

## Can Nonhuman Primates Read Minds ?

Joëlle Proust

► **To cite this version:**

Joëlle Proust. Can Nonhuman Primates Read Minds ?. Philosophical Topics, 1999, 27 (1), pp.203-232.  
ijn\_00139073

**HAL Id: ijn\_00139073**

**[https://jeannicod.ccsd.cnrs.fr/ijn\\_00139073](https://jeannicod.ccsd.cnrs.fr/ijn_00139073)**

Submitted on 29 Mar 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Can Nonhuman Primates Read Minds ?

Joëlle Proust  
(CREA, Ecole Polytechnique, Paris)

-----

Attributing mental concepts to non-linguistic animals poses well-known problems to ethologists and philosophers. It is all too easy to interpret a piece of animal social behavior (i.e. a behavior performed inside a group on the basis of information being displayed by behaviors from other members of the group) as involving representations of other individual's beliefs and motivational states as mental states. For this "intentional stance"<sup>1</sup> is indeed the most natural way for humans to explain and predict each other's behavior. Attributing mental states is a fast, easy and convenient way to predict behavior, even when it is known that the system under study cannot be granted intentionality - convenient in so far as it may allow instrumentally correct predictions, that sometimes are also strictly speaking false.

Intentionality refers to the property of some states (brain states, artefacts, or other material structures) to be about something else (like some repeatable event or property in the external world). The mental states that are typically intentional states are beliefs (that such and such is the case) and desires (that such and such should be the case). Now the capacity of forming beliefs and desires has to be distinguished from the ability to read mental states as such, i.e. to attribute mental contents to oneself or to other individuals. Having an intentional capacity does not any more secure a mindreading ability, than exerting a physical force provides the agent with a concept of cause. As one can exploit (extract and use) causal information without having the concept of cause, one can also exploit "mental" information without having the concept of mind nor of any mental state.

Ethologists and philosophers tend now to acknowledge that most animals are able to form representations, and to use them in controlling their own behaviors. Animals extract and store information, and retrieve it when necessary : although they are not endowed with an external language, their neural dynamics are partly dedicated to representing correlations in their environment. A more refined way of forming representations allows the animals to represent features in the external world that are observer-independent. This capacity to extract "distal information" is, according to some philosophers, distinctive of the disposition to form, and use, mental representations<sup>2</sup>.

Granted that a given species is able to entertain beliefs and desires, i.e. to have (epistemic and motivational) internal states with semantically evaluable contents, one can raise the question of whether the species under investigation is, in addition, able to represent properties and events that are not only perceptual or physical, but mental as well, and use the latter to guide their

actions, not only as reliable cues for achieving some output, but as mental cues; this assumption amounts to proposing that the animal indeed possesses some kind of a theory of mind. Since early work by Premack & Woodruff<sup>3</sup>, a theory of mind refers to the kind of knowledge that an organism has when it uses mental concepts such as those of belief and desire, and relies on the regularities that hold between those kinds of internal states, in order to understand and predict behaviour of conspecifics.

As David Premack notes<sup>4</sup>, there are two ways of understanding a piece of social behavior in non-linguistic animals, i.e. two types of reasons that can cause the actions performed by an individual in a social context. In one case, the animal can "act so as to affect what the other individual does ; in the other case, so as to affect what the other individual believes". In the second case only is it legitimate to take the relevant behavior as evidence for the animal having a "theory of mind".

Establishing this contrast may indeed prove difficult. For in the first case (aiming at affecting what the other individual does), often the only effective strategies - learned through their consequences or selected for through evolution - will be what is in fact deception or mimicry, concepts that may be well beyond the agent's conceptual repertoire. And in the second case (aiming at affecting what the other individual believes), the animal also aims at affecting ultimately what the other individual does. The difference between the two cases is not in their respective final goals, but only in the means used to achieve them. If a male chimpanzee is indeed able to refrain from vocalizing while copulating with a female when male conspecifics are in the vicinity, it may be because he believes that other males would learn about the presence of a receptive female if he produced his sexual cry, and would therefore want to approach her in turn. As he wants to keep her for himself, he thus prefers to refrain from vocalizing at the cost of reducing his present pleasure. But the same behavior could also happen simply because the copulating male knows from previous experience that suppressing his cry will have the favorable effect of his having exclusive access to the receptive female. No use of belief-desire attribution in this "kill-joy" story<sup>5</sup>. Again, there is no difference in the two types of behaviors as to their purposes. In all behaviors displaying social cognition, whether achieved by perceptual learning or by using a theory of mind, the purpose is fundamentally the same : manipulating others, or extracting information from the others' behaviors, in order to promote the capacities of the individual organism to fulfill his own basic needs. In social animals, one has to learn to manipulate others into acting according to one's own interest (safety, food sharing, mating, etc.); these actions may well be executed without being rationalized by "mental concepts" (to be analysed in a moment).<sup>6</sup>

There is a second source of indeterminacy. A "mentalizing" animal may well use mental concepts in an entirely non-reflexive way, that is, use his theory of mind in a tacit manner. Having a theory of mind no more requires knowing that one does, than competence in language production requires knowing which rules one follows, neither, indeed, knowing at all that one follows rules. One therefore cannot expect that an animal having a theory of mind in the sense that it uses some mental rules to infer what other individuals will do next, should also be able to categorize mental attitudes in himself in any principled way, or recognize that another individual is doing just what himself does, i.e.

attribute mental states. If one rejects an anthropocentric definition of mental attribution, the question of the kind of condition that is necessary and sufficient for attributing mental concepts to an animal becomes prominent. The criterion should not involve too much (reflexivity), while still including enough of the features that are linked to conceptual ability.

Finally, still a third problem emerges when one discusses non linguistic animal mentalization : not only should we distinguish levels or orders of theory of mind application ; we should also carefully establish which kind of theory of mind it is correct and adequate to attribute to an individual of a given age with a given background in a given species. Human developmental psychology suggests that, while infants may initially only understand the contrast between physical objects and self-propelled, goal-directed entities, children first acquire a primitive theory including only elementary epistemic and motivational states, like seeing, expecting and desiring, with no sense of the possibility of false belief ; the full complex of belief-desire folk psychology appears only later , and with it the idea that another may have a set of beliefs that is different from one's own on the same object<sup>7</sup>. It is then perfectly clear that human toddlers will not display the same mental capacities as older children and adults. Great apes could also have varying access to social and mental cognition according to their age and past experience. Another related problem has to do with the differences introduced by early exposure to human ways and in particular, to experimental training. It is assumed by many researchers that specific conditions -- such as a high level of early social interaction, and some kind of reinforcement of the intentional stance -- might constrain theory of mind acquisition, and that animals either too young, or raised in the wild, or isolated from their group, will fail to display mindreading abilities while older apes, hand-reared or laboratory animals might more often present this set of capacities.<sup>8</sup>

Given that mental attribution is an intricate kind of competence, with a range and a degree of complexity that are a priori indeterminate, one thing is clear : anecdotes are an inadequate way of establishing mental concepts use in non-linguistic animals. The proneness of humans to make sense in the richest mentalistic terms of just any kind of goal-directed behavior introduces a serious methodological flaw in efforts to capture the meaning of an animal's behavior by describing field observations -however carefully collected- in intuitive terms. This problem has been recognized and widely discussed by ethologists, experimental psychologists and philosophers.<sup>9</sup> "No single anecdote, Heyes writes, could provide convincing evidence that an animal's behaviour was influenced, not by a social interactant's appearance or behaviour, but by a mental state attribution based on those stimuli"<sup>10</sup>. Experimental research, although able to offer tighter control for antecedent learning, still often fails to eliminate the possibility that an animal performs behavioural prediction on the basis of observational cues or learned sequences rather than on the basis of mental attribution. In her influential paper, Heyes exposes the comparative merits of three different types of experimental paradigms testing theory of mind capacities.

The discrimination training method, used by Premack & Woodruff, aims at testing the mental capacities of chimpanzees in situations that seem to require recognition of mental properties. But these problems, immediately represented

by mental concepts in humans (e.g., "he believes what I tell him; I will let him believe something wrong") can also receive solutions exploiting exclusively contingencies between behaviours and outcomes. If the animal is able to show the wrong, non baited, container to the competitive trainer in order to deny him access to food, it may be because the animal extracted a regularity roughly like : "show the baited container only to cooperative trainer". According to Heyes, this type of experimental design is not adapted to testing mentalizing capacities in non-linguistic animals. It does not fare better than anecdotes that admit a mental interpretation (a female baboon grooms a male who has caught an antelope until he lolls back away from his food because she wants to distract him<sup>11</sup>), as well as a cue-learning one (the female steals the food at the first opportunity).

