philosophie.ch
SWISS PORTAL FOR PHILOSOPHY

Swiss Philosophical Preprint Series

 # 93

Florian Leonhard Wüstholz

Mindreading in Chimpanzees

added 03/04/2012

ISSN 1662-937X

Master Dissertation

Florian Leonhard Wüstholz

MINDREADING IN CHIMPANZEES — THEORETICAL AND PRACTICAL ISSUES IN THE ANIMAL MINDREADING DEBATE

University of Sheffield

Department of Philosophy

August 2011

Contents

Abstract

The animal mindreading debate is coined with theoretical and practical problems which have lead some to believe that it has reached a stalemate. Since Premack and Woodruff have asked whether the chimpanzee has a theory of mind over 30 years ago, the question is still on the table. Even though a lot of evidence has been brought forward by both sides, proponents and opponents of the view that animals can be mindreaders, nothing is definitive. The essential problems remain unresolved.

The two main problems are the difficulty to experimentally support the mindreading hypothesis in nonverbal animals against competing lower-level explanations, like complementary behaviour-reading hypotheses, and the difficulty of explaining the psychological mechanisms which produce the mindreading in question. The first issue is typically discussed in terms of 'the logical problem' while the second is known as the discussion between Theory Theory and Simulation Theory of mindreading.

In this dissertation, I will discuss these issues and offer possible solutions of how to go about solving the problems in a way that will hopefully leverage the discussion. Most importantly, I will show that it is essential to understand the nature of mental content that is required for mindreading. Furthermore, I propose tests which should be able to detect the ability of nonverbal animals to understand different mental state concepts which are required for mindreading.

Acknowledgments

I would like to thank George Botterill for assisting me in writing this dissertation. I am very grateful for his great support and supervision, and the thought-provoking discussions and feedback.

Also, I would like to thank all the different friends and people which have supported me, one way or the other, in the whole process of thinking and writing. Your help is greatly appreciated.

MINDREADING IN CHIMPANZEES — THEORETICAL AND PRACTICAL ISSUES IN THE ANIMAL MINDREADING DEBATE

1 Introduction

Why should we be interested in the cognitive abilities of nonlinguistic animals like chimpanzees? What is philosophically interesting about investigating animal minds? There are many different answers to these questions, but let me briefly outline two possible answers and then go on to give my own reasons for believing that discussing animal mindreading is important.

Investigating the functioning of our mind is an essential task in modern cognitive science and philosophy of mind. A fundamental component of tackling this task requires us to find out the connections between different capabilities, like for example the connection between language and cognition. It is a common view that language is prerequisite to many forms of thinking, and these are thus thought to be exclusively human. Abstract thinking, having higher-order metarepresentational mental states, full-blown rationality, and the possession of a theory of mind are held to be dependent on the ability to *think in words*. But it is far from obvious why this should be true.

While it might be true that the listed cognitive abilities are made possible in human beings by *our* public language, it is questionable why there shouldn't be other possible ways of achieving those mental skills in other animals and humans alike.

A second reason is the evolutionary history of the human mind, which can give us insights on the function and interdependence of the different cognitive parts and skills that have been acquired over time. Looking at evolutionary close animals like the chimpanzee and getting hold of their cognitive abilities may allow us to draw conclusions about the evolution of our own mind. If chimpanzees have a theory of mind, then our own ability to read others' minds might be much older than if it was dependent on language, which may have evolved later.

The importance of social cognition for the thriving of human society and the necessity to further our understanding of mental concepts like perception, belief, emotion, desire, intention, consciousness, self-consciousness, and mindreading give further reasons to find out the way mindreading works in humans and non-human animals alike.

However, there are two further reasons I want to point out which make it imperative to give attention to animal cognition research. Culturally, human societies have evolved as a system that uses animals without questioning the moral validity of that practice. Among the reasons that have rationalized this procedure is the naïve view that non-human animals are substantially inferior to humans which allows us to use them to satisfy our desires and interests. We used to believe that animals do not have the ability to think, feel pain, or be rational, and these beliefs have been used to rationalize our actions towards them.

However, there is a growing suspicion that most of those prejudices towards non-human

animals are accurate and how that relates to the moral question of whether it is allowed to use animals to serve our interests. However, the moral importance of cognitive abilities is highly questionable, since more basic capabilities like the capacity for suffering and feeling pain are what matters morally. But on a very pragmatic level, acknowledging the cognitive similarities between humans and other animals challenges our practices directly. If humans are very like other animals in some significant respect, why can we harm the latter and ignore their interests, but not those of other human beings?

In that respect, thinking about the cognitive abilities of chimpanzees serves the project of undermining the rationale behind many of the practices that involve humans causing pain and suffering of other animals. If chimpanzees turn out to be cognitively much like young humans, how can we then morally justify the permission to cause pain to the former in experiments but not the latter? In the end, there are few reasons other than that chimpanzees are *not humans*, which rightfully is no moral reason at all (Cf. Francione 1995; Singer 1975; Regan 2004).

In the following I will investigate the supposed ability of chimpanzees to understand and attribute mental states like beliefs and desires to other individuals, like conspecifics and humans. Do chimpanzees know what others *see* or only what they are *looking at*? Are they genuinely having thoughts about others' mental states or are they just reading and interpreting their behaviour and do not engage in *genuine mindreading* activities? Do they have a concept of a belief and are they capable of higher-order mental representations?

In order to find out if chimpanzees are *mindreaders*, I will first introduce the essential terms, distinctions and problems about mindreading in general in the next section. We will

see that there are different hypotheses that are competing to explain the phenomena, which some believe are best explained by attributing mindreading abilities to chimpanzees. It will be shown that there is a specific *logical problem* that needs to be overcome to strengthen the mindreading hypotheses against other competitors. How can we distinguish animal mindreading abilities from their behaviour-reading abilities, if all we can observe is their behaviour? It is doubtful whether this problem has been solved yet, or is likely to be solved in the near future.

Thereafter, I will present in the third part two different views of how mindreading operates and is cognitively structured in the mind, simulation theory (ST) and theory theory (TT), and link these two different theories to the supposed mindreading abilities in chimpanzees. Furthermore, I will investigate the problem of fixing the *content of mindreading* activities. If chimpanzees can read others' minds, what is the mental (representational) content of that cognitive process? I will argue that genuine mindreading requires *metarepresentational* content at some point or other.

Finally, in the fourth part, I will discuss the empirical and theoretical evidence for chimpanzee mindreading abilities. The investigation will be separated into different sections dealing with different mental states, *seeing* and *knowing* on one hand, *beliefs* on the other. It will become evident that the more complex and elaborate the mental state in question is, the harder it gets to find evidence for the ability of chimpanzees and other nonverbal creatures to understand those states.

2 Mindreading, behaviour reading and the logical problem

Social cognition is an important aspect of human lives and also those of social animals like birds (Clayton et al. 2007; Kacelnik 1984), bees, apes (Hare & Wrangham 2002; Addessi & Visalberghi 2006; Visalberghi & Addessi 2003), canines (Holekamp & Engh 2002), and dolphins (Connor & Mann 2006). Generally, it is the ability to mentally process information about conspecifics and use this to navigate in the social space of that species or environment. However it is highly debatable how far the abilities of non-human animals reach and especially, if they are capable of attributing genuine mental states to other conspecifics (Cf. Hurley & Nudds 2006; Lurz 2011)[1]. The ability to understand and interpret the behaviour of others is of great importance to the interaction amongst human beings (Goldman 2006: 3) and it is equally crucial to the social lives of other animals. However, while there might be overwhelming evidence that humans are generally capable of attributing mental states to others (i.e. being capable of *mindreading*), there is a lively debate on whether or not other animals can understand others' mental states and possess mental state concepts.

There is no doubt that animals do react to the actions of others, and also that they react, at least sometimes, *because* of those actions[2]. However, closing your eyes upon seeing an incoming punch to the face is essentially unlike giving a friend a present to lighten his mood. While the former is an immediate bodily reaction to the behaviour of someone else,

[1]Here I will focus on the ability to understand mental states like seeing, knowing, desiring, believing, etc. There are of course other mental states, the understanding of which is important in social cognition, e.g. moods and feelings (Tanner & Byrne 1993; De Waal 1986). But for simplicity's sake, I will largely restrict the discussion to the former category. There is of course a general worry here that those states are too sophisticated for non-human animals, and that we should instead focus on other, more simple, states, but that would be a hasty judgement of the cognitive abilities of non-human animals.

[2]I will use the term '*action*' in a loose way that includes both intentional and non-intentional behaviour in order to circumvent problems about intentionality in non-human animals and facilitate the discussion by focussing on what matters for the question about mindreading. Furthermore, it is unclear how to properly understand the term *because*, for it is an open question if non-human animals can *act for reasons* in our sense of the words, or if the connection has to be understood in different terms (Glock 2009, 2010).

the latter could be interpreted as a reaction to the fact that the friend is sad or wants some affection. The latter seems to involve the recognition of an other's mental state, while the former does not.

Even though it seems like this distinction is easily made, there are of course many different types of actions which are very hard to classify and explain correctly. However, we can, in theory, distinguish at least three types of mental mechanisms which bring about the actions in question:

(1) First off there are actions which are directly caused by the actions of others and do not involve thinking in a significant way. The mental mechanism underlying these actions are of a very causal and rigid input-output nature. Thus the cognitive mechanism that produces the action is a very simple and inflexible one.

(2) Secondly, there are actions which are a reaction to a specific behaviour shown by another animal (e.g. fleeing upon the sight of a predator, or anticipating friendly behaviour upon seeing 'play bows' in dogs (Cf. Bekoff 1977)). These actions are generally brought about by a mental mechanism which involves taking behavioral and environmental data as inputs and generating a specific output. I will use the term *behaviour reading* to designate the cognitive process which is involved in these types of actions. The animal's reaction is is thus generated by its recognizing and interpreting of the behaviour (and only that) of other animals. Of course, there are very different forms of behaviour reading mechanisms with varying levels of complexity and flexibility.

