Becoming Human

From pointing gestures to syntax

Teresa Bejarano

Advances in Consciousness Research

81

John Benjamins Publishing Company

Becoming Human

From pointing gestures to syntax

Teresa Bejarano
University of Sevilla

John Benjamins Publishing Company
Amsterdam / Philadelphia

Library of Congress Cataloging-in-Publication Data

Bejarano, Teresa.
Becoming human : from pointing gestures to syntax / Teresa Bejarano.
p. cm. (Advances in Consciousness Research, ISSN 1381-589X ; v. 81)
Includes bibliographical references and index.
1. Language acquisition. 2. Gesture. 3. Grammar, Comparative and general--Syntax.
4. Psycholinguistics. I. Title.
P118.B44 2011
153--dc22 2011009711
ISBN 978 90 272 5217 3 (Hb ; alk. paper)
ISBN 978 90 272 8679 6 (Eb)

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

SECTION ONE. **Evolutionary precursors**

SECTION TWO. The basic human ability

SECTION THREE. **Specifying some necessary requisites of language**

CHAPTER 7

CHAPTER 8

SECTION FOUR. **The origin of predication and syntax**

CHAPTER 19

Historical grammaticalisation: The answers are lacking, but the questions are good

SECTION SEVEN. **Syntax beyond predication**

CHAPTER 20

Interrogative communication

CHAPTER 21

Introduction

1. On the nature of an hypothesis on human abilities

The ability to perceive other minds is the starting point from which we shall attempt our explanation of different human abilities.[1] It is my contention that human beings are the only animal capable of conceiving the inner states of another individual looking at them. A second centre would thus be present in the human mind, and it is on this additional centre that the ability to perceive other selves depends. Of all the sources of arguments for this hypothesis, I will address just three.

Firstly, there is the link with antecedents in evolution. We would have to explain why this very special animal, the human being, appeared among primates. Likewise, we would also have to clarify how human beings differ from apes. Secondly, whichever hypothesis about the nature of human beings will have to account for a set of specifically human characteristics and behaviours. From the outset, and as a bare minimum, we should expect such a hypothesis to focus on the 'theory (which the subject has) of (another's and of one's own) mind', or theory of mind (ToM), for short, as well as on finger-pointing and different aspects of language. Thirdly, any theoretical approach put forward to explain these phenomena would be obliged to (at very least) look at the child development. Certainly, I am not advocating an absurd recapitulationism. There is obviously no reason why either the paths of evolution or of historical development should be reflected with any precision in ontogeny. Nevertheless, we might find some similarities between the postulated emergence of human characteristics and the data on child development.

Stating these requirements is easy; meeting them is another matter altogether. I have certainly tried very hard to do this, but I am well aware of the results of my labour. What we have here is a draft, still raw, a seed which will require endless correction and adding to, not to mention radical amendments to the whole. My hope, nevertheless, is that it will prove to be a timely study. There is no contradiction in this.

My optimism is based on the increasing demand right now for more and more studies of this kind. I, for one, see this as a task for precisely our generation. It is only now that contributions on human abilities from many different fields are finally available, and as a result that it has become urgent to ask how all these contributions might

---

[1] I thank to the director of the collection and to the two anonymous referees, who have substantially contributed to the improvement of this manuscript.

model will tell us how to build a brain that can do vocal learning), the primatocentric bias is the object of fierce criticism and ever increasing opposition. The decisive factors would be neither genes nor evolutionary proximity but habitat and niche, and hominids' habitats have been dramatically different from those of apes: these are the most recent academic trends and, in my opinion, they have made extremely valuable contributions. However, I think that we must still try to answer the above question. Or, more precisely, we must distinguish between two levels.

In my view, on the one hand, the indirect cause of human peculiarities is to be found in habitat-related changes and the ensuing changes in life style. On the other hand, the direct cause must be traced back to the cognitive novelties which, against a primate background, evolution moulded. There is no doubt that those cognitive novelties were shaped to cope with the urgent needs imposed by a new and more cooperative life style. However, the very nature of what eventually emerged would have been radically different if the starting point had not been the last common ancestor of chimpanzees and humans. It is essential to distinguish between these two causes. And since I am going to concentrate on the cognitive novelty or direct cause, the question of why it was among primates that humans appeared is still a useful and relevant question.

3. Outlining the proposal put forward here

The general aim of this book is to stress the crucial role played by the ability to perceive the minds of others. Undoubtedly, interpersonality is involved both in linguistic communication and in all other kinds of cultural transmission. And, also undoubtedly, it is these communicative processes which, by allowing the historic accumulation of successive findings, gave rise to the unfolding of human abilities. However, what will be proposed here is not this but that the human mind itself, and not just its fruits or results, would have originated in the perception of the minds of others.

How does all this relate to the simulationist*[2] subtrend of the studies of the 'theory of mind'*? In my opinion, simulation is often incorrectly invoked these days. As a result, monkeys' 'mirror neurons' have become the flagship of simulationism. However, I will suggest that the self-visible hand of primates would merely have given rise to a new kind of expectation. An improvement in this kind of expectation would also suffice to explain the ability (which chimpanzees probably have) to ascribe visual perceptions to a conspecific. These two capabilities, contrary to what many researchers have stated, would not require any latent imitation at all. I quite agree that in these abilities of non-human primates, a correspondence between a body seen (i.e., a body of a conspecific) and an internal state began to emerge. However, these two features of primates could be accounted for by expectation, that is, within the general mechanism of animal behaviour.

---

[2] Asterisks point out that the previous word is included in the Glossary.

But if this important new step in macaques and chimpanzees does not yet require simulation, the key question is this: when would the process of genuine simulation have come into being? My answer is that expectation is no longer useful when a relationship to oneself is attributed to the conspecific's interiority. If I attribute to someone else a visual percept that includes myself (or if, when I detect the kinaesthetic-postural interiority associated to someone else's movement, I interpret this movement as an attempt at communication with me), then I have had to replace the old procedure of expectation by that of simulation. The interiority of the conspecific's body would then be, at last and for the very first time, a radically other self, one which could only be conceived in a 'second centre' inside one's own mind. This second centre would be the site of true simulation.

There may be some who raise objections to this last sentence – e. g., "I can recognize that someone is angry without 'simulating' their anger". For this reason, I am going to give a brief summary of my more elaborate answer which comes later. Somebody's anger can be recognized at two different levels. On the one hand, an animal can adequately recognize and immediately respond to another animal's anger signals without ascribing any mental state to the angry conspecific. At this level there is no simulation at all. On the other hand, by simulation I can imagine another person's anger without myself experiencing that anger, but as the anger of someone else. In my view, it is thus that we should interpret the exclusively human process involved in simulation. But let us return to our topic.

Once this second centre had thus emerged, it could begin to sustain other processes different from those we have already mentioned. One of these processes would be the perception of beliefs different to one's own. The suggested branching of one's own mind into two different lines seems perfectly suited to preventing interferences between a given reality and false beliefs of others about such a reality.

It is well known that the perception of false beliefs of others has become the flagship of 'theory of the mind' research. In fact, the focus on this question has, on occasion, been excessive. However, despite all the research on the subject, I believe its crucial role has not yet been fully appreciated. What then is this crucial role?

Let us look at the perception of the belief of a 'second person', that is, the perception of the incomplete, incorrect or out of date knowledge a conversation partner has shown to possess on an issue. This kind of perception of beliefs of others is different from that observed in the classic tests (for example, in those tests such as the displaced toy test, where I have to infer the false belief of someone who, as a character in a sketch, is neither interacting with nor speaking to me). Compared with this, 'second person' perception would not only be easier, but also more decisive, inasmuch as it would have to do with predicative communication.

It is true that many authors have stated that the communicative function of predication would have come into being precisely in order to complete, correct or update the belief of addressees. I fully agree. However, what I am suggesting is not this, but that the belief of addressees would constitute the content which the element subject to predication (or rather, the theme*) has for the speaker. Correlatively, the predicate (or rather, the rheme*) would be the element which, according to the speaker, must be added to complete, correct or update the addressee's belief. As a result, the perception of beliefs of others would have given rise not only to the need for predicative communication, but also to the very form of predication, that is, to the hinge (or composition) between the element subject to predication and the predicate.

According to this, a key issue is how someone comes to understand that someone else's belief conflicts with reality. I suggest that that understanding is not inevitably the product of a predicative statement, that is, in no way has it to be derived from the thought 'That belief is false'. The opposite is more likely to occur, i.e., the perception of someone else's belief would actually be the triggering point for genuine predication, both at language level and at thought level. After presenting this hypothesis, I will defend the necessary theoretical complement, namely, the idea that the profusion of details peculiar to prelinguistic perception does not involve any composition of elements which are attended to separately.

According to this view, the first, pre-grammatical, syntax would have originated as a consequence of the perception of (false or incomplete or out of date) beliefs of others. But what about the inverse relationship? What might I say about the more classic debate about whether or not language is the original cause of the perception of beliefs of others? I am inclined to answer that it is. But this does not lead me into a vicious circle: my idea is that the perception of false beliefs of others would have at its origin a language which is prior to syntax and predication, that is, a language of holophrases* which ask and call.

The perception of beliefs of others would have given rise to the theme/rheme hinge, and consequently opened the door to compositionality. Is this hinge the original syntactic composition? It is certainly an empirically universal structure, one that is audible in intonation. There is, however, an enormous distance between this simple structure and the syntax observed in all languages. In all languages, the form of predication survives independently of its communicative uses, and there are grammatical links and syntactic categories. Could the genesis of the theme/rheme combination have been sufficient for all this to emerge from historical development?

I will mention one last issue in this summary. After looking at the origin of the syntax of predication, or simple sentence syntax, we will have to address the origin of subordinate sentences. We will suggest that subordination became necessary only with 'indirect reported speech'. The distinction between reported false beliefs (that is, the beliefs of a third person) and real information must be made clear since it cannot be left in the hands of the recipient and her optimisation of its relevance.

4. A brief description of the sections of this book

Section One will address the nearest evolutionary antecedents of the human basic ability we are proposing. In the chapter on mirror neurons I suggest that they are not involved in any kind of simulation. Starting from the emphasis that other authors have placed on self-perceptible movements, I propose that mirroring is only a very special kind of expectation. It may be special, but, like all expectations, it can only refer to contents which, under certain circumstances, would be experienced by the subject having the expectation. In the next chapter I take this hypothesis beyond the mirroring observed in monkeys in order to explain a special capability found in chimpanzees. Chimpanzees successfully reckon the visual field of conspecifics and probably also attribute visual perceptions to them. This is indeed a notable achievement. However, if we accept that the ability of the ascribing chimpanzee involves no real simulation – no emergence of a second mental centre – and only relies on a kind of expectation, then that ability will have the insurmountable limit mentioned above: it will never be able to refer to contents inherently impossible for the subject. Or, more precisely, from the very moment a chimpanzee perceives any gaze (or, more generally, any behaviour) directed at him, the animal would cease to ascribe any kind of internal states to the gazing individual (or to the individual whose conduct is directed at him).

In Section Two we begin to discuss the basic ability of the exclusively human mind, in other words, the second mental centre. We will have to study the different modes for processing eyes of others, beginning with the phylogenetically very ancient sensitivity to an eye fixed on us. The second, and much more recent, mode (for processing, in this case not exactly the eyes, but the visual perceptions of others) would be that of chimpanzees: These, as we said in Section One, reckon the visual field of conspecifics and probably also ascribe visual perceptions to them. With the third mode, that is, the ascription of visual perceptions to an eye that is looking at me, we would have before us the launch pad for exclusively human characteristics. Human beings would share with chimpanzees the ability to ascribe visual perceptions, but would have moved a step further. This step, which appears to add only a small difference, actually involves a highly demanding change which also has enormous consequences: the internal states of the fellow would now be perceived as a radically different self, and would, therefore, have to be located in a second centre of the mind. The observing subject could now conceive the self of the fellow as a centre in whose periphery he, the observer, is located. At this point, one inevitably thinks of the classic analogy with heliocentrism. Pointing gestures, including both the gestures made with the eyes or using a finger, as well as four-hand co-operative actions derive directly, I will suggest, from this basic ability (or more exactly, they constitute the original adaptive advantage which would have been responsible for it). Likewise, some children's play, for example play involving motor co-ordination with a playmate or the fun of mutual imitation, are exercises that aim to boost this ability.

Section Three considers the relationship between this basic ability and some requisites of language. It addresses, specifically, Saussurean parity or deep identity between production and reception of language, and also the ability to evoke absent objects as such. On the question of parity, I will bring in especially the 'motor theory of speech perception', or more concretely, its reliable core, 'the motor theory of the observational phase in the learning of complex motor patterns'. With respect to evocation, we will begin by asking if animals are able to evoke. We will then study symbolic play and its adaptive usefulness. In the end, it will be suggested that the big extension (or, in other words, the new, added function) of the 'second mental centre' would consist in the imitation of new and complex motor patterns. In short, what this section aims to show is how those requirements of language might have derived from the human basic ability (that is, from the second mental centre). However, learned signs, Saussurean parity and evocation are by no means resources powerful enough to produce true words. A word is inherently a part of syntax.

Section Four thus begins to turn toward the question of the origin of syntax. I ask how false beliefs of others could be perceived prior to syntactic language. To this end, I will analyse the protodeclarative holophrase*. This kind of communication, which is only useful for linguistic learning and would be, therefore, inconceivable if there were no signs learned through imitation, produces a sign with a precise referential connection in which there will no longer be any ambiguity between request and calling (or between giving a command and asking for an object). This is a dramatically new step when compared to animal communicative signals. But the subsequent consequences of protodeclaratives would have been of greater importance, at least in the historical origins of language. Let us note that the reception of an imperative-vocative message can reveal its producer's false belief if and only if that message has used a sign of the kind provided by the use of protodeclaratives.

In Section Five, I try to show that the *theme* of predication coincides with the (false, or insufficient or outdated) cognitive content the speaker ascribes to the hearer, or, using the terminology of the studies of the Theory of Mind, with the 'false belief' of the addressee. I will begin by stressing that a child's first predications are always reply predications. In addition, I will apply my proposal (that the *theme* is the perceived or conjectured mental state of the addressee) to two classic difficulties: firstly how there can be true and simultaneously non-redundant predicates, and secondly to provide a definition of *theme* suitable for all cases.

The subject of Section Six is the development of predication beyond its suggested origin – that is, beyond *theme* and *rheme* combinations. We shall attempt to dig deeper into the well-known thesis that word links are part of their very meaning, and, focussing on these links, we shall differentiate between three levels of abstraction: all the episodes or uses of the word, typical links, and grammatically syntactic links. With regard to the latter, we will reject the explanation of their historical origin as a decanting of the links in discourse syntax, or theme and rheme syntax. What, then, might this origin have been? In this regard, we will address the underuse of language constituted by the merely expressive, non-communicative, speech: what repercussion would this have on the behaviour of casual listeners? The end of this section continues to deal with the origin of grammaticalised syntax, but in a more general way. Can we place anything specific under this label?

SECTION ONE

Evolutionary precursors

This section aims mainly at providing an interpretation of the capability which chimpanzees probably possess to ascribe visual perceptions to other individuals. Although I accept that chimpanzees are capable of doing this, I will try to prove, nevertheless, that their ability is not tantamount to a simulation of states of others. Consequently, my argument on this issue will have two different parts. The first one will focus on the abilities possessed by non-human primates; the second one will focus on the differences between human beings and chimpanzees.

Before we go on, let's admit that we cannot assume a progression from monkey to chimp to human. Present-day chimpanzees and monkeys need not portray an accurate picture of our common ancestry with these species. In this situation, the study of chimpanzees and monkeys will be a possible road to take (or, at least, not completely illegitimate) if and only if we always remain aware of their limitations.

I consider the probable ability of chimpanzees to be an achievement beyond the mere tracking of someone else's gaze and also beyond the conditioned learning about the relationship of the direction of another individual's torso to its later behaviour. I suggest that this achievement derives from the power to perceive the matching between one's own body and another individual's body. The chimpanzee would reckon what he himself would see from the location, posture and orientation shown by another individual.

However, we humans can ascribe visual perceptions and mental states to individuals whose circumstances we, by definition, will never experience, namely, to individuals who are looking at us or interacting with us. If we assume that the chimpanzee is not capable of ascribing these perceptions that are radically and intrinsically different from his own perceptions, then we will be able to explain his inability to point (and, likewise, his inability for cooperative, 'four-hand' actions) in terms of cognitive incapacity, not only in terms of lack of cooperative motivation. Certainly, I fully agree that among chimpanzees cooperation never became sufficiently adaptive for evolution to give rise to pointing gestures. However, I also think that the increase in cooperation was only the indirect cause of the evolutionary emergence of pointing gestures. The direct cause for this emergence is to be found in a cognitive ability beyond the reach of chimpanzees.

The explanation of the capability which chimpanzees probably possess to ascribe visual perceptions will occupy the entire first Section. At this moment, a possible sceptical and, at the same time, pragmatic objection could be: is the effort really worth it? If, as I have said, my main concern here is to prove that chimpanzees' inability to point is directly and immediately caused by a cognitive incapacity, why do I try to show that chimpanzees are able to perform those ascriptions? ‘Why look for problems?’ might be the comment. But for me this is not an unnecessary problem. I think that chimpanzees really possess that ability. In addition, and at a deeper level, I find that the opportunity to compare pointing gestures with an ability so closely related to pointing (which is the case of the ability to ascribe one’s own visual perceptions to conspecifics) is extremely interesting. Needless to say, all other things being equal, the closer to one another the terms of this comparison are, the more accurate the description of the human peculiarity will be.

CHAPTER 1

Monkeys' mirror neurons

A quick glance at the title of this chapter may have provoked a reaction within the reader: why am I devoting all this space to monkeys' mirror neurons? This ability of monkeys is, or at least that is my opinion, quite far from what is exclusively human. Why focusing on them then? Certainly, what interests me, I insist, is the constitution of exclusively human abilities. However, there is a reason why it is pertinent to include macaques' mirror neurons in this first section. Studying them will allow me, above all, to strengthen a specific interpretation of chimpanzees' abilities. This interpretation places non-human primates completely outside genuine simulation (or, in other words, latent imitation).

1.1 Mirror neurons in macaques, a significant discovery and a controversial interpretation

At the beginning of the 90s, a team of researchers from the University of Parma discovered neurons in the cortex of macaques which were activated both when the animal (with its eyes closed or open, it made no difference) grasped an object with its hand and when it saw hands of others grasping an object (Gallese et al. [1996]; Rizzolatti et al. [1996]). They called these 'mirror neurons'. Mirror neurons quickly attracted the attention of philosophers and psychologists.

At first, some authors related them to imitative learning, or, in other words, to the imitation of complex and new motor patterns. (In Stamenov & Gallese, eds. [2002], we see works supporting these opinions as well as others rejecting them.) Although it is most likely true – I am convinced it is – that mirror neurons represent a landmark in the evolutionary line which will ultimately lead to human imitative abilities, this is no justification for attributing this function to the neurons of animals which are not at all capable of motor learning. If these neurons have an adaptive function, such a function would have to be adaptive for macaques themselves. As a result, those voices which immediately related them to imitative learning have tended to fade away. But when we remember these attempts we can derive a more general lesson. This lesson, which would perhaps still be useful today, is that, for now, it would be wise to clearly separate mirroring in monkeys from mirroring in humans. There would thus be no risk of the controversial question of their adaptive function in macaques being hidden behind solutions or suggestions that are only admissible for the study of mirroring in humans.

There is a tendency, nowadays, to associate these neurons with the subtrend in the 'theory of mind' which has been labelled simulationism, or with the interpretation and prediction of behaviour of others, or even with empathy. My own view differs from these also. I am much more in accord with authors who stress above all the role played by self-perceptible movements in mirroring. But first I want to critically analyse the widely spread view of the social utility of mirroring.

1.2 On supposed 'social' utility: Is the role of macaques' mirror neurons to understand and predict behaviour of conspecifics?

Contrary to what today is perhaps still the most widespread opinion, I believe that mirror neurons would not originally have the role of understanding or predicting the conspecific's behaviour. Clearly, if a macaque observes that a conspecific is grasping a fruit, the observing macaque is informed that that fruit is now less available than before. Equally clear is that this information will, on occasion, be useful to it. My point is, however, that in order to acquire that useful information, visual perception of movement and goal would be sufficient. In other words, the acquisition of this information would not require any association whatsoever with one's own grasping.