Another method discussed by Heyes is the "trapping" method <sup>12</sup>, i.e. an experimental test that does not require any previous training. This is exemplified by Cheney & Seyfarth's work on Japanese Macaques<sup>13</sup>, in which females were tested for the relative frequency of their food or predator calls when they could see, either in presence of their offspring or in their absence, that food or predator were hidden in an outdoor arena where the offspring (but not his mother) was subsequently allowed to enter. Although the frequency of calls was unchanged in the opposite conditions, Heyes shows that this negative result does not support any particular conclusion, given that a positive result could have been interpreted in terms of associative learning and not in terms of mental state attribution (the paradigm confounds thus between recent presence and knowledge attribution, and recent absence and ignorance attribution).

Now a third method, called "triangulation", consists in two phases : discrimination training is first developed, then transfer is tested on new items. Although this method can be misapplied in various, subtle ways, Heyes recognizes the ability of triangulation to distinguish observable cues from mental cues : the trainers that respectively know or guess where the food is located cannot, at least *prima facie*, be identified by the animals on the basis of observation ; individual trainers change roles (knower, guesser) -- none of them having done the baiting -- and also do not have the same spatial characteristics (presence or absence from scene) across trials. Since the animals are able, from one situation to the next, to transfer the discrimination between the trainer who can reliably indicate food and the one who cannot, it seems justified to conclude, with Povinelli et al., that chimpanzees can "model the visual perspectives of others"<sup>14</sup>.

Unfortunately, later work by Povinelli and his colleagues on knowledge about seeing in chimpanzees led to a different conclusion. They show that young chimpanzees do not perform above chance in discriminating two trainers, one able to see them, the other with his sight obstructed, in unfamiliar (not previously learned) situations (for example, one trainer had a bucket on his head, the other over his shoulder, or both trainers facing away from the animal, only one of them looking over his shoulder). Their experimental findings indicate that learning theory explains why chimpanzees have a preference for certain attitudes in communication, but it does not confirm that there is an intermediate variable that could play a role in their performance, having to do with the knowledge of what seeing entails. The authors establish that the stimulus that is actually used by the animals is the face, and not the

eyes, although the chimpanzees are also independently able to use the gaze direction of other individuals to extract the same visual information as the observed animal. In other words, chimpanzees understand seeing-as-attention, which involves only a motor routine (turning one's own attention to wherever the other looks at), and not "seeing as a knowledge acquisition device", which involves a primitive theory of mind<sup>15</sup>.

This particular example of a remarkable set of experiments shows clearly that the first kind of indeterminacy (does the animal use observational cues or mental concepts ?) is not the only difficulty that besets progress in theory of mind research. A pervasive difficulty in cognitive ethology is also that there is not any clear, well-developed conceptual framework for mental theorizing, universally accepted and well documented, that primatologists can use. Indeed Povinelli, Whiten, as well as other researchers in the field are confronted with general issues such as the modular or holistic character of mental concepts acquisition, or the role of metarepresentation in theory of mind competence, all questions that remain to be settled in developmental psychology; progress in primatology will follow from progress in child studies and in understanding human disorders of mentalization such as autism. Thus even though the selection of adequate tests for the existence of representations of mental states in non-linguistic animals is obviously an important condition for making progress, the main difficulty has not been solved : what exactly are the conditions in which one might say that an animal is able to read minds ? What are the kinds of competence involved in applying a theory of mind ?

Our strategy will be to first characterize abstractly mental concepts use, then to consider several types of competence present in nonhuman primates that might seem to qualify as mental-concept-using. Our aim in discussing potential mentalizing competence in the great apes will be to determine whether these animals exploit a particular kind of theory of mind, however minimal, in their interactions with conspecifics, and if it is the case, to find out which it is.

### Mental concepts use

Before going any further, we need to have some notion of what a mental concept is, in terms general enough to apply to animals that do not have any words to shape, still less to express their mental attributions. Some philosophers, among whom Donald Davidson<sup>16</sup>, have contended that non-linguistic animals simply lack the ability to apply concepts. I will not, in the present paper, try to substantiate why I take the question of concept application in animals to be not only meaningful, but also manageable<sup>17</sup>. It is safe to assume, given what we presently know on animal minds, that many animals categorize objects or events on the basis of non-purely perceptual features, for example in virtue of functional regularities. This helps the animal to memorize past experiences and make his action more flexible in novel contexts. Having a concept of X presupposes that the following conditions are fulfilled.

a) When an organism possesses concept X, it is disposed to decide, in most cases, whether something is or is not an X and act accordingly (this is close to

what is called, in classical philosophy, the "principle of complete determination"; it may fail to apply in certain cases, notably when concepts are "vague").

b) Having a concept (on the basis of various past experiences or due to some innate disposition) allows the organism to apply it in the same way to new cases, - what is called "generalizing".

Conditions a and b lead to the idea that conceptual thought is structured by what Gareth Evans called the "Generality Constraint".<sup>18</sup> In other words, a concept may always be used non-locally, i.e. systematically and in conjunction with all the other available concepts.

c) Being part of an inferential structure, concepts tend to cluster in theories and become sensitive to counterevidence; they may be modified by learning, in particular when a theory replaces another theory, i.e. when a more adequate inferential structure overcomes a less adequate one. The fact that concepts have to be organized in theories is just a consequence of their inferential character. The types of theories that encompass concepts can be quite simple, like the theories that young human children entertain about living versus dead or inanimate entities<sup>19</sup>. When one thinks about the kinds of theories that nonhuman social animals may have about social life, they include probably concepts of dominants, subordinated, children, adults, foes and allies, with all the associative and inferential links between these categories in terms of what kind of thing normally occurs when two individuals are in a certain situation (food sharing, presence of predator, mating, needing help etc.).

If this characterization of concept use is correct, mental concept use in a nonhuman animal will be distinguished by the fact that characterization of social events and objects, and the selection of adequate courses of action will be made on the basis of some mental inferential principles rather than on the basis of perceptual cues alone, or in virtue of other types of concepts -- for example non-mental social concepts -- that could apply to the same facts. In other words, inference rules will apply theory-laden mental regularities instead of associating perceptual, spatial, or other kinds of non-mental features. By a mental regularity, is understood some rationalizing connection between the mental contents of motivational and/or epistemic states.

Still there are many different ways of rationalizing others' behaviours. Let us suggest some of the basic competences that could qualify as "mental" :

- 1- Ability to recognize emotions in conspecifics and to respond adequately to them.
- 2 - Ability to understand motivational states and the objects that tend to cause them, and use these relations to predict behaviours.
- 3 - Ability to understand perception as a source of knowledge in attributing epistemic states to others.
- 4 - Ability to understand means-end structure in others' behaviours: recognizing that others have specific intentions that explain their current sequence of action.
- 5 - Ability to pretend that reality is different from what it is, and to understand perception of others from others' points of view.
- 6 - Ability to recognize genuine intentional actions of others from accidental doings.
- 7- Ability to attribute to others beliefs that are shared by self.

8- Ability to attribute to others beliefs that are not shared by self.

Many researchers consider that the capacity quoted last on the list is constitutive for having a theory of mind. Evidence collected so far seems to indicate that no non-linguistic animal, and no human child less than 3-and-a-half year old, are able to pass successfully the false belief task<sup>20</sup>. One of the reasons advanced is that only at the age of 3-and-a-half, is the full-fledged metarepresentational capacity put to work in the child's mind (although it may start developing earlier in capacities 5, 6 and 7). This emphasis on metarepresentational capacity is linked to the view that mindreading is the output of a specialized module having to do with higher forms of social cognition<sup>21</sup>. Other workers in the field try to make plausible the view that mindreading appears much earlier<sup>22</sup>.

Now that we have some explicit definition for mental concepts, we face the following problem : what kind of theory about internal states will count as mental ? Assuming that great apes, for example, are indeed able to store regularities and make inferences on the basis of prior learning, what kind of difference can help us establish that a prediction about behavior is done on the basis of a mental theory, and not on the basis of some other kind of theory ?

Joint attention and the idea of a precursor of the theory of mind.

