

# **The Role of Intersubjectivity in Animal and Human Cooperation**

Peter Gärdenfors  
Lund University Cognitive Science  
Kungshuset, Lundagård  
S-223 50 Lund, Sweden  
E-Mail: Peter.Gardenfors@lucs.lu.se

## **Abstract**

I argue that analyses of various kinds of cooperation will benefit from an account of the cognitive and communicative functions required for the cooperation. In particular I focus on the role of intersubjectivity (theory of mind), which has not been sufficiently considered in game theory. Intersubjectivity will here be divided into representing the emotions, desires, attention, intentions and beliefs of others. Then I analyze some kinds of cooperation – reciprocal altruism, indirect reciprocity, cooperation on future goals and conventions – with respect to their cognitive and communicative prerequisites. It is argued that uniquely human forms of cooperation depend on advanced forms of intersubjectivity.

## **Keywords**

Cooperation, intersubjectivity, reciprocal altruism, indirect altruism, convention, prospective cognition, future goals, symbolic language

# The Role of Intersubjectivity in Animal and Human Cooperation

## The Role of Cognitive Factors in Game Theory

*Homo sapiens* exhibits more forms of cooperation than any other species. Apparently, the evolution of the human brain has led to some unique forms of cooperation. The changes of the brain have generated new cognitive mechanisms that are exploited in establishing and maintaining human cooperative behaviors.

It has been an enigma for classical game theory that humans and animals do not behave as predicted by the theory. For example, humans cooperate more in ultimatum and prisoners' dilemma games than would be expected from a rationalistic analysis. Several explanations for this mismatch have been suggested, ranging from claims that animals and humans are not sufficiently rational to the position that the rationality presumed in classical game theory is not relevant for evolutionary arguments. The purpose of this paper is to argue that the behavior of different species must be judged in relation to their cognitive capacities.<sup>1</sup> In my opinion, these factors have not been sufficiently considered in game theory. I shall in particular analyze the role of various forms of intersubjectivity (a.k.a. "theory of mind") for cooperation. The focus will be on human cooperation, but I will make some comparisons with cooperation in other animal species.

There are many ways of defining cooperation. A broad definition is that it consists in joint actions that confer mutual benefits.<sup>2</sup> A narrower definition concerns situations in which joint action presents a dilemma so that in the short run an individual will be better off not cooperating (Richerson et al. 2003: 358). In this paper, the emphasis will be on the narrower definition.

Cooperation has been studied from two paradigmatic perspectives. Firstly, in traditional game theory, cooperation has been assumed to take place between individuals of the species *Homo oeconomicus*, who are ideally rational. *Homo oeconomicus* is presumed to have perfect intersubjectivity, i.e., the capacity to judge the beliefs, desires and intentions of other agents. In a Bayesian framework, this amounts to assuming that everybody else is a Bayesian decision maker. In the case of *Homo sapiens*, intersubjectivity is well developed, at least in comparison to other animal species, but it is far from perfect.

Secondly, cooperative games have been studied from an evolutionary perspective. The classical ideas are found in Axelrod and Hamilton (1981) and Maynard Smith (1982). Here, the players are the genes of various animals. A key concept is that of an *evolutionarily stable strategy* (Maynard Smith 1982; Dawkins 1976).<sup>3</sup> The genes are assumed to have absolutely no rationality and no cognitive capacities at all. However, their strategies can slowly adapt, via the mechanisms of natural selection, over a number of generations. In the evolutionary framework, the "theory of mind" of the players is not accounted for, or considered irrelevant.

The differences in cognitive and communicative demands of the two perspectives of cooperative games are seldomly discussed. For example, in Lehmann and Keller's (2006) classification of models of the evolution of cooperation and altruism, only two parameters related to cognition and communication are included: a one period "memory" parameter defined as the probability that "an individual knows the investment into helping of its partner at the previous round" and a "reputation" parameter defined as the probability that "an individual knows the image score of its partner" (Lehmann and Keller 2006: 1367). Here, I submit that the analysis of different kinds of cooperative games would benefit from a richer account of the cognitive and communicative functions required for the cooperation. The cognitive factor that will be considered is, above all, intersubjectivity, but other factors that also are relevant include recognition of individuals, memory capacity and prospective cognition (temporal discounting and mental time travel).