(3) Finally, there are mechanisms which involve the subject attributing mental states to others and acting, at least in part, as a result of those attributions and interpretations.

These are cases of *mindreading* and are very common in human social behaviour.

In what follows, I will focus on the distinction and interrelation between behaviour reading and mindreading, for these two abilities are commonly used to explain intentional social behaviour of chimpanzees and other non-human animals. Proponents of the view that chimpanzees are mindreaders are ready to explain their behaviour by admitting that they are able to understand at least some mental states of others. By contrast, opponents of this view generally offer an explanation that relies only on behaviour reading abilities[3].

2.1 Mindreading and behaviour reading

The scientific research of the ability of mindreading pretty much started with Premack and Woodruff's famous paper 'Does The Chimpanzee Have A Theory Of Mind?' (1978) where they conducted experiments on chimpanzees in order to investigate their ability to understand mental states of others. Their initial definition of the term *'theory of mind'* (TOM) was that 'in saying that an individual has a theory of mind, we mean that the individual imputes mental states to himself and others (...)'. They state that 'a system of inferences of this kind is properly viewed as a theory, first, because such states are not directly observable, and second, because the system can be used to make predictions, specifically about the behaviour of other organisms' (p. 515).

However, the term 'theory of mind' is sometimes exchanged for the more neutral term '*mindreading*', since the latter does not imply the possession of any implicit of explicit law-like theory (Goldman 2006: 10). For this reason, I will mostly use the term 'mindreading'

[3]It is tempting to think that the distinctions above are clear-cut ones. But in fact there are a great lot of subdivisions and interrelations which can only in part be taken into account here. The distinction between mindreading, involving attribution of beliefs and desires, and *mood-reading* is only one example.

to denote the ability of attributing mental states like beliefs, desires, intentions, and perceptual experiences to others (both conspecifics and members of other species).

Human beings use mindreading in a wide variety of situations when they are using behavioral information to predict and manipulate behaviour of others, but also in order to understand and explain (Lurz 2011: 2)[4]. So we explain an agent's buying an ice-cream by attributing a *want* to eat ice-cream to him. Similarly, we predict future behaviour of a thirsty looking person by thinking that she *wants* something to drink and *believes* that there is water in the fridge, which leads us to believe that she will move to the fridge and get some water. The mental processes which precede he explanation and prediction of actions is thus something that is going on in the mind of the mindreading subject. Even though we might be engaging less often in mindreading activities than is often thought (Cf. Bermúdez 2009; Povinelli & Vonk 2006; Baldwin & Baird 2001), the evidence that human beings are capable of mindreading is overwhelming.

However, mindreading in animals is heavily debated for some significant reasons. First, it is important to note that when discussing the possibility of mindreading animals, it is generally assumed that many animals, especially most mammalian species, and perhaps more simple-minded creatures (Cf. Carruthers 2004), have minds. These animals are generally believed to have first-order mental states like perceptions, emotions, beliefs and intentions of

[4]Prediction, explanation, and attribution are technically a linguistic activity. They are thus technically not available to nonverbal animals. However, since there is a cognitive mechanism that precedes the speech-act of predicting, explaining, or attributing, it is plausible to use these terms loosely to designate the respective mechanisms, which is not necessarily a linguistic process. In the case of predicting, there is something like an *expectation* of the subject about what will most likely happen next, and the subject will feel a need for explaining an action if these expectations are *violated* in some sense. Furthermore, verbally attributing something to someone goes in hand with a mental activity that involves *imputing* something to someone. For matters of simplicity and in order to use the same terminology as elsewhere in the literature, I will use prediction, explanation, and attribution to designate the psychological mechanism underlying those speech acts.

some kind. And secondly, since non-human animals can only be observed by the behaviour they show and not by verbal reports of their own mental states and thoughts, explanations of animal behaviour that only involve *behaviour reading* are readily at hand.

It is important to emphasize that the term 'behaviour reading' in general does not suggest that the animal's predictions and interpretations are only made on the basis of learned behaviour rules or pure associations. They might rely on innate mechanisms which lead the animal to anticipate specific behaviour when being confronted with a specific perceptual condition. It might even be possible that some types of behaviour-reading are based on a form of embodied simulation as will be seen later in this paper (Cf. Lurz 2011).

So how can we distinguish mindreading from behaviour reading in nonverbal animals? This problem can be nicely illustrated by looking at Premack and Woodruff's paper that was mentioned beforehand.

The two authors conducted a series of experiments with chimpanzees, designed to test their ability to understand different mental states like intentions, goals, or beliefs. In one of the experiments a chimpanzee, Sarah, was shown a videotape of a person facing different problems, like trying to reach for something that was out of reach, or trying to reach for objects that were obstructed. Before the solution of the problem was shown, the videotape was stopped, and Sarah was presented with two pictures of possible scenarios, one of which was constituting a solution while the other was not. Sarah then had to choose one of those pictures (Premack & Woodruff 1978: 516).

These experiments were meant to decide between different possible explanations of how Sarah would go about solving the problem: (a) *associationism*, (b) *theory of mind* and (c)

empathy.

According to associationism, animals recognize different types of actions and their respective routines. If a specific action is interrupted before the routine is finished, it chooses the solution that fits the type of action that is being represented. The animal therefore associates the action in the experiment with other versions of that action that it has encountered in her life and chooses the solution accordingly. However, Premack and Woodruff argue that this particular explanation has weaknesses when dealing with future events and encounters with new problems which have no association with past actions or routines. Nevertheless, they concur that it is possible to find associative explanations in all of the successful experiments they conducted, which makes it hard to rule out associationism in their case. This will be a common problem in many experiments. How can we provide reasons that something other than learned routines and making associations is responsible for the supposed mindreading behaviour in question?

According to the theory of mind explanation, which is favored by the authors, Sarah attributes mental states to the actor in the video sequence. On one hand she imputes intentions and goals to the actor (i.e. 'she *wants* to get the thing out of reach'), and on the other hand knowledge and belief (i.e. 'she *knows* the solution to this problem'). The chimpanzee makes sense of the actors behaviour by assuming that the actor wants the object out of reach and knows how to reach it, which is why Sarah picks out the correct solution when the videotape is stopped.

Finally, according to empathy explanations, the chimpanzee puts itself 'into the shoes of the actor', so to speak. It then chooses the solution that it would realize, if itself were in that situation. In this sense then, Sarah's choice is not an expectation of what the actor will

do, but a representation of what she herself would do, were she in the actor's situation. This is essentially the *simulation view of mindreading*. The authors acknowledge this by stating the similarities to the theory of mind explanation. In both cases, the chimpanzee identifies mental states of the actor (e.g. intentions and goals, knowledge and beliefs), which makes it a case of mindreading. But the explanations differ in how then these inputs are further computed and used to arrive at a decision or expectation.

The success with which Sarah chose the correct solution to the problems led to the conclusion that the chimpanzee was capable of attributing mental states (e.g. wanting the objects out or reach or believing that there is a specific solution) and made her decision *based on these attributions*. They argue that admitting the possession of a theory of mind offers the simplest kind of explanation to the phenomena, whereas the other explanations are weaker, since they are based on *ad hoc* hypotheses about the associations that are drawn out by the chimpanzee, or lack empirical evidence. This already shows the common way of deciding between behaviour-reading and mindreading explanations, which often involves an *inference to the best explanation* kind of argument (Cf. Tomasello & Call 2006).

However, among the commentators of the article were three philosophers that raised an interesting point. Dennett (1978), Harman (1978) and Bennett (1978) showed their doubt about the conclusion made by the original authors of the study. While a mentalistic explanation is coherent and fits the results of the experiment, it is doubtful whether the experiment in fact favors it over a non-mentalistic explanation which does not rely on the ability to attribute mental states to others. The problem they pointed out was that the experimental paradigm of the study is incapable of empirically distinguishing between a

mentalistic and non-mentalistic process. Therefore, Premack and Woodruff fail to show that mindreading is the best explanation of the chimpanzees' behaviour. Everything they were able to show by their experiment was that chimpanzees can master these kinds of tasks.

The problem of experimentally distinguishing cognitive processes that involve mental state attribution from cognitive processes that only involve observing of behavioral and environmental indicators has been labeled the '*logical problem*' and is considered to be a fundamental problem of research on animal mindreading abilities (Lurz 2011: 9; Hurley & Nudds 2006: 58). The logical problem therefore consists in the question of how best to explain some of the phenomena of social cognition. Is it best to claim that the animal possess mindreading-abilities, or is this an unnecessary addition of a mental mechanism which is not warranted by the empirical data?

Furthermore, the three commentators pointed towards the necessity of designing a test which can clearly indicate the possession of the *concept of a belief* by testing for the understanding of *false belief* in others (Harman 1978). Harman proposes one such test that could identify the difference between recognition of behavioral and environmental indicators and attribution of mental states. In this test, an observing chimpanzee (O) is shown a scene where another chimpanzee (S) sees a food item placed under one of two containers. O then observes that the location of the food is changed by the experimenter without S being able to see it. If O expects S to look under the container where the food was originally placed, we can assume that he attributes to S the belief that the food is where S has seen it being placed. This would then show that the chimpanzee has an 'conception of mere belief' (p. 577).

Possession of the concept of belief is often considered to be fundamental to the possession of other mental state concepts like desires, intentions and goals (Davidson 1975; Davidson 1980). However, this view has widely been challenged (Bermúdez 2003: Chapter 6; Lurz 2011: Chapter 4), especially by researchers in the field of animal mindreading who suggest that mindreading is not an all-or-nothing affair (Tomasello 1999). Instead, understanding of different mental states should be explored individually and not based on the success in false belief tasks.