Researchers in the theory of mind have noted that joint attention might be a precursor of mentalization, and might even offer a first case of mental concept application<sup>23</sup>. Joint attention is a three term relation, between two organisms and an object or event, the first directing its gaze to some definite portion of space, the second following the gaze of the former to the relevant location. In most animal species, a fixed stare at oneself is routinely interpreted as possible predator hazard. Such an interpretation does not have to rely on any mental concept. Gaze aversion is among primates the normal response to another's stare. Gaze nevertheless carries information of a different sort, namely it indicates which portion of the environment another selects as potentially relevant in the present context. Again, mental concepts do not have to be used for an animal succeeding at extracting the information that is of interest to him, i.e. not on what the other finds relevant, but on what is relevant. Still great apes could after all take advantage of relevance of gaze orientation in a mental way. Juan Carlos Gomez suggests that gaze could also be used by apes (but not by monkeys) in ostensive communication, "as an expression of their intention to communicate and to address other individuals".<sup>24</sup> In another work, Gomez argues that [hand-reared] gorillas can implicitly understand that requesting involves monitoring the attention of the person whose help is requested<sup>25</sup>. Now if joint attention in human infants plays an important role as a precursor of theory of mind, this is a very interesting claim. Would great apes, at least in certain circumstances, be able to understand "seeing as knowing" ?

Communication in humans and nonhumans.

Let us pause here a moment to comment on how communication is to be understood both in the context of animal coordination and in the present discussion of joint attention as a mental concept. Such considerations, as will appear later, are essential for understanding properly the contrast between mental and social explanations of behaviour. In the Gricean paradigm, communicating depends on the presence of several intentions : the speaker S intends

- a) that by uttering  $x$  S will produce a response  $r$  in an audience A;
- b) that A recognizes this first intention.
- c) that the fact that A recognizes this first intention is in part a reason for A's response  $r$ .

Condition c, which stresses the reflexive character of the communicative intention, (the understanding of its own recognition being presented as crucial for achieving adequate communication) seems to discourage any attempt at attributing communicative intentions to "non-sophisticated intentional systems"<sup>26</sup>, i.e. to organisms that cannot metarepresent others' mental states - i.e. represent them as mental states [intentions] whose satisfaction depend on their being adequately represented by the receiver.

Sperber & Wilson suggest some useful clarifications in Grice's analysis, that effect a welcome weakening of the concept of communication. First, the "overtness" of the intention to communicate can be replaced by a less demanding and more precise condition of "being mutually manifest"-- implying that communicating individuals share a cognitive environment. Second, Sperber & Wilson suggest that there is a continuum of cases between showing something and making an utterance, all of them belonging to ostension. Ostension consists in two layers : "first, there is the information which has been, so to speak, pointed out; second, there is the information that the first layer of information has been intentionally pointed out"<sup>27</sup>. In highly social animals as humans are, as Sperber & Wilson rightly stress, one needs to know the intention that governs ostension to understand what is relevant in it, i.e. to carry out the inferences that will allow grasping the intended informational content. Who the sender is becomes highly relevant, as well as the particular characteristics of the message token.

Still, sometimes it may be enough to pick up the first layer. This applies in particular to typically rigid kinds of coded communication. In non verbal animals, the work often simply consists in making something known to a receiver, who simply uses the information without further processing. It may of course also be essential for each individual to be able, not only to pick up information delivered by other members of the group, either on external or social events (predator presence, social hierarchy in the group, availability for mating, etc.), but also to detect the intention of the signaller when signalling. Such an intention will not be identified by the signallee as an internal state of the signaller, not even as a disposition to act or as a particular behavior, but as a motivating stimulus, triggering a sexual approach behaviour, for example. A number of species have evolved specialized codes to convey a specific information. As a consequence of evolutionary "arm-race" described by Dawkins and Krebs<sup>28</sup>, other species have evolved a similar code to take advantage of the hard-wired responses to the primary message. In those kinds of cases, the individuals communicate on the basis of nonflexible stimulus-response associations. The sender has some kind of expectation for a certain

kind of response, in the sense that its own behavior depends upon succeeding at conveying some relevant information. The arms-race puts new evolutionary pressure for more fine-grained intention-detectors, which leads to mentalizing animals.

Communication can thus be realized between animals who do not have the concept of an intention, but who do have the capacity to respond in a goal-oriented way to imperative signals. Although they do not communicate in any "mental" way, i.e. by relying on the identification of the communicative intention in the strong sense of condition c, they succeed at "transferring information to an appropriate audience", in the terms of Philips and Austad<sup>29</sup>. Coded signals play a prominent role in animal communication, yet animals from many different species can also invent new means of conveying information - using behaviours or features of the environment to make others receptive to their motivations (a domestic dog may bring a ball to indicate play mood, or go to the door to indicate desire to get out). Thus they can use not only (hard-wired) signals, but also what Philips and Austad name "signs". Let us then come back to ostensive communication in great apes.

#### Ostensive communication and mental concepts

It has been suggested that "in the great apes there is evidence that eye contact has evolved into an ostensive behavior associated with the expression and assessment of communicative intent"<sup>30</sup>. As was indicated above, an ostensive behavior includes two "layers" : showing some piece of information and showing that it is shown intentionally. To secure the second step, the sender has to help the recipient to make the correct inference. One of the most usual ways of achieving this consists in addressing the recipient, for instance exposing one's own gaze in a particular way (for example looking successively at the recipient's eyes and at the object and back again at the recipient's eyes). The idea is, in Gomez' words, to call "someone's attention not just upon oneself, but upon one's own attention"<sup>31</sup>. Meeting the gaze of the target recipient might be understood as meaning "wait for a message coming from me to you". It is a preliminary step to mark all the coming information as a piece of intentional communication. (In humans, it may also have many other inferred meanings depending on context). One question has to be raised to understand the intentional range of this addressing behavior. Is the kind of mechanism that allows to understand a gaze as an addressing behavior coded or inferential ? In the first case, the addressee uses a conventional signal for attracting attention. In the second case, the addressee uses contextual cues to assess what is the information conveyed : he notices that someone is looking at his own eyes, understands that the addresser is not watching him gratuitously, and concludes that he is expected to watch upcoming, addresser-related, events. The difference is crucial for assessing the communicative intention that causes the behavior. If the direction of gaze works as a coded signal, it can easily be picked up without any deeper understanding of the relevance of someone's gaze. By contrast, if it is inferred by using relevance criteria, then the animal should understand that looking implies acquiring perceptual knowledge.

According to Gomez, great apes, in contrast with monkeys, seem able to pick up someone's attention by using an "addressing" behavior through eye-contact<sup>32</sup>. De Waal reports that apes look at each other's eyes while reconciling<sup>33</sup>. Hand-reared gorillas have a repertoire of attention-getting gestures that they use with their human caretakers when they want to express requests. Observations by Menzel on chimpanzees suggest that they look at conspecifics's faces before showing them the location of a hidden object<sup>34</sup>. Many other similar observations were made on hand-reared bonobos and orangutans. In all observed cases though, with possibly the puzzling exception of Menzel's observations, the apes use addressing and redirecting of the addressee's attention exclusively in the context of requests, whereas they fail to issue protodeclaratives as do all human infants from about one 9 months on. The basic difference between a protoimperative (a request) and a protodeclarative is that, while the first is an act soliciting the other's help for changing the world (e.g., getting a desired object), the second attempts to monitor the attention of an addressee to some piece of information : it is not attempting to change an observed condition O, but only the mental attitude of somebody else in having him perceive O.

How are we to model such a request-related ostensive communication behaviour ? It could be seen as involving a metarepresentational capacity, insofar as it includes a representation of the fact that something is going to be used intentionally as a representation. Even discarding the infinity of successive reflexive levels as involved in Grice's condition of mutual knowledge in communication, and accepting instead Sperber & Wilson's weaker notion of mutual manifestedness, still this presupposes that both the sender and the receiver are able to represent both that F and that F is being represented intentionally. So maybe do we have here the distinctive property of mindreading organisms : being able, generally, to construct representations on representations, allows interpreting mental states in others -- i.e. storing and retrieving informational structures about first-order epistemic and motivational contents in others or in oneself. The required generality in applying mental concepts such as desire or belief would thus be explained by a metarepresentational competence that allows to categorize representations according to their mental functions. The capacity to categorize mentally would only be made possible by a capacity to represent metacognitive properties. Now this hypothesis does not seem consistent with all the comparative work done on chimpanzees' and human children's mentalizing performances. Whereas a metarepresentational task is achieved at 3 and a half in humans, it is not unambiguously and generally passed by adult nonhuman primates.<sup>35</sup>

First-order mental representations ?