In addition, if the role played by mirror neurons were to understand or predict behaviour of others, then a piece of data which has been firmly established since mirror neurons were first discovered would remain unexplained. I am referring to the fact that mirror neurons – or, at least, the mirror neurons which have been discovered up to now – are never activated when the hand that is seen is not grasping an object. No matter how much movement this hand makes, the observer's mirror neurons will not fire. Were we to accept that the information that can be associated with mirror neurons is the meaning of the behaviour of others, this well-established piece of data would, I repeat, be inexplicable. A hand which moves forward to hit, or (in chimpanzees) to beg from, whoever is observing it would undoubtedly be performing a behaviour which is relevant for the observer. However, the observer's mirror neurons are still not activated in this case, no matter how interesting the perception of the behavioural meaning of these hand movements may be to the observer.

We have, thus, followed two routes with the aim of raising doubts about whether the role of mirror neurons is to interpret or predict behaviour of conspecifics. On one hand, we have seen that the firing of mirror neurons – or, at least, the mirror neurons which have been discovered up to now – does not occur when faced with some behaviours of others which are indisputably relevant to the observer. On the other hand, we have argued that, when the firing of mirror neurons actually does accompany some behaviour of others, it cannot be said that this firing is useful in interpreting the behaviour nor in predicting the following step. On the basis of these points, we have some support to reject this as the function of mirror neurons.

In summary, it does not appear to be at all clear that, at their origin, mirror neurons have the adaptive function of understanding or predicting the behaviour of others. Fortunately, there is no need to insist on this *pars destruens* of my proposal. Hickok (2009) has sufficiently insisted on it.[1]

In addition, we should bear in mind that the hypothesis of the social origin of primate intelligence, which, beginning with Jolly (1966) and Humphrey (1976), had completely dominated until now, is beginning to be viewed more reticently. If wolves and hyenas also live in packs, why would the simple fact that primates live in groups have special consequences for them? This objection, about which many of us have given thought, has been set out authoritatively by Holekamp (2007). Do not get me wrong. I am not denying the crucial importance of social living for the evolutionary emergence of the human basic ability. But for that emergence to take place, other factors must have concurred. The following chapters deal specifically with this issue.

1.3 Mirror neurons, a secondary effect of self-perceptible movements

As a result, the view, for the moment a minority one, which sees the origin of mirror neurons as merely a secondary effect of self-perceptible movements, seems to me to be preferable – preferable, at least, to the famous 'social utility'. Only later would there have been an exaptation* of this secondary effect for a useful purpose. So far as I am aware, Oztop & Arbib (2002) were the first to class the later playing of 'social' functions by mirror neurons as 'exaptation'. As Hurley (2005) said, monkeys' mirror neurons would originally be only a 'secondary effect' in evolution. The most developed hypothesis within this line is by Keysers & Perrett (2004).

1.3.1 Self-visible hands: Connecting Keysers & Perrett with Piaget

In Keysers & Perrett's proposal, the self-visibility of the hand is a decisive factor. Motor orders to the hand are normally simultaneous with the viewing of hand movements. The direction of this association, from the motor to the visual, may be inverted later, and this leads to mirror neurons' characteristic function. Keysers and Perrett invoke a Hebbian type of learning ("Cells that fire together, wire together"). By turning to the visibility of one's own hand, their hypothesis succeeds in explaining why the function of mirror neurons relates in the first instance to hand movements. In the same way, Keysers and Perrett might succeed in justifying, at least in some degree, the fact that

---

[1] Although they study human subjects (and although, as I have said, it is wise to clearly separate mirroring in monkeys from mirroring in humans), see also Brass et al. (2007, p. 2117): "We show that brain areas that are part of a network involved in inferential interpretive processes of rationalization and mentalization but that lack mirror properties are more active when the action occurs in an implausible context. However, no differential activation was found in the mirror network."

reliable than the other. Before further inquiry into the mouth-related mirror neurons of macaques, we should perhaps have a look at other data.

The internal division that I have referred to in Keysers and Perrett's proposal echoes a debate which has been going on since the early 80s, when the discovery that newborn babies are able to imitate facial gestures upset the Piagetian model of motor imitation development. According to Piaget, the correspondence between mouths would emerge from two different routes, both related, nevertheless, to self-perceptible movements. On the one hand, the sounds of actions such as licking, which can be self-audible, i.e. audible to the agent performing them, not only to an observer, would be extremely important. On the other hand, the matching between the mouths could also be derived from the matching between self-visible hands: One's own hand carries the food to a hole that is felt but not seen, and the hand of others carries the food to a hole that is seen but not felt. In this respect, Piaget speaks of intelligent mistakes which reveal how hard it is for children to acquire the matching of the mouths: numerous observations confirm that the child sometimes tries to coordinate someone else's mouth opening and closing with the opening and closing of his or her own eyes, or conversely.

Let us note that there is an important difference between Heyes, who (in 1.3.1) we introduced above as close to Piaget, and Piaget himself. Ray & Heyes (2011, p. 101) say: "Imitation of mouth opening and lip smacking *when accompanied by sounds* (...) is a perceptually opaque action." I do not agree with this classification. Mouth opening and lip smacking *when accompanied by sounds* are self-perceptible (more concretely, self-audible) movements, or, in other words, perceptually transparent actions.

But the whole edifice of the Piagetian model was shaken when Meltzoff & Moore (1983) got a very high percentage of new-borns, only a few hours old, to imitate movements as opening their mouths or sticking out their tongue. The question then arises: in the face of these data, can we still maintain that the child's understanding of matching between mouths takes place only later and relies on either of the two routes (self-audible sounds and relation to self-visible hands) or both, as suggested by Piaget?

(Again I stress that, in my view, the child's perception of the matching between mouths is the key issue. I am focusing on the ability to see somebody else' mouth as a counterpart to one's own mouth, not on behaviour. By contrast, Ray & Heyes [2011] say: "Infants do not need to detect or recognize that they are being imitated in order for this experience – which appears to be plentifully available in typical development – to support the learning of matching vertical associations."[3] Thus, I ask the reader to keep in mind that my use of 'correspondence' or 'matching' is different to that of these authors.)

---

[3] In order to play down the importance of understanding of homology as the base for imitation, Heyes et al. (2005) showed that a brief period of incompatible sensorimotor training – in which participants responded to hand opening stimuli by closing their hands, and to hand closing stimuli by opening their hands – abolished automatic imitation, e.g. the involuntary tendency to make an open hand response faster to an opening than a closing hand stimulus. But I reply that the strenght of verbal instructions in adult humans (or of conditioning in animals) is beyond all doubt. Therefore, in my view, those experiments do not refute at all the idea that such

Since the experiments of Meltzoff & Moore, a lot of research effort has been concentrated on this issue. After Anisfeld (1991) (1996, p. 60: "only tongue protrusion modeling is matched by neonates"), two attractive and mutually compatible explanations have arisen. Jones (1996) shows not only that tongue protrusion occurs to some extent in response to any interesting visual display, but also that the 'virtual' exploration with the tongue stops as soon as infants begin to reach with their hands: "Infants produced tongue protrusions in response to objects within reach before but not after reaching developed. Our results suggest that infants' tongue protrusions in response to a tongue-protruding adult reflect very early attempts at oral exploration of interesting objects" (p. 1970). Nagy & Molnar (2004) focus on the very frequent 'delayed imitations' observed in neonates, and they suggest that such 'delayed imitations' were actually serving as provocations for interaction. These 'delayed imitations' were accompanied by a totally different heart rate pattern than the 'imitations': the heart decelerated, indicating greater preparatory attentional focus on the acts of the other, in contrast to the clear acceleration characterising 'imitations'. Obviously, Nagy & Molnar are close to what Ferrari et al. (2006, p. 302) suggest about macaques ("Macaques' neonatal imitation may serve to tune infants' affiliative responses to the social world"). However, in spite of all these research efforts, the question has still not been fully answered.

Neonatal imitation is not the only problematic issue, however. Other facts could easily raise a similar question. Let us consider, for example, the suggestions made by authors like Hurford (2004, p. 305) or Hurley (2005), who (although they also stress the contrast between activation of mirror-neurons and open behaviour) link mirroring with the almost certainly innate mechanism responsible for birds banding together in flocks, or fishes in shoals. Here there is no self-perceptible movement at all. Is this behaviour, nevertheless, akin to mirroring? Or is it something radically different?

Against this background, we should ask ourselves a more radical question. Is monkeys' mirroring caused by self-perceptible movements? We need to ask with Zentall (2003, p. 94): Are mirror neurons' visual-motor (or audio-motor: infra, 1.5) connections "prewired neural pathways or, on the contrary, do they have to be trained?" In this regard, the following experiment would be helpful. From the moment of birth, a macaque is prevented from seeing its own hands. Without modifying this prevention in any way, it is given the opportunity to see hands of a conspecific. Would the mirror neurons be activated in this case?[4] Admittedly, there is no clear-cut evidence on this issue so far.

---

understanding is actively present in primates (although in a different way and degree in macaques, chimpanzees or humans).

[4] Certainly some may say: It is well established that mirror neurons are the result of learning in that, e.g., there is mirroring for paper tearing and tool use (see these data in Kohler et al. [2002]; Umiltà et al. [2008]). However, this answer does not persuade me. The fact that the peculiarities of the paper or the tool must obviously be learned does not necessarily imply that the basic connections cannot be innate. Consequently I think that the suggested experiment would be helpful.

In short, despite much research, we do not yet have answers for our questions or, more precisely, we still cannot decide definitely which alternative is the valid one. According to one of these alternative responses, mirroring is close to social innate coordinations and not necessarily linked to self-perceptible movements. In other words, according to this first alternative response (I am going to refer to it in this way, regardless of the chronological emergence of the theories), unification takes place because self-perceptible movements are no longer viewed as the core and paradigm of mirroring. By contrast, the second, alternative response suggests that genuine mirroring begins with the learning prompted by self-perceptibility and, on the other, that neonatal imitation does not connect with the perception of the matching between one's own mouth and someone else's mouth until a later stage in ontogenetic development. More precisely, neonatal imitation would only become genuine mouth-related mirroring after one or both of the two routes suggested by Piaget were fully covered. In this second response, the key element of mirroring rests, as in the first part of Keysers and Perrett's proposal, on self-perceptible movements and on the need for consistent learning. This also implies a unification of mirroring, but this time unification takes place because neonatal imitation is neither interpreted as genuine mirroring nor as being sufficient to cause true mirroring. (Keysers & Perrett [2004] cannot be ascribed to any of those choices. As I said before, I think that there is an inner division in these authors' proposal.)

I insist that, in my view, so far there are not sufficient data to make a decision. Both rival hypotheses can be defended. Obviously, this has a restrictive consequence: neither hypothesis can be taken as evidence. But it also has a consequence in the opposite direction. As part of a hypothetical-deductive approach a researcher can take either of the two as his or her starting point. Thus, I will opt for the hypothesis which relies on self-perceptible movements. More precisely, I will be (relatively) close to the first part of Keysers and Perrett's proposal. Consequently, my explanation of mirroring, relying heavily on the importance of self-visible hands, could be described as primatocentric. (But self-audible learned songs might be a different source of mirroring: see infra, 1.6.)

Let us sum up. I have rejected the adaptive utility that Rizzolatti, Gallese and the other members of the Parma group ascribed to macaques' mirror neurons, and, by contrast, I agree with Keysers and Perrett in ascribing an important role to the self-perceptibility of the hand. However, the question which most interests me, as I have said, is not this, but the question of the mechanism which underlies the activation of mirror neurons when faced with action of others. Until now, the peculiar function of monkeys' mirror neurons has been understood as off-line or latent motor imitation (or, in other words, as simulation). However, I believe there are reasons to doubt this.

1.4 Simulation or expectation? The crucial question about the abilities of non-human primates

The purpose that will guide us in the rest of this chapter has already been clearly outlined. I want to look for arguments against the possibility that there is simulation involved in macaques' mirror neurons.[5] Although I accept that mirroring involves some kind of connection with someone else's inner state, I do not consider that genuine simulation or, in other words, latent imitation is involved. As can be seen, I am about to embark on a journey in a direction opposite (and complementary) to the one imposed by my choice at the end of Subsection 1.3.2. Then, I separated mirroring from other types of behaviour unrelated to the hand, behaviour which is earlier in the course of ontogeny. By contrast, now I am going to describe mirroring as being different from true simulation, which probably is peculiar to humans. In Section Two we will see why this difference from simulation is an important issue. However, the immediate task can be defined more specifically. We shall set the foundations of an alternative explanation piece by piece.

1.4.1 Animal behaviour and expectation

To begin with, we must stress how the expectation of results is a key concept for animal behaviour. The brain emerges primarily to provide the animal with the best choice of behaviour at every moment. Once we accept this, the expectation of results is foregrounded. It should be remembered that the brain has outlined the effects it wishes to achieve not only before it undertakes any behaviour, but also before it selects the appropriate movements and means at every moment along the way: it is due to a precise expectation of results that animals can do without the inefficient preset motor plan resource (cf. e. g. Thelen & Smith [1993]).[6] This expectation of effects is at once the motor and the guide for behaviour; in other words, it not only generates but also selects behaviour. The expectation wishing to be satisfied marks the beginning of any behaviour, and the satisfaction of such expectation marks its end.[7]

---

[5] A possible objection could be that, since monkeys do not imitate (to any great extent), these arguments seem unnecessary. I disagree, however. We must bear in mind that any imitation of simple movements (something completely different from motor learning, i.e., from imitation of complex and new motor patterns) is of no use whatsoever, and I will comment on this later in 2.2. Consequently, the fact that no external imitation has ever been observed in macaques does not necessarily dismiss the interpretation of mirroring as internal simulation. A defender of such an interpretation could argue that no researcher can ever make macaques understand the command to perform a completely useless external imitation. (See also Paukner et al. [2009], who suggest that capuchins monkeys display affiliation toward humans who imitate them.) Consequently, if we do not accept, and I certainly do not accept, the interpretation of mirror neurons in terms of simulation (i.e. latent imitation), then we must provide another reason. More precisely, we must provide an alternative interpretation for mirroring.

[6] An extreme case in this sense (i.e. an extreme argument against innate, preset motor plans) can be found in some 'freaks of nature' (Blumberg [2009]). Of course, we humans perform preset motor plans when we imitate new and complex motor patterns. But these complex plans are not the issue here.

[7] An extended concept of homeostasis could involve all animal goal-driven behaviours (Richter [1943], cited e. g. in Bechtel [2009, p. 165–166].

Expectation – of effects not sought, in this case – is also invoked to explain so-called 'attenuation'. The concept of attenuation was coined in relation to the two types of movement possible in retinal images: movement caused by the environment, and movement which depends on the movements of one's own body. As has been known for a long time (Von Holst [1954] or Sperry [1950]) the brain only takes note of the first; the second is thus attenuated.[8] Attenuation has been applied most recently to the tickle sensation, or, more specifically, to explaining why it is not possible to tickle oneself: Blakemore et al. (1998) and (1999). In every case, the sensations caused by movements of one's own body are attenuated because they respond to a previous expectation in the brain, to a mechanism prior to the real movement (see also Bompas & O'Regan [2006a], [2006b]). As regards all types of attenuation, Wolpert's group has undertaken in recent years to disprove the hypothesis of a postdictive mechanism, i.e. a mechanism subsequent to movement: Voss et al. (2006), Bays et al. (2006). As a result, we can conclude that expectation of the movement's effects, aside from its principal function as a guide for behaviour, is also capable of explaining the phenomenon of attenuation.

"Test – Operate – Test – Exit" was a motto of the first cognitive revolution (Miller et al. [1960]). I don't like what I have; I act; I like what I get; end. Whichever terminology we may adopt, this sequence of steps describes goal-driven animal behaviour. Naturally, however, we must place the profile of the state we are seeking in front of these, as the first step. If this profile did not exist in the animal beforehand, there could be no behaviour. (For a brief review of this matter, see Carver [2005].) Expectations which seek satisfaction (or, in other words, empty profiles which urge the organism to fill them) are absolutely necessary elements both for opening and for closing – or, rather, for satisfactorily closing – any behaviour. Lastly – that is, at the beginning, and, therefore, also at the end –, would come innate consummatory patterns, i.e., the "teaching mechanisms" about which Lorenz (1966) spoke. But with learning and conditioning, many other expectations, no longer innate, take root in the animal (see, for example, Bar [2007]); these learned expectations cannot be the end goals for animal behaviour, only subgoals.

Moving beyond this general framework, and on finally to what we are interested in, let us look at the step posed, within any behaviour, by a simple movement. Here, too, the starting point would be an expectation outlining the effects sought. This expectation, which, like all expectations, would only be deactivated when it was satisfied,

---

[8] Expectation and attention both facilitate the interpretation of perceptive contents, but attention strengthens sensations while expectation *at times* attenuates it. Cf. Summerfield & Egner (2009, p. 405): "Given that attention and expectation have similar facilitatory effect on visual object recognition, one might anticipate that expected (relative to unexpected) stimuli would also be associated with enhanced sensory responses. However – strikingly – the opposite is in fact typically the case: expected stimuli tend to elicit reduced visual responses, relative to their unexpected counterparts, and an extensive literature has documented the corresponding phenomenon in the auditory domain."

such results are reached: etymologically speaking, this is what is suggested by the term expectation. On the other hand, in mirror neurons in their mirroring function, according to our suggestion, the expectation of results would be being activated after the observed grasping has been performed. We need, therefore, to propose a time-relative inversion. Keysers & Perrett, in contrast, have no need for this. For them, as for many authors, the firing of mirror neurons in response to a merely observed grasping means the latent, inhibited, activation of a movement, not an expectation of postural results. And, therefore, although Keysers & Perrett speak, of course, of an inversion of direction (the motor to visual direction would become, after Hebbian learning, visual to motor), they do not need to postulate any time-relative inversion. The performance of the movement and its visual perception occur practically simultaneously. In contrast, with the expectation of results things are very different. In my suggestion, the expectation of postural results would be located differently for each of the two roles played by mirror neurons. In the case of the expectation of grasping which the subject will immediately perform, the expectation is prior to movement: this is what occurs in mirror neurons' 'motor' role. In contrast, when the grasping is merely observed, the postural expectation is *a posteriori* to movement: this is the interpretation of mirroring that I am suggesting.

Observe that the link or the shared element in these two different situations in which there is firing of the mirror neurons is by no means less clear than in the other interpretation. It may even be the opposite. Let us compare both interpretations. If we interpret that the firing of mirror neurons corresponds to motor commands, we must accept that that motor command would in some cases produce latent motor activation (the mirroring role of mirror neurons) while in other cases it would produce unfolded motor activation (in the so-called motor role of mirror neurons). If, by contrast, we interpret the firing of mirror neurons as corresponding to expectations of postural results of a movement, it is true that we must inevitably distinguish the cases in which the expectations are going to be fulfilled by the relevant somatosensory sensations, i.e. in the so-called motor role of mirror neurons, from the cases in which they are not going to be fulfilled, i.e. in the typical mirroring role. Nevertheless, the expectation as such would remain the same. In short, the expectation of postural results is a good candidate for that element – that factor shared by the two different situations of the firing of mirror neurons – which any explanation of the mirror neurons' function must discover.

The objection that could be made of my hypothesis is, of course, that it needs to postulate a very peculiar type of expectation and perhaps a strong new development in evolution. Throughout animal evolution, the expectation of results would always have come before movement pursuing those results. But with mirror neurons a revolutionary, *a posteriori*, expectation would have appeared. These postulated expectations are so special that the criteria of parsimony or simplicity seem to provide a ruling contrary to my hypothesis.

My reply would be that, should there turn out to be other reasons in favour of this interpretation, we would then be fully within our rights to postulate this strong new development in evolution. Everything therefore rests on other reasons being found which might support the hypothesis. What might these reasons be?

The main argument I can offer lies in stressing the extreme complexity and high demands which a latent motor imitation (or, more generally said, a simulation) would require. This complexity is hinted at in some areas of the bibliography, but we have not yet come to the heart of it, I believe. Let us consider this carefully.

Focusing on mirror neurons, Gallese (2003), Gallese et al. (2004) and also Hurley (2005) have insisted that 'we' existed prior to 'I'. This immediately raises the question of how movement of others is then differentiated from own movement. By contrast, if we accept the interpretation of mirror neurons in terms of the expectation of postural results, the problem of how to differentiate between another's and one's own states disappears. If (as with the typical situation of mirroring) the grasping is merely observed, i.e. if it is somebody else' grasping, then the expectation of postural sensations will not be satisfied.