In this section we have seen the different approaches which are possible to explain specific types of animal social cognition. While behaviour reading only depends on the recognition and interpretation of behavioral and environmental indicators, mindreading requires the subject to impute mental states to others. First, it is important to note that mindreading depends heavily on behaviour reading and that there are different possibilities of interaction between the two of them (Whiten 1996). Secondly, social cognition is not exhausted by the kind of tasks which are in the focus here (e.g. explanation and prediction of others' behaviour). In fact, social cognition might very likely involve quite diverse cognitive mechanisms and systems which are specifically designed for picking up different things (e.g. moods, feelings, judgements, and what not) in different species. And thirdly, it shouldn't be thought that the focus on mental state attribution, including beliefs and desires, is meant to restrict the discussion on the 'folk psychology' aspect of social cognition (Carruthers 1996). Mindreading can happen in absence of anything like a folk psychological theory of mind.

2.2 The logical problem

I will now discuss the challenge to mindreading explanations which we called the 'logical problem' in more detail. We have seen that the reaction to Premack and Woodruff's paper showed some doubt over their favoured (mentalistic) interpretation. However, it is plausible to think that this doubt stems in part from the fact that the experiment used a novel protocol and its methodology was rather weak for it to support the wanted conclusion. In any case the empirical data and methodology has greatly improved in recent years, especially after the comments made by Harman, Dennett, Bennett and others (Lurz 2011: 10). This lead to the possibility of making stronger claims and commitments about the mindreading-ability of non-human animals.

Nonetheless there have been some influential opponents who always favor a strong behaviour-reading hypothesis over a mentalistic explanation. According to this hypothesis, non-human animals can not understand mental state concepts, but instead make sense and predict others' actions only by looking at the behavioral and environmental facts and indicators. They do this by learning from experience, infererence, and maybe by possessing innate structures and mechanism (Lurz 2011: 22). Among the chief defendants of these kinds of behaviour-reading explanations is Daniel Povinelli, who strongly argues for the *in principle incapability* of current experimental paradigms to determine the animal's understanding of mental states (Povinelli & Vonk 2006; Vonk & Povinelli 2006; Penn & Povinelli 2007; Penn et al. 2008; Tomasello & Call 1997).

For example, Vonk and Povinelli write that 'there is ample evidence to conclude what we already knew: chimpanzees are keen observers of readily perceivable features of the social world, and make predictions about future states of that world on the basis of such features'

(Vonk & Povinelli 2006: 372). This in turn means that the authors feel no urge to employ mentalistic explanations of the observed phenomena, but in fact think that a purely non-mentalistic explanation, only involving behaviour-reading mechanisms, is sufficient, and for theoretical reasons superior.

The best way to characterize the logical problem is by contrasting two different *psychological systems*, S_b and $S_b + ms$. The former is a system that is 'dedicated to social cognition, but one which forms and uses concepts about only 'behaviours' which can, in principle, be observed' (Povinelli & Vonk 2006: 391). That is, a system which is only capable of behaviour-reading. The latter however is a system capable of both behaviour-reading and mindreading and thus can, in principle, attribute mental states to others. It is 'a system which must perform joint computations about both the behaviour and the mental states of other organisms in order to successfully predict future behaviour' (p. 391). The logical problem is then stated as follows:

> The general difficulty is that the design of these tests [which investigate mindreading capabilities] necessarily presupposes that the subject notices, attends to, and/or represents, precisely those observable aspects of the other agent that are being experimentally manipulated. Once this is properly understood, however, it must be conceded that the subject's predictions about the other agent's future behaviour could be made either on the basis of a *single step* from knowledge about the contingent relationships between the relevant invariant features of the agent and the agent's subsequent behaviour, or on the basis of *multiple steps from the invariant features, to the mental states, to the predicted behavior*. Without an

> analytical specification of what additional explanatory work the extra cognitive step is doing in the latter case, there is nothing to implicate the operation of $S_b + ms$ over S_b alone (p. 393, emphasis mine).

It is a matter of debate what exactly specifies the 'additional explanatory work' which is necessary. From a theoretical point of view, the question of *simplicity of explanation* is important in deciding between two competing theories, in our case the mindreading and behaviour-reading hypothesis. However, simplicity is polysemous and there are at least two different principles which are in conflict here. On one hand, Morgan's canon, which states that we should prefer explanations of animal behaviour that make use of fewer, and less complex, mental states, gives us reason to avoid postulating higher-order mechanisms if the phenomena can be explained by lower-order mechanism. There is no question that mindreading involves higher-order mechanisms when compared to mere behaviour-reading, and also that it is often possible to explain the phenomena using only lower-order mechanisms. However, in using only behaviour-reading abilities to explain the phenomena, many additional hypotheses are added which amounts to a less simple and elegant overall explanation of the phenomena (Cf. Hurley & Nudds 2006; Tomasello & Call 2006). Nonetheless, if the phenomena can in principle be explained by lower-order mechanisms, there is an apparent theoretical problem for the mindreading hypothesis, even if it might be simpler and more economical in a quantitative sense.

Additionally, Povinelli claims that there is a *complementary behaviour-reading hypothesis* for *every* mindreading hypothesis in nonverbal animals. The reason is that in conducting experiments which are designed to find out whether a nonverbal animal has mindreading

abilities, the only indicators for the experimentor are behavioral cues and facts that can be observed in the experiment. All supposed mental state attributions in nonverbal animals have to be based on the observable features of the experiment which automatically yield a behaviour-reading hypothesis that only relies on those facts, which in turn does not impose additional non-observable mental state attributions.

When we attribute mindreading abilities to a nonverbal animal based on these kinds of experiments, Povinelli claims that one unnecessarily adds a further theoretical step (i.e. the mental state attribution step) for unjustified reasons. If an animal's behaviour can be explained by taking the observable facts and the animals reaction, there seems to be no need to induce an additional step wherein the animal has thoughts about the mental states of others. Therefore, according to the behaviour-reading hypothesis, 'identifying and then analyzing exactly the behavioral/environmental variables that best predict future behavior of other agents because they reliably constrain the future behavior of these agents, is, functionally, what a system that is specialized for behavior reading does' (Lurz 2011: 28).

The general argument that is incorporated in the logical problem is thus the following (Cf. Lurz 2011: 76-7):

1. There is a mindreading hypothesis according to which an animal A anticipates that another agent B will perform behaviour *r* on the grounds that B is in some mental state *m*, and that B's being in *m* is likely to lead him to do *r*.

2. Mental concept applications are based on the observation of behavioral indicators. Therefore A has to apply the mental state concept *m* on the grounds of some behavioral indicator *s* about B's behaviour or environment.

3. There is a complementary behaviour-reading hypothesis which replaces A's mental state attribution of *m* to B with A's acknowledging the fact that the behavioral indicator *s* obtains.

4. If there is reason to believe that A has observed *s*-type indicators in connection with *r*-type behaviour in other animals, or himself, then it is plausible to suppose that A anticipates B doing *r* on the grounds that *s* obtains. This means that A does not additionally need to attribute *m* to B in order to anticipate B's *r*-behaviour.

5. There is no experimental paradigm which is able to eliminate the plausibility of 4.

∴ There is no experimental paradigm that can distinguish between the mindreading hypothesis and the complementary behaviour-reading hypothesis.

This argument however does not show that the logical problem cannot be overcome (Cf. Lurz 2011). By designing a test that tackles the problem of finding evidence that can rule out a complementary behaviour-reading hypothesis in a specific task (i.e. finding evidence that shows the falsity of 5.), the logical problem can be ruled out and the corresponding mindreading hypothesis given the needed empirical weight. However, as we will see in chapter 4, it is doubtful whether such a test has been designed as of yet, and if it is in principle possible. The theoretical argument presented as the logical problem thus needs further theoretical and empirical evidence to prove that it is *in principl* impossible to show that nonverbal animals are capable of mindreading.

In this section, I have illustrated the logical problem of experimentally distinguishing mindreading from behaviour-reading which has been stated by different people but is most

strikingly defended by Povinelli and colleagues. The fact that it seems possible to form a complementary behaviour-reading hypothesis for every test on nonverbal animals which relies on behavioral indicators, makes it hard for defendants of the mindreading hypothesis to give their mentalistic interpretation enough theoretical and empirical support. It also underlines the need for a specific experimental paradigm that is capable of solving the logical problem, if the mindreading hypothesis is to be upheld for nonverbal animals.

3 Theories of Mindreading and their representational content

As we have seen, the mindreading discussion has been initialized by Premack and Woodruff's paper on chimpanzees' mindreading abilities. It quickly extended to the domain of developmental psychology and cognitive science in general, where two different general theories about how mindreading might cognitively work, Theory Theory (TT) and Simulation Theory (ST)[5], have been developed and fine-tuned. Most of the research that fuels the discussion amongst the proponents of both theories is found in child developmental research, however, for various reasons, insights from nonverbal, non-human animals like apes and others are often used to support claims for either theory[6]. Furthermore, there are, as always, different versions of both theories and also some attempts to combine the two in order to get a fuller explanation without suffering from the faults that come with both theories on their own.

In general, TT assumes that mindreading is achieved by an implicit or explicit theory of how the mind works, and there are variations of how theoretical they are, i.e. how coherent and tight the underlying principles need to be (Carruthers & Smith 1996). The most common example of a version of TT is the belief-desire folk psychological theory of mind. According to this, most human adults have an underlying theory according to which actions are guided by the beliefs and desires of the agent, and accordingly, their mindreading interpretations and predictions are based on these principles (Cf. Davidson 1963; Botterill 1996)[7].

[5]These two theories roughly map onto Premack and Woodruff's characterization of 'theory of mind' (as a rough form of TT), and 'empathy' (as a rough form of TT) (Cf. Goldman 2006).

[6]Especially research on simulation theory that involves intrusive experiments (e.g. experiments that discuss the cognitive structure and functioning of *mirror neurons* (Cf. Addessi & Visalberghi 2006: 323; Pepperberg 2006: 477)) is often restricted for (doubtful) ethical reasons to non-human animals and then used to draw conclusions about the supposed human equivalent mind structure (Cf. Lurz 2011: 4).

[7]This does not imply that animal mindreading has to be anything like belief-desire folk psychology. It is only necessary to be implicitly theory/principle-driven.