Gomez and his colleagues have an interesting suggestion that would allow to sidestep the metarepresentational obstacle. They make two claims. First, non-linguistic animals or human infants, in preverbal request behaviors, have a way to understand attention that is "based upon the first-order representation of the external manifestations of other people's attentional processes"<sup>36</sup>. Attention being correlated with observational cues, an animal might represent

the mental state of attention in a purely implicit way, by using the causal link connecting eye-contact or direction of gaze with subsequent behaviour. He could know how to monitor attention in the addressee in order to secure the latter's further cooperation without having yet any metarepresentational ability. Once attention contact is established and used in a practical way, metarepresentations could be later constructed, and enrich the understanding of the cognitive implications of attentional states. 37 Therefore, in this view, protoimperatives (as well as protodeclaratives in infants) are first expressed in a first-order representational format. Instead of forming the output of a modular metarepresentational capacity, as in Leslie's view, they are assumed to later give rise to metarepresentational constructions, which in turn allow a theory of mind to develop.

Second, the authors suggest that what distinguishes protoimperatives from protodeclaratives is not so much a difference in representational structure, as a difference in motivational properties. Protodeclaratives have as their goals "the external manifestations of attention and emotion", whereas protoimperatives "use those manifestations as means to regulate a variety of behaviours in other people"<sup>38</sup>. To issue a protodeclarative, one has to be interested in people, not in things. Or, more exactly, one must find pleasure at extracting from others, at will, diversified, object- and event- related emotions. Nonhuman primates in fact are not interested "intrinsically" in their conspecifics' emotions.

The authors conclude that anthropoid primates do have access to mental information of the attentional kind, use it in their behaviour, but simply do not care about their conspecifics' emotions for their own sake. In analogy with autistic children, who do not engage in protodeclaratives for lack of sensitivity to expressions of emotion<sup>39</sup>, great apes do enjoy some kind of "primitive" mental understanding of conspecifics, while failing to develop a metarepresentational theory of mind.

The strategy of advancing these two claims against a strong modularist metarepresentational theory has the merit of stressing the role of empathy - transmission of emotional states with conspecific - in animal communication and social coordination. Developmental psychologists have insisted that the initial phase of communication in human infants crucially involves affective communication between the caregiver and the child.<sup>40</sup> Autism could be partly explained by a deficit in empathy in the autistic child, preventing him from taking advantage of ostensive behaviors for lack of interest in emotional stimuli.

Still such an explanation leaves unclear exactly how far "true requests" involve mental knowledge. Gomez and Tamarit contrast three levels of protoimperatives made through contact gestures, that can be exemplified in a number of animal behaviors towards their human caretakers. The animal can take someone as a physical mean to achieve some end, to be pushed or dragged; he can see him/her as an agent able to act by him/herself ; finally he can take someone as a subject : reasons to act are now included in the overall representation of the request situation. Looking at the eyes when soliciting is now understood by the gorilla as being 'as important as the gesture itself for the latter to become effective'<sup>41</sup>.

To see why this interpretation unduly inflates the role of a mental component in communication, it can be useful to recall that it is an essential property of a

concept that it be intensional; in other words, each particular concept expresses one particular way of accessing a reference, and it cannot be replaced by any other concept with the same extension. In Frege's terms, a concept is a "way of donation of reference". Granted that some internal representing state in the ape does refer to someone's attentional availability, it does not follow that this state instantiates a mental concept; there can be many ways of referring to a mental event, some mental, some observational. As we saw earlier, animals can take advantage of perceptual cues to identify motivational states (smells, for example, or various signals, like the "play face" in the ape for intention to play, or the bow in canids). Ostensive communication can qualify as a mental routine only if it is adequately connected with mental concepts, in which internal states are categorized with the help of mental marks. If ostensive communication is secured simply by using coded signals, the animals engaged in that communication do not need to represent any kind of knowledge as being possessed by another individual in virtue of its looking. The ape just knows that X looking to Y is a precondition for X helping Y.

It may be worth noting, at this point, that the role of empathy in mental concepts acquisition is far from clear. To empathize is to feel the same as someone else. There is in this case a distinct qualitative impression that is entertained by the subject. This impression could qualify automatically as mental if, by that term, was meant the fact that something appears under a particular quale or set of qualia<sup>42</sup>. However, if such a subjectivist view is taken on the nature of minds, one is lead to a dead-end when it comes to explaining the minds of nonhuman animals. There is no way we can ever feel as an ape or as a dog, and as a consequence no way to understand the ape's or the dog's mind. Moreover, the qualitative character of felt emotions in empathy might in this perspective warrant the attribution of a mind, but not the attribution of a mentalizing capacity to the organism.

If one adopts a functional instead of a subjectivist view on the mental, one may consider that empathy fails to qualify as involving mental concepts. True, it endows the individual with affective experiences that provide the perceptual inputs with the values that conspecifics associate to them. The animal could form motivational categories such as "pleasant", "dangerous", etc. Empathy seems to have the function of effecting social coordination, by automatically tuning various individuals to the same kind of responses. But empathy does not provide by itself the individual animal with an ability to apply to his emotional flow boundaries as to types of emotional qualia, that seem properly connected to other mental concepts. Empathy does not even provide the notion that some experience is being shared by two (or more) individuals. Although empathy is a rich source of various emotional states, it does not need to be brought into a conceptual format.

We may then conclude that, whether they are receptive or not to other's emotional states and intrinsically interested into them or not, apes may achieve social coordination in request without using mental-concepts. It is compatible with the data collected by Gomez that the gorilla picked up the human behaviour of attention-capture and attention-monitoring in social contexts in order to have its requests fulfilled, without understanding the causal efficacy of attention as a mental instrument for directing deliberate knowledge acquisition in other minds. There is no reason to consider that this acquired know-how implicitly taps on any mental knowledge. Even though theory of mind

acquisition depends upon the existence of some mechanism for joint-attention, it does not follow that joint-attention presupposes or involves in any way theory of mind capacities, even "implicit" ones.

Imagining how things could be

Another way of suggesting that great apes could have mentalizing capacities consists in showing that there is an intermediary phase between simply representing the world, and metarepresenting it. The crucial feature of this approach consists in realizing that mindreading presupposes an antecedent ability to represent non-present situations, i.e. past, future, or more generally counterfactual situations as such. The key to epistemic attribution in nonverbal animals would lie in imagining hypothetical situations. This idea is made plausible by work on theory of mind acquisition in human children.

Human children above 2 years of age, according to Josef Perner, are "situation theorists"<sup>43</sup>, insofar as they have trouble differentiating a picture and the situation that it represents; thus they cannot yet understand a picture, or any other representational item, "as a representation" (which is, in Perner's perspective, a necessary condition for having metarepresentations). But they still can grasp the representational content of a picture. Only later (around 3 and a half) will the children be able to form an adequate concept of representation, as a relation between a medium and a content. At this point the child can form the concept of belief, a concept that cannot be reduced to some hypothetical aspect that the world could turn out to have. Now Perner's hypothesis is that autistic children, although they fail in the false belief task, do succeed at other tasks involving counterfactual states of the world, such as the Zaitchik task<sup>44</sup>. Autistic children could after all be sophisticated situation theorists. They would not be totally deprived of mental concepts, but they would lack non-situational mental concepts.

Accordingly anthropoid primates, as well as other nonhuman vertebrates, could be taken to lack a concept of representation as representation as is the case for younger human children -- they actually have chance performance at the false belief task --, but still be "situation theorists", and therefore be able to handle some kinds of mental concepts - those that are possible-world oriented. Andrew Whiten compares three types of cognitive capacities that could involve this type of representational ability: imitation, pretense and mindreading<sup>45</sup>. According to Whiten, all three cases require that a second-order representation should be constructed on some first-order representation. In imitation, one should form a primary representation of the act performed by the model, and a secondary representation of the act in the imitating self. In pretense, a primary representation -- e.g. of a banana -- helps to construct a secondary representation -- of a banana used as a telephone. Finally in mindreading, there is the primary representation that F is- or is not - the case, and the secondary representation that F might not be believed to be - or not to be- the case.