## Intersubjectivity

One form of cognition that, by comparison with other animals, is well developed in humans is intersubjectivity, which in this context means *the sharing and representing of others' mentality*. The term "mentality" is taken here to involve not only beliefs, but all sorts of forms of mental states

including emotions, desires, attentional foci and intentions. In the philosophical debate, intersubjectivity is commonly called having a “theory of mind” (see e.g. Premack and Woodruff 1978; Tomasello 1999; Gärdenfors 2003). I want to avoid this term since it often presumes that one can represent the beliefs of others – something that, on the account presented here, is but one aspect of intersubjectivity.

A prerequisite of an animal (or a human) entertaining any form of intersubjectivity is that it has an *inner world*, i.e., an ability to imagine objects and situations also in their absence.<sup>4</sup> The crucial issue as regards cooperation is whether an individual has any representation of other individuals’ inner worlds.

The question whether an animal or a child exhibits intersubjectivity does not have a simple yes or no answer. To provide a more detailed analysis, intersubjectivity will here be decomposed into five capacities. In Gärdenfors (2001, 2003), I discussed four: representing the emotions of others (empathy), representing the attention of others, representing the intentions of others and representing the beliefs and knowledge of others. To these I now want to add one more: representing the desires of others. From the analysis of these five factors it will be clear that humans exhibit more intersubjectivity than other animals. In particular, we have a well-developed competence for representing the beliefs of others. We also excel in forming joint intentions (Tomasello et al. 2005) and joint beliefs. As I shall argue, these differences are critical for advanced forms of cooperation.

### **Representing the Emotions of Others**

Given this capacity one can, for example, represent that someone else is in pain. This is what is usually meant by *empathy* (Preston and de Waal 2003). The capacity to represent others’ emotions, however, does not mean that one represents what they believe or want.

Bodily expressions of emotions have a communicative purpose. Such expressions are most obvious among social animals. The evolutionary explanation for this seems to be that a capacity for empathy leads to greater solidarity within the group.

### **Representing the Desires of Others**

This involves representing that the other may not like the same things as you do. Representing emotions concern the inner state of an individual, without reference to an external object. A desire is a positive attitude to some external object or event. Representing that somebody has a desire for something therefore involves more than representing that he or she is experiencing an emotion of some kind. Results from child development studies (e.g. Repacholi and Gopnik 1997; Wellman and Liu 2004) suggest that 18 months old children can understand that others have desires other than those they have themselves. However, the presence of this capacity seems to be little investigated in animals.

### **Representing the Attention of Others**

This means that one can represent, for example, that someone else is looking at some object or event. Humans, other primates, and some other mammals are good at following the direction of another individual’s gazes (Emery 2000; Kaminski et al. 2005). Even very young children can understand where other people are looking. A more sophisticated form of attention is to succeed in drawing *joint attention* to an object. If I see that you are looking at an object and you see that I see the same object, we have established joint attention. The achievement of joint attention is a necessary condition of deliberate cooperation. The ability to engage in joint attention and reciprocal behavior vis-à-vis a third element so far has only been shown in humans.

### **Representing the Intentions of Others**

It is important to notice that even though one can interpret someone else’s behavior as goal-directed and can follow gaze to a target, this does not necessarily mean that one represents the other’s intention. It is sufficient to create a representation of the goal of the action. To represent goal-directed action thus requires less than representing intentional behavior.

It follows that sharing intention is not the same thing as sharing an instrumental goal, and intentionality is not similar to teleological action or goal-directedness (cf. Csibra 2003).

There exist cases of goal-directed cooperation in some animal species. For example, Boesch and Boesch-Achermann (2000) make a distinction between four kinds of co-operation based in studies of the hunting behaviour of wild chimpanzees. The most advanced kind is the collaborative hunt in which the hunters perform different complementary roles directed towards the same prey. However, this cooperation can be analysed in terms of the chimpanzees representing the goal of the others and does not require representing their intentions.