The problem of this differentiation appears in other authors. Thus, when, addressing simulation (or off-line imitation), Brass & Heyes (2005, p. 493) asked "why do we not imitate all the time?", they replied that the key must be found in the distinction between the subject and the observed individual (see the 'Who?' system proposed by Georgieff & Jeannerod [1998]; see also Decety & Chaminade [2003]). I believe that this latter type of response – the distinction between subject and observed individual – is close to being correct when we are interested in this question only in human beings. But we should return to macaques and see what would be implied by interpreting their mirror neurons in terms of latent imitation.

In my opinion, granting motor simulation to macaques' brains is probably to underestimate the difficulty and complexity involved in such simulation. We must not forget that the simulating organism has to keep itself informed about its own kinaesthetic and postural states. I will formulate this idea through a critique of Glenberg (1997). This author, commenting on how past episodes are evoked, says that the subject's awareness of its real situation is suppressed during these evocations. I believe this to be an incorrect way of explaining what takes place. Such suppression occurs only during sleep, that is, while the subject's motor activity is switched off and the subject is in a relatively safe place. Under different circumstances to these, suppressing the awareness of real circumstances might be extremely dangerous. What is the consequence of all this for our discussion? The consequence, in my view, is that genuine motor simulation must involve a double line of information. On one hand, a line for the state of the simulator's body; on the other, a line for the state of the observed body. This duality, I suggest, is so extremely demanding that only the human brain can sustain it.

(But, when one animal fights another, must it not exhibit something of this duality?: this could be a plausible objection. However, I think that the prediction of an opponent's behaviour needs no association whatsoever with one's own similar movements. In other words, as I said in 1.2, the visual ('from the outside') perception of movement and goal would be sufficient to make such predictions. Thus, if we admit that these predictions are independent from mirroring, they will be *a fortiori* independent from the duality that which we are talking about.)

In contrast, the expectation resource in animals does not require this duality at all. It is true that normal expectations have to do with a later moment. But this does not at all mean that it needs to be addressed as an environment additional to the current real environment.[12] Quite the opposite, expectations are built-in constructions (more specifically, empty but, nevertheless, well-defined profiles) which often exert their influence precisely by guiding attention toward the current real environment.

In short, we have two contrary interpretations of mirroring. Whereas I explain this through a new type of expectations (i.e. *a posteriori* expectations), many proposals, in contrast, state that mirroring involves some form of latent motor imitation. Different authors classify mirroring in highly varied ways: direct resonance and 'we' perception *versus* (Heyes, or Keysers & Perrett) a (relatively) more Piagetian process; stressing competition *versus* stressing empathy; understanding of behaviour of conspecifics *versus* (Csibra [2007]) prediction of the immediate future behaviour of conspecifics. Nevertheless, none of these controversies is of any interest to me now. It is the interpretation of mirror neurons as involving latent motor imitation (or "emulative action reconstruction": Csibra [2007]) that I wish to question for monkeys' and apes' mirror neurons.

My effort to differentiate simulation from a posteriori postural expectations, my obsession with that difference, may seem to be splitting hairs. If the reader has had this impression, I beg for his patience until we use that difference, or more specifically its derivation (that is, that chimpanzees would have the ability to understand the correspondence between their own body and that of a conspecific, but they would lack simulation) to build a more general proposal.[13]

As I have suggested, that ruling provided at the outset by the criteria of parsimony or simplicity (a ruling, it should be remembered, which originally seemed contrary to my hypothesis) may now change. Postulating a simulation (i.e., a latent imitation or an emulative action reconstruction) in a macaque's brain may perhaps be much less parsimonious than postulating an evolutionarily new type of expectation for self-perceptible movements. In addition, the speculation I shall now put forward – if in the future it were to receive some confirmation – might support the idea of *a posteriori* expectations.

---

[12] See again the paragraph (in 1.4.1) about forward-models. With regard to the two alternatives (mock input versus a more simple and peripheral mechanism), it is the second one that nowadays is most likely true for biological systems.

[13] In fact, in the elaboration of my proposal the issue of mirror-neurons was included very late. And my curiosity regarding them was whether they entailed stimulation or not.

1.5 An adaptive but 'non-social' role? A speculation which would act as an argument in favour, were it to enjoy slightly more support

An arboreal primate moving from tree to tree finds it very useful to see which branch its own hand is grasping. Only with this type of visual perception would it be possible to prove the solidity of the branch, as well as to choose which nearby branch to move to next. However, although visual perception of one's own hand is extremely advantageous, it could also entail a grave danger. If, erroneously, the visual perception of the hand of somebody else is taken as the visual perception of one's own hand, making use of this information may be catastrophic. If the branch grasped is not sufficiently solid, yet the animal trusts that grip because it trusts a visual perception (or equally, if, trusting the visual perception, it decides to move towards a place where there is no branch), the likely result will be that the animal will fall from the branches. Consequently, there would be strong selective pressure at some very early point in the evolution of primates in favour of a mechanism which would allow them to distinguish between their own hand and the hand of somebody else. This would be, I suggest, the role originally played by visual mirror neurons – or, more precisely, by these neurons in conjunction with the somatosensory postural area. The animal has to compare the postural expectation that it detects in the seen hand with its own somatosensory information.[14] Only if the two postures coincide will the subject be dealing with its own hand.

However, is it really necessary to postulate an *a posteriori* postural expectation in order for the monkey to be able to avoid these dangerous confusions? I would respond as follows. Certainly, this confusion would be much more easily overcome for a subject that can freely move its hands. As soon as the subject makes the slightest movement with its hand, it will succeed in clarifying whether or not the seen hand is its own hand.[15] This 'move-and-see' strategy deserves to be called 'the easy route' towards that

---

[14] Postural expectation *and* somatosensory information. Cf. Heilman et al. (1998), who have recently added an interesting aspect to the traditional explanation of the anosognosia. According to the traditional explanation, anosognosia results from an inability to represent current body states automatically and through the appropriate signaling channels, i.e., the somatosensory systems. Heilman et al. suggest that the patients *also* lack an intention to move (or, as I would say, lack postural expectations) and are thus robbed of a means to check their defect easily.

[15] Let us examine some data regarding macaques' brains. These data, it should be noted, refer to hand movements that are *different* to the movements involved in grasping an object, and likewise refer to an area of the brain very *different* to that of mirror neurons. "We have measured responses of visual movement sensitive neurons in the anterior part of the dorsal Superior Temporal Sulcus of monkeys to stimulation caused by the animal's own active movements. These cells responded to any stimuli moved by the experimenter, but gave no response to the sight of animal's own limb movements" (Hietanen & Perrett [1993, p. 117]). Of course, it is not a case of the movement of one's own hand being invisible; clearly, the animal will see its own hand move. However, since this has been self-generated, the movement appearing in the retinal image will

clarification. (Cf. Hogendoorn et al. [2009].) However, it should be remembered that a hand holding an object is less able to move.

We should comment in a more general way on the speculation that this subsection has suggested. Our initial point was that discovering whether a seen hand is another's or one's own hand is not an adaptively useful task *except* when this hand is grasping a branch. We can now add that the sophisticated route provided (under the present speculation) by mirror neurons is unnecessary *except* when the hand is already still and prevented from making any new movement. In all other situations, it is enough to make a movement with one's own hand to determine if the seen hand is another's or one's own hand ("Agency structures body-ownership": Tsakiris et al. [2006. p. 423]). The two exceptions mentioned coincide with one another – that is, the only type of situation where that utility occurs and the only type of situation where that need occurs. Let us see this in a more detailed way. When does the compulsory stillness occur which prevents 'the easy route' from functioning and which, thus, makes the sophisticated route (i.e., the intervention of mirror neurons) necessary? This occurs only when the hand is already grasping an object, which is an indispensable requirement for mirror neurons associated with the hand to be activated. We have, thus, a triple coincidence: the situation responsible for the utility, the situation responsible for the need, and the situation where the mirror neurons are activated.

Thus, according to my suggestion, the absence of manual movements would be involved in mirroring in two different ways. Firstly, this absence is the key of the type of situations where the adaptive advantage of mirroring originally arose. Secondly, it is a necessary element in the mechanism of mirroring (and consequently it must be present in any activation of mirror neurons in their mirroring role). The animal has to compare the postural expectation that it detects in the seen hand with its own *current* somatosensory information. Note that, should there be any manual movement of the animal, this somatosensory information would then change, and consequently 'the sophisticated route' towards the useful comparison would become an impossible route. (Please note that it will become impossible even in those cases in which it would have still been useful – i.e., in cases in which the hand previously perceived is hidden behind an object or when in a runaway or fight situation it is not advisable to waste time by looking again at the hand perceived).

Kraskov et al. (2009, p. 922) have discovered that "many pyramidal tract neurons in area F5 showed complete suppression of discharge during action observation, while firing actively when the monkey grasped food rewards". This finding can be interpreted in two different ways. According to an interpretation, this "inhibition" of self-movement during action observation could answer the famous simulationist question (why do animals with mirror-neurons not imitate all the time?). By contrast, in my

---

be subject to attenuation, that is, it will be excluded from the set of information about the world (remember 1.4.1, about attenuation).

expectation of the results of a movement is previous to the movement. These previous (that is, classic) expectations are of no usefulness in differentiating whether the seen hand holding an object is another's or one's own hand. In this situation, the hand has already finished its movement and has already satisfied previous expectations. It is therefore in this situation where the task of discrimination between one's own seen hand and the seen hand of the conspecific must turn to the sophisticated method, i.e. to mirror neurons in their mirroring role. Now for the problematic fact: very shortly after mirror neurons were discovered, it was noticed that they were also activated in the case of a hand which was hidden immediately before it reached the graspable object toward which it was moving (Umiltà et al. [2001]). Here, however, I have explained mirroring based solely on the already performed grasping function.

How can we face the challenge this datum poses to our suggestion? Clearly, our suggestion would be unsupportable if this were the only situation where mirror neurons are activated. In this situation, the sophisticated mirror neuron mechanism is entirely unnecessary in order to detect whether the hand is another's or one's own. The immediate movement or absence of movement in the subject would be sufficient. However, since this is by no means the only situation, the datum is explainable within the lines of our suggestion. Here, the prediction of grasping would have been produced by solely visual means, i.e. without any motor simulation. More concretely, this prediction would be supported by dedicated neural substrates within the visual system, most noticeably the superior temporal sulcus (STS). Let us pay attention to the discovery of STS cell populations coding for actions in relation to contextual cues rather than for actions per se (Jellema & Perrett [2005]). This prediction, which at first would have been derived, I repeat, from mere visual perception of trajectory, could then unleash an identical reaction to the reaction provoked by the visual perception itself. We already know what such a reaction to visually perceived manual grasping consists of: according to my hypothesis, there will be an a posteriori postural expectation reaction.

As is evident, we have had to accept an assumption in making this explanation: the prediction of the grasping would first have emerged as a merely visual prediction. (*Supra*, in 1.2, I have focused on this type of prediction.) Clearly, this is no more than an assumption. However, it is one that, it turns out, is quite probable. We know that mirror neurons will not be activated in this situation if the trajectory of the hand is not adjusted to the necessary requirements for 'good visual coherence'.

A different attack on the usefulness we have speculated may come from the discovery that mirror neurons also respond to grasping made with instruments: Ferrari et al. (2005). Here, as the reader may already anticipate, I will insist on the datum that this habituation needed many attempts and lasted for months. Likewise, I will reiterate that the originary utility of a resource does not have to appear in all its uses.

But we must not stick to just this defensive comment. Such comment is certainly appropriate in the context of arguing in favour of 'adaptive but non-social role' of original mirroring. However, Ferrari et al. (2005) and, even more, Umiltà et al. (2008)

may be interpreted as highly compatible with my more nuclear proposal, i.e. with a posteriori expectations. In this latter study, monkeys were trained to grasp objects using two types of pliers: normal pliers, which require typical grasping movements of the hand, and 'reverse' pliers, which require hand movements executed in the reverse order (that is, first closing and then opening the fingers). The results showed that mirror neurons discharged during the same phase of grasping in both conditions, regardless of whether this involved opening or closing of the hand. So, there would not be a faithful simulation of movements. This type of data clearly favour Csibra (2007, p. 436)'s proposal ("action mirroring in the observer is achieved not by direct matching but by emulative action reconstruction"), and, likewise, would have led to the emphasis on the part of Rizzolatti & Sinigaglia (2010, p. 268) in that "parietal-frontal mirror neurons – owing to their motor nature and the fact that they encode the goal of motor acts – can be triggered by different visual stimuli that have a common goal (for example, grasping)". But with equal, or even greater, clarity they favour my proposal that in mirroring there would not be latent motor acts, but only a posteriori expectations.

Let us return to the adaptive but non-social role. Hickok (2009) concludes his critique of the dominant theory about mirroring with the following words: "the action understanding theory (...) has distracted the field away from investigating other possible (and potentially equally important) functions". I do not think that he is referring to the above speculation. As far as I know, nobody has yet taken that possibility into account. However, this, namely, "another possible (and equally important) function", is exactly what this speculation of mine has suggested for mirror neurons.

As I have already said, if this speculation were to be confirmed, it would be a support for the hypothesis we are really interested in making. Note that it is only through *a posteriori* postural expectations (in conjunction with the absent or present somatosensory information, of course) that another's and one's own grasping can be differentiated. In contrast, motor simulation, far from being the cause, would in fact depend on this differentiation (Let us remember Brass & Heyes [2005, p. 493]: "why do we not imitate all the time?".)

A further question remains unanswered after all this: Why would mirror neurons associated with the hand have led to the emergence of mirroring (true mirroring: supra, 1.3.2) associated with the mouth? Why would this extension have arisen? A first possibility is to see this as merely a consequence of the association between one's own hand and somebody else' hand. (Keep in mind the two Piagetian paths signaled above, in 1.3.2. Both paths are associated to self-perceptible movements.) The extension would thus originally have taken place outside any useful function. A second answer would be that the expansion towards the mouth would have had a useful function which was not shared by mirror neurons associated with the hand. But the latter probably has to be discarded. Certainly, in chimpanzees, we will propose (in the next chapter) that the understanding of the matching of mouths would have a useful consequence, that is, the understanding of matching between own body and foreign body. However, in macaques, it is much more doubtful that such matching between whole bodies takes place. Therefore, it seems that we have to dissociate from any kind of utility the expansion towards the mouth.[18]

1.6 The relationship between the central hypothesis of this chapter and the above speculation

But we must concentrate on the key issue. The central hypothesis of this chapter is independent from the above speculation. More precisely, the central hypothesis can admit other possibilities without any problem. One of these possibilities is that the novel type of expectation, i.e., the *a posteriori* expectation would have originally emerged without any adaptive utility (i.e. as "a mere Hebbian learning") and only afterwards did exaptation take place.

Another possibility is that mirroring – that is, *a posteriori* expectations, according to my hypothesis – originally emerged in self-audible movements (birds' learned songs) and not in self-visible movements and, consequently, was originally unrelated to the hand (something which would not prevent the hand from being a crucial element in the origin of visual mirroring). Against this background, self-perceptibility would be a block shared by different situations. The adaptive advantages would be different for each situation. Birds have an interest in comparing the singing they hear in their environment and the singing they produce (this issue will be seen *infra*, in 6.2 and 6.3.2). For monkeys, by contrast, the advantage would consist in a distinction between one's own hand grasping and that performed by other individuals. However, despite this variable functionality and despite the long evolutionary distance between birds and primates, the basic building block could be homologous*. See Fitch et al. (2010, p. 796): "Capabilities that are convergent at one level (e.g., behavioral) may employ mechanisms that are homologous at another level (e.g., genetic)". See also de Waal & Ferrari (2010, p. 201): "There is increased appreciation that the basic building blocks of cognition might be shared across a wide range of species". (Nevertheless I will continue focusing on primates; as I said above, the very nature of human cognitive novelties would have been radically different if the starting point had not been the last common ancestor of chimpanzees and humans.)

All this, I insist, is perfectly compatible with my central hypothesis. If we reject the previous speculation about the original adaptive advantage, the hypothesis only loses

---

[18] But there might be another utility: For example, observing adults' feeding habits might be useful to macaques' young. Certainly, what, in this field of eating habits, has come to be known as 'program-level imitation', is limited to apes: see Byrne & Russon (1998) and Whiten et al. (1996). However, it is likely that, by observing their parents, the young of monkeys learn, at least, which part of fruit is edible.

one of its supports and not exactly the most reliable one. Thus, we should move on finally from the question of the original utility of mirror neurons.

1.7 Summarizing the hypothesis defended in this chapter

Let us return, therefore, to the central hypothesis of this chapter and insist on its central question. The firing of mirror neurons in their two roles would not correspond to motor commands, but to expectations of postural results. There are no postural expectations prior to the observed movement of the conspecifics, as they would be impossible; there can only be *a posteriori* expectations, that is, postural expectations which follow observed movement of the conspecifics. Here, in this time-relative inversion, lies the great novelty of mirror neurons in their mirroring role. The ancient type of expectation can only point to a (possible) future moment, that is, to the moment when expectation will finally be satisfied. With mirroring, on the other hand, there are expectations that do not progress towards the future, but are derived from their visually observed satisfaction.

For this reason, mirror neurons constitute a perception of internal states of others (a perception of others "from the inside", as Rizzolatti & Sinigaglia [2010, p. 264] say). However much I have critiqued the interpretations that have been made of mirror neurons, I fully share the idea that mirroring is a very important landmark in evolution. The expectation of those internal postural states that is activated in the macaque corresponds to an action of the conspecifics. In some sense, we would already have come across, in the observer macaque, the beginnings of the ability that is absolutely central for human beings, namely, the ability to perceive internal states of others. However, even if we accept all this, we must heed the other side of the coin. That perception performed by monkeys' mirror neurons is very primitive.

I would enumerate three characteristics that attest to this primitivism. Firstly, macaques' mirror neurons relate only to a very limited area of their bodies (only hand and mouth). Secondly, internal state of the conspecific originally is not yet interesting in itself. (Originally, the perception of this internal state would be a merely secondary effect, that is, Keysers & Perrett's 'Hebbian learning', or, as in my risky speculation, it would have an adaptive function, but one which would be limited to confirming – or, in their case, rejecting – that the seen hand is its own). Thirdly, and this is the result that really interests me, there would be no simulation or latent imitation involved at all, only expectation (albeit *a posteriori*) of internal states. The role played by this chapter in the book as a whole derives exclusively from this third characteristic, I insist. The difference between simulation and expectation, or more precisely, the limitation of any expectation, is that expectations are always possibilities of the expecting subject. Consequently, only simulation can implant in a subject contents radically different from his own contents, that is, contents that belong to a self that is currently communicating and interacting with the subject. (*Infra*, in Section Two, this difference will be the nuclear issue.)

The last of these three limitations or absences would only disappear, by my understanding, with human beings. I conceive the step from confinement within the resource of expectation to the possibility of true latent imitations or simulations to be extremely demanding: we would have to consider simulation as a second line of awareness that would be added to the first line (that is, to the awareness of own, real current states) only in humans. Would there not then be simulation in chimpanzees? It is precisely this question that the next chapter will address.

of these two motor learning routes will not be really similar if the environment does not provide sufficient relevant points to be chosen as subgoals. (Note that circus animal trainers typically resort to building an environment with numerous and well-distributed relevant points.) In short, results will become really similar only on particular occasions.[1] In addition, beyond the difference in results, there is a more crucial issue in order to differentiate between program-level imitation and genuine imitation of new and complex motor patterns. Genuine imitation of new and complex motor patterns is much more demanding and powerful than any 'program-level imitation' (I will discuss this in Chapter 8). Returning to our argument, we can say that chimpanzees' ability to imitate simple movements has nothing to do with any 'program-level imitation'.

As soon as we accept the uselessness of the imitation of simple movements, we are forced to think of a new suggestion regarding such imitative abilities of chimpanzees. There we would find a side effect of another ability, an ability which in this case truly does have adaptive advantages. That adaptively advantageous ability would be, as I have said earlier, the ability to match one's own body and another's body and thus to reckon visual fields of the conspecific.

2.2 Are chimpanzees merely exploiting visual findings of the conspecific? Introducing the current debate

But it is now time to ask ourselves the question which is the object of a debate which has reached boiling point in the last five or six years. In what sense is the ability to detect visual fields of conspecifics useful for chimpanzees? Is its only benefit to enable them to use others' visual findings (as Povinelli and his associates claim)? Or, on the contrary, is it also of benefit so they can ascribe (i.e., attribute) visual perceptions to conspecifics and thus predict their behaviour (as Tomasello, Call and Hare claim)?

