that these phenomena of social cognition are not explained by individual capacities that would precede them. On the contrary, it is the dynamics of social interaction that explains the emergence of these phenomena and, possibly secondarily, the formation or modification of individual capacities.

The initial individual capacities that are mobilized do not have any social meaning in advance. They consist simply in the maintenance of individual autonomy by the search for sensorimotor invariants in a stable coupling. Following on, the logical order of our explanatory schemes consists of starting with the *de facto* social interactions resulting from the coupling of these simple individual capacities, by means of a minimal structure of the physical and social environment that is already there (ch. 6.2.1). We then observe the emergence of an organization of interaction which concretely instantiates the phenomenon of social cognition which is of interest (recognition of other subjects, perception of their intentions, imitation of their expressions). Then, the individual capacity to mobilize the result of the collective dynamics can be achieved by a link between this dynamics and perceivable properties or forms of the body-object of the partner (ch. 6.2.2), a link which then creates the conditions for associative learning (ch. 6.2.3). In this way one can understand the stabilization of new individual structures which will allow for new types of social engagements. We shall spell out in detail these three steps for each of our three situations of social cognition, and indicate their general consequences (ch. 6.2.4.); and we will then examine the conditions for the application of this approach to social cognition (ch. 6.3 and 6.4).

6.2.1. Initial capacities and dynamics of interaction

For the **recognition of others** we saw that it was accomplished collectively even though each partner is incapable of recognizing the presence of another behind a particular stimulation. The initial non-social capacities that are necessary are really very simple. It is essentially a question of maintaining a perceptual coupling, by reversing the direction of movement following a variation in sensory input. That is sufficient to explain the emergent dynamics of perceptual coupling, and hence the collective resolution of the task (identify one's partner).

For the **perception of the intentions** the initial individual capacities consist in the know-how for following a target in a two-dimensional space. The coupling of the individual perceptual activities produces *de facto* a shared intentional movement.

For the **imitation of facial expressions**, the necessary initial capacities are barely more complex. It is a question of being able to maintain a perceptual coupling, and to perceive its drift, in order to take the appropriate corrective measures. A very simple, basic "imitation" results directly from the dynamics of the interaction, even while participants ignore what they provide to the perception of their partner. They do not proceed by comparison between a perceived shape or behaviour and the shape or behaviour that is produced. The "imitation" is produced in the search for an equilibrium in the perceptual coupling, achieved by the twin spatial adjustment of the links between receptor fields and body-object.

This experimental setup may seem to be so minimal and artificial that it is difficult to relate it to the strong correspondence problem of neonatal imitation. However, if one employs this explanatory scheme to describe the games of imitation in proto-conversation, we understand that the new-born infant may have the capacity to recognize an agreement in the interaction *without* needing to know whether her facial expressions are more or less correct reproductions of those of the adult (Reddy 2003, Trevarthen and Reddy 2007).

6.2.2. Link between the social dynamics with determinate properties of the body-object of the partner

For the **recognition of others**, we have seen in two complementary experiments that the individual capacity to recognize the other subject requires a sufficient determination of the body-object of the partner. Either (ch. 3.2) because a pattern of turn-taking allows each partner in turn to present herself as an object in a sufficiently determinate position to be designated at the end of the trial; or (ch. 3.3) by differential auditory features of the object-bodies. In both cases, the subjects are able to recognize and designate the other subject by linking this determinate body-object to the interactional dynamics of the perceptual crossing.

If we deploy this explanatory scheme to account for a rich and complex real-life situation such as mother-child proto-conversation (Trevarthen 1979, 1993), we can understand that it is indeed a form of perceptual crossing, and that conversely its absence (as in the case of an interaction with a recording) can be recognized and felt as disturbing. In this natural situation there are of course a multitude of properties for the infant that she perceives in a determinate manner, in particular, the distinction between the mother's face and other objects in the environment. There is every chance that the dynamics of perceptual crossing be linked with the competing perception of this face.