By contrast, ST holds that mindreading does not need an underlying theory containing principles of the human decision-making process. Instead, mindreaders can simply run a simulation of this process by inputing others' mental states and using the output of their own decision-making mechanism as basis to explain and predict others' behaviour. This is to say that 'mindreaders use their own minds to "mirror" or "mimic" the minds of others' (Goldman 2006: 20). In order to prevent the result of the 'mirroring' decision-making process leading to an action, defendants of ST generally hold that these simulations are run *off-line* or detached, so that the output does not result in an actual action.

In this section, I will discuss the pros and cons of both TT and ST and relate them to the question of the possibility of mindreading chimpanzees and other animals. Furthermore, the question of what content drives the respective mindreading process according to TT and ST will be evaluated. It will be argued that both theories need a mental state with metarepresentational content at some point. If genuine mindreading is supposed to happen, thinking about mental states is necessary.

3.1 Theory Theory and Simulation Theory

It is not the aim here to give a full overview of all the arguments for and against either theory and all their respective versions. Rather, the relevance to explaining animal mindreading and the question of the representational content of both theories will be important[8]. In this section, the first issue will be discussed, while the second question will be tackled in the

[8]In fact, it is very likely that a mix of both ST and TT is needed to fully explain animal mindreading mechanisms (Perner 1996). The diverse mental tasks involved in mindreading require an equally diverse cognitive architecture that is best accommodated by a mix of ST and TT. However, for simplicity's sake, I will here only discuss the two theories in isolation.

following section.

As we have seen, TT relies on the assumption that there are some underlying, sometimes implicit, principles that govern the explanation and prediction of others' behaviour by looking at their supposed mental states. The mental state concepts, like *belief*, *desire*, *intention*, *seeing*, *being in pain*, and others, are part of a *theoretical framework*. This framework incorporates the principles of how these mental states interact and relate to actions. Important thereby is the necessary interrelation between the mental states, so that understanding one of them seems to require understanding some or all of the others (Stone & Davies 1996). In order to grasp the functional role of beliefs and thus grasping the concept of a belief, it is necessary to understand how they relate to other mental states like seeing or desiring. For example such a theory of mind would incorporate a principle like: 'Persons who want that P, and believe that Q would be sufficient to bring about P, and have no conflicting wants or preferred strategies, will try to bring about that Q' (Churchland 1988: 58-9).

However, it is highly questionable if animal mindreading capabilities fulfill this condition. While there is good evidence that suggests mastery of the concept of seeing and other perceptual mental states (Cf. Call & Tomasello 2008), there is little evidence for other states, like beliefs, which are commonly held to be vitally important to grasping the whole mindreading theory. This has given various researchers reason to believe that mindreading could have evolved in a series of individual steps and not as a all-or-nothing affair (Povinelli & Eddy 1996).

So if TT implies that mindreading is an all-or-nothing affair, it seems to be an insufficient framework for understanding animal mindreading. However, there are two reasons why this is not the case. First, TT need not imply that there is only one type of full-blown theory of

mind. And secondly, failure to find empirical evidence for, say, the possession of a concept of a belief in chimpanzees, does not imply that they do not in fact have such a concept. And furthermore, the logical problem remains, which suggests an alternative explanation of pure behaviour-reading, if no substantial evidence for a theory of mind in non-human animals can be found.

Before discussing the first reason, let me elaborate on the last point. If we want to find out how mindreading in animals work, TT offers a solution that is very straightforward and based on the similarities to human mindreading. Animals can predict behaviour of others by using implicit principles that describe the interrelation between mental states and their relation to actions and behaviour. Currently, this is still a viable option for explaining mindreading in animals. There are two possibilities at hand, either, it is possible to devise a test that can overcome the logical problem, or this is not possible. If the latter is the case, which is in doubt (Cf. Lurz 2011), TT cannot offer an explanation of animal mindreading since substantial parts of the picture lack empirical evidence[9]. However, if the logical problem can be overcome and it is possible to show that some animals understand others' mental states, then TT could be used to explain their mindreading abilities. If it is not possible to show that chimpanzees understand beliefs, TT could still be used as a theory of animal mindreading, since it need not imply that there is only one type of full-blown theory of mind that is heavily dependent on the possession of a concept of belief.

The way TT has been described so far seems to imply that it can only give an explanation of a theory of mind that is very like our own, depending crucially on some form of belief-desire psychology and incorporating the principles that govern the interrelation of

[9]Of course, this is not a problem specifically for TT but applies equally to all sorts of theories which attempt to explain animal mindreading.

those important mental states. However, this need not be the case. In a very broad sense, TT only says that there are *some* principles which govern the explanation and prediction of others' behaviour (Stone & Davies 1996)[10]. How exactly those principles are structured and what content they require is an open question and may differ between different species or developmental stages. One proposal is that there is a '*body of knowledge*' (Stich & Nichols 1995) which contains beliefs about the psychological facts that govern behaviour. If this is true, there is no strong restriction to belief-desire psychology, but instead other, maybe less concise, principles are capable of doing the job for mindreading. For example, Tomasello and Call (2008: 191) suggest that the best way to explain the empirical evidence for chimpanzee mindreading is in terms of a form of *perception-goal* theory of mind which could easily be explained with a version of TT.

However, ST offers a challenging alternative to TT in explaining animal mindreading which in fact seems much more intuitive and less theoretical. The most important difference between the two is that ST is *process-driven* whereas TT is *theory-driven* (Goldman 1989). Mindreading as explained by ST is therefore essentially an ability which is enabled by a specific process which is not in need of a special theoretical basis. However, this requires the mindreader to *embody* the mental states of others in order to get the simulation-process to run. The decision-making process is then simulated *off-line* with the mental inputs derived from others. Essentially, this amounts to finding out which available options are *attractive from the point of view of the other agent*.

[10]Furthermore, TT seems to imply some heavy dependency on linguistic abilities, recognition of the principles in question, and the conscious attempt to the theoretical framework underlying mindreading. However, this is not necessarily implied at all.

In this sense then, embodied off-line simulation is a more basic and lower-level seeming cognitive ability. It does not rely on the mental representation of other's mental states in propositional format, as is often suggested by the TT approach (Gallese 2007). Even though it is possible for us to form such sophisticated mental representation and formulate a theory that accounts for their interrelation and connection to actions, it is neither our only way of reading others' minds, nor is it something we should be over-ready to attribute to non-human animals. ST does not require us having an implicit theory because it relies on the mindreader's ability to act based on their own mental states. Therefore, it does not impose a new sophisticated ability, which seems very attractive if we are interested in animal mindreading.

Because we have agreed that chimpanzees do have mental states that are responsible for their actions and possess basic rationality, ST only has to explain how this ability can be used to embody mental states of others and explain and predict behaviour by off-line simulation. Gallese and Goldman (1998) further suggest a division between *low-level* and *high-level* simulation. Whereas low-level simulation is based on strong internalized reactions to specific behaviour (e.g. as seen within social monkey feeding behaviour (Addessi & Visalberghi 2006)), high-level simulation is less internalized and not dependent on invariable, possibly innate, mechanisms. The question remains whether low-level simulation can still be considered genuine mindreading, and I will discuss this in the next section.

The empirical evidence that supports the ST approach to mindreading in animals mostly focusses on the existence of *mirror neurons* in the premotor cortex of macaque monkeys and in homologues brain areas of other species (Lurz 2011; Goldman 2006; Addessi & Visalberghi 2006; Pepperberg 2006; Gallese 2007). The feature that makes those neurons inter-

esting for mindreading research is the fact that they show activity when the subject performs a specific action and also when it observes the same type of action in others. This 'direct matching' of the neural activity of observing and acting is believed to 'underpin a direct form of action understanding by exploiting embodied simulation, a specific mechanism by means of which the brain/body system models its interaction with the world' (Gallese 2007: 660). Different experiments further support the hypothesis that the same mental *action representations* are guiding both action production and action understanding.

This again would mean that the same inputs and mechanisms are at work when an animal is producing goal-directed action as when it is predicting or understanding others' goal-directed actions by means of embodied off-line simulation. This supports the view that ST can in fact explain action understanding and predicting in animals by pointing towards the activity of mirror neurons. However, it is unclear if the application of such a mechanism is rightly a case of '*substantive mindreading*', which is an activity that is systematically dependent, *but not strictly causal*, on the creature's representation of the mental states of others (Bermúdez 2009: 148). For example, Addessi and Visalberghi (2006) describe the feeding behaviour of macaque monkeys and the role mirror neurons play in the understanding of others' behaviour and guiding of the subject's actions. However, the fact that macaques seem to respond to the actions in a very systematic and causal way (i.e. the more conspecifics are eating, the more they eat), suggests that the existence of mirror neurons may be supporting low-level simulation but not necessarily the capability of high-level simulation. And thus, even though ST can account for mindreading in animals, the existence of mirror neurons alone does not account for the ability of substantive mindreading. They do, looked at in isolation, only show that the mirror neuron system is sophisticated enough to enable social

interaction, but not necessarily mindreading abilities.

In this section, I have briefly discussed the two main approaches to mindreading, Simulation Theory and Theory Theory, and their relevance to mindreading in animals. Even though a rudimentary form of TT seems implausible to explain animal mindreading, an adapted and more liberal version of TT is capable of providing a reasonable explanation. For example, Tomasello and Call (2008) suggest that chimpanzees are mindreaders with a perception-goal theory of mind. On the other hand, ST offers a more basic way of understanding the functioning of mindreading in animals. It relies on the fact that animals are capable of producing goal-directed actions and are thus already equipped with the mental modules that are needed for embodied simulation of others' mental states and actions. This is empirically supported by the existence of mirror neurons which seem to explain how offline embodied simulation is working on a neural level. However, there is doubt about the relevance of such evidence for the claim that animals are substantive mindreaders and not just good behaviour-readers with a good capability of understanding very specific actions and behaviours of others (like feeding in groups of monkeys).