2.2.1 Ascribing visual perceptions: The experiments carried out by Hare, Call & Tomasello

As a previous paragraph announced, I support the most generous position as regards chimpanzees' abilities. I think that these animals would be able to perform that 'ascription and prediction'. When I read Tomasello, Call & Hare (2003), I find their arguments persuasive. Given the intense social life of chimpanzee clans, we must think that the detection of visual fields of the conspecifics would have immediately brought about the more complex ability, provided the brain requirements involved in 'ascription and prediction' were not too much for apes. What is more, if the exploitation of others' visual findings were not associated with the ascription of that visual perception to a

---

[1] Leighton, Bird & Heyes (2010) express their opposition to the general idea that goals mediate true imitative behaviour.

conspecific, this exploitation could often give rise to disadvantages. Suppose a chimpanzee exploits the visual finding of a dominant conspecific, and suppose that visual finding relates to food. If there had been no ascription of the visual perception to the dominant animal, our chimpanzee would immediately go after the food he had just discovered and then suffer the punishment inflicted on him by the dominant animal. Of course, this conclusion can obviously be avoided if we add that the use of others' visual findings is restricted to detecting dangerous situations. So far, however, there are no data to support such a restriction.

Let us now look at last at the experiments carried out by Hare, Call & Tomasello (2001) and Tomasello, Call & Hare (2003), which support the idea that chimpanzees ascribe visual perceptions to conspecifics. The first thing recorded by these authors is that when a human looked at an object that was placed behind a barrier the chimpanzees moved accordingly to get a suitable visual angle on the human's visual field. The chimpanzees even looked back at the researcher if they inspected the corresponding visual field and found nothing there. But none of these observations were relevant to the question of whether or not chimpanzees ascribe visual perceptions to other individuals.[2]

Thus, Tomasello, Call & Hare designed a further, genuinely interesting, experiment that makes use of the rivalry between a dominant chimpanzee and a lower-rank individual. In half of the tests, the subordinate animal could see a piece of food that the dominant animal, because of a barrier, could not. There was no barrier in the other half of the trials. The (statistically strongly predominant) result was that the lower-rank individual took the pieces of food when the dominant animal could not see them, whereas it refrained from doing so when the food was visible to the dominant subject.

In a new set of experiments, an even more sophisticated ability was tested. The lower-rank individual was held back at a distance from which it witnessed the higher-rank animal either look or not look at a piece of food. When the lower-rank animal was released a moment later, it found that the piece of food had already been hidden from both chimpanzees by a barrier. Nevertheless, the behaviour of the lower-rank chimpanzee still continued to adapt flexibly to the circumstances. The (statistically strongly predominant) result was that the subordinate animal took the food when the dominant one had not seen it, and refrained from doing so when it had. In my view, in this new set of experiments chimpanzees ascribed to conspecifics, at the moment *m*, the goal that the perception of a visual field would have provided in the moment immediately preceding *m*.

The question here is whether or not that sensible behaviour on the part of a lower-rank animal can be adequately explained as a result of mere conditioning between

---

[2] Gaze-following is common among all primate species and possibly shared by a wide range of animals: See Teufel et al. (2010) about monkeys, and also Kaminski, Riedel, Call & Tomasello (2005) about domestic goats, Miklósi, Polgárdi, Topál & Csányi (1998) about dogs, and Schloegl, Kotrschal & Bugnyar (2007) about common ravens.

other individuals' behaviour and their torso direction. In other words: is it correct to suggest that chimpanzees sensibly guide their behaviour after having first noticed the matching between their own body and that of other individuals and accordingly ascribed visual perceptions to them?

2.2.2 Povinelli's argument: The conspecific's blindfolded eyes

The author who most vigorously rejects those results, or, in other words, who maintains what (above in 2.1) I have called the second option, is Povinelli (2004) (see also Povinelli & Vonk [2004]). For years this author has been opposing what he calls the 'argument by analogy'. Although two behaviours, one of them human and the other one animal, may be analogous on the surface, this does not allow us – Povinelli claims – to interpret them as the expression of the same process. Of course, such a claim is, in principle, not disputed. The question is where we should reject the analogy of the underlying processes, and, in particular, whether the detection of conspecifics' visual fields by chimpanzees must be interpreted 'behaviouristically' or, on the contrary, 'mentalistically'.

In general, it might be said that the last few years (say, from the mid 90s onwards) have seen something of a revival of anti-mentalism. This is probably a pendulum reaction. Mentalist abuses in fields such as cognitive ethology had become notorious.[3] It is in this context, it goes without saying, that we need to place Povinelli.

According to Povinelli, several failures by chimpanzees point to the rule followed by these animals being a behaviourist one. The rule would be something like the following: 'The direction of the torso of a conspecific may reveal where there are interesting objects'.[4] According to this author, no ascription of visual perceptions would, therefore, occur in chimpanzees.

What are the failures that Povinelli has detected? When it comes to supposed ascribing of visual perceptions to a conspecific, chimpanzees make no distinction as to whether the subject has his eyes blindfolded or not. Since the animals had previously experienced the effect of an opaque cloth on their eyes or a basket placed on their head, Povinelli concludes that this lack of sensitivity favours a non-mentalist interpretation of chimpanzees' behaviour where they look towards where a conspecific is looking. Is this conclusion correct? Is this argument forceful enough to object to the results of Tomasello, Call & Hare's experiments?

---

[3] A sarcastic criticism of those excesses can be seen in van Rooijen (2010).

[4] Let us examine some data regarding macaques' brains, more concretely, their superior temporal sulcus (STSa). "Two-third of STSa cells selective for whole body motion prefer a combination of body view and direction where the body moves in the direction that the head is pointing (walking forward following the nose)" Jellema & Perrett (2003, p. 1735). In other words, these cells pay attention to the direction of the torso. However it is clear that this datum does not answer our question of whether or not chimpanzees ascribe visual perceptions to other individuals.

Certainly this question is what interests us, but let me insert a digression. Penn, Holyoak and Povinelli (2008) emphasize their explicit purpose of stressing the human peculiarity. In this respect, their approach is certainly very close to my own. However I think that Povinelli's view – which rejects the idea that chimpanzees' abilities represent a preliminary or primitive stage, however remote, of the human skills concerning a theory or perception of the minds of others – is less appropriate for clarifying this peculiarity.

2.2.3 Blindfolded eyes and adaptive usefulness

The question is why the specific detail of the blindfold or basket is not included as part of the posture shown by a conspecific. Is this non-inclusion proof that chimpanzees do not ascribe visual perceptions to conspecifics? For an observing chimpanzee, blindfolded eyes would no doubt be absolutely decisive if the objective were to detect eyes of the conspecific fixed on him. This would be the case even for non-primates. However, as far as the ability of ascribing visual perceptions is concerned, things may be different.

From the perspective of adaptive advantage, including the detail of a blindfold over the eyes could perhaps be an unnecessary complication. Cases where an animal in the wild was unable to see would have been too rare to have influenced the evolutionary configuration of the ability to ascribe visual perceptions. Let us remember that, according to previous suggestions (in 2.1.1), this ability would be an extension (and also a change in function) of monkeys' ability to detect the *a posteriori* postural expectations corresponding to an observed hand or mouth. That extension would have given rise to the matching of bodies because this was a requisite for the fulfilment of the new function. Under normal conditions this was not only necessary but – and this is the key point – also sufficient. Why should that extension become more refined? Why should it pay attention to such strange details as blindfolded eyes or closed eyes in a perfectly upright body?

Thus, in my opinion, Povinelli's results do not refute a 'mentalist' interpretation. The inability to discern the detail of a blindfold would be perfectly compatible with an ascription of visual perceptions to a conspecific. While here I am accepting Povinelli's point about the chimpanzees' inability to discern the relevance of a blindfold on someone else's eyes, Premack & Premack (2003, p. 141) do not. They argue that at least one of the chimpanzees with which they worked for years would remove the blindfold from the researcher's eyes, but not the nose or the top of the head, for the researcher to go to the box containing the reward and unlock it. The Premacks interpret Povinelli's results as the consequence of previous training of the chimpanzees by Povinelli, training in which, according to the Premacks, the animals had become used to obtaining food without concerning themselves with where the donor was looking. But let us return to what interests us. An observing chimpanzee would actually ascribe the perception of a visual field to his conspecific, and on this assumption, as the experiments by Tomasello, Call & Hare have suggested, he would successfully anticipate the reactions of that conspecific.

2.2.4 Ravens and chimpanzees

Nowadays many comments made on the results of Tomasello, Call & Hare are very different to those made by Povinelli. Thus, instead of questioning whether chimpanzees are able to ascribe visual perceptions to conspecifics, what seems increasingly doubtful is that this ability is restricted to chimpanzees: See Fitch et al. (2010). In this way, macaques have been shown to discriminate between human experimenters who can and cannot see food (Flombaum and Santos [2005]), as well as individuals who can and cannot hear the removal of food (Santos et al. [2006]), indicating multimodal sensitivity to others' perception. (This latter observation seems contrary to Bräuer, Call, Tomasello [2008], who claim that chimpanzees do not take into account what others can hear in a competitive situation.) In addition, there is evidence of similar mechanisms in both scrub jays and ravens: See Emery and Clayton (2001) or, even more crucially, Bugnyar and Heinrich (2005).

At this point, I would like to be more specific about my suggestion. Bugnyar & Heinrich (2005) and Tomasello et al. (2003) have showed that ravens and chimpanzees modify their tactics in relation to whether or not competitors previously had the opportunity of observing food. In addition, these authors observed that animals know that obstacles can obstruct the view of others. But I would like a research to be made regarding a slightly different ability.

An animal knows where an object is thanks to the fact that it has seen it recently. But in the moment that we take interest in it, this animal no longer sees that object. What he sees in that moment is a conspecific placed in a location, orientation and posture which are adequate for the perception of the object in question. Will the animal be able to predict the subsequent behaviour of its conspecific? As I say, this experiment should be carried out with the two species, ravens and chimpanzees. Will the two species be able? Will both be unable? Will only the chimpanzee be able?

I think that this ability that we have just described is much more demanding than the one studied in the experiments by Tomasello et al. 2004 or by Bugnyar & Heinrich, 2005. The one that all these experiments study is 'logically compatible' with the interpretation proposed by Heyes (1998). This is admitted by Bugnyar & Heinrich (2005). Gaze-following, detection of visual barriers that are placed between the conspecific and the object, grasping of the affordances of the object: The total of these resources and learnings could, perhaps, be enough to predict the conspecific's behaviour. Likewise, that total of resources could be enough to explain that, when a human looks at an object that is placed behind a barrier, the chimpanzees, and also the ravens (Bugnyar et al. [2004]), move accordingly to get a suitable visual angle on the human's visual field. However, if animals succeeded in the task that I have just described, that result could not, in my opinion, be interpreted in the deflationary way proposed by Heyes. In

SECTION TWO

The basic human ability

CHAPTER 3

The three modes of processing the eyes of others

3.1 The progressive convergence of this issue and the 'theory of mind'

In 'theory of mind' research, it took some time for the question of processing of the eyes of others to be addressed. As we know, almost all interest, at first, focused on the understanding of beliefs and desires. For a long time, cognitivism saw visual perceptions as being too far from the mind. This would only change when the pointing gesture, which Bruner (1977) (or Scaife & Bruner [1975]), and Bates et al. (1976) had already emphatically stressed, was introduced into the agenda of the 'Theory of Mind', as well as Trevarthen's concept of secondary intersubjectivity or triadic communication (also in the 70s: See a very late description in Trevarthen [1998]). Among the factors responsible for such a change in the agenda, we must also mention Butterworth's work on the long and progressive development of pointing gestures in children (Butterworth, [1991, e.g.]) and, on the other hand, the more general issue of autism.

Baron-Cohen was the first to systematise the processing of another's eyes. This author's starting point was an interest in autism, in which he saw a sort of 'negative' of the 'theory of the mind'. The symptoms that allow an early diagnosis of autism are not related to typical 'second-order mental states', that is, with the perception of beliefs of others, but with skills which appear much earlier in a child's development, such as communication through sight direction or pointing gestures. This led Baron-Cohen (1999, for example) to propose three modules whose staged maturity would lead to successive forms of processing the eyes of others: I(ntentionality) D(etector), E(ye) D(irection) D(etector) and S(hared) A(ttention) M(echanism). I will certainly take Baron-Cohen's modules on board in this chapter, but I will also move some distance from them. My hypothesis focuses only on the so-called Direction Detector, and identifies three very different evolutionary levels of this detector.

Kobayashi & Koshima (2001) are notable along a different line. These two primatologists studied the presence and absence of the 'white of the eye' in different primate species. Certainly, their final statement – only humans possess a conspicuous and contrasting 'white of the eye'– only transcribes a datum that had always been in clear view for everyone. However, this statement was extremely important as it meant the incorporation of the white of the eye into what we might call the official list of exclusively human features.

3.2 What are the three different modes of processing the eyes of others to be proposed here?

3.2.1 The difference between the first mode and the second one

The first mode would have started early in evolution, long before primates appeared. An animal detects an eye staring at it, and responds by increasing its alert. That is why some butterflies have eye-shaped markings on their wings. Even a mark only remotely resembling an eye scares away potential predators; for that reason, that resemblance was progressively selected. But it is not only the alert response. An eye staring at oneself would always indicate the opportunity for any kind of interaction. We know that, right from birth, humans are fascinated by other people's eyes. Already as neonates, we prefer to look at faces compared to non-face stimuli (Johnson et al. [1991]), and we favor faces with open rather than closed eyes (Batki et al. [2000]) as well as faces with mutual rather than averted gaze (Farroni et al. [2002]). Similar preferences in nonhuman primates suggest that the tuning of the neonatal visual system to face- and eye-like stimuli is a general primate heritage (Myowa-Yamakoshi et al. [2005]). In addition, Rosa-Salva et al. (2010, p. 565) have concluded that "Faces are special for newly hatched chicks (...), and the eye region of stimuli is crucial in determining the expression of spontaneous preferences for faces". In my view, the first mode would typically involve no ascription of visual perceptions to other individuals.

To be more precise, let me yuxtapose two ideas. The first idea is that, as the current followers of Trevarthen's approach (Reddy [2008]) suggest, two month-old babies are capable of genuine emotional interactivity. The second idea is the assumption that two month-old babies are not yet capable of ascribing visual perceptions to other individuals. What I want to make clear is that for me these ideas are not incompatible. That is, the first mode of processing the eyes of others could certainly imply interactivity, but, according to my definition of the phenomenon, it would also lack the power to ascribe visual perceptions to any seen eyes. Likewise, I think that the two following ideas are completely compatible. One, "nonhuman primates can and do make judgments about being watched by others" (see, for example, Tomonaga & Imura [2010]). Other, this sensitivity to being watched would involve no ascription of visual perceptions to other individuals.

In this respect, we must also focus on Ferrari et al. (2009, p. 1768), who (besides recording frequent episodes of mutual gazing between days-old macaques and their mothers: see above, in 1.3.2) also make the following remark: "The mother held the infant and actively searched for the infant's gaze, sometimes holding its head and gently pulling it toward her face". In my opinion, there would be no ascription here of mental perceptions on the part of the mother to her infant. If there were such an ascription, the scene, as we will soon see in this same chapter, would be a case of the 'third mode'. However, there is no need, in my view, to resort to such an ascription to

3.3 Why would the 'third mode' be so demanding?

3.3.1 The peculiarity of visual perceptions ascribed in the 'third mode'

As soon as we have formulated this option or theoretical bet, hypothetical-deductive reasoning requires the process of combining the first two modes to be extremely difficult (Perhaps experimental work by Conty et al. [2010] might provide a small amount of evidence in favour of this difficulty.) If it were not, chimpanzees, which possess both modes, could not be deprived of the third mode. From here, a question forces itself on us. Why would transferring chimpanzees' new ability to the field of the old ability have to be so difficult? There would be no real value in continuing to set out the suggestion if there were no answer to this question. So, before anything else, let us try to respond.

Certainly, we know that any ability which evolution has attained will hardly ever be lost later in evolution (although there are cases). However, the difficulty cannot lie here. Note that this transferral, far from implying any loss, means only that the two different resources, ancient and new, would both be applicable to the field of the old resource.

Why, then, would chimpanzees be unable to ascribe visual perception to the eye staring at them? Our first answer is that this type of ascription is very different from any other. The perception ascribed there is one that in no event could be that of the ascriber. Seeing oneself as a distal stimulus is a perception radically and intrinsically different from one's own perceptions. Let us examine this point.

Chimpanzees, as we have said, detect the front/rear axis of the body of the conspecific, and thus calculate the direction in which the conspecific was looking. Then, (by moving in whatever way necessary and which does not at all imply imitating the movements or postures of the conspecific) they supply themselves with the visual field that their conspecific was looking at.[2] In other words, chimpanzees obtain from reality the contents of visual perceptions that they are ascribing to conspecifics. Certainly, this does not prevent the ascription from being correct: it is just the same collection of objects that they and their conspecifics are looking at. However, the content of the perception ascribed to conspecifics is primarily the ascriber's own perception. This is precisely what changes when I ascribe to the other a visual field in which I am included. I will never be able to obtain this content with my own eyes.

---

[2] Piagetian spatial *décentration* would therefore be accessible to chimpanzees. They may not be able to overcome the complexities of the three mountains test (Piaget & Inhelder [1956]), but the nucleus, at very least, of what decentration there was in that task is perfectly accessible to chimpanzees. The fact that Piaget made no distinction between this spatial decentration and the mental decentration (i.e. thinking 'according to someone else') to which he points in his 'Commentary on Vygotsky' was probably one of the most significant curbs on the Piagetian model. I have harboured this idea for over twenty years (Bejarano, 1985). In contrast, I have only recently noticed another deficiency in Piaget's approach: the failure to differentiate between the two kinds of spatial decentration which correspond respectively to what I have here labelled as the second and the third modes of processing the eyes of others.

3.3.2 Radically not-own visual field, expectation, simulation

How can we reach that radically and intrinsically not-own visual content? The only way of understanding that content would lie in imagining a radically not-own interiority. We can thus conclude that such content would be as cognitively demanding as a metabelief (or, according to the terminology of studies of 'Theory of Mind', a 'false belief'). It is obvious that I must detach false beliefs from me and set them aside, as my knowledge of the world cannot include them. Such beliefs must be attributed to an interiority different from my own. But the conception of that interiority, I stress, would not have to wait for the understanding of false beliefs of others. Less is necessary for it to emerge. It is only necessary for a visual perception that includes the ascriber to be ascribed to the conspecific. That particular ascription of visual perception is the truly ideal term of comparison we were longing for in the previous chapter (2.3). The type of ascriptions found in chimpanzees can be differentiated from a metamental state without having to step outside the ascription of the visual perception.

The third mode of processing eyes of others would therefore be highly demanding. We have already said this, but we shall try to explore it more deeply. The key to its demands lies in that here (that is, in this 'third mode'), the old resource of expectation – which had sustained all animal behaviour effectively – breaks down for the first time. Expectation is no longer useful. By contrast, the 'second mode', that is, the ability to ascribe visual perceptions to conspecifics, would require nothing more than an expectation (even if it is an expectation which emerges, not before the result of the movement, but in view of the result already achieved in the conspecific's body, that is, an *a posteriori* expectation).

Think back to how expectation was defined in Chapter 1. All expectation is expectation of a state that is, in principle, reachable for the subject. When chimpanzees gauge a visual field from the location, posture and orientation of a conspecific, they would actually be gauging the visual field that they, the observers, would obtain if they were in the same circumstances. In addition, as we stressed in that same chapter, expectations (both a priori and a posteriori expectations) are always empty profiles that the subject could satisfy or fill. This is what we often find in the 'second mode'. After observing their conspecifics, chimpanzees (by moving in whatever way necessary) supply themselves with the visual field that their conspecific was looking at, or, in other words, satisfy the expectations that their observation had raised in them.

This is all over the top and obsolete when facing the 'third mode'. Visual perceptions that are now ascribed to the seen eye – or more in general, contents that are ascribed to a self who is looking at me, addressing to me or communicating with me – are intrinsically unreachable for me. Thus, this ascription was an impossible task if expectation is the only resource. But the impossible task became possible with simulation. Simulation, that is, a second and novel centre within the mind, would necessarily be involved in the third mode of processing eyes of others. This is what I shall hypothesise. But we should first re-examine more calmly what we have just called the 'radically not-own visual field'.