Such a comparison of secondary representations use in the three cases is obviously theory-driven. Whiten's comparison favors a theory-theory of mind approach, according to which mindreading crucially involves internal, [domain-general or domain-specific] innate representational capacities. An

alternative theory of mindreading acquisition, named simulation theory, could offer an equally convincing parallel account of what those three cases have in common. Imitation consists in acting "as if I am you doing F". Pretending is "acting as if I was an X doing F". Mindreading is doing "as if I was in your situation, able to think what you think".

What is interesting to note is that, whatever the theoretical framework we select, success in these kinds of activities supposes two major subcapacities : the first is that the organism must be able to "go beyond representing things as they are", and "represent them as they might be", as Whiten correctly points out<sup>46</sup>. In other words, the organism should take situations as variable inputs to be dealt with, and not be confined to representing occurrent stimuli. This amounts to saying that the organism can use concepts referring to situation components and dynamics, and combine them at will to imagine other possible situations. The second is that the organism must maintain his "secondary" representation - or his simulated content - active in his working memory long enough to allow all the consequences to be drawn in the possible world under construction. While engaged in this kind of mental activity about possible worlds, it must resist solicitations from the environment. For example, while pretending that a banana is a telephone, he must not eat the banana. While voluntarily imitating someone<sup>47</sup>, he must refrain from turning his head to see an incoming individual. While mindreading he must keep in mind what the other's perspective on the facts is, even when it differs from his own. This capacity of resisting prepotent stimuli is not itself dependent on the ability to mindread, to imitate or to pretend. It certainly is their common condition. We shall later see that this may help understand why external word-language makes such a difference for mindreading in primates. Let us first explore what is the range in which nonhuman primates, as situation theorists, engage in imitation, pretense, and mindreading. If great apes are situation theorists, one might expect that they have some significant imaginative capacities, while failing to develop fully capacities to imitate, to pretend and to read minds.

## Imitation

Although primates are traditionally considered as prone to "ape", cognitive imitation is far from being present in all species. Levels of imitation should be carefully distinguished before evaluating primates' achievements. It is known that successful imitation of a course of action can result from stimulus enhancement or observational conditioning. Cognitive imitation that is of interest for mental concept attribution involves several potential capacities : noticing the results of goal-directed behaviour, understanding goal or intention of action, and understanding the means-end relation in a goal-oriented behavioural sequence.<sup>48</sup>

Nonhuman primates seem to have limited ways of imitating their conspecifics in various tool-use and food-preparation techniques. Some authors assume that great apes can imitate at the "program-level" of task organization and goal-structure<sup>49</sup>. In other words, the animals are supposed to copy the "logical structure" of the task, and to identify the various subgoals that have to be satisfied for the final outcome to be reached. It is widely recognized that they nevertheless fail to copy the sequence of motor performances that the others perform in order to complete the action. The important question for our

present inquiry, is to know what kind of representation is involved in non-human primate imitation. It seems that such "program-level imitation" can be achieved by representing intermediate subgoals, while leaving the means to reach them unspecified. Apes seem to engage in what has been called emulation learning about the objects<sup>50</sup>: while unable to reproduce the model's behaviour (impersonating the model), they try to reach the result that the model's action produced.

This kind of imitation-as-emulation appears to involve a capacity for representing a possible situation where the goal is satisfied. But it seems that a chimpanzee, for example, cannot represent and memorize the details of a means-end structure<sup>51</sup>. Maybe what is missing is constructing a hierarchy of motor images as a paradigm for one's own action. What the animal simulates is only the motivational structure and its final state: goal achievement. The apes can learn what the aims of the model are, but they cannot simulate another creature developing steps of action over time.

Imitation in apes suggests that apes might at least understand what someone tries to do. This seems consonant with experimental work done by Premack, in which a female chimpanzee, Sarah, is proven able to differentiate several kinds of problem-solving situations and even to distinguish relevant from irrelevant solutions<sup>52</sup>. Although imitation and ability to understand purpose are generally considered as precursors of a theory of mind in human children<sup>53</sup>, and not as constituents of such a theory, one still may insist that understanding purpose can give rise to a kind of economical, restricted theory of social behavior, in which an animal can understand that acts lead to results, and that conspecifics usually do things in order to secure certain ends.

Imitation in the anthropoid primates strongly suggests that understanding purpose is dissociable from understanding belief. Being able to grasp purpose, apes can try to reach the result achieved by someone else by performing an action with the same goal. They also can predict that given some motivational object, it will become a goal for another individual. This kind of capacity explains most behaviors often interpreted in richer epistemic terms: chimpanzees can hide their sexual arousal from other males as a subgoal in the larger sexual intercourse plan of action. "Not showing sex condition to males" might be an important social condition, independently of the belief that can be attributed to the watching males to rationalize that step.

Being unable to understand belief possession, the apes then understandably lack the motivation for learning epistemic conditions for succeeding at a task. They may represent various counterfactual situations, -- including goals -- but not counterfactual sets of beliefs. Therefore they cannot fully represent ignorance in themselves and in others, and not be tempted to teach, or want to be taught.

Evidence on ape imitation reviewed above seems at first blush compatible with two different hypotheses explaining the limits of imitation behaviors and their relation with theory of mind limitations. One might suppose that the animals do understand the means-end sequence in each particular step of a complex task, understand that to each subresult corresponds a particular adapted behavior that should be learnt, but fail to store the corresponding data and/or to retrieve them effectively. On the other hand, one might suggest that the animals fail to use the know-how of other animals because they fail to understand the contribution of knowledge to successful action. The first

hypothesis thus tends to blame limitations in imitation on insufficient memory and executive functions ; the second on defective mentalization. I will try to show below that in fact the two hypotheses can be combined to understand a deficit in mentalization and in imitation as dual effects of a same executive feature.

### Pretense and deception

Pretend play, also called symbolic play, is an activity that occurs spontaneously in normal human children. It can take several forms : substituting an object for another (telephoning with a banana as a receiver), attributing pretended properties to an object (taking a match box to be a car) or imagining non existant entities (filling a toy cup with imaginary milk). For Leslie, shared pretense - which appears around 18 months - has the same cognitive format as mindreading: it involves a general datastructure inside which an agent is attributed a propositional attitude - pretending - about a given real object [the content of a primary representation], to the effect that it has some imaginary property [content of a secondary representation]. The idea is that pretending involves decoupling two representations, which allows a metarepresentation of someone else's attitude to be constructed. In pretending, one can attribute truly non factually true mental contents ; it generates the kind of opacity that is characteristic of mindreading (what the other believes is not necessarily what is real).

Now Perner's account can also explain why a toddler of 18 months can pretend play : he can represent a counterfactual situation in which one makes a call by using bananas. In fact, the child does not need to attribute mental states to accomplish this. He only needs to vary the concept extensions in order to freely invent possible worlds. Coordinating imaginings with another individual does not involve representing mental content in the other. Doing the same thing or (in social play) complementary acts with a sense of common understanding of how things are in the imaginary world is enough for the play to go on in a satisfactory way. No metarepresentation of mental state contents seems necessary.

In nonhuman animals, one finds only few cases of pretend play (or symbolic play), the only uncontroversial cases being provided by hand-reared apes<sup>54</sup>. But there is another common achievement in primates that could be compared with pretend play for its cognitive requirements , i.e. tactic deception.

Just as is the case in pretend play, deception could be seen to involve decoupling a real situation and an imaginary one, and attributing the decoupled situation to the other individual (representing the imaginary situation as real). In contradistinction to pretend play, the deceiving animal does something in a context that does not justify what it does in the other's perspective (he does not communicate his context shift, as is the case in play). His context shift remains private, with results that are exclusively related so self-interest. For example Byrne describes the tactics of a young baboon who screams as if abused to get his mother to chase away a subordinate female, Mel, from a corm she had just painfully dug out<sup>55</sup>. Again, this can be understood in the following way : "Were it the case that I was screaming for abuse, my mother would chase Mel away. Mel's leaving the scene might help me get her corm. Scream for abuse !"

A situation theorist can model the same kind of case, without having to represent the intermediate causal link -- the mental content in his mother's mind, [that her child is abused by Mel] -- between his scream and his mother's stepping in. Decoupling can be explained in counterfactual terms, without having to attribute a complex metarepresentational structure to the nonhuman animal or to the human toddler.