We may note that this capacity does not correspond to the result of an individual "contingency detector", but simply the association between the dynamics of the perceptual crossing and clues (the sound-coded properties of the objects). The emergent collective dynamics can be appropriated by the subject, by associating the indetermination of the position of the partner with a sufficiently distinctive sensory property. In this situation again, it seems that the perceptual crossing does play a constitutive role for the perception of the other subject to the extent that what the perception of the other subject is, in the limits of this minimalist experiment, is indeed the presence of the characteristic

dynamics of perceptual crossing linked with the perceived property.

For the **perception of the intentions** the dynamics of interaction directly link the movements of the two partners. The recognition of the gesture of the other subject has to be linked to the recognition of the gesture produced in order to perceive.

For the **imitation of facial expressions**, we have seen that the agreement between the partners can arise on the basis of a *de facto* imitation which has an infinite number of solutions. If one wishes that the individuals can appropriate the imitation of a particular expression it would be necessary, in addition to the actual conditions of the experiment, to give them perceptual access to the particular form of the "face" of the partner (here the length of the link between the receptor field and the body-object). In natural situations that means linking the interactional dynamics instantiating *de facto* mimicry, to the perception of a particular facial expression.

6.2.3. Associative learning in a situation of interaction

The link established in the social situation between the emergent organization of the dynamics of interaction on one hand, and determinate properties of the body-object of the partner on the other hand, should then allow for an individual associative learning of this correlation.

For the **recognition of others,** the stabilization of the interaction in the perceptual crossing allows the establishment of a link with the position of the body-object of the partner, or with a sensory feedback sound that can take the meaning of an encounter with another subject. Thus, an individual learning of the association between certain properties of the objects and a behaviour of perceptual crossing becomes possible. That means in natural situation an individual learning of the association between the stimulus of a face and the behaviour of perceptual crossing. This novel individual capacity for recognizing the presence of another subject in a perceived face finds its origin and its intentional content in the social interaction that it makes possible.

For the **perception of the intentions**, it is widely supposed that internal structures such as the Mirror Neuron System (MNS) make it possible to understand a perceived movement as being equivalent to a movement that one could produce oneself. Now in the very simple experiment that we have proposed, it is the perceptual activities themselves that are dynamically linked. This opens a path for explaining the *origin* of such internal structures. The perceived gesture is associated to a gesture one has performed oneself because it very precisely perceived *by* being performed. The neuronal arrangements which control the accomplishment of a specific gesture are functionally the very same arrangements which are mobilized to perceive the performance of this gesture by another subject. To the extent that the interaction stabilizes a given gesture, we can understand how it is possible to construct structures which associate the concomitant multimodal sensations that are received, whether they be of exteroceptive origin (the gesture of another subject), or proprioceptive (one's own gesture); and this should remain true even when, with a parallelism at the level of the sensory input (a retina), the perceptual activity tends to become separated from the movements which are observed or performed tends to establish itself. We therefore have here a potential way of *explaining*, by the functional meaning of interactions, the *formation* of internal structures such as the MNS, which will subsequently make it possible to recognize the intention of an observed movement without now having to actually perform it oneself. This new capacity finds its origin and intentional content in the engagement that it allows in interactions with the observed person.

For the **imitation of facial expressions**, there again, the collective dynamics of the perceptual crossing brings about a situation of sensorimotor interactions that are sufficiently stable to serve as the basis for associative learning. We have seen that in the natural situation of proto-conversation, the perception of an "agreement" precedes an explicit awareness about what that

agreement is about (Reddy 2008). Mimicry spontaneously results from the mutual regulation of collective interaction dynamics, and it is this social interaction which provides the newborn with the motivation and means for linking her perception of the other with her proprioceptive sensations. It is only later that the child will discover that what she is doing is in fact an imitation. And from then onwards, on the basis of an agreement in the perceptual crossing, she may presume that her own facial expression, that she cannot see, actually resembles that of her partner that she does see.

What is particularly interesting here again, is that it is a question of an association between the face of the partner and the dynamics of interaction which is socially meaningful (the perceptual crossing). The classical logic of "imitation" is inverted. It is the *de facto* "imitation" resulting from the collective dynamics, which then provides the means for linking the perceived image to proprioceptive sensations. The existence of structures such as "mirror neurons" for facial expressions might be genuinely *explained* by such a social situation, which allows for an association between different synchronized fluxes of multimodal and proprioceptive sensory inputs, sensory data which comes both from the behaviour of the other subject and from the subject's own actions (De Jaegher et al. 2010, 7).