3.2 The mental content of animal mindreading

So far, the mindreading abilities have been looked at only very broadly and most importantly, the question of the content of mindreading has been left in the shadows. However, it is crucial to discuss the nature of the content of mindreading mechanisms in animals as this will give rise to experimental paradigms which allow to test more exactly whether or not

animals are capable of mindreading[11].

Generally, decision-making processes start of with specific desires and beliefs as *inputs* which are fed in a *decision-making mechanism* and produce a decision as an *output* (Cf. Goldman 2006: 27). Therefore, we can assume that animals, which have mental states and are capable of decision-making have something like desires and beliefs as inputs of their decision-making mechanism[12]. In the case of simple decision-making, the contents are first-order mental states which are fed into a decision-making mechanism of some sort.

However, the story is more complex if we look at the mindreading-mechanism. According to TT, the mindreader needs a representation of the other's mental states (i.e. their beliefs and desires), and a specific principle which explains how mental states relate to decisions, as inputs. These are then processed in a specific *factual reasoning mechanism* and generate a belief about the predicted decision of the other (Goldman 2006: 28). The chimpanzee thus reasons something like this:

> $\ulcorner \text{belief}_1 \urcorner$: The dominant wants/has as a goal/desires/intends to gather as much food as possible.
>
> $\ulcorner \text{belief}_2 \urcorner$: The dominant beliefs/can see/knows that there are two items of food in the room.
>
> $\ulcorner \text{belief}_3 \urcorner$: Some principle or body of knowledge about how desires and beliefs relate to actions (e.g. 'agents who want that P, and believe that Q would be sufficient to

[11]Furthermore, this is also important from the philosophical point of view. Mental state concepts are generally held to be *intentional*; they have something as their content. However, it is debatable which contents are accessible without language.

[12]I say 'something like' in order to account for the possibility that some animals have mental states which are roughly like desires and beliefs but not exactly the same. For example, explaining an action in terms of what an animal's goals and perceptual states are would not be an explanation in terms of desires and beliefs. However, I will speak of desires and beliefs in a broad sense to facilitate the discussion. Furthermore, Bermúdez (2003) has given strong arguments for the view that we should think of animal mentality in terms of beliefs and desires.

bring about P, and have no conflicting wants or preferred strategies, will try to bring about that Q' (Churchland 1988:58-9)).

These beliefs are then fed into the factual reasoning mechanism and generate a belief about a state of affairs (i.e. the action the dominant will most likely perform). It is clear that the inputs for mindreading of this type are of *metarepresentational* nature. They are beliefs about mental states, which themselves are representations of state of affairs. The output could be both a first-order belief (e.g. 'I believe that he will do X') or a second-order belief (e.g. 'I believe that she is feeling sad about X'), depending on the nature of the reasoning process and the inputs.

If we contrast this with a chimpanzee who is only engaging in behaviour-reading activities, we can see that only first-order representational mental states are needed.

$\ulcorner \text{belief}_1 \urcorner$: The dominant shows behavioral cues and indicators X directed at food gathering.

$\ulcorner \text{belief}_2 \urcorner$: The dominant had a clear line of sight to two items of food in the room.

$\ulcorner \text{belief}_3 \urcorner$: Whenever dominants show behavioral cues and indicators X and have clear line of sight to two items of food, they will get both items.

When processing these three first-order beliefs about states of affairs, the chimpanzee arrives at the prediction that the dominant will get both items of food to which he had a clear line of sight (Cf. Lurz 2011, Chapter 2). If we therefore want to explain mindreading in animals with using some variety of TT, it will be necessary that the animal is capable of forming higher-order metarepresentational beliefs about mental states of others and use

those beliefs to understand and predict actions of others.

However, the picture is different for ST. Because embodied simulation does not require something like a factual reasoning mechanism, but instead depends only on the already present decision-making mechanism, there are more possibilities for different representational contents. When engaging in mindreading by means of embodied simulation, the beliefs and desires of the other are used as *pretend inputs* in the personal decision-making process. This then generates a *pretend decision* which will again be attributed to the other (Goldman 2006: 30). Therefore we get a process that looks something like this:

⌜pretend desire$_1$⌝ : I want *g*.

⌜pretend belief$_1$⌝ : If I do *m*, I can get *g*.

These two mental states are then fed into the decision-making mechanism which will give us something like 'I will do *m*' as an output. However, in order to achieve off-line embodied simulation, it is necessary to somehow attribute the pretend decision and pretend mental states to others. One such way is by saying that effectively the pretend mental states have metarepresentational content, since they are different from *genuine first-order mental states*. Although, while TT explains mindreading as an activity that uses metarepresentations, the content of which (i.e. a first-order mental state) is ascribed to the target, ST seems to assume that mindreading only needs to take first-order mental states and then *project* the output to the target (Goldman 2006: 40).

This however presents a problem. The projection ('T will do *m*') is a genuine first-order belief, derived, but different, from a pretend decision ('I will do *m*') which is ascribed to the

target T. However, this has to be some kind of belief if it should not be an actual decision. But how can the decision-making mechanism generate a belief about a decision by only using pretend desires and beliefs as inputs, when they are not in fact also higher-order beliefs with those pretend desires and beliefs as their content. This again would mean that ST has to assume that the mindreading subject is capable of having higher-order mental states with metarepresentational content[13].

Therefore it seems that mindreading, whether explained by TT or ST, always has to include some form of metarepresentational content. This in effect means that mindreading is a metacognitive activity as it necessarily requires having mental states with other mental states as their content. In the case of TT, this was pretty clear from the start, because TT relies on beliefs about mental states whenever we are dealing with genuine mindreading, and not mere behaviour-reading. But in the case of ST the above argument shows that having pretend desires and beliefs effectively implies having further beliefs about those mental states (i.e. that they are in fact not *my* desires but those of T). I will now explore a possibility that tries to explain genuine mindreading by means of simulation, but without need of metarepresentations to do the job of distinguishing between *genuine mental states* and *pretend mental states*.

3.3 Simulation without metarepresentation

If we want to explain the possibility of simulation which does not rely on metarepresentational content, we need to find a way of explaining the mindreading in question that satisfies the following three conditions:

[13]This is not an argument against ST, but an argument against the claim that mindreading, as explained by standard ST, does not require metarepresentational content by the mindreading subject.

(1) The process or mechanism uses only first-order representations of states of affairs as inputs and outputs.

(2) The process or mechanism does not rely on other metarepresentational capacities during the decision-making or projection process.

(3) The whole activity is still genuinely mindreading.

I will argue that there is no possibility to satisfy all the three conditions at once. Additionally, it seems to be obvious that these three conditions are jointly necessary in order to speak of mindreading without metarepresentation. Let me briefly explain why. (1) seems to be pretty straight-forward; if we want to explain mindreading without the need for higher-order representations, the inputs and outputs should clearly be first-order mental states. Additionally, we also want to exclude any other necessary higher-order mental states which would be needed to get to the mindreading stage at all. This is what is meant by (2). Furthermore, we are interested in the ability to think about mental states of others and not just their behaviour. This is why (3) needs to be satisfied in order to make a strong case for the possibility of mindreading without metarepresentation.

How could simulation work without need of higher-order mental states like beliefs about beliefs or desires? First off, it requires that the subject does have beliefs about the behaviour of others. This is straight-forward, as I have already indicated how this might look. A chimpanzee is able to form a belief about the line of gaze of another chimpanzee and could use this belief to predict and explain behaviour (Lurz 2011: Chapter 3). Furthermore, there are proponents of the view that thinking about one's own mental state does not need

metarepresentational content but can be achieved by employing only first-order mental states (Proust 2007, 2010; Hieronymi 2009; Peacocke 2008, 2009). So, by evaluating one's own thoughts by means of mental simulation and unconscious control, it is possible to think about thoughts without the need for higher-order mental representation.

The question remains though, how we can fit these different facts into one theory that can explain mindreading, i.e. a capacity that not only involves having thoughts about thoughts, but projecting those mental states on to others, which involves having some concept of the mental state in question. Reflecting upon one's own beliefs might not involve metarepresentational content (Cf. Arango-Muñoz 2010), but projecting a belief onto somebody else necessarily involves having some concept or other of the state in question and this implies the capability of having mental states with metarepresentational content.

If we want to achieve genuine mindreading, there needs to be a point where the subject has a belief-like state about a mental state of someone else. The chimpanzee utilizing the concept of *seeing* to understand a dominant's behaviour, needs to have some belief of the form

BELIEF [dominant SEES X] .

And this is clearly a belief with metarepresentational content. Even if we use ST and are sympathetic to the view that only pretend beliefs and desires are required, there is still the need to ascribe to the mindreading subject some beliefs about the mental states of the other at one point or the other. If we think that this is not necessary, we would be in serious need of a clear explanation on why this is genuine mindreading and not just behaviour-reading. If the subject only utilizes first-order representations of state of affairs, and never actually ascribes a mental state to another being, it is hard to accept that we are in fact talking about

*mind*reading at all.

I believe that these reasons are sufficient to undermine the possibility of mindreading without metarepresentations. On the TT-view we are clearly in need of beliefs about mental states of others, and the same applies, upon closer reflection for ST. However, this does not imply that mindreading cannot be achieved by nonverbal animals like chimpanzees. It would require further arguments to support the view that metarepresentational content can only be possible in the presence of language (e.g. Bermúdez 2003). While I am very suspicious about the success of any such arguments, I cannot discuss them at this point and will instead grant that some form of (metarepresentational) mindreading is in principle accessible to nonverbal, non-human animals. Therefore, I will now discuss the empirical evidence that suggests the ability of chimpanzees to understand and use different mental state concepts.

4 Are chimpanzees mindreaders?

There is a substantive amount of empirical data that is discussed on the subject of chimpanzees' ability to attribute mental states to others. This of course means that there is simply not enough space here to discuss all the different experimental protocols and findings which are related to the different mental states in question. Therefore, I will only discuss some of the more striking and well-discussed empirical data which has been gathered to give an overview over the different experiments which are used to support the mindreading hypothesis for chimpanzees. Furthermore, I will point out the problems which are faced by those experiments and judge whether or not it is possible to say that the experiments do in fact show a capacity to understand others' mental states.