### Simulation as exercised in great apes

To see why decoupling does not have to be metarepresentational, it may be useful to contrast two ways of simulating. Let us note first that simulation is in the humans a powerful way of achieving all the cognitive abilities that have been examined. Imitating is acting here and now as the other does, or did - simulating fully and "on-line" his actions, or his movements. Pretend playing is simulating in a sketchy way and "partly on-line" that you are doing something - - you can manipulate real objects, props or imaginary objects. Mindreading beliefs is pretending "off-line", in the sense that you just feed imaginary possible input states and just process what you would normally do, or feel, if these states were actually yours (56).

Now although Perner does not defend a simulation-theory of mindreading, one can use his view on situation theory to represent a very simple form of simulation that is clearly non-metarepresentational, and contrast this simple simulation with more elaborate forms that allow belief representation. Simulating in its core could consist in restricting the scope of a mental process or assertion to some particular, unusual context.

François Recanati 57 suggests the following representation :

$$(1) [s] \mid \stackrel{=}{=}_w \sigma$$

Proposition (1) expresses the fact that a subject represents a certain imaginary situation  $s$  in a world distinct from the actual world, such that some fact  $s$  holds in that world. According to Recanati, we engage in simulation each time we form a representation that concerns an imaginary situation. In other words, we treat in thought this imaginary situation as if it were real, and consider its various relevant consequences. Exercising simulation in this way is what allows great apes, and younger human children, to understand utterances on past, future, distant or fictional events like :

- (2) There would be bananas for lunch. I would eat plenty.
- (3) In those days there was a dragon. He loved to feed on people.
- (4) A long time ago, Granma was my age.

To achieve this cognitive understanding, a subject must be able to represent  $s$  and to hold in mind that property  $s$  is true in  $s$ .

According to Recanati, exercising simulation should not be confused with exploiting simulation. In the first case, one simply makes a supposition and considers some salient consequence of that supposition, as an output immediately delivered by the very process of simulating. In the second, one asserts inside the real world that making the supposition entails some particular proposition. The structure of that mental operation is truly

metarepresentational, insofar as the imaginary situation is now relocated within the real world (noted @), and its content made explicit relative to the representational process - pretending, believing, imagining - that made it possible. The quotation marks express the fact that the structure in quotation is mentioned (and not used anymore) :

(5)  $s' \mid \equiv @ \quad \ll s \mid \equiv_w \sigma \gg$

(5) thus expresses a metarepresentation, to the effect that in some part at some time in the real world, there is a representation of a situation in a supposed world  $w$  where  $s$  holds.

Applying this distinction to the case of mentalizing performances in nonhuman and human primates, it is quite different a matter to represent a counterfactual such as

(6)  $s_1$  screaming next to  $A \mid \equiv_w A$  leaving scene

(where  $A$  refers to an adult animal collecting some desired food)

and to represent a conditional assertion true in a type of situations :

(7) "if in the real world, the situation is such that a scream next to  $A$  is produced by me, it will be represented in my mother's world as the situation such that  $A$  abuses me, and as a result she will chase the abuser away".

$s' [ \text{child screaming next to } A ] \mid \equiv @ \ll s_1 [ \text{abusive adult} ] \mid \equiv_{mw} \text{mother to chase } A \text{ away} \gg$  58.

In (7),  $s'$  is the real situation, in which no abuse is presently taking place, but in which the animal nevertheless produces a scream in the vicinity of  $A$ ;  $s_1$  is the situation which is represented as being the one that the mother will take to be the case. In this counterfactual situation whose plausibility hinges only on the actual scream production, the mother should act in conformity with the thought that her child is abused, i.e. chase the [pretended] abuser away. In contrast with (6), in which no metarepresentation was necessary for the young to get rid of the adult, it is now in virtue of his capacity at exploiting the regular relationship between a particular counterfactual situation and the resulting motivated acts that the animal produces his call and reaches the intended result.

While the counterfactual simulation (6) is effected, so to speak, inside the imagined world, the metarepresentational simulation (7) is effected inside the real world where all the inferences from various epistemic contents can be made. Metarepresentation is thus a consequence of exploiting simulation in a deductive way. It offers a structure into which one can attribute a belief, a desire, and derive the various consequences of these representations according to their contents. If one engages in (7), one needs to make an assertion - or to construct a mental symbol - that mentions a represented situation, and to assert that, as represented, such a situation "supports such and such facts".

Following a number of ethologists, we could describe tactic deception, as other cases of pretending in the apes, in terms of (7) : an animal develops skills for exploiting simulation , and obtains on the basis of a complex metarepresentation of situations seen from the others' points of view some individual advantage. But such a rich interpretation of animal deception is not needed : given what is known on similar processes in imitation and pretense, it is not likely that the baboon exploits its simulation in the sense that it metarepresents, from the real world's perspective, how it could be beneficial to pretend that a certain situation obtains. It seems more plausible that the animal has a capacity for entertaining counterfactual thoughts about possible situations, i.e. of simulating that he is in such and such a situation. Entertaining counterfactual thoughts may be a necessary step for acquiring mental concepts, as is seen in the very presence of a "situation-shaped" thought like (1) within a metarepresentational thought like (5). But in itself, it does not allow the animal to represent mental states in others. It only allows him to extract some interesting regularities and to use them even in the absence of their triggering stimuli.

#### How executive functions could play a major role in mentalizing

As we saw earlier, there is ample evidence that hand-reared primates are much more advanced than wild animals in using mental concepts. They seem to be able to cope with deception, to pretend, to imitate and to understand tool use, not to mention ostensive communication gestures, whereas wild animals generally fail to present all these abilities. It might be the case, as it seems to be for self-recognition, that mental concepts emerge developmentally as well as phylogenetically. One possible reason of this difference could lie in language use, and in the strengthening of executive processes that such a language use indirectly allows.

To show how symbol use can have executive implications, it is important to contrast two types of behavior-control. Exogeneous control is what happens when an organism responds to some salient feature of the environment, either because it has learnt to do so, or because it applies an innate fixed action pattern. Endogeneous control, on the other hand, characterises behaviors that have been selected by the individual for contextual reasons. One of the consequences of symbol use could be to help an organism maintain endogeneous control against powerful exogeneous stimuli.

Recent experimental work on primates seems to show that language-trained animals could present better performances in tasks involving executive functions. Sarah Boysen has trained chimpanzees to learn numerosity in sets of objects and to learn corresponding numerical symbols<sup>59</sup>. The animals then are given a task in which their numerical knowledge is put to use. Among two plates X and Y with unequal numbers of sweets, a chimpanzee A has to point to the plate X that another animal B will get as a result of A's pointing to it; A will receive himself the remaining plate Y. Now two conditions are distinguished. In condition 1, the selecting animals see the plates with their contents. In condition 2, the animals see numbers indicating how many sweets the plates will contain. The interesting contrastive result is the following : in the first condition, the animals generally fail to refrain from pointing to the

plentiful plate. In spite of the growing rage and frustration at the unwelcome results of their own choices, they persist in their "prepotent" pointing. In the second condition however, the same animals are able to refrain from obeying stimulus-driven gestures : they do not point to what they want to have, but to what the other will get.

Although this second performance does not directly involve mental concepts, there are good reasons to insist that it has an important feature in common with belief attribution. For the chimpanzee who selects the plentiful dish fails to construct the situation as embedded within a larger rule, while the successful performer (choosing the symbol for the smallest number) is able to make his choice relative to his long-term goal as defined by that rule<sup>60</sup>. It is important that the same animal may, according to the nature of the stimulus - concrete or symbolic - fail or pass the test. This suggests that performance depends not only on intrinsic capacities for representing a rule, but also on the effects that the coding has on execution. If we come back to our previous distinction between exercising and exploiting simulation, we see that the animal remains in condition 1 a situation theorist, as represented by (1): there is something up for grabs, and he immediately activates the corresponding goal. He "forgets" the rule specifying that he will not get for himself what he first points to, a rule whose form is represented by (5). What Perner called a situation theorist might thus be someone who can apply several rules, but one at a time, in some exogenously controlled way. The problem in condition 1 would be not adequately described by saying that the ape cannot represent the embedding rule (for he does in condition 2); the problem is rather that he cannot act on the basis of the representation of the larger rule, because of the prepotence of the exogenous stimulus.

The same contrast between rules of different scopes is at work in belief states. Belief is essentially "perspectival": it is not dictated exclusively by the corresponding objective situation in the world, but by the world as apprehended by a believer. In order to attribute a belief to someone else, one has again to master two rules. The first rule ("one believes the world to be how it is") has to be embedded in a larger rule ("one believes the world to be how it is, except when one has misleading representations").