We may note that the modification of the bond between receptor field and body-object, are modifications of an embodiment of the action strategy (Maillet et al. 2008). Indeed, the distance between the receptor field and the body-object define the manner in which each subject intervenes in the dynamics of the interaction and the collective drift which follows if the distances are not matched. Following this explanatory scheme, a facial expression corresponds to a configuration of the relation between perceiving-body and body-object, configuration which can be associated with a given form of the face for each type of relational engagement. In the framework of a relational conception of emotions (Dumouchel 1995, Damiano et al. 2015) it is possible to conceive how the emotional meaning of expressions could be learned. The form of the face of the partner is associated in real time with a particular meaningful relational structure (for example a moment of joy and approaching, or a moment of disgust and retreat); there can thus be an association between this social structure and the corresponding facial expression.

6.2.4. Novel capacities for social cognition

In its generality, this kind of explanatory scheme does not decide in favour of either a hereditarian conception or an environmentalist conception of human social cognition. It does however militate strongly in favour of an interactionist approach, and thus a fully social approach to social cognition. The important point is that the dynamics of inter-individual interaction constitutes a situation which link on one hand a collective dynamics, and on the other hand a particular perceptual content. The association between this social signifying dynamics and a determinate property could equally well be the result of individual associative learning, or of the selection of hereditary characters which accomplish this association. The logical point which is crucial here is that the individual neuronal structures which participate in the association can be partly the *result* and not the primary cause of the dynamics of inter-individual interaction.

The perceptual learning that we envisage consists of profiting from stable configurations in the dynamics of interaction, in order to favour the formation of internal (neuronal) structures which associate (a) the fluxes of actions and sensations which result from this dynamics of inter-individual interaction and (b) the coupling with determinate elements of the situation (the body-object of the partner).

In coherence with the enactive approach adopted here, we do not require that these internal structures function as "stand-ins" or "representations" of the situations which served for their learning, as if they could substitute for the actual presence of the social interactions. These structures are to be understood rather

as the bases for strategies of action which stabilize the situations and which permit engagement in richer forms of social coupling. For example, the association between the determinate form of a face, and the dynamics of perceptual crossing, makes it possible in what follows for an individual to base herself on the sole perception of the face in order to engage in an interaction, supposing in advance the existence of a perceptual presence who will be able to respond. Similarly, counting in advance on an intentionality in an observed movement one can engage in other movements; or perceiving a facial expression one can respond by another expression.

At the same time, the link, between determinate perceived properties and the social meaning of the dynamics of interaction, makes it possible to understand that one can engage in a coupling with traces or signs, finding a social meaning in this coupling even in the absence of actual interaction with other subjects. This is a possible route for accounting in an enactive framework for capacities of imagining, reasoning or planning. It is sufficient to understand these activities, not as the manipulation of mental representations, but as the result of a game of coupling with traces of a social presence.

6.3. The distinction perceiving-body / body-object

Throughout these explanations it has been necessary to make a distinction between the body that perceives and the body that is perceived. This distinction first became obligatory at the level of technological mediation; in the Tactos system, for an inter-individual interaction to be possible, it was necessary to define for each individual on one hand her receptor fields (i.e. where the encounter with objects triggers a sensory stimulation for the subject), and on the other hand the body-object which can affect another subject. There must then be some relation between these two aspects of the body; this relation can be constant (as in ch. 2 and 3) or variable (ch. 3).

In a way, this distinction corresponds to the phenomenological distinction between one's own lived-body, and the body for others (Merleau-Ponty 1945); there can be no question of opposing them as in a dualism between body and mind. Here, in a pragmatic framework, it is simply a distinction between the body that makes it possible to perceive and the body that enters into social interactions, i.e. the body that affects others and can be perceived.