For reasons I have stated earlier, it makes sense to look at the different mental states individually, since it is very doubtful if mindreading is an all-or-nothing affair (Povinelli & Eddy 1996). Therefore, I will first discuss the mental states of *seeing* and *knowing*, for they are very directly connected to behavioral indicators. After that the understanding of *beliefs* is discussed, which will bear much weight for it is considered a very important mental state, necessary to fully grasp other mental state concepts. Finally, it would also be interesting to discuss chimpanzees' understanding of *intentions* and *goals*. However, the empirical data in this respect is unfortunately very sparse, which makes a discussion unproductive (Cf. Tomasello et al. 2003, Call & Tomasello 2008).

4.1 Seeing and knowing

Seeing and knowing are two mental states which are closely related. However, understanding of the two has to be tested in quite different ways. I will discuss the *Competitive Paradigm*

Experiments by Hare et al. (2000, 2001) which deal with chimpanzees' understanding of *seeing*. As well, I will look at findings from various experiments which use the *knower-guesser protocol* (Premack 1988; Gómez & Teixidor 1992; Hare et al. 2001; Kuroshima et al. 2002, 2003) in order to examine the chimpanzee's understanding of the concept of *knowing*.

4.1.1 Seeing

When testing for chimpanzees' understanding the concept of seeing, Hare and colleagues (2000) have designed an experiment which involves two chimpanzees, a subordinate subject and a dominant, both competing for food items located in a room between them[14]. The subject learned that the dominant would get all the food items which are out in the open and visible to him. There were then three different test-scenarios. In the first, both food items were visible to both chimpanzees. In such cases, the subject would not engage and refrain from taking either piece, since it was aware of the dominant seeing both food items. In the second scenario, one item was placed visible for both chimpanzees, but another item was placed behind an opaque barrier, only visible to the subject. It was reasoned that the subject would realize that the dominant could not see the second food item, and thus retrieve it when released from it's room. In order to rule out the possibility of the subject waiting for the dominant to make the first move and then react on it's behaviour, it was released slightly earlier so that it had to make a decision first. Finally, they conducted a further test where

[14]Hare (2001) argues that due the competitiveness of primate social lives, experiments based on cooperation, like the one conducted by Premack and Woodruff in 1978, are not the best way to explore their cognitive abilities. Instead, since their social lives are shaped by competition and not cooperation, their cognitive abilities are very likely to be adapted to situations which involve competition, but not cooperation. In order therefore to further validate the significance of research on primate social cognition, experiments should involve the competitive aspect of social cognition.

one of the food items was behind a transparent barrier, so that the dominant could see both. It was suggested that the subject would then realize that the dominant could see through the barrier and thus see both food items, and therefore the subject would not engage.

The tested chimpanzees in fact behaved as predicted and usually tried to retrieve only food that was not seen by the dominant, i.e. food behind an opaque barrier or inside of the subjects own room, but not food that was out in the open or behind a transparent barrier. This lead to the conclusion that chimpanzees can understand and attribute the mental concept of *seeing* to conspecifics. Because the subject could not see through the opaque barrier, it reasoned that the dominant could also not see through. On the same reasoning it attributed to the dominant the state of seeing the food when it was placed behind transparent barriers and out in the open.

However, this mentalistic interpretation was challenged on the basis that in these experiments attributing of seeing to chimpanzee rests solely upon certain behavioral indicators (Penn & Povinelli 2007). 'Although one can see another subject *looking at* an object, one cannot see his or her *seeing* the object' (Lurz 2011: 34, emphasis mine). Mental state attributions cannot be observed directly and the experimental protocol used by Hare and colleagues cannot circumvent this problem. The complementary behaviour-reading hypothesis suggests that the subject chimpanzee knew only which items the dominant was looking at and believed that those were off limits. One such hypothesis which does not rely on mindreading simply states that the dominant has put his 'evil eye' on the items it was looking at, thus making them off limits for the subject (Povinelli & Vonk 2006).

Even though Tomasello and Call (2006) rightly point out that the mindreading hypo-

thesis is a more economical theory that explains how chimpanzees master so many different tasks which supposedly involve attribution of seeing to others, they fail to propose a test that can decide directly between the mindreading hypothesis and the complementary behaviour-reading hypothesis for the concept of seeing.

Recently, Robert Lurz (2011) has suggested a theory that could theoretically and practically decide between those two hypotheses and answer the question of whether or not animals are capable of attributing states of seeing to others and thus possessing that mental concept. The general idea behind this suggestion is that possess the concept *seeing* has a specific purpose (i.e. understanding the difference between *appearance* and *reality*). Lurz writes (p. 84) that

> [u]nderstanding that a dominant conspecific, for example, does not see the camouflaged insect before it as an edible insect but *as a leaf* would enable a subordinate mindreading conspecific to predict that the dominant conspecific in this setting will not try to eat the insect [...]. Were a complementary behavior-reading subordinate [i.e. one that is not able to attribute mental states] placed in the very same situation as this mindreading one, it would be capable of understanding only that the dominant conspecific is looking directly at an edible insect.

Thus, this theory is called the *appearance-reality mindreading* (ARM) theory of mindreading. He then suggests different experimental setups which could test for understanding the difference between appearance and reality and are related to the setup by Hare and colleagues

(2000). They all involve different kinds of visual barriers which change the appearance of the item behind it, thus allowing the subject to differentiate between what the dominant is merely looking at and what he is seeing. These experimental protocols are able to rule out any complementary behaviour-reading hypothesis because of the fact that the subject has to attribute the mental state of *seeing-as* to the dominant in order to successfully master the task. If for example a barrier changes the color of a food item to something undesirable for the dominant, the subject could see that (a) the dominant is looking at the food item which is behind a colored barrier, (b) the barrier makes everything behind it look, for example, red and thus undesirable in this case, (c) the dominant *sees* the food item behind the barrier in undesirable colors, and thus will not get it. It seems to be clear that (c) can not be achieved through a behaviour-reading hypothesis that involves line of gaze, since it essentially contains a manipulating factor which inserts a difference between *appearance* and *reality*.

Thus, while current protocols have failed to show that chimpanzees understand seeing, ARM proposes a way of distinguishing the mindreading aspect of social cognition by implementing the importance of the appearance-reality distinction for the mental concept of *seeing*.

4.1.2 Knowing

In order to find out if chimpanzees have a concept of *knowing*, experiments often make use of the *knower-guesser protocol*. This setup relies on the distinction between a knowing agent and a guessing agent, and the subject's supposed understanding of this difference. They are however, normally conducted in a way that tests for recognition of *past* states of *seeing*.

Therefore, the same general problems which apply to the mindreading hypothesis of seeing apply to the attribution of the concept of knowing. According to the protocol, chimpanzees attribute to another agent a state of *knowing* the location of an object, if they also attribute a state of *having seen* the location of the object in the past. Similarly, the chimpanzee attributes the state of *ignorance* to another agent, if there is an absence of past states of seeing in the other agent (Cf. Lurz 2011).

In this spirit Hare and colleagues (2001) conducted an experiment very similar to their 2000 study, again using the Competitive Paradigm with a subordinate subject and a dominant chimpanzee. However, in this experiment the dominant either witnessed the hiding of food in one of two opaque containers or he did not see the hiding. The subordinate subject was able to observe the hiding and thus see whether the dominant could witness the hiding process. The upshot of the experiment was that the subject went for the food items more often when it had observed the dominant *not having seen* the hiding and thus not *knowing* where the food was.

Similarly, Kuroshima and colleagues (2002, 2003) conducted experiments with capuchin monkeys where the animals were confronted with a *knowing* experimenter (i.e. an agent which the monkeys had witnessed seeing the baiting process) and a *guessing* experimenter (i.e. an agent which the monkeys had witnessed *not* seeing the baiting process). Apparently, the monkeys learned to prefer the containers which were indicated by the knower over those touched by the guesser. This lead Kuroshima and colleagues to conclude that the monkeys understood the relationship between seeing and knowing. However, the logical problem applies to this study as well. The design is not able to eliminate a complementary behaviour-reading hypothesis according to which the attribution of knowledge is merely an

attribution of *having had a direct line of gaze in the past*, which is a non-mentalistic relation between an observer and an object.

In order to exclude such complementary behaviour-reading explanations for the concept of *knowing*, a test would be needed which works similarly as the one proposed by ARM. The relation between *knowing* and *having seen in the past* should allow a design that uses the protocol which tests for the ability to attribute states of seeing and adapt it to test for the attribution of states of knowing. A proposal might be to use the setup of Hare and colleagues (2000, 2001) but use a transparent, but appearance-changing, barrier, as suggested by Lurz (2011). In one case the subordinate subject observes that the dominant sees the original attractive item being place behind the appearance-changing barrier, thus knowing that the dominant knows that there is an attractive item behind the barrier, even though it appears to be unattractive. In the other case the subordinate witnesses that the dominant does not see the baiting process. If the chimpanzee understands the state of knowing, it should not go for the item when it had observed the dominant seeing the baiting process. However, it is unclear, whether such a complex experimental setup is possible to conduct without thus inducing other problems[15].

4.2 Beliefs

Following the same logic as in the last section, we now come to the discussion of chimpanzee's understanding of the concept of *belief*. There are very influential philosophical

[15]Among these other problems are issues concerning memory capabilities of chimpanzees and questions about how much mental processing power we can assume that chimpanzees possess. Furthermore, it has to be possible for the subject to learn the routines of the experiment without loss of concentration of attentiveness, etc.

arguments that deny the possibility of having beliefs without language (Cf. Davidson 1975, 1982) which unfortunately cannot be discussed here. Nowadays, it is a widespread view that the theoretical arguments brought forward in this respect fail to do justice to the empirical facts, which strongly support the view that at least some animals are capable of having genuine beliefs of the form 'S beliefs that p', where p stands for a specific state of affairs[16]. They are thus capable of representing others as being in a relation to actual state of affairs (e.g. having a direct line of gaze) and they can themselves represent nonactual state of affairs (e.g. trying out solutions of a problem in their head, as seen in ravens (Heinrich 1995; Heinrich & Bugnyar 2005)). However, there is a lack of evidence that supports the claim that chimpanzees and other animals are capable of *representing others as having a relation to nonactual states of affairs*. For this would be necessary if they were able to attribute belief-states to others (Lurz 2011: 138).