If this is true, one should expect to find the same kind of overlap in executive tasks and in mentalizing tasks in human children. It seems indeed to be the case. James Russell and his collaborators had devised earlier a task with the same structure as Sarah Boysen's condition 1. Called the Windows Task<sup>61</sup>, it consists in two boxes whose content can be perceived by the child through a window: one contains a chocolate, the other is empty. The child must point to the box whose content will be given to the opponent. As was the case for the chimpanzees, 3-year old children and autistic children cannot help pointing to the baited box, whereas 4-year olds point to the empty box. James Russell and coll. later explored the way to make the task easier by presenting a representation of the chocolate instead of the actual prepotent stimulus, and found the children were half as many to perseverate showing the baited box when they were shown brown circles instead of chocolate buttons. As the author notes, there is an empirical parallel, as well as a conceptual one, between success at the windows task and passing the false belief task<sup>62</sup>.

Now the reason why the chimpanzee can benefit from the language form of the stimulus may also explain why hand-reared animals may be better at strategic deception, and other mentalizing tasks. In condition 2 of Sarah Boysen's experiment, the animal seems to be able to regiment his pointing behavior inside a contextual arbitrary rule. He may at this point construct something like a metarepresentation of the whole situation in the sense of proposition (7) above. The numerical symbols that he previously learned may prompt, or make easier, a metarepresentational approach of the situation.

In a similar way, theory of mind might be both simulatory and subject to learning, more exactly learning how to use a symbolic representation to resist the impact of prepotent responses to objects and situations. Apes could have mental precursors similar to humans' (a non-mental gaze-direction detector, an attention sharing mechanism, a face detector, empathetic reactions), but fail to maintain active rules on embedded contexts for lack of a well-trained working memory responding to internalized symbolic rules.

## Conclusion

This article attempted to explore in what sense a theory of mind could be attributed to non-human primates. It was argued that present data do not allow attributing mental concepts to these anthropoids; although they may be said to use mental information in inferring behavior patterns, they do not do so by using mental representations. Hand-reared gorillas and chimpanzees, as well as other domestic animals, may be able to pick up human attention-capture and attention monitoring behaviors without grasping the causal efficacy of those behaviors as mental instruments for directing other individual's awareness. Ostensive communication, in particular, is shown to develop independently from any representation of mental states in others. Animal communication as such does not imply generally any deep understanding of intentions to communicate. It might nevertheless exert phylogenetic and ontogenetic pressures towards representing mental states, as an outcome of the "arms-race" that it opens up for individuals and species.

The main aim of this paper was to suggest that theory of mind acquisition in the apes depends on two prior capacities : exercising simulation, as when one actively disengages from the present environment to imagine a counterfactual situation, and exploiting simulation, which implies that an imaginary situation is relocated within the real world. Although the apes seem to be good "situation theorists", they might be poor "simulation exploiters". It was suggested that the latter ability could have an essentially executive dimension, in that it requires that working memory activates an abstract (superordinate) general rule in which another local (subordinate) rule is embedded. Symbol use could play a decisive role in helping an organism to keep endogeneous control on its actions, by releasing the vividness of prepotent stimuli at the subordinate level. Non-human primates have developed a number of abilities that are related to the needs of social life, and that superficially appear to be mental concepts-driven. They can recognize emotions and use them to predict behavior - as many other social animals do. They can interpret others' goals, which allows them to reach a certain amount of goal imitation, and interpret others'

intentions. They probably fail to attribute beliefs about the world to others, and still can achieve complex social organization. This suggests that social cognition is largely independent from mental cognition. Indeed social cognition may use mental information in a perceptual format.

### *Footnotes*

1 See Daniel C. Dennett, "Intentional Systems", *Journal of Philosophy*, 8, (1971): 87-106 ; reproduced in *Brainstorms*, (Montgomery: Bradford Books, 1978).

2 See Ruth Millikan, *Language, Thought and other biological categories*, New Foundations for Realism, (Cambridge: MIT Press, 1984) ; Fred Dretske, *Explaining Behavior, Reasons in a World of Causes*, (Cambridge: MIT Press, 1988) ; Joëlle Proust, *Comment l'esprit vient aux bêtes*, *Essai sur la représentation*, (Paris: Gallimard, 1997).

3 See D. Premack & G. Woodruff, "Does the chimpanzee have a theory of mind ?" *The Behavioral and Brain Sciences*, 4, (1978): 515-526.

4 In D. Premack, "'Does the chimpanzee have a theory of mind ?' revisited", in R. Byrne & A. Whiten (eds.), *Machiavellian Intelligence : Social expertise and the evolution of intellect in monkeys, apes and humans*, (Oxford: Clarendon Press, 1988): 160-179, p. 162.

5 This expression is used by Daniel Dennett in an analogous context. See his "Intentional systems in cognitive ethology: the 'Panglossian paradigm' defended", *The Behavioral and Brain Sciences*, (1983), 6: 343-350. Partly reproduced in "The intentional stance in theory and practice", in R. Byrne & A. Whiten (eds.), *Machiavellian Intelligence : Social expertise and the evolution of intellect in monkeys, apes and humans*, (Oxford: Clarendon Press, 1988), 180-202.

6 C. Heyes makes a similar point in "Anecdotes, Training, Trapping and triangulating : do Animals attribute mental states ?", *Animal Behaviour* 46 (1993): 177-88. See also A. Whiten, "Imitation, pretense, and mind-reading : secondary representation in comparative primatology and developmental psychology?" in A.R. Russon, K.A. Bard & S.T. Parker, (eds.): *Reaching into thought : the minds of the great apes*, (New York: Cambridge University Press, 1996), 300-324 .

7 See D. Premack, D. & A. J. Premack, "Intention as psychological cause", in D. Sperber, D. Premack & A. J. Premack, (eds.), *Causal Cognition, A multidisciplinary debate*, (Oxford: Clarendon Press, 1995): 185-199.

8 On hand-reared animals, see D. Premack, D. & A. J. Premack, *The mind of an ape*, (New York: Norton, 1982) ; J.C. Gomez & J. Tamarit, "The comparative study of early communication and theories of mind : ontogeny, phylogeny and pathology", in S. Baron-Cohen, H. Tager-Flusberg, & D.J. Cohen, (eds.), *Understanding other minds, Perspectives from autism*, (Oxford: Oxford University Press, 1993), 397-426; J.C. Call & M. Tomasello, "The effect of humans on the cognitive development of apes, in A.R. Russon, K.A. Bard, & S.T. Parker, (eds.): *Reaching into thought : the minds of the great apes*, (New York: Cambridge University Press, 1996): 372-403.

9 See for example D. Premack, "'Does the chimpanzee have a theory of mind ?' revisited"; D. Dennett, "Intentional systems in cognitive ethology: the 'Panglossian paradigm' defended", Heyes, op.cit., D. Povinelli, "Chimpanzee theory of mind ?: the long road to strong inference", in P. Carruthers & P.K.

Smith, (eds.), *Theories of theories of mind*, (Cambridge: Cambridge University Press, 1996): 293-329.

10 C. Heyes, "Anecdotes, Training, Trapping and triangulating : do Animals attribute mental states ?", 179.

11 Anecdote commented on in A. Whiten & R.W. Byrne, "Tactical deception in primates", *Behavior and Brain Sciences*, 11 (1988): 233-273. See also C. Heyes, "Social learning in animals, categories and mechanisms", *Biological review* 69 (1994): 207-31.

12 term borrowed from Dennett, *op.cit.*.

13 reported in Cheney & Seyfarth's *How monkeys see the world : inside the mind of another species*, (Chicago: University of Chicago Press, 1990): 220 sq..

14 See D.J. Povinelli, K.E. Nelson, S.T. Boysen, "Comprehension of role reversal in chimpanzees: evidence of empathy ?", *Animal Behaviour* 43, (1992): 633-640.

15 See D. J. Povinelli, "Chimpanzee theory of mind ?: the long road to strong inference", 321.

16 See in particular Donald Davidson's "Rational animals, *Dialectica* 36 (1982), 318-327.

17 On this problem, see C. Allen & M.C. Hauser, "Concept attribution in nonhuman animals: theoretical and methodological problems in ascribing complex mental processes", *Philosophy of Science* 58 (1991): 221-240; reproduced in M. Bekoff & D. Jamieson (eds.), *Readings in animal cognition*, (Cambridge: MIT Press, 1996): 47-62 .