3.3.3 Self-recognition in the mirror and perception of a radically not-own visual field: Facing a potential objection

We will pause here from our string of arguments to ask the following question: would self-recognition in a mirror be sufficient to provide the subject with that visual field which we were viewing as impossible for him? We have said that the radically not-own visual field includes myself as a distal stimulus. Can I not, however, perceive such a visual field in the mirror?

Certainly, the mirror is a cultural artefact of which we can only find very poor approximations in nature. However, leaving aside the question of whether it played a role in evolution, let us address only the question mentioned. Is self-recognition in the mirror sufficient to obtain a radically not-own visual field? Let us imagine the answer was affirmative. The suggestion we have been defending (requirements of the 'third mode' can only be satisfied by the human brain) would then be in trouble. Remember that self-recognition in a mirror is an ability that chimpanzees (and probably also other species) do possess. The importance of this question to us is therefore clear: does the self-recognised image in the mirror equal the visual field of those who are looking at me?

Certainly, the image in the mirror in which I recognise myself is similar to the visual field of the other who is looking at me face to face. The similarity is complete, except for the right/left axis.[3] However, once the self-recognition of the image in the mirror has been achieved (and this self-recognition must be assumed in order even just to pose our question), this image would be a very peculiar one. Each subject would perceive it as undistally in relation to himself as he would perceive his own hand. Attempting to confirm that it is one's own hand can sometimes cause a subject (certainly, as long as the subject is not grasping an object or hanging from the branch of a tree: remember 1.5) to execute manual movements in order to check whether he sees them or not. The first self-recognition in the mirror also implies this kind of check. But, I stress, once the understanding that it is one's own image has been attained, that image will no longer be able to be considered distal, and the subject will no longer be able to consider it as the body of another.

On the contrary, within the visual field of the eyes of others that includes myself, my image will be considered truly distal. This marks a profound difference between the two perceptions, mine and not-mine, however much they share the same visual content. So I am inclined to conclude that self-recognition in mirrors, precisely because

---

[3] That axis belongs to the corporal scheme and, therefore, it could well coincide with the gravitational up/down axis when we are looking at ourselves in the mirror while lying down. "*In our ecology*, the axis perceived as inverted in the mirror turns out to be the horizontal one" (my emphasis): Navon (2002).

it is self-recognition, does not offer at all the radically not-own visual field that we have linked to the 'third mode'.

There is another different way of reaching the very same conclusion about self-recognition in the mirror. Several authors – mainly, Mitchell, in press (Mitchell [1993] is a classic study on chimpanzees' self-recognition in the mirror) and Bräten (1998, p. 115)– have highlighted that contrast, to which we have earlier referred to, between one's own image in the mirror and a not-own body facing us. In the not-own body (but not in the image in the mirror), the hand that I see facing my right is the left hand. Until recently, I had never considered this contrast to be important. However, now that I have seen how much attention they are given by these authors, I have begun to think that those facts about the lateral axis fit very well with my suggestion about the basic human ability. The moment a subject can understand the break-up of his right/left body axis in the not-own body would be good evidence that simulation has replaced expectation. We shall now examine this in more detail.

When do I need to make room for inversion of the right/left axis? Precisely at the moment I pay attention to the interiority of a not-own body that I perceive to be approaching me. But this is the same as saying precisely at the moment when I must dispense with expectation and begin to use genuine simulation. Expectations – which, according to our hypothesis, can in no way be made of an interiority which I notice looking at me, communicating with me or addressing me – do not need to take account of this rupture of axis. The detection of the visual field of the conspecific through the 'second processing mode', as we said above, would actually be reduced to the calculation of own expectations – what is it that we would see if we were in the other location, posture and orientation. It is always our own body with its own axis that intervenes there.

On the other hand, in order to detect an interiority which approaches me, or which is staring at me, I will have to conceive it as a self radically different from me. Any simulation of a movement toward the right would have to be detachable from the real space to my right. (In other words, 'anatomical matching' – not 'mirror matching'– of the conspecific that looks at me requires genuine simulation.) Thus, from that moment on, two different implementations would become necessary for the lateral axis, one spatial and behavioural and the other for motor simulation.

I suspect that this may have quite a lot to do with the fact that the hemispheric specialisation of the animal brain is modified in the human brain.[4] But, how could that idea be displayed? That is unfortunately much more difficult than suspecting.

---

[4] In my view, we should investigate the changes occurring in brain activation in situations such as the following. A subject has been looking at himself in a mirror, the light slowly fades and the subject is deluded into believing that he is still in front of the mirror. But then, suddenly he discovers that he is no longer in front of the mirror, but in front of someone who is staring at him. Surely there must be evidence of many brain changes at that very moment, for example, changes corresponding to the subject's surprise or a state of sudden alert. However, some neurophysiological change attributable to simulation could perhaps also be seen. Could

Animal hemispheric specialisation involves contralaterality and (see Meguerditchian et al. [2010]) some behavioral and brain asymmetries. But the human left hemisphere typically controls language and the rest of the movements that are culturally learnt. How would this particular separation of types of movement link with the two different implementations for the lateral axis? It is only in combination with these questions that the above referred suspicion could really become relevant. Thus, let us leave this issue of hemispheric specialisation for the time being.

None of this – none of this duality of implementations for the lateral axis – would be at all required for self-recognition in the mirror. This self-recognition, in order for it to be self-recognition, can in no case view the kinaesthetic interiority of the image as approaching the subject. For this very reason, there is no need for a right/left axis different to our own body's axis. So we conclude that simulation would not be necessary for self-recognition in the mirror. Or, in the words of our previous question, self-recognition in the mirror does not equal a radically and intrinsically not-own visual field. (See Rochat & Zahavi, in press, whose conclusion about self-recognition in the mirror is close to this.)

3.4 The communicative use of sight direction

We have just suggested that ascribing visual perceptions to the eyes staring at me would be exclusively human. What data can we offer to support it? We might begin by finding the adaptive advantage of that 'third mode', that is, by addressing the communicative use of eye direction.

Since just before they are a year old, children know that someone is asking them to look at an object when that person looks at the object and at them in turn. This communicative resource, which (almost) always accompanies the hand-pointing gesture, can frequently occur without it.[5] What is involved in understanding this communicative use of eye direction?

Of course, chimpanzees' own ability is clearly involved. The perception of the object needs to be ascribed to the conspecific. Chimpanzees are capable of all this, according to Tomasello, Call and Hare. But now this ability would be included within another completely new ability. This ascription must continue to be made when the eyes of the conspecific move on to look at oneself. If that immediately previous

---

that possible neurophysiological change have something to do with hemispheric specialisation? Needless to say, the same experiment would have to be carried out with chimpanzees. What are the cerebral changes which, in that situation, are common to both humans and chimpanzees, and what are, on the other hand, exclusive of humans?

[5] More concretely, small children are unable to understand the meaning of any hand-pointing gesture in the absence of this gaze-shifting. Admittedly, in the middle of a conversation, adults' understanding can dispense with the gaze-shifting, but small children definitively cannot do so.

white sclera, is “a cooperative eye” (cf. Tomasello et al. [2007]). However, according to my hypothesis, it was only after the emergence of the ‘third mode’ (only after and as a result of that) that the white of the eye could become useful for cooperative or mutualistic behaviour. In the next chapter, I shall focus on this question.

Before moving on to look at another point, we shall comment further on the question of the white of the eye. The adaptive usefulness of this feature would come from the third mode of processing. It is thus plausible that the white sclera would not be prior to such a processing mode. As a result, it would be extremely interesting to find out which was the first species to have the white of the eye. Did Neanderthals have it? DNA analysis may soon be able to give us the solution to those questions.[11] Guillermo Lorenzo (see his elaborate generativist view of the genesis of language in Lorenzo [2003]) believes (personal communication) that Neanderthals’ rich cultural heritage invalidates the possibility that they lacked pointing gestures. That culture is indeed a rich one. We must escape the temptation to look down on it. However, this is not the question; instead, what we need to ask is whether attention towards the adult’s task could be provoked without pointing gestures. In my opinion, the response to that is far from obvious, and all we can do is wait.

We might bring in other arguments, apart from the already mentioned white of the eye that is exclusive to humans, in favour of my description of the ‘third mode’. It is obvious that ‘self-conscious emotions’ – embarrassment, pride, shame and guilt – are relevant in this sense. “It is not the simple act of reflecting on our own appearance, but *the thinking what others think of us*, which excites a blush” (Darwin [1965, p. 327]; my emphasis). See also Lewis (1992) and (2000).[12] But, although it is a subject to which I am greatly drawn, I shall not address it in this book: What interests me here is the series of abilities, that is, of biological innovations, out of which language could historically emerge. Thus, let us focus on the idea that the same specifications that we have proposed for the ‘third mode’ would also be necessary for human understanding of the finger-pointing gesture, as well as for four-hand tasks. I shall address this in the following chapters.

---

[11] There is an exceptional report of a wild male chimpanzee who had exposed white sclera (Goodall [1986]). Finding further apes with such characteristics could be useful to guide the search in the genome of Neanderthals.

[12] See also Leary [2004], Rochat [2009] and Zeedyk [2009]). In addition, it will be useful to read Jackendoff (2007, Chapters 10 and 11). Tomasello (2008) and (2009) focuses on shame and guilt in a context perhaps more similar (that is, more similar than the one used by the above mentioned authors) to the one of current Section Two. However, Tomasello´s invocation to rules, indisputably adequate as far as the study of ontogenesis is concerned, seems to me no longer appropriate when we deal with the evolutionary origin. In short, in my view, the power of rules, the ‘construction of social reality’, could not come into being until later and as a result of the emergence of the ‘third mode’ or basic human ability. (Certainly, it has been repeated that language could not function without a ‘social contract’ because words are cheap. However, in my view, in the origins, each collaborative task, from pointing gestures on, would bring immediate results for each of the collaborators.)

CHAPTER 4

Pointing gestures

4.1 Pointing gestures in children

In children, finger-pointing gestures would be an extension and emphasis of the communicative use of the eye. Pointing gestures are always accompanied in children by the communicative use of the eye. This use of the eye can occur on its own, but in most cases it will be accompanied by a pointing gesture.

There have been some anti-mentalist interpretations of pointing-gestures in children. We shall start by looking to Vygotsky's theory, later elaborated on by Bates, on the origins of the pointing gesture. Children, at first, would make an effort to reach an object, stretching their arm and body out as much as possible, but without succeeding. The adult accompanying them realises what the child wants and gives him the object.[1] This is repeated on several occasions. And then, the child's behaviour changes: while he initially made the gesture to reach for the object without any communicative purpose at all, he now begins to do it without any serious effort, just so the adult will see. As we have seen, in the original formulation, the child, at first, wants to reach the object. By contrast, in Delgado, Gómez & Sarriá (2009) and also in Carpendale & Carpendale (2010), infants first use their extended index finger as a manifestation, and probably also an enhancement, of their own attention. But this difference is irrelevant. It is the very nucleus of Vygotskian theory that we can not accept. We should note that this theory clashes with a well established datum about children's acquisition of the communicative pointing gesture, namely, the precedence of understanding the gesture over producing it (see, for example, Corkum & Moore [1995]).

In recent years, there have been some anti-mentalist interpretations of this understanding. Moore (1999) is a good example. According to him, 12-month-olds would merely understand that an adult's pointing gesture normally precedes an interesting experience. As you will see, this hypothesis, unlike Vygotsky's, refers to the genesis of understanding and not of production. Correspondingly, Moore's theory envisages a

---

[1] What Vygotsky actually says is that the mother would inadequately interpret the child's effort to grab the object as a requesting gesture: this is what has been called 'the mother's deception' or 'the illusion of intentionality'. This interpersonal mediation would be, Vygotsky continues, an example of the General Principle that he formulated ("superior processes would originate interpersonally and only later would they become intrapersonal"). As it will be seen in other chapters, I agree with the Principle. However, I do not agree with that specific example.

less instrumental (or less actively instrumental) conditioning than the one contemplated by Vygotsky.

But some research placed this behaviourist interpretation in difficulty. Certainly, infants can be conditioned to follow changes in direction both by people and by objects. However, “the temporal co-ordination between pointing, gazing, and vocalisations occurred at a much higher rate in the Person than in the Object condition. That infants produce these behaviours more often to people than to inanimate objects reveals that infants have different conceptions about people and objects, namely, that one communicates with people and not with objects” (Legerstee & Barillas [2003]).

From this point on, therefore, I will assume that children understand the pointing gesture completely, at least from their first birthday onwards. Of course, this does not imply that the child understands the ulterior motives, wishes and intentions that underlie the pointing gesture that he observes. Children would only perceive that the adult is pointing out an object to them, but this would be precisely the essence of the pointing gesture.

For quite some time, pointing gestures have been split into two types: imperative (also called directive) and declarative (Bates et al. [1975]). With the latter, the producer tries to attract the recipient’s attention toward an object, but not so that the recipient will give them an object. Why then? It is clear that the producer-child tries to get the recipient to talk about the things being pointed at, i.e. to name or talk about them. Language learners’ need of such linguistic nourishment is enormous, and declarative pointing is a crucial resource in supplying it. On occasions, the child will accompany the gesture with a term. That term – the verbal protodeclarative – will help the child to check that the meaning he is giving to the term is correct, or, in other words, to ‘negotiate meaning’, as well as serving exactly the same purpose as the simultaneous gesture. Both the verbal and the pointing protodeclarative do not only seek indiscriminate linguistic stimulation, but also create the ideal context to conveniently exploit that stimulation. The child will receive words that she does not yet know, but she will know that such words would have to do with the object that has been pointed at. The conclusion of all this is that protodeclarative pointing-gesture is a type of communication whose main usefulness is linked to linguistic learning. (See Southgate et al. [2007]). This usefulness has been tested by Bigelow et al. (2004). But the protodeclarative gesture can also fulfil the communicative function of getting the recipient to notice the object. Liszkowski (2006): “At 12 months old some of children’s pointing gestures intend to provide information (telling an adult the location of something the adult is looking for)”

This second type of declarative pointing, i.e., Liszkowski-style pointing, has very frequently been associated with altruism or cooperation, whereas imperative pointing has been asssociated, by contrast, with selfishness. But however valid the second correlation may be for the individual, it cannot account for the establishment of the communicative resource. The imperative type requires altruist or cooperative recipients. It is precisely the cooperation issue which has recently gained considerable attention in attempts to explain why chimpanzees in the wild fail to perform pointing gestures.

4.2 Why don't apes point? Distinguishing the indirect cause from the direct cause

Why don't apes point? Let us begin with the idea pursued by Hare & Tomasello (2004): chimpanzees' abilities unfold in a framework of competition, not collaboration. In a chimpanzee clan, each individual looks for its own food. A piece of communication that intentionally seeks to transmit useful information would thus make no sense in chimpanzees. This call for attention certainly merits being taken into consideration. We must bear in mind the characteristics of each species' lives. However, the importance of the competitive framework in the clan could perhaps be compatible with some degree of altruistic or, at least, mutualistic, collaboration.

Since in a chimpanzee clan one gang of allies often confronts another gang, there are both confrontation and alliance. It is true that Tomasello (2008) distinguishes between 'helping allies in a fight' and 'cooperating'. This distinction, however, would, at most, reveal different – low or high – levels of efficiency, or different – less or more general – types of altruism, but has nothing to do with motivation, which is a shared feature on both cases. As can be seen, I have refrained from invoking Boesch (2005) and the allegedly cooperative strategies used by chimpanzees in their hunting parties. I have focused only on the indisputable aid behaviours during fights. But there is surely evidence to accept a fair amount of cooperation: See de Waal (2006), who ascribes a large amount of altruistic behaviour to chimpanzees.[2]

But do not misinterpret my reluctance in the face of the connection between absence of cooperation and absence of pointing gestures. There is no doubt that the human way of life is much more dependent – incomparably more dependent – on cooperation. There are three main types of altruism as defined by the 'commodity' involved: goods, services and information (Warneken & Tomasello [2009]), and it is only the second type – helping behaviour – that Warneken et al. (2007) in their experiments have observed in both human-raised and mother-raised chimpanzees. I also completely accept the idea that some changes in habitats or niches would have

---

[2] Chimpanzees share food, and foraging is a highly social affair. Like many other primates, chimpanzees possess specific vocalisations to announce that they have found a new food source. They cooperate regularly during hunting, territory defence, anti-predator behaviour and intragroup aggression, facts long known since the 1960s when studies of wild chimpanzees began in Africa. In addition, there is evidence that capuchin monkeys will help a human experimenter to obtain an out-of-reach object, irrespective of whether or not they are offered a reward afterwards (Barnes et al. [2008]). Marmoset monkeys spontaneously provide food to non-reciprocating and genetically unrelated individuals (Burkart et al. [2007]).

required higher levels of mutualism eventually producing the evolutionary emergence of hand-pointing or gaze-pointing. However, in my view, this higher demand of cooperation would have only indirectly caused that emergence. By contrast, the crucial step or direct cause would have been the emergence of the basic cognitive ability that I have called the 'third mode of processing the eyes of others' or 'second mental centre'.

Many authors take into account changes in habitats and ensuing changes in life style. Tomasello speaks of a more cooperative or mutualist way of life. Hurford (2007, p. 219) prefers to speak of a process of "self-domestication". (See Hare et al. [2005] and Hare & Tomasello [2005]; see also Wellman et al. in press.) But it is perhaps in Bickerton (2009) where, in the light of the Niche-Construction Theory, the issue becomes particularly prominent. As I have just said, in no way am I opposed to this whole range of proposals. But this is not my concern here. After distinguishing between the direct cause (that is, the cognitive changes) and the indirect cause (that is, the increasing necessity to cooperate), I will concentrate on the cognitive changes.

I agree with Tomasello, Call & Hare in their generous interpretation of chimpanzees' abilities. Chimpanzees are presumably capable of ascribing visual perceptions to conspecifics, that is, they would have reached the second mode as I have already defined it. But I disagree with the idea that it is only out of a lack of motivation to cooperate that chimpanzees fail to point. In my view, the direct cause for the pointing ability of humans is the exclusively human cognitive ability that I call the 'third mode of processing others' eyes' (or, more in general, the 'duality of mental centres'). In short, according to Tomasello, the difference between humans' abilities and chimpanzees' abilities for a theory-of-mind is only a question of degree. However, I see a clear-cut line between abilities based on an *a posteriori* expectation (monkeys' mirroring and the chimpanzees' ability to ascribe visual perceptions) and abilities requiring genuine simulation.

As for Povinelli (see mainly Penn et al. [2008]), I agree with him in stressing human exclusivity. But I disagree with his ensuing radical confrontation (i.e., his confrontation without bridges or interconnections) of mere conditioned learning in animals, and intellectual, exclusively human capabilities. In my opinion, if we pay due attention to the changes that the self-perceptible hand and the subsequent matching of one's own body with a conspecific's body provoked among primates, we are not depreciating human exclusivity, rather the contrary. In the terms used in this Section we can articulate it as follows: the best way of accurately understanding the complexity of the third mode is by comparing it with the second mode.

But, leaving these general comments aside, let us go back to our main point and see whether or not a concrete conclusion of my hypothesis, namely, that the presence or absence of pointing has a direct and cognitive, not-motivational cause, can be refuted. Let us start by noting that apes held in captivity multiply their request behaviour in the presence of benevolent carers as a result of the cooperative environment they live in. I will try to show that this request behaviour presents observable features that are very different from those of human pointing.

entirely within the reach of the abilities widely acknowledged in apes. Thus, this complies with a desideratum for any spontaneous communicative behaviour, namely, that, to be understood by its recipient, the behaviour involved must only require abilities possessed by the species of the individual performing it.

Let us compare this description of the behaviour of the animals of Leavens or Gómez with a classic statement made by Seyfarth & Cheney (2003, p. 159) on animal communication: "listeners (or recipients) obtain information from a caller (or sender) who may not have intended to provide it." As can be seen, the movements that animals of Gómez or Leavens directed towards the food or the closed door clearly exceed that situation, since the sender intends to provide information. However, they are still confined to an old type of communication, since their communicative success does not depend on the recipient's ability to perceive them as being communicative.