An individual does not perceive his own body-object as it is presented to others. This is what makes it possible to characterize a concrete difference between various situated points of view, distinct from one another. It is a constitutive feature of the social that we never perceive exactly the same things as others (otherwise we could not perceive each other mutually). In the case of the face, there is a distinction between the face as an organ of perception and the face as an organ of interaction. This is a distinction that functions without being directly accessible at the individual level. Its role is profoundly social.

The body-object does not participate in the initial individual perceptual activity, but it participates *indirectly* in social perception since it is necessary to the establishment of dynamics of interaction which allow each subject to give a social meaning to her activity. In fact, this distinction is functionally necessary to the formation of the social interactions that we have examined. In the various experiments, the perceptual crossing corresponds to a dynamics that escapes each participant because it depends on perceptual actions bearing on the object-bodies that the participants do not perceive themselves and which are nevertheless related to their perceptual activity. There is thus something radically indeterminable in the body-object of another subject. However, at the same time, we have seen that the individual appropriation of the organization of the dynamics of interaction passed by moments where each individual could sufficiently determine the properties of the body-object of her partner in order to subsequently link them to this organization. Thus the distinction between the perceiving-body and the body-object rebounds at the level of the body-object as a distinction between the body-object which serves to enter into a

dynamics of interaction (in which it cannot be precisely determined) and the body-object which bears determinable properties. This constant tension does indeed seem to be constitutive of the face of another: at the same time something which resists complete determination, and which nevertheless persists in being present and that one can partially determine.[4]

6.4. Enaction and technology

The collective, the dynamics of interaction which escapes from each individual, can in return confer on this individual new capacities for acting, interacting and hence capacities for sense-making which did not exist before this social engagement (Di Paolo and De Jaegher 2016). The social transforms the field of individual autonomy by giving it access to a novel domain of action and meaning. However, if we say that the dynamics of collective interactions fashions the individual structures of social cognition, does this not admit a social determinism which would put into question the autonomy of the individual? It is not the case if we interpret this feedback of the social on the individual as a mechanism of individual appropriation, which can of course be a source of alienation as well as empowerment.

In order to understand this constitutive feedback of the collective on the individual one can take the model of a tool. Made socially, a tool is given to an individual as a new power to perceive, to act and to interact. The social gives to each individual a technological and linguistic environment which is constitutive of individual activities. Now the tool which is received poses first of all the problem of its appropriation. It forces the learning of specific strategies, i.e. neuronal configurations allowing for mastery of its use.

For the mechanisms of social cognition that we have presented, the body-object plays the role of a tool. It is given by others (since it is not perceived directly by the individual concerned), and it serves to enter into the dynamics of social interactions. The body-object plays a constitutive role not directly (like the perceiving-body), but indirectly as a body constitutive of interactions which give their meaning to various forms of social cognition. The stages of appropriation that we have described correspond to the individual learning of how to use this primal "tool" of social interactions since it serves to recognize another subject, to perceive her intentions or to imitate her expressions.

In an enactive perspective, the body-object cannot directly be part of individual sense-making. It only has meaning by way of interactions. If we wish to understand social cognition it is thus necessary to admit a reconfiguration of the operational closure of individual autonomy by considering a form of sense-making which is accomplished in a coupling mediated by this body-object, a coupling with meaningful forms of the social and first of all with another subject, her intentions and emotions.

Hence, just like a tool that one keeps and can carry around with one, each individual keeps her body-object even in the absence of the social environment which created it. Like other traces of social presence the body-object, the prime basis for reflexivity, serves to repeat, even in solitude, the meaning of inter-individual interactions.

[4] It seems to us that there is here something which corresponds to the description of the face as proposed by Emmanuel Levinas in *Totality and the infinite (*Levinas 1979), the other as something that does not let itself appear as a phenomenon, yet exists precisely at the phenomenal level (Lenay & Sebbah 2016). We may note that this reading of Levinas is quite different from that proposed by Gallagher (2014) which aims rather at locating the transcendence of otherness in the interaction itself.

7. Conclusion

We have proposed three explanatory schemes, in order to provide an interaction-based account of three prime forms of social cognition: the recognition of another subject, the perception of intentions and the imitation of facial expressions. For this we have proposed a minimalist method which makes it possible to disentangle the various components of the dynamics of inter-individual interaction. We have thus shown that initial non-social capacities allow for the emergence of collective organizations which are socially significant; situations of interaction which provide the opportunity for the formation of individual structures associating determinate properties of object-bodies and these situations of interaction.