The test for possession of the concept of belief is usually conducted by testing the understanding of *false belief* (Cf. Harman 1978). In these tests, subjects witness how an agent is induced a false belief about the location of an object. If the subject has the concept of a belief it should understand that reality and the other agent's representation of reality are not matching, thus attributing to the agent a false belief. When testing chimpanzees, Call and Tomasello (1999) ran a *cooperative 'bait-and-switch'* test. In this test chimpanzees observed how an experimenter baited one of several containers while a communicator watched the

[16]Among the evidence that language is not necessary for belief-attribution to others is the study by Onishi and Baillargeon (2005), which shows that even 15-months old human infants seem to have a concept of belief and are able to attribute beliefs to others. However, it is unclear how this empirical evidence relates to the animal issue. While humans are in principle capable of language and thus have the cognitive setup to achieve this task, the same is not true for chimpanzees and others. So it is an open possibility that nonverbal humans do understand beliefs of others but nonverbal, nonhuman animals do not.

baiting. The chimpanzees were not able to see which container was baited but they could see that the communicator was facing the scene. After that the communicator marked the baited containers and the chimpanzees learned to associate the mark with the reward that was given to them when they chose the correct container.

In the test trial, the communicator turned around, while the experimenter switched the location of the food. Thus, the communicator's belief about the location of the food no longer matched reality. If chimpanzees understood the concept of belief, they should understand that the communicator no longer had a correct belief and therefore ignore the communicator's mark. However, this did not happen, suggesting that chimpanzees were incapable of attributing genuine beliefs to others.

As a result of the failure of this experiment, Kaminski et al. (2008) conducted a similar experiment with the difference that two chimpanzees were competing for food items (hence calling it the *competitive 'bait-and-switch'* test). They reasoned that the chimpanzees failed in the original test because they had to engage in cooperative action which is not in the natural repertoire of chimpanzees. Their own test uses a rather complex setup of two chimpanzees (one of which was the subject subordinate and the other the dominant) and a sliding table with three containers. While both chimpanzees were watching, one of the containers was baited with a highly desirable reward. After that, only the subject chimpanzee witnessed one of the following two actions. Either an *unknown-lift*, where the baited container was lifted and put back, but the location of the food was not changed. Or an *unknown-shift*, where the baited container was lifted and the location of the reward was changed. In the next step, the dominant was allowed to choose one container, while the subject was unable to see anything. If the chosen container was baited, the competitor received the reward and the

container was put back into place, in order to not indicate to the subject which container was chosen by the dominant. As soon as this happened, the subject was allowed to either choose one of the containers or go for a less desirable item which was not on the table.

Kaminsky et al. suggested that the subject should show a preference for the more desirable reward in the unknown-shift case, but not in the unknown-lift case. If it understands the concept of belief, it should also understand that the competitor's belief about the location is correct in the unknown-lift case, but false in the unknown-shift case. Therefore it should have reason to believe that the competitor would choose correctly in the former, but incorrectly in the latter. However, the chimpanzees did not show this preference, but instead the subject chimpanzees were just as likely to choose the less desirable reward in the unknown-lift case as in the unknown-shift case.

This then is further evidence that chimpanzees might not have a concept of belief and are incapable of attributing this mental state to others. However, we cannot conclude from these experimental findings that chimpanzees do not have the concept of a belief. In fact, this is an open empirical and theoretical question which requires further, well designed experiments which are able to support the mindreading hypothesis for belief-attribution. Three possible designs are suggested by Robert Lurz (2011: Chapter 4), which are very complex and test for the animal's ability to distinguish appearance from reality by testing their success in *deceptive amodal completion tasks*. In these tasks, chimpanzees would be required to distinguish between the appearance of partly occluded objects and their real look. However, there is not enough space to discuss these very complex tasks in more detail, in part because of the fact that they only constitute a theoretical possibility and have not been conducted on chimpanzees.

It has been shown in this section that experiments testing the possession of the concept of belief in chimpanzees have almost always returned negative results. However, this does not mean that they in fact are incapable of attributing beliefs to others. In fact, there are a number of experimental designs to test for this mindreading ability, which however have not been conducted yet. Additionally, it has become evident that testing for belief-attribution in nonverbal animals requires very complex setups. Because the concept of belief comes with specific conditions of understanding, these have to be tested for in the experiments without thus inducing the logical problem or other issues.

5 Conclusion

It has become obvious that finding out if chimpanzees are mindreaders is an intricate quest. There are not only practical issues about how best to design experiments that help discover the cognitive abilities of nonverbal animals. Also, there are theoretical questions of how to decide between different theories of mindreading, how the different mental states are interrelated and bear on the chimpanzee's understanding of individual ones.

I started off with introducing the different terms of *mindreading* and *behaviour-reading* and the logical problem which describes the experimental difficulty of distinguishing the two activities in nonverbal animals. While mindreading involves the ascription of mental states to others, behaviour-reading only relies on recognizing others' behaviour and anticipating their next moves. The logical problem revealed that by conducting experiments which can only rely on observing the behaviour of nonverbal animals, it is difficult to support the mindreading hypothesis, which states that some nonverbal animals sometimes attribute mental states to others. This again cannot be observed directly, and only with cleverly set up experiments could it be possible to indirectly observe the mental attributions.

After that, different theories explaining mindreading have been examined and put into relation with animal mindreading. Theory Theory proposes that mindreading is a *theory-driven activity* based on possessing implicit principles which state the interrelation of different mental states and their conjunction to actions and decision. Simulation Theory on the other hand states that mindreading is a *process-driven activity* which relies on the agent's intrinsic capability to make rational decisions. By using the mental states of others as pretend inputs to the own decision-making mechanism, they are simulated and generate a pretend output

which then again is projected onto the other agent.

I have then argued that whatever theoretical framework we use to explain mindreading in chimpanzees and other animals, the mindreading subject has to be able to generate mental states with metarepresentational content. Only if this is achieved can we attribute mindreading abilities and understanding of mental state concepts. However, I have also stated that further arguments are needed to support the view that having mental states with metarepresentational content is only possible with language, and expressed my doubt based on the experimental evidence that suggests some mindreading capability in nonverbal animals and humans.

Finally, I have discussed some of the many experimental protocols that discuss chimpanzees' understanding of *seeing*, *knowing* and *believing*. Even though the experiments show that chimpanzees are extremely skilled in solving the different problems they were facing in the tasks, the experiments mostly fail to show that mindreading the best explanation of their success. The logical problem strikes again and shows us that there is an urgent need to design experimental protocols which can indirectly observe the mindreading activity by imposing tasks which can only be solved by imputing mental states to others. However, a lot of experiments which involve tasks of that type have returned negative results.

This also indicates the further direction in which research in animal mindreading should go. First, there is a need to solve the logical problem for different mental states and distinguish mindreading-abilities from behaviour reading-abilities. Secondly, it would be interesting to get further empirical evidence that can decide if Theory Theory or Simulation Theory is the best way to explain animal mindreading. My suggestion is that we need a mixture of both in order to explain all different domains of mindreading, for the diverse tasks involve

distinct mental capacities. And this again suggests that different mental processes are needed to successfully master all these tasks. And last but not least, there is a need to investigate further into the animals' abilities of understanding different mental states and how those are interrelated.

Florian Leonhard Wüstholz

University of Sheffield, August 2011

Word count: 13'613

References

[1] Addessi, Elsa & Visalberghi, Elisabetta (2006): 'Rationality in capuchin monkey's feeding behaviour?'. In: Hurley & Nudds (Eds.): *Rational Animals?*. Oxford University Press. 313-328.

[2] Arango-Muñoz, Santiago (2010): 'Two Levels of Metacognition'. *Philosophia* 39. 71-82.

[3] Baldwin, Dare & Baird, Jodie (2001): 'Discerning intentions in dynamic human action'. *Trends in Cognitive Sciences* 5(4). 171-178.

[4] Bekoff, Marc (1977): 'Social Communication in Canids: Evidence for the Evolution of a Stereotyped Mammalian Display'. *Science* 197(4308). 1097-1099.

[5] Bennett, Jonathan (1978): 'Some remarks about concepts'. *Behavioral and Brain Sciences* 1. 557–560.

[6] Bermúdez, Jose (2003): *Thinking Without Words*. Oxford, Oxford University Press.

[7] Bermúdez, Jose (2009): 'Mindreading in the Animal Kingdom'. In: Lurz (Ed.): *The Philosophy of Animal Minds*. Cambridge University Press. 145-164.

[8] Botterill, George (1996): 'Folk Psychology and theoretical status'. In: Carruthers & Smith (Eds.): *Theories of theories of mind*. Cambridge University Press. 105-118.

[9] Call, Josep & Tomasello, Michael (1999): 'A nonverbal theory of mind test. The performance of children and apes'. *Child Development* 70. 381-395.

[10] Call, Josep & Tomasello, Michael (2008): 'Does the chimpanzee have a theory of mind? 30 years later'. *Trends in Cognitive Sciences* 12(5). 187-192.

[11] Carruthers, Peter (1996): *Language, Thought and Consciousness: An Essay in Philosophical Psychology*. Cambridge University Press.

[12] Carruthers, Peter (2004): 'On Being Simple Minded'. *American Philosophical Quarterly* 41(3). 205-220.

[13] Carruthers, Peter & Smith, Peter (1996): *Theories of theories of mind*. Cambridge University Press.