18 See Gareth Evans , *The Varieties of Reference* (Oxford: Clarendon Press, 1982) : "An idea of an object, then, is something which makes it possible for a subject to think of an object in a series of indefinitely many thoughts, in each of which he will be thinking of the object in the same way" (104).

19 See Susan Carey's *Conceptual change in childhood*, (Cambridge: MIT Press, 1985).

20 This task consists in predicting behavior of someone on the basis of the misleading perceptual data that he/she collected from a scene. See H. Wimmer & J. Perner, "Beliefs about beliefs, Representation and constraining function of wrong beliefs in young children's understanding of deception", *Cognition*, 13 (1983): 103-128.

21 □ See A. M. Leslie, "Pretence and representation, The origins of 'theory of mind'", *Psychological Review*, 94 (1987): 412-426 ; S. Baron-Cohen, A. Leslie, U. Frith, "Does the autistic child have a theory of mind ? " *Cognition*, 21 (1985): 37-46.

22 A. Gopnik & H.M. Wellman, "Why the child's theory of mind really is a theory", in M. Davies & T. Stone, (eds.), *Folk Psychology*, (Oxford, Blackwell, 1995): 232-258.

23 W.H. Wellman, "Early understanding of mind : the normal case", in S. Baron-Cohen, H. Tager-Flusberg, & D.J. Cohen, (eds.), *op.cit.*: 10-39.

24 See J.C. Gomez, "Ostensive behavior in great apes: the role of eye-contact", in A.R. Russon, K.A. Bard, & S.T. Parker, (eds.), *op.cit.*, 137.

25 J.C. Gomez & J. Tamarit, "The comparative study of early communication and theories of mind : ontogeny, phylogeny and pathology", 409. From the existence of joint attention behaviour in this hand-reared gorilla, the authors conclude "Thus, a closer look at requests reveals that they are complex communicative behaviours that may involve a sophisticated understanding of people. (..) If this analysis is correct, a consequence would be that requests or

- protoimperatives would seem as good candidates to be theory-of-mind precursors as protodeclaratives."
- 26 See Dennett's hierarchy of intentional systems in "Intentional systems in cognitive ethology: the 'Panglossian paradigm' defended".
- 27 D. Sperber & D. Wilson, *Relevance*, (Cambridge: Harvard University Press, 1986), 50.
- 28 See R. Dawkins & J. R. Krebs, "Animal signals : information or manipulation ?", in J.R. Krebs & N.B. Davies (eds), *Behavioural Ecology*, 1st ed., (Oxford: Blackwell Scientific, 1978), 282-309.
- 29 See M. Philips & S.N. Austad, "Animal communication and social evolution", in M. Bekoff et D. Jamieson, (eds.), *op.cit.*, pp. 257-267.
- 30 See J.C. Gomez, "Ostensive behavior in great apes: the role of eye-contact", 131.
- 31 *Ibid.*, 133.
- 32 *Ibid.*, 143 sqq.
- 33 F. de Waal, *Chimpanzee Politics: Power and Sex among Apes* (London: Jonathan Cape, 1982).
- 34 E.W. Menzel, "A group of young chimpanzees in one-acre field", in M. Schrier & F. Stollnitz (eds.), *Behavior of Nonhuman Primates* vol. 5 (New York: Academic Press, 1974): 83-153.
- 35 See D. Premack, "'Does the chimpanzee have a theory of mind ?' revisited".
- 36 J.C. Gomez & J. Tamarit, "The comparative study of early communication and theories of mind: ontogeny, phylogeny and pathology", 410.
- 37 See J.C. Gomez, "Mutual awareness in primate communication: A Gricean approach", in S.T. Parker, R.W. Mitchell, & M.L. Boccia, (eds.), *Self-awareness in animals and humans* (Cambridge: Cambridge University Press, 1994), 73.
- 38 J.C. Gomez & J. Tamarit, *op.cit.*, 414.
- 39 See Peter Hobson, "Understanding persons : the role of affect", in S. Baron-Cohen, H. Tager-Flusberg, & D.J. Cohen, (eds.), *op.cit.*, 204-227.
- 40 See C. Trevarthen, C., *Communication and cooperation in early infancy : a description of primary subjectivity*, in M. Bullova ed., *Before Speech: the beginning of interpersonal communication*, (New York: Cambridge University Press, 1979).
- 41 J.C. Gomez & J. Tamarit, *op.cit.*, 409.
- 42 For a subjectivist approach to the mind, see T., Nagel, "What is it like to be a bat ?", *Philosophical Review* 83 (1974); reproduced in *Mortal Questions* (Cambridge: Cambridge University Press): 165-180 ; C. McGinn, C., *The Problem of Consciousness*, (Oxford: Blackwell, 1991).
- 43 See J. Perner, J., "The Theory of Mind deficit in autism : rethinking the metarepresentation theory", in S. Baron-Cohen, H. Tager-Flusberg, & D.J. Cohen (eds.): 112-137.
- 44 See D. Zaitchik, "When representations conflict with reality: The preschooler's problem with false belief and 'false' photographs", *Cognition* 35 (1990): 41-68. Zaitchik's idea is to compare the understanding of epistemically outdated information affecting behavior, characteristic of the false belief task, to the understanding of a physically outdated information affecting the content of a photo, in the false photograph task.
- 45 See A. Whiten, "Imitation, pretense, and mindreading : secondary representation in comparative primatology and developmental psychology?" in A.R. Russon, K.A. Bard & S.T. Parker, (eds.), *op. cit.*, 300-324.

46 Ibid., 314.

47 There are cases of imitation in which the individual cannot refrain from imitating : a newborn imitates a smiling face, human subjects with the imitation syndrom cannot help imitating the behaviours of people around them. All the present considerations on imitation apply only to voluntary imitation.

48 See A. Whiten, A. & R. Ham, "On the nature and evolution of imitation in the animal kingdom : reappraisal of a century of research", *Advances in the Study of Behavior*, 21 (1992): 239-283.

49 See R. Byrne, *The Thinking Ape, Evolutionary Origins of Intelligence*, (Oxford: Oxford University Press, 1995).

50 See M. Tomasello, M. Davis-Dasilva, L. Camak & K. Bard, "Observational learning of tool-use by young chimpanzees ", *Human Evolution* 2, (1987): 175-83; J.C. Call & M. Tomasello, "The effect of humans on the cognitive development of apes".

51 M. Tomasello, M. Davis-Dasilva, L. Camak & K. Bard, "Observational learning of tool-use by young chimpanzees ".

52 See "'Does the chimpanzee have a theory of mind ?' revisited", 176.

53 See S. Baron-Cohen's *Mindblindness, an Essay on Autism and Theory of Mind* (Cambridge: MIT Press, 1995) and S. Ozonoff's " Components of executive dysfunction in autism and other disorders". In J. Russell (Ed.) *Executive dysfunction in autism*. (Oxford: Oxford University Press, 1997).

54 J.C. Call & M. Tomasello, "The effect of humans on the cognitive development of apes", 391.

55 R. Byrne, *op.cit.*, 124.

56 On the simulation theory of mindreading, see A. I. Goldman, "Interpretation psychologized", *Mind and Language*, 4, 1989, 161-85, and "In defense of the Simulation Theory, *Mind and Language*, 7, 1992, 104-119.

57 See François Recanati, "The Iconicity of Metarepresentations", (*Rapport du CREA 9707, 1997*), 46 sqq..

58 To facilitate understanding, I don't represent tensed verb phrases in a formal way, although situation theory allows it. See J. Barwise & J. Perry, *Situations and Attitudes*, Cambridge, MIT Press, 1983.

59 See S.T.Boysen, "Rule-governed resource distribution in chimpanzees", in A.R. Russon, K.A. Bard & S.T. Parker, (eds.), *op. cit.*

60 James Russell first made a similar point concerning autistic children. See his "How executive disorders can bring about an inadequate 'theory of mind'", in J. Russell (ed.), *Autism as an Executive Disorder*, (Oxford: Oxford University Press, 1997), 256-302.

61 See J. Russell, N. Mauthner, S. Sharpe & T. Tidswell, "The Windows Task as a measure of strategic deception in preschoolers and autistic subjects", *British Journal of Developmental Psychology*, 9, (1991): 331-349; J. Russell, *Agency. Its role in mental development*, (Hove: Erlbaum/UK Taylor & Francis, 1996): 232 sqq.

62 See J. Russell, *Agency. Its role in mental development*, 233.