Let us look at the difference with the case of a genuine pointing gesture. If a pointing gesture is not interpreted as communicative behaviour, then it cannot be understood. An arm extended in the air would be seen as to be an absurd, rather meaningless piece of behaviour.[3] Curiously enough, Tomasello, who stresses how strange any mimicry can be for a recipient if the gestures involved are not interpreted as being communicative –"the recipient will see my iconic gestures as some kind of strangely misplaced instrumental action" (2008, p. 149 and also p. 203)–, never says anything of the sorts about pointing. In my view, however, mimicry and pointing are equally strange and absurd if they are not interpreted as being communicative. By contrast, the requests performed by Gómez's or Leavens's apes appear, I insist, as entirely sensible behaviour even for recipients who do not understand it as communicative. These communicative behaviours are 'environmentally shaped' actions, or, in other words, they are not 'communicatively shaped' actions.

In short, I am hypothesising that it is the understanding of communicatively shaped actions that chimpanzees lack. In order to interpret an experiment conducted by Hare & Tomasello, 2004 (see also Tomasello [2008, p. 40–41]), it could be useful to bear in mind this lack of understanding. More precisely, it could be useful in order to interpret the results of the experiment in a very different way from that suggested by these researchers. Hare & Tomasello (2004), in support of their explanation based on the absence of cooperative motivations, point out the astonishing contrast between chimpanzees' inability to understand a 'benevolent' (i.e., cooperative) carer and their

---

[3] Do you say that much of the animal behaviour performed to alter the behaviour of other conspecifics or animals is equally 'absurd'? My reply is that we must distinguish between two kinds of communicative behaviour. On the one hand, there are innate communicative signals whose production and reception have been phylogenetically linked and, consequently, can never be absurd for the individuals of the species performing them. On the other hand, there are the requests of the apes of Gómez or Leavens, which are, by contrast, improvised (Tomasello [2008, p. 20] insists on the point that the gestures of apes are much more flexible than their shouts). It is in relation to these requests that I maintain that for them to be understood by apes, they must be an 'environmentally-oriented type of behaviour'.

extremely skilful knack of understanding a ‘competitive’ carer. According to the hypothesis of these authors, the key point marking the difference between the success or the failure of the chimpanzees is just that: the carer’s attitude. Tomasello and Hare conclude that chimpanzees do not understand cooperative behaviours. But, if we read the proceedings of the experiment with care, we see that the lack of understanding took place in the face of a hand-pointing gesture, while the animals’ understanding, by contrast, was triggered by an attempt to reach a container (despite the obstacle presented by a hole in a curtain that was far too small), that is, by environmentally-shaped behaviour. Consequently, results would be more clarifying if the contrasting pairs could be dissociated, that is, the pair ‘competition and environmentally-oriented behaviour’, on the one hand, and the pair ‘cooperation and hand-pointing gesture’, on the other.[4] Contrary to Tomasello and Hare, I think that it is only the second element of each pair what makes the difference between success and failure. ‘Hand-pointing gesture’ (or, in other words, ‘communicatively shaped action’): Chimpanzees´ failure. ‘Environmentally-shaped behaviour’: Chimpanzees´ success.

Comprehension of the communicative use of eye direction, on the one hand, and comprehension of the pointing gesture, on the other, would both have the same crucial requirement. That practically all pointing gestures are accompanied by alternating looks is, as I see it, a good clue. Just as both behaviours are related to each other, the respective underlying processes would also be related as well. Infra, in 4.9, we will try to go deeper into the analysis of those underlying processes. But beforehand, it is convenient to underline certain points.

4.5 Commenting about Grice and also about triadic communication

Here we must remember Grice’s description of what speakers (utterers, for Grice) mean. The speaker intends (1) that the audience believe something or go on to do something, and (2) that the audience recognize the intention (1).[5] Focussing on comprehension of pointing gestures, from Grice’s research we can derive the formulation that recipients must perceive the producer’s intention of making them look at an

---

[4] Admittedly, I cannot think of an experiment designed to test the new pair competition/pointing gesture. What about an experiment combining an unfriendly carer, that is, a competitor, who hand-points towards another likewise ‘unfriendly’ carer? Needless to say this experiment would not do, because the understanding of a gesture pointing to a third person would very likely be different from the habitual understanding of the gesture. (But a gesture to a third party is possibly an interesting fact for primates: See Teufel et al. [2010]: “One specific facial expression that is given in response to social interactions between third parties was particularly efficient in eliciting gaze-following responses”.)

[5] As you can see, in order to stick to what is of interest right now, I have cut down to just two points Grice’s formulation, which even in its more primitive version (Grice [1957]) was made up of three points.

object. This is tantamount to what I have been calling 'the third mode of processing the eyes of others'.

In that case, a possible objection could be: is there any point in proposing new terms to designate something which already has a familiar, almost classic label, namely 'the understanding of utterers' communicative intention'? My answer is that Grice's description adequately depicts only one of the possible applications of the ability concerned. Note that Grice's description fails to account for self-conscious emotions' – embarrassment, pride, shame and guilt – (see above, 3.5) or 'four-hand-actions' (our issue in the next chapter). In these cases it is not a communicative intention of others that the subject perceives. Therefore, I prefer to continue using my formulation, that is, the concept of 'third mode of processing others' eyes' (or, more in general, the duality of mental centres).

Anyway, it is very important to distinguish explicitly the communicative intention I am talking about from another communicative intention on the part of the producer, whose perception would involve much less demanding requisites. Note that it is extremely easy to perceive a readiness for interaction or communication in the eye which (as happens, for example, at the end of a pointing gesture) fixedly stares at the recipient. This perception could require only the 'first mode of processing another's eye', or, more precisely in this case, primary intersubjectivity. But this would not be enough to understand a pointing gesture. Consequently, it is important to make perfectly clear what the difference between these two perceptions of communicative intention is.

Let us move on to a similar question. On one hand, I have described (above, in 3.3) the basic ability as being able to 'attribute visual perceptions to the eye which is looking at me'. In this ability, sometimes there is no 'third element', i.e., there is no referent characterising triadic interaction.[6] On the other hand, however, in order to show that basic ability, I am turning to pointing gestures or triadic communication. Let me explore this apparent contradiction in my position.

As long as there is, in the subject, no triadic interaction or attempt to communicatively guide another's attention toward some referent or other, we can have no basis for believing that this subject's dyadic interaction actually involves the basic ability. Who can guarantee to us that those dyadic interactions are surpassing the phylogenetically very old 'first another's eye processing'? However, when the subjects have shown their 'third processing mode' in pointing gestures (eye- or finger-pointing), then we would certainly be authorised to interpret their participation in 'mutual attention' episodes as the 'perception of an interiority which is looking at me'.

---

[6] In this regard I seem be moving closer to Reddy (2005), who insists that there is no need to wait for pointing gestures to appear if we want to observe the comprehension of another's attention. But, in my view, we would have to clarify what is being understood by another's attention in Reddy's formulation: see supra, 3.2.

4.6 Some unavoidable issues which must be dealt with

4.6.1 Wild chimpanzees that extend their arm in the direction of an object: How could those gestures really happen and yet be so scarce?

It is likely that this lack of understanding explains a controversy that has recently arisen. Some primatologists claim to have observed chimpanzees in the wild make gestures where they extend their arm and reach out their hand: In Vea & Sabater-Pi (1998), there is a very detailed description of a bonobo pointing twice to human observers who were hiding in some shrubbery. Other researchers are unsure, however. If these gestures did in fact happen, why have so few been observed? Why are such observations so utterly unusual? I believe that we can explain why such observations, being real, would be so rare. We shall start by looking back to Vygotsky's theory, later elaborated on by Bates, on the origins of the pointing gesture.

Children, at first, would make an effort to reach an object, stretching their arm and body out as much as possible, but without succeeding. The adult accompanying them realises what the child wants and gives him the object. As I said in 4.1, we can not accept this theory about the genesis of the pointing gesture in children. We can, nevertheless, bring the idea of conditioning to the controversy we have been addressing.

A wild chimpanzee stretches its arm in the direction of an interesting object. Its movement might arise from the conflict of opposite motivations; or it is also possible that its movement has the function of re-picture it as an objective in its mind (cf. Delgado et al. 2009 about private pointing in children). The question is that, even though a chimpanzee produces such a movement, it will never find in another chimpanzee a recipient like the one in Vygotsky's theory. Consequently, those observations about wild chimpanzees could be real. Their exceptional rarity would be entirely explicable.

4.6.2 Dogs and chimpanzees compared to the human pointing gesture

At this point, we need to create space for another issue. Hare et al., 2002, have shown that untrained dogs are sensitive to human gestures that point out a direction. He did not find this sensitivity, however, when he later studied wolves. As a result, he was able to conclude that dogs had been selected according to that criterion right from the very first time that humans had begun to make use of them. In a human environment, the best hunting or shepherd dog would have more opportunities to transmit its genes. This is not surprising: think back to Darwin's initial inquiries with breeders of domestic animals. We assume, therefore, that sensitivity to human gestures of pointing would have been selected in dogs.

Yet, I would like to continue my questioning. Originally, that is, in wolves, there had to be a base from which to develop toward the ability of our best dogs. Wolves hunt in herds and it is very possible that the head of that herd would decide which was the weakest of the group of potential prey. This decision would need to be known by the other wolves. All in all, wolves would possess, in the context of the co-ordinated hunt, a natural sensitivity to specific gestures by the dominant individual. It is on this base that the selection of tamed dogs would have operated. This generous position as regards wolves' abilities seems to have been confirmed and even strengthened by Udell et al. (2008).

Once we have reached this point, we can ask the question that we are interested in: how close are dogs to the human understanding of the pointing gesture? Would they be closer than chimpanzees? I could only say that, in essence, there is no need to believe that dogs are ascribing even a visual perception to the producer of the gestures. Thus, chimpanzees and ravens might be much closer than those canine abilities to the understanding of the gesture of pointing.

4.7 True pointing in chimpanzees brought up by humans?

Chimpanzees brought up by humans do eventually learn the pointing gesture. We must mention here Savage-Rumbaugh, on human-reared bonobos (see, for example, Benson et al., 2002). The success of this learning seems, in principle, to refute my explanation of the cognitive inability of chimpanzees for these comprehension and production. Am I not acknowledging, after all, that chimpanzees are able to learn how to hand-point? And, on top of that, don't we teach children how to do something similar? This issue is the object of a hot debate.[7]

As I have already said, I align myself with the defenders of the difference between these learning processes. I believe that Vygotsky's and Moore's antimentalist explanations (which in 4.1 I have dismissed as an explanation for pointing in children) would be, however, adequate for enculturated chimpanzees. In short, a more or less instrumental conditioning could explain the production of gestures (remember Vygotsky's proposal) or the reception (remember Moore's proposal) by enculturated apes.

In other words, on this issue I agree with Povinelli's criticism of the argument by analogy. Above, in Chapter 2, I did not agree with his deflationary opinion on the abilities of chimpanzees in connection with a conspecific's visual field, and, consequently, contrary to his stand, I found it correct to apply the argument by analogy. However, in the current issue, I definitely do not think that that argument should be applied.

This is not an arbitrary whim of mine. In my view, there are good reasons for differentiating between the two learning processes of pointing, that of children and that of enculturated chimpanzees. It is not only that the behavioural repertoire of

---

[7] I must say that Tomasello & Call, who were previously very keen on the influence of enculturation, have modified their view to some extent. Tomasello & Call (2004, p. 214): "It is likely that human experience only serves to modify existing social interactional and attentional skills, rather than creating new ones".

Therefore, if Nissen's anecdote and its interpretation were true, it would not hurt my proposal. In contrast, were this anecdote and interpretation true, it would become necessary to deduce a conclusion which was contrary to Tomasello's hypothesis – it would become necessary to grant chimpanzees motivation to share intentional states. There is only, therefore, one clear conclusion: we need reliable data regarding this behaviour as soon as possible.

Before we move on, let us look at the question of falsifiability. Although it can admit those apparent data, the proposal I am making holds back from predicting them, as I have already said. My point relates only to a cognitive inability in chimpanzees. Can chimpanzees desire the situation they would find if, contrary to what occurs in reality, they were able to carry out pointing gestures? Could they then, driven by that desire, try to accomplish it by using other, less economical means (along the lines of Nissen, for example)? I have no opinion about these questions; instead, I will make only the trivial comment that we must not attribute this desire to chimpanzees if the use of such less effective means was never observed. The ability to have a type of desire is an adaptive ability only if there is some instrumental means connected to the realisation of that desire. But beyond that point, there is no further prediction on my part.

Falsifiability would not, thus, come to my hypothesis by this route. Only Tomasello's lack of motivation theory would be falsifiable this way. But do not automatically condemn me. In many other paragraphs, both before and after this one, I scrutinise possibilities which, should they be true, would cause me to withdraw my hypothesis.

4.9 What are the requirements for a genuine understanding of pointing gestures? A closer look at the expectation/simulation dichotomy

4.9.1 Going back to the expectation/simulation dichotomy

Now it is time we also addressed the dichotomy between expectation and simulation in this field of finger-pointing gestures. We saw above, in relation to the ability to ascribe visual perceptions to a conspecific, that the expectation resource works up to a point. Beyond that point, expectation becomes bankrupt. The key to this limitation is that the inner contents corresponding to the fellow staring at me can never be an expectation of mine. What happens, then, in the case of kinaesthetic-postural and visual interiority of the conspecific and comprehension of the pointing gesture?

If the perception of another's interiority remained at the level of mere expectations, then the communicative reception of the pointing gesture would never be attained. Expectations of any kind involve the possibility of their fulfilment. For that reason, it will be impossible for A to have an expectation of inner states which seek to communicate with A. But if that communicative intention is not included in the perceived expectations, then the movements extending the arm and moving the hand forward, which were perceived in the conspecific's body, will be absurd. As I said in

4.4, extending one's hand in mid-air is meaningless if it is not related to communication. Consequently, the human-type comprehension of pointing gestures cannot be based on the old expectation resource, but on the exclusively human resource of motor simulation. This simulation must be ascribed to a radically not-own self.

In the chapter on monkeys' mirror neurons (above, 1.3) some readers may have felt uncomfortable at my selection of received gestures of hitting or begging as examples to be contrasted with the only manual movement able to activate mirror neurons in their mirroring-role, that is, the grasping movement. These readers might have felt that mirroring needs a certain 'contemplative distance' which does not occur when the not-own behaviour so directly and immediately affects the animal being considered as the subject; or, in other words, when the not-own behaviour is interacting with the subject. Now, here in Chapter 4, we can completely back up and justify that reticence, which was in the past entirely real in some of my students, in addition to being possible in some of my readers now. The mirroring applied to not-own movements that are interacting with the subject would be a more costly and demanding process than a non-human brain can afford.

The key point here is to stress what is achieved with the third processing mode of others' eyes (or, in other words, with true simulation, or the perception of a radically different self). With this acquisition a brain could, for the first time ever in evolution, attend to two lines of awareness (the primary mental centre and the second, simulatory centre). Obviously, the second line is not only partial and discontinuous, but also much weaker than the first line.[9] However, despite its weakness, the second line is enormously important.

---

[9] The second line is weaker because it is sporadic and cannot guide behaviour. (Flanagan [1992, p. 103]: "If it means having the experiences exactly as the experiencer has them, then this never happens; but if it means understanding another or conceiving of what things are like from another's point of view, then it often happens.") This weakness of the not-own interiority inside one's own mind contrasts with the subject's recognition of the fact that, in principle, the contents of the two lines are at a similar level. I would suggest that this contradiction, inherent to the second mental line, between an actual lower-rank status and a theoretical equal-rank status, is one of the defining features of human beings. Neither interests of others nor the remote future are as intimately related to biological forces as are the contents of the first line. Cf. Albrecht et al. (2010): "Our results support the hypotheses that participants show less affective engagement (i) when they are making choices for themselves that only involve options in the future or (ii) when they are making choices for someone else." (It is well known that humans discount the value of future rewards over time.This discount – althought, in my view, it is a less interesting issue – can be more easily studied than the role of interests of others: Metcalfe & Mischel [1999].) Despite this fact, however, a human being knows objectively that both interiority of others and his own interiority (or both the remote future only envisaged by evocation and the immediate future connected to expectation) are similar. This problem, this contradictory duality of opposed estimations, appears alongside the ability for a 'theory of the mind'. This contradictory state can only be alleviated by means of an effort aiming at focusing on the complete known reality (e. g., Boyer [2008] has suggested that the capacity for episodic future thought – also referred to as prospective mental time travel – may underlie the human ability to make choices with high

4.9.2 The different manners in which somebody else's body may be informative

I would like at this point to refer to a position from philosophical anthropology that I find particularly sensible.[10] This position has been frequently neglected and ignored throughout history, but reappears from time to time. The human body would be a mixture of opacity and transparency. On the one hand it hides its thoughts and, in general, its mind or personal interiority; in this sense, the body acts as a genuine barrier. On the other hand, however, the body is a bridge (the only possible bridge) toward that interiority, and is thus the condition that enables us to open up to others.

Needless to say, it is not only when confronting a human gaze that the body is informative. Such a restriction would be obviously false. Communicative innate signals or, likewise, the circumstances allowing the ascription of a given visual field to another individual's body, are also perceptible in the body itself. They are, obviously, only for perceptive abilities permitting it in each case. However, given those abilities, the conspecific's body is informative.

That informative function of the body can be found, I repeat, in many animals. But now I want to deal with its role as a bridge towards the perception of a radically not-own self, a perception which, according to my hypothesis, would be exclusively human. I think we should analyse further the relationship between another individual's body and the perception of a radically not-own self. If we focus on the exact moment in which the individual who is pointing (after having looked at the object) turns his or her head to look at the recipient, we will realize that the only information that the producer's body can provide at that very moment consists of signalling a readiness to interact with the recipient. Nothing else – nothing relevant for the comprehension of the pointing, I mean – can be extracted from the perception of the producer's body. At that very moment the producer's body goes on stage only to interact with the recipient. Now his or her body is neither looking at the object in question nor ready to establish any relation with that object: the producer's body seems only ready to interact with me. In short: if we do not ascribe to the producer interiority in clear contradiction with what his or her body at that very moment is proclaiming, we will not be able to understand the pointing gesture. The genuine complexity involved in triadic communication derives from the fact that only one single relation can be directly perceived in a body at every moment. Either you perceive in that body the innate communicative signals (of readiness to interact with you), or you perceive the relationship between that body and a given object, but never both things at the same time. The only way to go beyond this is to conceive an interiority different from one's own, that is, a second

---

long-term benefits). It is an issue which interests me deeply (Bejarano [2010b]), but one for which there is not space in this present book.

[10] Hierro Sánchez-Pescador (1990) pointed out that, in order to leave their disorientation, the philosophy of language should be connected to anthropology. That recommendation, which I considered to be so lucid back then, now seems to be completely indisputable.

mental centre in one's own mind. That radically different interiority would have his/her own accumulative sequential line, a line where the current moment is the heir of the previous moment.

However, is this accumulative sequence not in fact the case of a chimpanzee's ascriptions? Remember the third experiment mentioned in Tomasello, Call & Hare (2003). A lower-ranked chimpanzee, having seen that at the moment *m* the dominant individual was looking at the food, is able to foresee that at the moment *m + 1* the higher-rank animal will go for that food (*supra*, 2.2.1). Is it not true that this chimpanzee's ascription involves the same accumulative sequence that I have presented as typical of human pointing gestures? No, I think there is a crucial difference. Admittedly, the lower-ranked chimpanzee's forecast relies on the visual perception that it has previously ascribed to the higher-ranked animal. At that very moment *m*, the ascription of a visual perception *ipso facto* generated the forecast of future behaviour directed at the food, and that forecast remains active at the moment *m + 1*. No objection to that, I insist. But at the moment *m + 1* the higher-ranked animal's body reveals nothing that contradicts the forecast. It is just that – i.e., to contradict the forecast – that the producer's looking at the recipient provokes at the second moment of his/her pointing gesture. At the previous moment, the producer's look and hand directed both towards the object might be remarkably clear and perceptible signals of the producer's probable future actions on the object. However, immediately afterwards, the producer's body interrupts his relationship with the object, puts an end to those apparently announced actions and reveals a completely different readiness, a readiness to interact with the recipient. (In other words, in my production of the pointing gesture, "*for you* signals", as Tomasello [2008, p. 96] calls them, cannot be simultaneous with the signals of my focus on the object.) Consequently, only by perceiving the self, or mental line, of the producer, can the recipient avoid being deceived by the apparent cancellation of the previous indications. Only by conceiving an interiority different from his, that is, a second mental centre in his own mind, can the recipient truly understand pointing gestures. The recipient must understand that the inner state at the second moment aims at communicating something about the object looked at at the first moment. Or, in other words, he must understand that the gazing at the object at the first moment was only preparatory to the second moment.