In the three cases *intercorporeity*, which can be understood as the factual dynamics of inter-individual bodily interactions, is the *explanans* and not the *explanandum* of specific neural structures. Rather than the neural mechanisms that make human capacities of social cognition possible, it is a matter of showing how the dynamics of interactions that makes sense in the human social world, make possible these neural structures. In the three cases, explanation of social activities is based on the meaning of these activities. It is not based on meta-cognitive internal representations, nor on infra-personal structures. It is itself social.

It is now an empirical question to verify whether such explanatory schemes are useful tools for accounting for phenomena in the more complex situations of real everyday social life.

References

Auvray, Malika, Charles Lenay, and John Stewart. 2009. "Perceptual interactions in a minimalist virtual environment". *New ideas in psychology* 27 (1): 32–47.

Auvray, Malika, and Marieke Rohde. 2012. "Perceptual crossing: the simplest online paradigm". *Frontiers in human neuroscience* 6.

Baron-Cohen, Simon. 1997. "How to build a baby that can read minds: Cognitive mechanisms in mindreading". *The maladapted mind*, 207–239.

Brass, Marcel, and Cecilia Heyes. 2005. "Imitation: is cognitive neuroscience solving the correspondence problem?" *Trends in Cognitive Sciences* 9 (10): 489–495.

Clark, Andy. 1998. *Being there: Putting brain, body, and world together again*. The MIT Press.

Clark, Andy. 2004. *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford University Press.

Damiano, Luisa, Paul Dumouchel, and Hagen Lehmann. 2015. "Towards human–robot affective co-evolution overcoming oppositions in constructing emotions and empathy". *International Journal of Social Robotics* 7 (1): 7–18.

De Jaegher, Hanne, Ezequiel Di Paolo, and Shaun Gallagher. 2010. "Can social interaction constitute social cognition?" *Trends in cognitive sciences* 14 (10): 441–447.

De Jaegher, Hanne, and Ezequiel Di Paolo. 2007. "Participatory sense-making". *Phenomenology and the Cognitive Sciences* 6 (4): 485–507.

Deschamps, Loïc, Charles Lenay, Katia Rovira, Gabrielle Le Bihan, and Dominique Aubert. 2016. "Joint Perception of a Shared Object: A Minimalist Perceptual Crossing Experiment". *Frontiers in psychology* 7.

Dewey, John. 1896. "The reflex arc concept in psychology." *Psychological review* 3 (4): 357.

Di Paolo, Ezequiel, Marieke Rohde, and Hiroyuki Iizuka. 2008. "Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing". *New ideas in psychology* 26 (2): 278–294.

Di Paolo, Ezequiel, and Hanne De Jaegher. 2016. "Neither individualistic, nor interactionist". *Embodiement, Enaction, and Culture*.

Di Paolo, Ezequiel, and Hanne De Jaegher. 2012. "The interactive brain hypothesis". *Frontiers in human neuroscience* 6.

Dumas, Guillaume. 2011. "Towards a two-body neuroscience". *Communicative & Integrative Biology* 4 (3): 349.

Dumas, Guillaume, Julien Laroche, and Alexandre Lehmann. 2014. "Your body, my body, our coupling moves our bodies". *Frontiers in human neuroscience* 8.

Dumouchel, Paul. 1995. *Emotions: essai sur le corps et le social*. Institut Edition Synthelabo.

Froese, Tom, and Ezequiel Di Paolo. 2011. "The enactive approach: theoretical sketches from cell to society". *Pragmatics & Cognition* 19 (1): 1–36.

Froese, Tom, Charles Lenay, and Takashi Ikegami. 2012. "Imitation by social interaction? Analysis of a minimal agent-based model of the correspondence problem". *Front. Hum. Neurosci* 6: 202.

Froese, Tom, Hiroyuki Iizuka, and Takashi Ikegami. 2014. "Embodied Social Interaction Constitutes Social Cognition in Pairs of Humans: A Minimalist Virtual Reality Experiment". *Scientific Reports* 4 (janvier).