[14] Churchland, Paul (1988): *Matter and Consciousness*. MIT Press.

[15] Clayton, Nicola, Dally, Joanna, Emery, Nathan (2007): 'Social cognition by food-caching corvids. The western scrub-jay as a natural psychologist'. *Philosophical Transactions of the Royal Society* 362(1480). 507-522.

[16] Connor, Richard & Mann, Janet (2006): 'Social cognition in the wild: machiavellian dolphins?'. In: Hurley & Nudds (Eds.): *Rational Animals?*. Oxford University Press. 329-367.

[17] Davidson, Donald (1963): 'Actions, Reasons, and Causes'. In: Donald Davidson: *Essays on Actions and Events*. Oxford, Clarendon Press, 2nd ed. (2001). 3-20.

[18] Davidson, Donald (1975): 'Thought and Talk'. In: Donald Davidson: *Inquiries into Truth and Interpretation*. Oxford, Clarendon Press, 2nd ed. (2001). 155-170.

[19] Davidson, Donald (1982): 'Rational Animals'. In: Donald Davidson: *Subjective, Intersubjective, Objective*. Oxford, Clarendon Press (2001). 95-105.

[20] Dennett, Daniel (1978): 'Beliefs about beliefs'. *Behavioral and Brain Sciences* 1. 568–570.

[21] De Waal, Frans (1986): 'Deception in the Natural Communication of Chimpanzees'. In: Mitchell & Thompson (Eds.): *Deception, perspectives on human and nonhuman deceit*. State University of New York Press, Albany.

[22] Francione, Gary (1995): *Animals, property, and the law*. Temple University Press, Philadelphia.

[23] Gallese, Vittorio (2007): 'Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition'. *Philosophical Transactions of the Royal Society* 362. 659-669.

[24] Gallese, Vittorio & Goldman, Alvin (1998): 'Mirror neurons and the simulation theory of mind-reading'. *Trends in Cognitive Sciences* 2. 493-501.

[25] Glock, Hans-Johann (2009): 'Can Animals Act For Reasons?'. *Inquiry* 52(3). 232-254.

[26] Glock, Hans-Johann (2010): 'Can Animals Judge?'. *Dialectica* 64(1). 11-33.

[27] Goldman, Alvin (1989): 'Interpretation Psychologized'. *Mind & Language* 4. 161-185.

[28] Goldman, Alvin (2006): *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press.

[29] Gómez, Juan & Teixidor, P. (1992): 'Theory of mind in an orangutan: a nonverbal test of false-belief appreciation?'. Paper presented at the *XIV Congress of the International Primatological Society. Strassbourg. August 1992.*.

[30] Hare, Brian, Call, Josep, Agnetta, Bryan & Tomasello, Michael (2000): 'Chimpanzees know what conspecifics do and do not see'. *Animal Behaviour* 59. 771-785.

[31] Hare, Brian, Call, Josep & Tomasello, Michael (2001): 'Do chimpanzees know what conspecifics know?'. *Animal Behaviour* 61. 139-151.

[32] Hare, Brian (2001): 'Can competitive paradigms increase the validity of experiments on primate social cognition?'. *Animal Cognition* 4. 269-280.

[33] Hare, Brian & Wrangham, Richard (2002): 'Integrating Two Evolutionary Models for the Study of Social Cognition'. In: Bekoff et al.: *The Cognitive Animal*. MIT Press, Cambridge, Massachusetts. 363-370.

[34] Harman, Gilbert (1978): 'Studying the chimpanzee's theory of mind'. *Behavioral and Brain Sciences* 1. 576–577.

[35] Hieronymi, Pamela (2009): 'Two kinds of agency'. In: O'Brien & Soteriou (Eds.): *Mental actions and agency*. Oxford: Oxford University Press. 138-162.

[36] Heinrich, Bernd (1995): 'An experimental investigation of insight in common ravens (*Corcus Corax*)'. *The Auk* 4. 994-1003.

[37] Heinrich, Bernd & Bugnyar, Thomas (2005): 'Testing Problem Solving in Ravens: String-Pulling to Reach Food'. *Ethology* 111(10). 962-976.

[38] Holekamp, Kay & Engh, Anne (2002): 'Fields Studies of Social Cognition in Spotted Hyenas'. In: Bekoff et al.: *The Cognitive Animal*. MIT Press, Cambridge, Massachusetts. 371-378.

[39] Hurley, Susan & Nudds, Matthew (2006): *Rational Animals?*. Oxford University Press.

[40] Kacelnik, Alejandro (1984): 'Central Place Foraging in Starlings (Sturnus vulgaris)'. *Journal of Animal Ecology* 53(1). 283-299.

[41] Kaminski, Juliane, Call, Josep & Tomasello, Michael (2008): 'Chimpanzees know what others know, but not what they believe'. *Cognition* 109. 224-234.

[42] Kuroshima, Hika, Fujita, Kazuo, Fuyuki, Akira & Masuda, Tsuyuka (2002): 'Understanding of the relationship between seeing and knowing by tufted capuchin monkeys (*Cebus apella*)'. *Animal Cognition* 5. 41-48.

[43] Kuroshima, Hika, Fujita, Kazuo, Adachi, Ikuma, Iwata, Kana & Fuyuki, Akira (2003): 'A Capuchin monkey (*Cebus apella*) recognizes when people do and do not know the location of food'. *Animal Cognition* 6. 283-291.

[44] Lurz, Robert (2011): *Mindreading Animals*. MIT Press, Cambridge, Massachussets.

[45] Onishi, Kristine & Baillargeon, Renée (2005): 'Do 15-Month-Old Infants Understand False Beliefs?'. *Science* 308. 255-258.

[46] Peacocke, Christopher (2008): *Truly Understood*. Oxford, Oxford University Press.

[47] Peacocke, Christopher (2009): 'Mental Action and Self-Awareness (II): Epistemology'. In: O'Brien & Soteriou (Eds.): *Mental actions and agency*. Oxford: Oxford University Press. 192-214.

[48] Penn, Derek & Povinelli, Daniel (2007): 'On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind''. *Philosophical Transactions of the Royal Society* 362. 731-744.

[49] Penn, Derek, Holyoak, Keith, Povinelli, Daniel (2008): 'Darwin's mistake: Explaining the discontinuity between human and nonhuman minds'. *Behavioral and Brain Sciences* 31. 109-178.

[50] Pepperberg, Irene (2006): 'Intelligence and rationality in parrots'. In: Hurley & Nudds (Eds.): *Rational Animals?*. Oxford University Press. 469-488.

[51] Perner, Josef (1996): 'Simulation as explication of predication-implicit knowledge about the mind: arguments for a simulation-theory mix'. In: Carruthers & Smith: *Theories of theories of mind*. Cambridge University Press. 90-104.

[52] Povinelli, Daniel & Eddy, Timothy (1996): 'Chimpanzees: Joint Visual Attention'. *Psychological Science* 7. 129-135.

[53] Povinelli, Daniel & Vonk, Jennifer (2006): 'We don't need a microscope to explore the chimpanzee's mind'. In: Hurley & Nudds (Eds.): *Rational Animals?*. Oxford University Press. 385-412.

[54] Premack, David & Woodruff, Guy (1978): 'Does the chimpanzee have a theory of mind?'. *Behavioral and Brain Sciences* 1. 515–526.

[55] Premack, David (1988): 'Does the chimpanzee have a theory of mind? Revisited.'. In: Byrne & Whiten (Eds.): *Machiavellian intelligence*. New York, Oxford University Press. 160-179.

[56] Proust, Joëlle (2007): 'Metacognition and metarepresentation: is a self-directed theory of mind a precondition for metacognition?'. *Synthese* 159. 271-295.

[57] Proust, Joëlle (2010): 'Metacognition'. *Philosophy Compass* 5(11). 989-998.

[58] Regan, Tom (2004): *The case for animal rights*. University of California Press, Berkeley.

[59] Singer, Peter (1975): *Animal liberation: A new ethics for our treatment of animals*. New York Review.

[60] Stich, Stephen & Nichols, Shaun (1995): 'Second Thoughts on simulation'. In: Stone & Davies (Eds.): *Folk psychology: The theory of mind debate*. Oxford: Blackwell Publishers. 87-108.

[61] Stone, Tony & Davies, Martin (1996): 'The mental simulation debate: a progress report'. In: Carruthers & Smith: *Theories of theories of mind*. Cambridge University Press. 119-137.

[62] Tanner, Joanne & Byrne, Richard (1993): 'Concealing facial evidence of mood: perspective-taking in a captive gorilla?'. *Primates* 34. 451-457.

[63] Tomasello, Michael (1999): *The Cultural Origins of Human Cognition*. Harvard University Press, Cambridge Massachussets.

[64] Tomasello, Michael & Call, Josep (1997): *Primate Cognition*. Oxford University Press.

[65] Tomasello, Michael, Call, Josep & Hare, Brian (2003): 'Chimpanzees understand psychological states – the question is which ones and to what extent'. *Trends in Cognitive Sciences* 7(4). 153-156.

[66] Tomasello, Michael & Call, Josep (2006): 'Do chimpanzees know what others see—or just what they are looking at?'. In: Hurley & Nudds (Eds.): *Rational Animals?*. Oxford University Press. 371-384.

[67] Visalberghi, Elisabetta & Addessi, Elsa (2003): 'Food for Thought: social learning about food in feeding capuchin monkeys'. In: Fragaszy & Perry (Eds.): *The Biology of Traditions*. Cambridge University Press. 187-212.

[68] Vonk, Jennifer & Povinelli, Daniel (2006): 'Similarity and Difference in the Conceptual Systems of Primates: The Unobservability Hypothesis'. In: Zentall & Wasserman (Eds.): *Comparative Cognition*. Oxford University Press. 363-87.

[69] Whiten, Andrew (1996): 'When does smart behaviour-reading become mind-reading?'. In: Carruthers & Smith: *Theories of theories of mind*. Cambridge University Press. 277-292.