Focussing again on the informative role which the perceived human body plays for a human perceiver, I would stress that the informative body is, first and foremost, the body in all its movements, i.e. its visible as well as its audible movements. It is in not-own movements that not-own interiority can be perceived. If we keep this in mind, the almost esoteric aroma of terms such as 'mind reading' disappears and gives way to fairly concrete explanations. We find comments on this term, as well as, of course, the hope that somehow adequate explanations may be found, in a number of authors: see, for example, Sabbagh & Baldwin (2005).

Needless to say, therefore, that my view has absolutely nothing to do with the debate between 'embodied cognition' and 'theory of mind' (see, e. g., Spaulding [2010]). In my view, this alleged opposition ('embodied cognition' *versus* 'theory of mind') rests on a very constrained understanding of the 'theory of mind', an understanding that identifies it with a single one of its subcurrents. The distinction I am interested in is not the one at the core of this debate, I insist, but that of the three modes of processing a conspecific's eyes, or, more concretely, that between expectation and genuine simulation.

But, let us continue with the relationship between the third mode and the movements of someone else's body. Language, only perceptible in the results of its motor component, would be, in my opinion, the paradigmatic example of this opportunity to grasp mental contents of others. Analogously, some 20th-century conceptions about language are also typical of the persistent reluctance to accept the above mentioned philosophical position: from considering any mental reference as a taboo, research shifted to concentrating on artificial intelligence and computational cognitivism. Neither of these extreme approaches paid due attention to the peculiar bridging role played by the human body. However, it is not in language, but in other less complex abilities, that the origin of this role is to be found.

I have dealt here with the human ability to point using a finger and the eyes. It is in these skills that we must look for the first showings of that bridging role towards a radically not-own interiority. In addition, as I have tried to suggest, pointing gestures would be similar to several features that distinguish primates, and at the same time subtly different from those features. Studying pointing gestures could thus bring us closer to two understandings at the same time. On the one hand, we would be close to understanding why in those gestures the informative role of the producer's body is peculiar, that is, why it is possible only with the human observer. In my view, the peculiarity that we have just seen at its initial stage, that is, the ability to understand a radically not-own interiority, is the grain of truth enclosed in the classic concepts of spirit or soul (cf., e. g., Barresi, 2010, or also Barresi & Moore, 1996). On the other hand, if the ascription of visual perceptions to conspecifics were to be confirmed in chimpanzees but not in other animals (see above, 2.2.3 and more importantly 2.2.4), we could also be moving closer to an explanation of why it was precisely among primates that this special animal, the human being, emerged. But let us leave aside all these premature comments and continue with the task in hand.

Thus far, we have brought together two abilities. The comprehension of the communicative use of the alternating gaze and the comprehension, likewise, of pointing gestures would share a single crucial requirement. Now, in the next chapter, we shall also add four-hand co-operative action.

precisely when the respective suitable context has been finally achieved. This would not be sufficient for our plan for four-hand action. Here, the movements have to be shared out between two subjects.

We should, perhaps, comment further on the objection and response at which we have just hinted. It is true that the movements forming the second step could be learned as context-related instrumental conditioning (with the goal, obviously, of getting water). Specifying that this conditioning would have to be linked to a context does not constitute any special requirement; any conditioned instrumental learning is context-related. As a result, learning the second step (i.e. the water-extracting movements) could, therefore, be very simple. This is true; however, we should note that this simple learning is not of the kind required for planning. In a plan, the context associated with the water-extracting movements is a result that will be attained when the first individual performs his task. In the planner's mind the suitable context for them is the goal that a different agent would have to achieve.

5.1.2 A comparison with co-operation among chimpanzees

Let us be a little more specific about the key element of this kind of co-operative action (which, for reasons of brevity, I have called 'four-hand action', despite it not necessarily involving the intervention of four hands[3]). In so doing, it will be useful to compare it with data on co-operation among chimpanzees. Of course, research on this issue has been going on for a long time, although it is still a matter of contention.

Crawford attempted an experiment on this issue as early as 1937, although how this particular experiment has been interpreted is generally exaggerated. Many years later, the reliability of the available data on the issue of co-operation is still highly controversial. There is no general agreement regarding the observations made of the hunting parties of groups of chimpanzees trying to catch a monkey (although Boesch [2005] tries to present these expeditions as a task planned by the group).

But let us move on to encultured chimpanzees, and look specifically at the experiment carried out by Savage-Rumbaugh (more than twenty years ago) with Austin and Sherman, (two chimpanzees that had learnt to use a lexigram board). "Sherman was placed in a room in which were a number of boxes baited with different kinds of food. Each box could only be opened by using a different kind of tool, and Austin was placed in another room with all the tools. Through a window Austin could see all the different foods in the boxes, and he would signal to Sherman which it was that he wanted. Sherman responded by using the lexigram board to tell Austin which tool he needed to open that box. Austin would select the appropriate tool (e. g., a key or a wrench) and pass this through a small hole to Sherman. Sherman would then open the right box and pass the food through to Austin (eating a small portion of it along the way)" (Rumbaugh et al. [2000, p. 121]).

---

[3] In a similar way, pointing gesture can be called 'four-eye perception'.

As we can see, this old experiment focussed on the use of lexigrams. However, we are interested only in collaboration. Thus, from this point of view, the chimpanzees Austin and Sherman are very similar to Gómez's gorillas who asked their carer to open the door. However, in the chimpanzee experiment an additional point is apparent. The animal not only knows to ask for an intermediate step in the task to be resolved, but also plays the role of recipient in this type of requests.

Another question is whether the abilities revealed in this example can be legitimately considered as merely a result of enculturation. On this matter, I tend to think that training and learning are simply making the most of chimpanzees' natural abilities. In human beings, as we shall see in other chapters, interpersonal relations and cultural learning do indeed have an impact on mental abilities. However, this would be because exclusively human processes, both on the part of the teacher and of the learner, play a part when children learn human culture. The situation would, for this reason, be different with encultured apes. This, as I have said, is what I tend to think, although on this issue we cannot as yet categorically either prove or reject any opinion. Indeed, as I said in 4.6.3, Tomasello & Call, who were previously very keen on the influence of enculturation, have modified their view to some extent. At any rate, I consider my own opinion on encultured chimpanzees as a theoretical wager.

After all these clarifications, let us consider the following question: is there any difference between this co-operation and our four-hand action? The 'tool-opening-food' plan could have been carried out by one single chimpanzee. The plan finally put into practice did not differ from the plan each individual had learnt by himself. If chimpanzee A were not locked in the cage, he himself would go for the tool, but he cannot do so in his current situation. Likewise, if chimpanzee B had been able to, he would have carried out the subsequent behavioural steps (opening, food) by himself. Only the circumstance that the box was inside the cage and, consequently, inaccessible to him, forced him to leave the next steps in the hands of his companion.

By contrast, one single individual could never have performed the planning involved in our four-hand water-extraction task. From the very moment the plan was formed, it required the involvement of two individuals. In short, whereas chimpanzees A and B each had one individual plan which was repeated in both of them, the planner of the water-extraction task has a plan requiring two individuals. There is an enormous difference between a plan that is repeated in two individuals and a plan that intrinsically involves two individuals.

But the data on which this explanation relies are too comfortable. Let us make things slightly less comfortable, and imagine that only one of the chimpanzees knew how to operate the tool. Here, the least able chimpanzee would be like Gómez's gorilla, and the other chimpanzee would play the role analogous to the gorilla's carer. Does this change anything? Certainly, it could not now be said that it is only external circumstances which prevent the chimpanzee performing the entire task by himself. Collaboration is now intrinsically necessary for him. However, we would not be looking at a genuine four-hand action in this case either. The request for help is the instrumental means for the sub-goal of opening the box. When this sub-goal is achieved, the individual will return to working alone (although, clearly, in order to ensure success when the task is repeated, he will have to compensate his companion). On the other hand, in four-hand tasks, the companion must maintain his collaboration while the subject is acting. Consequently, this simultaneity of the two actions has to be considered in the unitary plan.

Summing up, what would be the crucial requirement of this four-hand plan? The crucial requirement, I stress once more, would be the same requirement we suggested for understanding alternating gazes or pointing gestures: I have to envisage a not-own centre within my own mind. In other words, what is needed is to perceive an interiority so radically different from me that it is able to address me.

Let us open out this contrast once more. What happened when chimpanzee B handed the tool to the caged chimpanzee A, or when Gómez's gorilla, in a kind of implicit request, put his carer's hand on the closed door? In these cases, it is certainly possible that there was not only expectation of the desired results, but also even, moments later, an *a posteriori* expectation about the inner state (visual perception of the box or the door) corresponding to a not-own position. However, even accepting this, the interiority thus perceived may continue to lack the feature of being conceived of as radically not-own. In such situations there would be no need to perceive the not-own interiority as interacting with the subject. By contrast, when we think of the role of the not-own interiority in the full plan of a genuine four-hand action, that interiority must be defined as a radically not-own self.

(Above, in 3.5, to explain why we did not study 'self-conscious emotions', I said that amongst the consequences of the duality of mental centres, what interests me here is the series of abilities out of which language could historically emerge. Therefore, I will be asked why I have studied here four-hand tasks. I reply that the ability for those tasks implies taking a first step towards pre-motor planning. Certainly, that step is still little. However, from there on, we will reach the latent imitation of motor sequences which, as we will see in Chapters 8 and 9, will be crucial for the emergency of language.)

So far we have given only a single example of four-hand co-operative actions. However, with primitive technology this kind of co-operation was clearly a constant need. If we were looking for an explicit statement of the schema which would be so often repeated, it would be something like 'Hold this steady while I do that' (once again, I am quoting Reynolds [1993]).[4] What interests me here is that the ability to do this kind of four-hand task became more and more essential. The development of that ability in a child is something that it was worthwhile to ensure. And thus we come to the final issue in this chapter: interpersonal motor co-ordination games.

---

[4] This collaboration would not have to imply either 'altruism in front of genetically related individuals' or even 'reciprocal altruism'. Each collaborative task would bring immediate results for each of the collaborators.

5.2 The interpersonal motor co-ordination game

5.2.1 From the adaptive advantages of play in general to the interweaving of evolution and culture

The pleasure of game playing did not appear, of course, only with human beings. Among mammals, playing games is widespread. The chasing and fighting observed among young animals are, firstly, probably great fun for them, and secondly, very useful given their contribution to developing the skills adult animals need for survival. The novelty in human play is simply that the skills exercised and promoted by the child are the specific skills needed by human beings. This view of play, evidently, is not far from the 'pleasure of function' in early 20th century authors. But there is a difference: nowadays, we take evolution, adaptive advantages and the selection of favourable features very seriously (for a recent exposition of this issue, see Ellis & Bjorklund [2005]).

In short, the pleasure of playing works like any other pleasure. Through it, the subject is not only taught some adaptively useful behaviours, but, in addition, often has them activated in him. However, in the particular case of game playing, the adaptive usefulness of pleasurable behaviours, although undoubtedly enormous, is indirect. It has to do with learning and updating skills that cannot be completely specified on the genetic level.

A question arises, since, in the case of humans, some of the skills exercised and promoted by game playing will be requirements for cultural learning. On what level should we locate the usefulness of these games? On the level of evolution or of culture? My view is that the requirements of cultural learning became adaptively useful, from a given moment onwards. They became so adaptively useful that rather than saying that their presence provided the individual with advantages, it would be truer to say that their absence meant disaster. For that reason the biological pleasure mechanism was able eventually to become associated with play that exercised and promoted those requirements.

Human abilities that would be exercised in this way by game playing are of different types. At this point, we are interested in the ability to co-operate in four-hand actions. Specific games related to these actions almost certainly exist in all societies.

5.2.2 An interpersonal motor co-ordination play

In the part of Spain where I live there is a play called 'Té, chocolate, café' ('tea, chocolate, coffee'): "Tea" (adult and child each slap their own knees with the palm of their hand), "chocolate" (adult and child each clap their own hands), "coffee" (adult and child clap each other's hands). The question referred to above regarding the interweaving of evolution and culture arises here in an extreme form. The adaptively advantageous exercise of the potential for four-hand co-operation is embodied here in a cultural tradition taught by adults to children.

What was it that evolution supplied in order to establish such a game? Almost certainly, evolution merely needed to design that particular pleasure or innate consummatory pattern. The attentive and joyful reaction of children to any event involving something similar to four-hand co-ordination would have prompted adults to try to prolong that joyful moment; thus, each conventionalised game of this kind would eventually have come into being.

However, we must also explain co-operation on the part of adults. What can we say about adults? On one hand, adults would always experience pleasure whenever they catch a baby's attention. This pleasure would have a role in adults' general behaviour towards children. (This would be similar to what occurs in the case of lullabies and *motherese*. Fernald [1989] showed that the musical component or intonation pattern used in lullabies and motherese is ideal to attract a baby's attention). On the other hand, and more particularly for this game, we might also point out that adults would still be sensitive to the innate pleasure which guided them during their own childhood. Although those games are adaptive especially during childhood, it was beneficial that the pleasure associated with them did not fully disappear in adult life. Indeed, it is true that continuing to find pleasure in games during adult life can lead adults to perform activities which would not be in the least adaptive; however, this risk is extensively compensated for by the tremendous usefulness derived from games which adults play with children.

5.2.3 The 'tea, chocolate and coffee' game: The learning process

Let us analyse how a child learns to play the 'tea, chocolate and coffee' game. From a long time before, the child has known how to imitate the clapping hands movement, or the movement of slapping something with his hands. Hand movements are the epitome of self-visible movements and are, therefore, the easiest to imitate. What is of interest here is that the motor imitations involved in the first two steps of the game (the 'tea' and 'chocolate' phases) can be performed using the old resource of kinaesthetic-postural *a posteriori* expectations. The imitator could simply detect the kinaesthetic-postural expectations corresponding to the movements seen in the model and then fulfil them on his own person. Nothing else would be required to successfully imitate these movements. But it is obvious that the *quid* of this game lies in the third step (the 'coffee' phase): without it, the game would be reduced to imitations similar to the ones the child had been performing for many months. What does the third step involve?

When, in the third step, the child observes the model's movements he must also perceive that the model is addressing him both physically and communicatively. And here, we come to our constant refrain once more. Merely detecting a kinaesthetic-postural expectation is not enough in the third phase. The inner state detected in the model turns out to be addressing the child himself. Consequently, that state can never be encompassed in one of the child's expectations. True motor simulation and the radically not-own centre would necessarily be involved in the third step or, in other words, in the step which forms the *quid* of this game.

detects kinaesthetic-postural *a posteriori* expectations in movements of others. However, although this procedure actually enables the internal analysis of not-own movements, it will never enable me to conceive that this self is looking at me, is addressing me or copying me. Consequently, understanding that the other is imitating me would involve the crucial requirement of a double line of awareness. By analysing the pleasure experienced by the child who had been acting as a model, we have also explained the variant in which the adult plays at imitating a child.

Turning now to the child imitator, at what point does he experience enjoyment? Obviously, he has had to be copying the model's successive movements. Even so, he is still waiting for the moment of pleasure. This is why he continues with his silent imitation. The imitator's enjoyment will only occur when the model realises he is being copied and looks up with an amused expression at the imitator. As you can see, we find here the same cause as before. The pleasure only occurs when the interiority that I am perceiving in the other becomes, by directing itself at me, a radically not-own self for me.

At this point, let us begin to get interested again in chimpanzees. Nielsen et al. (2005) provide the first evidence that a chimpanzee is able to detect that he is being imitated when "the experimenter replicated all movements and body postures of the chimpanzee as he exhibited them". This ability was entirely to be expected, bearing in mind the well-established fact that chimpanzees can imitate simple movements (see supra, 2.1.2). However, according to my suggestion, this detection on the part of chimpanzees would be different from the playful and pleasurable experience we find in children. Children would be to able to perceive a not-own interiority directed at themselves and would, therefore, have to abandon the resource of expectation and turn to simulation. It is this required establishment of simulation and of a second centre in their minds that would explain the enjoyable character of these children's exercises. By contrast, nothing of the kind would occur in chimpanzees, who, after themselves having been repeatedly imitated, would already have established the prediction that the movement which they are about to perform will also be observable in the body which they are observing. However, as long as chimpanzees keep that kinaesthetic-postural interpretation of the not-own movements active, they will not be able to understand them as communicative signals: This is what would be deduced from the hypothesis set out here.

In short, that this pleasure ever emerged in evolution was undoubtedly because there were adaptive advantages to splitting of one's own mind in two in that way. The advantages would come through the communicative use of alternating gazes and the finger pointing gesture, as well as through the ability to perform four-hand actions. This is what we have said so far. Evidently, however, the basic human characteristic suggested here will be truly significant only if we can show its connection with language. This will keep us busy in Sections III and IV. In the first of these Sections, we will deal principally with how what we will call the symbolic ability might have emerged. In Section IV, it will be full, syntactic language that we will attempt to derive.

SECTION THREE

Specifying some necessary requisites of language

This section deals with some necessary, but still not sufficient, conditions for language. We will give attention, first, to what has come to be known as Saussurean parity, or symmetry between the production and reception of linguistic meaning. The second requirement we will examine is symbolic ability. In both cases, we will try to show the connection with the basic human ability.

CHAPTER 6

Saussurean parity and the perception of a radically not-own self

We will now, for the first time in this book, begin to look directly at language. What will occupy us here is but a small part of language, an aspect which, despite already involving a socio-culturally advanced learning, looks like an almost immediate, scarcely-developed, derivative of the pointing gesture. But we are getting ahead of ourselves; we must proceed more slowly.

When a recipient hears an order or a request, he understands that he is being ordered or requested. By contrast, when he utters exactly the same word with the same intonation, it is he who is ordering or requesting. In spite of this difference, the word in question is the same for him in both cases.

This happens in all human languages. We have here an absolute linguistic universal which, despite its simplicity, may be of interest. This characteristic of signs has come to be known as 'Saussurean parity' or 'producer/recipient symmetry', or in Tomasello's (2008, p. 103) words, 'role reversal imitation in communicative conventions'.

6.1 Toward a formulation of the problem

6.1.1 Production and reception in animal communication

Does the same thing happen in animal communication? At first glance, the parity between the signal produced and the signal received may seem a natural and inevitable element of all communication. It is, however, a question we must tackle.

There is an inquiry into parity, which has become a classic, in Hurford (1989). Likewise, Arbib (2005) or Rizzolatti & Arbib (1998) hypothesise that mirror neurons would evolve to support parity. However, the aspect focussed on there is different from the one occupying us here.

For the time being, we can state that this parity (the awareness of the identity between the signal produced and the signal received) is not essential for animal communication to occur. In order to show an aggressive attitude, a fish raises itself up and exhibits the red spot on its belly. Another fish sees that red spot and reacts accordingly, whether by retreating or attacking. The recipient has understood perfectly the communicative value of the red spot. Even when confronted with a supranormal stimulus, a remarkably big red spot on a wall, the fish will react in the same way, although more intensely. However, when it, in turn, becomes the producer of the same signal, it will not know that it is showing the red spot on its belly. The fish is simply in an aggressive state at that point and performs the corresponding innate motor pattern. In short, these fish have absolutely no need to equate the aggressive signal they produce with the aggressive signal they receive. When we humans observe how fish behave, we can easily identify both signals as equal, but fish themselves cannot. It is clear that fish can do without that identification. If we leave the world of fish and look at a cat's swept back ears or its entire body in a demonstration of submission, for example, we will have to draw the same conclusion. A cat, unable to recognise itself in a mirror, does not know what it looks like when it adopts a submissive posture. However, when it takes on the dominant role and observes those same postures in another cat, it understands them perfectly. Evolution must have made the necessary arrangements for this co-ordination to occur. ("In the evolution of any communicative system, whenever change of any sort occurs, there must be a change in two respects: the signal and the receiver": Alexander [1962, p. 465][1].) The individual animal does not need to consciously equate the signal received with the one produced.[2]

Thus far, we have limited our examples of animal communication to a particular type. We have only mentioned visual signals that are not perceptible in one's own body. The fish cannot see its belly, nor the cat see its ears. However, there are also very frequent auditory signals in animal communication, signals which are perceptible both for the receiver and the producer.[3] Consequently, the animal would have the possibility of identifying a signal she receives with the signal she produces. What we must now

---

[1] The manner in which the change in the two plans will be executed (that is, in the production and also in the reception) is a more difficult question. Scott-Phillips (2010, p. 79) says: "Since mutually dependent behaviours are unlikely to emerge simultaneously, this gives rise to a prediction that communication will only emerge if cues or coercive behaviours do so first".