Froese, Tom. 2016. "Interactivity Should Aim to Extend, Not Reject, the Conceptual Foundations of Enaction". Édité par Tom Froese, Sebastjan Vörös, and Alexander Riegler. *Constructivist Foudations* 11 (2): 247-49.

Gallagher, Shaun, and Andrew N. Meltzoff. 1996. "The earliest sense of self and others: Merleau-Ponty and recent developmental studies". *Philosophical psychology* 9 (2): 211–233.

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

Gallagher, Shaun. 2008. "Intersubjectivity in perception". *Continental Philosophy Review* 41 (2): 163–178.

Gallagher, Shaun. 2013. "When the problem of intersubjectivity becomes the solution". *The infant mind: Origins of the social brain, Guildford Press, Toronto*, 48–74.

Gallagher, Shaun. 2014. "In your face: transcendence in embodied interaction". *Frontiers in human neuroscience* 8.

Gallese, Vittorio. 2001. "The 'shared manifold' hypothesis. From mirror neurons to empathy". *Journal of consciousness studies* 8 (5-6): 33–50.

Gallese, Vittorio, Christian Keysers, and Giacomo Rizzolatti. 2004. "A unifying view of the basis of social cognition". *Trends in cognitive sciences* 8 (9): 396–403.

Gallese, Vittorio. 2007. "Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition". *Philosophical Transactions of the Royal Society B: Biological Sciences* 362 (1480): 659–669.

Gallese, Vittorio. 2009. "The two sides of mimesis: Girard's mimetic theory, embodied simulation and social identification". *Journal of Consciousness Studies* 16 (4): 21–44.

Gallese, Vittorio, and Valentina Cuccio. 2015. "The paradigmatic body". *Open MIND*, 1–23.

Goody, Jack. 1977. *The domestication of the savage mind*. Cambridge University Press.

Hanneton, Sylvain, Olivier Gapenne, Christelle Genouel, Charles Lenay, and Catherine Marque. 1999. "Dynamics of shape recognition through a minimal visuo-tactile sensory substitution interface". In *Third international conference on cognitive and neural systems*. Boston, USA.

Iizuka, Hiroyuki, and Ezequiel Di Paolo. 2007. "Minimal Agency Detection of Embodied Agents". In *Advances in Artificial Life*, édité par Fernando Almeida e Costa, Luis Mateus Rocha, Ernesto Costa, Inman Harvey, and António Coutinho, 485-94. Lecture Notes in Computer Science 4648. Springer Berlin Heidelberg.

Lenay, Charles, Stéphane Canu, and Pierre Villon. 1997. "Technology and perception: the contribution of sensory substitution systems". In *Cognitive Technology, International Conference on*, 44-53. Los Alamitos, CA, USA: IEEE Computer Society.

Lenay, Charles, Olivier Gapenne, Sylvain Hanneton, Catherine Marque, and Christelle Genouëlle. 2003. "Sensory substitution: Limits and perspectives". In *Touching for Knowing, Cognitive psychology of haptic manual perception*, édité par Y. Hatwell, A. Streri, and E. Gentaz, 275–292. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Lenay, Charles, Malika Auvray, François D. Sebbah, and John Stewart. 2006. "Perception of an intentional subject: An enactive approach". In *Third International Conference on Enactive Interface*, 37-38. Montpellier, France.

Lenay, Charles, and Pierre Steiner. 2010. "Beyond the internalism/externalism debate: the constitution of the space of perception". *Consciousness and Cognition*, n° 19: 938-52.

Lenay, Charles, John Stewart, Marieke Rohde, and Amal Ali Amar. 2011. "You never fail to surprise me: the hallmark of the Other: Experimental study and simulations of perceptual crossing". *Interaction Studies* 12 (3): 373–396.

Lenay, Charles and John Stewart. 2012. "Minimalist approach to perceptual interactions". *Frontiers in Human Neuroscience* 6.

Leroi-Gourhan, André. 1993. *Gesture and speech*. MIT Press.

Levinas, Emmanuel. 1979. *Totality and infinity: An essay on exteriority*. Vol. 1. Springer.

Maillet, Barthelemy, Charles Lenay, and Anne Guenand. 2008. "Designing for interpersonal tactile interaction over distance". In *HFT2008*, 391-98. Kuala Lumpur.

Meltzoff, Andrew N., and M.Keith Moore. 1977. "Imitation of facial and manual gestures by human neonates". *Science* 198 (4312): 75.

Meltzoff, Andrew N. 2005. "Imitation and other minds: The "like me" hypothesis". *Perspectives on imitation: From neuroscience to social science* 2: 55–77.

Merleau-Ponty, Maurice. 1996. *Phenomenology of perception*. Motilal Banarsidass Publishe.

Merleau-Ponty, Maurice. 1968. *The visible and the invisible: followed by working notes*. Northwestern University Press.

Michael, John, and Søren Overgaard. 2012. "Interaction and social cognition: A comment on Auvray et al.'s perceptual crossing paradigm". *New Ideas in Psychology* 30 (3): 296–299.

O'Regan, J.Kevin, and Alva Noë. 2001. "A sensorimotor account of vision and visual consciousness". *Behavioral and brain sciences* 24 (5): 939–972.

Pacherie, Elisabeth. 2005. "Perceiving intentions." In *A Explicação da Interpretação Humana*, edited by J. Sàágua, 401-14. Edições Colibri.

Reddy, Vasudevi. 2008. *How infants know minds*. Harvard Univ Pr.

Rizzolatti, Giacomo, Luciano Fadiga, Leonardo Fogassi, and Vittorio Gallese. 2002. "From mirror neurons to imitation: Facts and speculations". *The imitative mind: Development, evolution, and brain bases* 6: 247.

Sartre, Jean-Paul. 2012. *Being and nothingness*. Open Road Media.

Schilbach, Leonhard, Bert Timmermans, Vasudevi Reddy, Alan Costall, Gary Bente, Tobias Schlicht, and Kai Vogeley. 2013. "Toward a second-person neuroscience". *Behavioral and Brain Sciences* 36 (04): 393–414.

Sribunruangrit, Nitiphan, Catherine K. Marque, Charles Lenay, Sylvain Hanneton, Olivier Gapenne, and Clotilde Vanhoutte. 2004. "Speed-accuracy tradeoff during performance of a tracking task without visual feedback". *Neural Systems and Rehabilitation Engineering, IEEE Transactions* 12 (1): 131–139.

Stiegler, Bernard. 1998. *Technics and time: The fault of epimetheus*. Vol. 1. Stanford University Press.

Stewart, John, and Olivier Gapenne. 2004. "Reciprocal modelling of active perception of 2-D forms in a simple tactile-vision substitution system". *Minds and Machines* 14 (3): 309-30.

Thelen, Esther, and Linda B. Smith. 1996. *A dynamic systems approach to the development of cognition and action*. MIT press.

Tomasello, Michael, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll. 2005. "Understanding and sharing intentions: The origins of cultural cognition". *Behavioral and brain sciences* 28 (5): 675–690.

Trevarthen, Colwyn. 1979. "Communication and cooperation in early infancy: A description of primary intersubjectivity". *Before speech: The beginning of interpersonal communication* 1: 530–571.

Trevarthen, Colwyn. 1993. "The self born in intersubjectivity: The psychology of an infant communicating." In *The perceived self: Ecological and interpersonal sources of self-knowledge*, 121-73. New York USA: Cambridge University Press.

Trevarthen, Colwyn, and Vasudevi Reddy. 2007. "Consciousness in infants". *The Blackwell companion to consciousness*, 37–57.

Varela, Francisco J. 1997. "Patterns of life: Intertwining identity and cognition". *Brain and cognition* 34, n° 1: 72–87.

Varela, Francisco J., Evan Thompson, and Eleanor Rosch. 1991. *The embodied mind: Cognitive science and human experience*. MIT press.

Wilkerson, William S. 1999. "From bodily motions to bodily intentions: The perception of bodily activity". *Philosophical Psychology* 12 (1): 61–77.