[2] This same consideration can also be applied to behaviours called imitative behaviour, such as birds grouping together in flocks or fish in schools. Certainly, here, contrary to what occurs in the cases we are dealing with, there is no inversion (threaten/be threatened, e. g.) of the producer and recipient of the relevant stimuli. Here, the result is that the behaviour of both is identical. However, despite this difference there is still a strong similarity between those communicative cases and these 'imitative' behaviours. In these birds (or fish) grouping together, it is again only the human observer that perceives that the produced and received stimuli are equal. No bird joining the flock perceives its behaviour to be identical to that of the other members of the flock. But the issue of imitation will have to wait until later chapters.

[3] How is it that hearing his own cries does not induce the corresponding responses in the animal making those sounds? How is it that communication by means of self-perceptible signals does not have the same impact on the producer as on the recipient? Answering this question is straightforward: it has been well known for years that there is a kind of phylogenetically very old brain mechanism which 'attenuates' or suppresses some expected perceptual consequences of one's own movements (see above 1.4.1). In addition, we do not hear our own vocalizations as we hear others. There is distortion that occurs from bone conduction when perceiving one's own voice: Maurer and Landis [1990].

(Go away!, Get out!) will then be experienced as being very different from the threat produced by oneself with precisely the same words; the same thing happens for 'Give me!, Water!', for example.

The problem remains unsolved, however, even after this alleviation and in spite of it. In the receptive processes for these articulatory-phonetic signs, the core is just the same as if the hearer himself were producing those same signs. Since the articulatory-phonetic patterns are interpreted in production-format, reception of them inevitably refers back to the production process. As a result, we will have to keep looking for a solution to this problem. Saussurean parity is paradoxical and requires explanation.

6.2 Saussurean parity and the second mental line: Our suggestion for resolving the problem

Why would interpreting in production-format the received signal not be harmful in human communication? In my opinion, this question takes us directly to the central hypothesis of this book. We humans are able to perceive a radically not-own self, to be exact, a self that is currently looking at or addressing us. This not-own centre that emerges inside our own mind would be the centre of true simulation. In the third processing mode of the eyes of others we would ascribe visual perceptions to eyes that were looking at us. In four-hand tasks we would detect an internal aspect to movements being directed at us. In our reception of speech we would analyse (analyse in production-format) signals that turn out to be directed at us. In all these cases, expectation would not work; simulation located in our not-own centre would be necessary. For this reason (namely, that simulation would be linked to the not-own centre), what is irreconcilable in animal communication can now be reconcilable in human language. A conscious identification of produced and received motor patterns would now take place. However, thanks to the human ability to conceive a not-own centre inside one's own mind, that identification does not obstruct difference or opposition between states accompanying production and states accompanying reception.

Let us sum this up. Why does parity between production and reception become a problem? Only because reception entails interpretation in production-format. Where is the solution? At the very same point which causes the problem. The kinaesthetic-postural interpretation of not-own movements, if applied to a communicative movement directed at oneself, must necessarily be performed in the second, or simulatory, mental centre. Thus, although the reception involves a production, it is a production assigned to this centre.

We now need to face some data which seem to threaten the proposal which has just been presented. According to recent investigations, something like mirror neurons has been found in songbirds (Prather et al. [2008]). Thus, there are animal communications whose reception involves their kinaesthetic interpretation. Does this force our proposal to attribute also to birds the duality of mental centres? I would stress that bird singing does not involve dyadic communication. By singing, birds mark their territory, proclaim what species they belong to and probably send an honest signal of their good health. However, this hormone-driven singing does not involve communication deliberately directed at another individual. We know that birds reared experimentally in complete isolation begin singing at exactly the same age as their wild conspecifics (even if the former only sing the basic melodic pattern, without the dialectal embellishments of the latter: Marler [1991]). In this sense, it is convenient to remember that specialists in vocal behaviour of birds establish a strong difference between male singing and the calling coming from any individual within the flock. In other words, although songbirds can obviously be heard, there is no clear addressee's role involved: no recipient understands the message as being produced for him or her. Consequently, in birds, although they carry out the kinaesthetic-postural interpretation of the singing heard, that is, even if their reception of the singing takes place in a production format, the problem detected in the Saussurean parity does not occur here.

An argument in favour of these suggestions lies, of course, in the more comprehensive explanation which they provide.[6] In this way we can bring together several human abilities, more precisely, abilities that originate in children at relatively similar ages, as is the case of early linguistic communication and pointing with a finger or indicating with the eyes. Saussurean parity would already be present – this is the key point – in the pointing gesture. The receiver of the pointing gesture would have to estimate the kinaesthetic-postural and visual interiority that he or she would have if he or she were in the body of the producer, but, at the same time, the recipient realizes that what he or she, the receiver, is doing is not pointing, but receiving a pointing gesture.

Now, we must set out the reasons there are for believing that interpretation in production-format occurs in human speech. Where should we look for them? Two directions appear promising. First, Piaget's theory of imitation of new and complex motor patterns, or, in other words, the theory of 'motor reception during the learning stage'. Second, the question of the meaning of deictics, or of some deictics, to be more precise.

---

[6] I have said that in this comprehensive explanation would lie an argument in favour of my hypothesis. We should, I think, at this point delay no further a question that has already raised itself on several occasions. What is it I am doing in this book? As I said in the introduction, the hypothesis being put forward in this book is a synthesis. I have collected data from here and there, and simply make use of them as arguments in favour of my theoretical framework or synthesis. In this way, the mutual support between the different arguments emerges. But, and this is the key question, should we be taking into account support like this which is based on the decision to put such data together? This will of course depend on the theoretical wager. I will return to this question in later chapters.

6.3 ‘Motor reception during the learning stage’, the reliable core of the ‘motor theory of speech perception’

6.3.1 Liberman’s theory: The ‘motor theory of speech perception’

Alvin Liberman devoted the whole of his long career to developing the discovery he made in the 1950’s while doing some work for telephone companies. His discovery was this: the confusions most frequently observed in recipients cannot be described as confusion between auditorily similar sounds (or, in other words, between sounds regarded as similar on a spectrogram), but rather as confusion between sounds which are articulatorily similar. The relevant data had to do with a double dissociation: some sounds which a spectrogram describes as close, but which are very distant from an articulatory-motor point of view, were never confused. On the other hand, some very distant sounds according to the spectrogram, but which are close from an articulatory-motor point of view, were very frequently confused.[7]

Liberman formulated a theory from this data: the ‘motor theory of speech perception’: see Liberman & Mattingly (1985) (or also Liberman & Whalen [2000], the last, almost posthumous, paper that Liberman wrote). Speech reception would involve latent articulation. Motor simulation, which many years later would ride the crest of the wave, along with mirror neurons and the simulationist trend of the theory of mind, was clearly visible in Liberman’s contributions of the 60’s.

Liberman very soon moved on to hypothesise that the phonemic categorisation of sounds would depend on this articulatory reception. However, the discovery that many animals were able to perform some categorisation of sounds (Kuhl & Miller [1975], Kluender et al. [1987]) hit this on the head, and brought considerable discredit to the theory. Thus, Galantucci et al. (2006) conclude that “the claim about the phonemic categorisation of sounds is likely false”. See also Ohms et al. (2010, p. 1003): “Zebra finches (*Taeniopygia guttata*) can discriminate and categorize monosyllabic words that differ in their vowel and transfer this categorization to the same words spoken by novel speakers independent of the sex of the voices. The birds, like humans, use intrinsic and extrinsic speaker normalization to make the categorization. This finding shows that there is no need to invoke special mechanisms, evolved together with language, to explain this feature of speech perception”.

In fact, categorisation, despite Liberman’s efforts to use it for his own purposes, is more closely related to general perceptual abstraction. What occurs repeatedly in several perceptions is differentially reinforced. Of course, phonemic codes differ depending on the particular language. (To illustrate, let us take the example that appears in

---

[7] Let’s recall a point by Liberman et al. (1967) which is usually less quoted. In dichotic listening experiments (i.e., studies in which headphones were used to play one speech sound to the right ear and a different speech sound to the left ear), listeners showed an advantage for phonemic categorisation in speech played to the right ear, that is, to the ear that has stronger connections to the left hemisphere of the brain. Cf. Shankweiler & Studdert-Kennedy (1967).

Kuhl [2010, p. 717]: "Adult speakers of English and Japanese produce both English r- and l-like sounds, even though English speakers hear /r/ and /l/ as distinct and Japanese adults hear them as identical. Japanese infants are therefore exposed to both /r/ and /l/ sounds, even though they do not represent distinct categories in Japanese".) The change in meaning determines the maximal limits acceptable up to which and only up to which a sound may vary without changing its category.[8] Anyway, the perceptual abstraction process would work identically in all languages. The only variable element would be the exact profile of the arrangement from which the perceptual abstraction is later performed.

Let's go back to the nucleus of Liberman's theory. Is there anything in this theory that would still stand today? Before answering, we must pay attention to some of Piaget's proposals.

6.3.2 Piagetian premotor plan

Piaget (1959) argued convincingly that, in children, latent motor imitation must always precede the first copy of all new and complex 'motor patterns', among which he included articulatory-phonetic patterns. From his data, he concluded that children made two achievements virtually simultaneously. At a certain age, the child, during her first reproduction of a learned pattern, manages to move beyond trial and error (she performs it without "tâtonnements"), and, virtually within days, also successfully performs a delayed 'first reproduction'. These two simultaneous achievements can be explained – this is Piaget's hypothesis – by means of a pre-motor plan or latent imitation that the child would have carried out while observing the model.

Can Piagetian 'latent imitation' be extended to the learning of singing dialects by birds? We know the answer to this. Marler (1991) showed that young sparrows stored the adult singing in an auditory format during their mute and learning stage. To be precise, the animals stored the songs as an enrichment added to the innate auditory

---

[8] Fraser (2004, p. 274) has revisited and stressed this indisputably structuralist issue: "Any sublexical conceptualization depends upon prior conceptualization of words" (or, in other words, knowing the phonemes requires knowing the words). Certainly, we must remember 'statistical learning' that makes infants segment word-like units from ongoing speech – Saffran et al. (1996). Statistical learning is computational in nature, and reflects implicit rather than explicit learning. It is based on the notion that the units (phonemes, syllables, chunks of whatever size) that make up a word remain in fixed positions relative to each other whenever the word occurs. Thus, it relies on the ability to automatically pick up and learn from the statistical regularities that exist in the stream of sensory information we process. See also Scott et al. (2007) and Mehler & Dupoux (1994). However, as Kuhl (2010, p. 715) proposes, "social factors 'gate' computational learning". Social cues, such as eye gaze and pointing to an object of reference, would help infants segment word-like units from ongoing speech. Thus, Fraser's structuralist point would be not completely wiped off the map. As Cutler (2008) has said, speech perception per se must be investigated in connection with speech perception in the service of spoken word cognition.

singing pattern, that is, to the foundation to the songs of each individual in the species. How was it discovered that in sparrows' brains both the learned dialect and the simple foundation initially adopt an auditory format and not a motor format? Marler deprived some birds of hearing just when they were about to start singing, and thus proved that learner birds, in order to become singing adults (either with a dialect, or, if the birds had been reared in isolation, only with the simple foundation) must first hear their own singing attempts, detect to what extent these are similar to or dissimilar from the model and rectify accordingly.[9] This, incidentally, could be extended to the abilities of parrots. Marler's experiments reveal that dialectal learning has nothing to do with 'motor reception during the learning stage' in these birds. See also Gobes et al. (2010). It seems that the mirroring discovered by Prather et al. (2008) is confined to adult males who already know how to sing.

Among humans there is probably something relatively equivalent to this learning in sparrows. This (relative) equivalent might be the stage when children start to make their first sounds. Children make one attempt after another until in one of those attempts they hear a sound similar to the one they have heard.[10] However, once this stage of first sounds is over, the child succeeds in reproducing in at least a recognisable way a word the first time she attempts to reproduce that word. Thus she gives up Marlerian way.

Piagetian 'latent imitation' can definitely be relabelled as 'motor reception during the learning stage'. In other words, we have a convergence between Liberman and Piaget. This convergence is surprising given the enormous differences between these two authors: Piaget was never interested in adult speech reception and Liberman took only peripheral interest in child learning. However, both would suggest a motor format for the perception of learned motor patterns.

6.3.3 What happens once acquisition has come to an end? A proposal for a reformulation of the motor theory of speech reception

The learning stage would be the most reliable core for the motor perception of speech. The hypothesis of 'motor perception during acquisition' has some supportive evidence.

---

[9] All this clearly resembles the descriptions provided by connectionists: See a summary in Churchland (1988). Marler's research probably offers a proof that these descriptions match with a real brain process.

[10] But, in children, this is likely to be preceded by the process described by Vihman (2002, p. 310): "The infant practices canonical babbling and produces CV sequences at 6–8 months of age. This practice in production sensitizes the infant to similar input patterns, which are now easily recognized because they pop out of the acoustic stream". Later, after having acknowledged in the acoustic stream 'words similar to babbling' ('mamma', 'baba'..., which in all languages design one or other element in children's world), children would feel more motivated to pay attention to speech and to the similarities and differences of the patterns that they hear regarding the ones they already know from their babbling.

or fictional scene which is the object of attention. (I have written 'there is no need to specify'. But it would be preferable to say that the speaker does not make an effort to specify anything. This lack of effort sometimes leads to misunderstandings, and it is likely that it is such misunderstandings that have originally given rise to the motivation so that language would develop the resource of deixis.)

However, it would be impossible to avoid deixis in a message invoking a particular referential anchorage. Thus, the sentence "Julius Caesar crossed the Rubicon in 49 BC, or in the year 704 after Rome was founded", despite using proper nouns to refer to the subject, place and time (the calendar, which is a system for producing designations, pivots around a proper noun), cannot completely avoid deixis. The temporal morpheme of the verb 'crossed' clearly indicates that the speech act happened after the year mentioned. The day before he crossed the Rubicon, the verb would have been in the future, and at the very moment he was crossing the speaker would have opted for 'is crossing'. As can be seen, the two extreme types of speech (metalinguistic or implicit) that do not need deixis constitute a tiny portion of all linguistic messages. Deixis is, consequently, almost consubstantial with language.

6.4.2 The egocentrism of deixis, and deictics 'which cannot be repeated as an echo'

When a speaker says 'today' or 'yesterday', the hearer can repeat these words immediately afterwards without the term's specific anchorage being changed. The day they are talking is shared by the speaker and the hearer. However, not all deictics can be repeated as an echo with such impunity. Some deictics would acquire a completely different sense if the hearer immediately repeats them in a conversation. This, of course, is the case of deictics related to the pronouns "I" and "you". This effect can be observed not only with "mine" and "yours", but also, depending on the contexts, with "this" and "that", or "here" and "there", or with "in front" and "behind". These 'deictics that cannot be repeated as an echo', as we might call them, are what will occupy our attention now.

The meaning of all deictics depends on who is speaking, and on when and where. The zero point of the spatial and temporal co-ordinates that will determine the referential anchorage of a deictic is always the speaker acting as such. The egocentrism – the pivoting around the speaker's ego – peculiar to deictics is indisputable. However, thus far we have paid attention only to the producer. What about the reception of deictics? Or, in order to restrict the question to what we are interested in, what about the reception of deictics which cannot be repeated as an echo?

Obviously, the recipient cannot apply his egocentrism to the reception of the pronoun 'I'. (Later, in Chapter 21, we will reformulate this in a broader context, but for the moment this comment will suffice). However, since any linguistic term is consciously recognised as identical both when produced and received, there must be a resource where the identity of the "I" produced and heard can be reconciled to the completely different referential anchorage of the one and the other. What will this resource be?

Let us address first the resource involved in metalinguistic definitions. In order to define what the term "I" means, it is said that it refers to the speaker in question in each case. As can be seen, that definition can be used both by the speaker and by the hearer, and so our problem disappears. The question now is whether or not this definition is really used in mental processes. It is true that that neutral and aseptic definition, from which all trace of egocentric feeling has been erased, can perfectly account for the case of 'I'. However, this neutrality and asepsis will not work so well if we now consider other deictics belonging to the I-series. Imagine that a speaker says: "The thing (a cobra, for example) is right behind me". The two speakers in the conversation will probably not be shoulder to shoulder; they will, instead, be face to face, although not necessarily precisely geometrically opposite each other but at an angle somewhere between shoulder-to-shoulder alignment and an exact geometrical face to face. The metalinguistic definition of 'behind me' will be 'behind the speaker in question'. However, the definition makes it necessary to calculate what 'behind someone' means. The only way to perform that calculation is to perceive the 'face/back' body axis of the body of the speaker. Of course, in order to perceive that axis explicit instructions can be given. However, if we take into account that the ability to perceive this is a skill inherited from apes, the neutral linguistic definitions become more and more grotesque and unbelievable. Finally, we should remember that deictics such as 'right here', or 'there' are usually accompanied by gestures or glances. Thus the comprehension process involved in the reception of these deictics may be very similar to the comprehension process involved in the reception of pointing gestures.

We can now put forward our suggestion regarding the comprehension of deictics which cannot be repeated as an echo. This comprehension does not ignore egocentrism. Egocentrism is our inherited biological procedure for conceiving any distality (near/far; up/down; in front/behind):[13] consequently, any other procedure designed to replace it will not only be horribly complex but also, in the end, inefficient. However, in reception, the egocentrism involved in the meaning of 'I' is associated with a self that we have conceived inside our minds and that is different from our own self. A similar proposal to mine – although applied to the comprehension of scenes and, of course, without using my terminology of a second mental centre – can be found in Lozano, Hard & Tversky (2007): when observing an action, the observer tends to adopt the agent's perspective. In a similar manner, we can also say that when listening to a linguistic message, the hearer tends to adopt the speaker's perspective.

Once again, the key point is this: when the hearer receives a linguistic sign, he does not see it as a received sign, but as a produced sign, although definitely produced by another individual, i.e. as a not-own production. With this explanation of deictics

---

[13] More generally, this 'egocentric' distality is the key to animal perception and to animal self.

which cannot be repeated as an echo we have reached the second argument in favour of my explanation of Saussurean parity. But let me repeat what I have said above: the suggestion that this feature of the linguistic sign is possible only as a result of the perception of a radically not-own self would have to be seen within a more general framework in which, for the moment, pointing gestures (with the finger or with the eyes) and truly co-operative actions are included.

CHAPTER 7

About evocation

7.1 What is it we mean by “evocation”?

Linguistic signs have the ability to make us evoke objects not within reach of our perception at that particular moment. This is an ability they share with other kinds of symbols. This is extremely clear. What is not at all clear, on the other hand, is whether evocation is something animals can access.

Can the term ‘evocation’ be replaced by ‘displacement’? This latter term, at least in the sense given to it by Hockett (1960), or also Bickerton (2009), refers to a feature of communication: human language, just like bee-dancing, is an example of displaced communication. Given that the term refers to communication, it is clearly unsuitable to formulate two of the questions I am interested in -*i.* whether the goal pursued by an animal is evoked or not by that animal; *ii.* whether children are able or not to evoke by means unrelated to both language and communication.

But let us define more exactly what evocation is. Can we talk about evocation in dreams? Certainly, when a horse chases me in a nightmare, the horse is not actually present. In the midst of my nightmare, however, there is no way I believe that the horse is not real. But, during those moments, what has no reality whatsoever for me are, in fact, my bed and pillow (Cf. above, 1.4.2). This leads us to the following clarification: evocation is a question of having the image of an object while at the same time being aware that the object is not present.

Note that, by denying that the contents of dreams are evocations, we have blocked one route to answering the question of whether animals possess symbolic capability. Nowadays there is a method available to verify “from outside” if a sleeping person is dreaming. Rapid eye movements and so-called delta waves appear to offer reliable evidence. When a person exhibiting these patterns is woken up, he will say he has been dreaming. As a result, we have some grounds for ascribing dreams to animals (specifically mammals, with the exception, it seems, of the platypus) exhibiting these patterns. However, as I have said, all this cannot help us to answer our question.

On occasion, evocation has also been defined as voluntary, that is, produced in us at will, in contrast with involuntary perception.[1] Nevertheless, I have my objections to this ‘at will’. Remember that a hearer often inevitably evokes the meaning of a received

---

[1] For example, Myin & O’Regan (2009, p. 196): “A perceiver will have more a feeling of control, less a feeling of imposition, when he or she is thinking or remembering than when he or she is engaging in sensory interactions”.