There is a stronger sense in which intentionality may be shared that is crucial for complex forms of cooperation, and that concerns *joint intention*.<sup>5</sup> This involves an individual representing the plans of another and coordinating his or her own intentions with the goals of the other. This requires that “I intend that you intend” and that “you intend that I intend”; it also requires both individuals to be aware of these second-order intentions. When one can coordinate roles in working towards a goal, joint intention is achieved. For example, in building a tower of blocks, a child may understand that an adult will hold the tower steady while the child places a new block on it. This capacity occurs after 24 months (cf. Brownell and Carriger 1990). To coordinate complementary roles in working towards a future goal that is not present in the shared context is even harder, because which action to take next will not be evident from the context.

### **Representing the Beliefs and Knowledge of Others**

The most advanced test of the intersubjectivity of humans and animals is to find out whether they represent what others believe or know. Tomasello and Call (2006) review the experimental evidence concerning whether chimpanzees know what others have seen. The most interesting results come from a series of experiments by Hare et al. (2000, 2001) where a subordinate and a dominant ape were placed into rooms on opposite sides of a third room. Each of them could see food placed in the open or behind barriers in the middle room. The problem for the subordinate is that the dominant will take all food she can see. Sometimes the subordinate could see one piece of food on the subordinate’s side of a barrier that the dominant could not see. It was found that the subordinate went for the food that only they could see much more often than the food they could both see.

On the basis of this and other experiments, Tomasello and Call (2006: 375) conclude that the chimpanzees “know not only what others can and cannot see at the moment, but also what others have seen in the immediate past.” This phenomenon can be phrased as “seeing is knowing” and it suggests that the apes have some limited representation of the beliefs and knowledge of others.

It is easier to test whether young children can understand that “seeing is knowing”, since one can communicate with them through language from a fairly early age. The most common method uses so called false belief tests (see e.g. Perner et al. 1987; Gopnik and Astington 1988; Mitchell 1997). In a nonverbal version of the test (Call and Tomasello 1999), children performed as well as in the verbal form, but all apes that were tested failed the test.

Wellman and Liu (2004) argue that children can represent other persons’ “diverse beliefs” before they can judge false beliefs. They found that many 3-year-olds who cannot manage the false belief test still answer the target question opposite from the own-belief question, which suggests that they understand that people have diverse beliefs that influence their actions.

Humans can not only know that someone else knows – that is, have second-order knowledge – but also have higher orders of knowledge and belief, as witness “Of course I care about how you imagined I thought you perceived I wanted you to feel.” This capacity forms the basis of *joint beliefs*, which collectively are often called *common knowledge*.

I shall argue that some forms of uniquely human cooperation that will be presented in the following section presume joint intentions and joint beliefs. Since current evidence concerning the intersubjectivity of other animal species does not indicate that they can achieve these levels, this will explain why only humans have developed the more advanced forms of cooperation.

## **A Classification of Animal and Human Cooperation**

In this section, different kinds of cooperation will be outlined. I will take my point of departure from the forms of cooperation that have been studied within game theory, in particular evolutionary game theory, but also consider some themes from animal behavior. The paradigmatic game in my analysis

will be prisoners' dilemma (PD), mainly in its iterated form, although I will sometimes refer to other games. A reason to focus on games of the PD type is that such games frequently arise in ecological settings. A finding that generates much of the interest in studying these games is that biological players often cooperate considerably more than the total defection that is predicted by the strictly rationalistic analysis of the PD. A problem for a biocognitively oriented analysis is to identify the factors that promote cooperation. I shall argue that, above all, they require varying forms of intersubjectivity.

I shall present the various kinds of cooperation roughly according to increasing cognitive demands. My list is far from exhaustive – I have merely selected some forms of cooperation where the cognitive and communicative components can be identified, at least to some degree. For several of the items on my list it is also possible to make finer divisions concerning the forms of cooperation.

### **Flocking Behavior**

A cognitively low-level form of cooperation is *flocking behavior*. Flocking is found in many species and its function is often to minimize predation. Simulation models of flocking (e.g. Reynolds 1987; Lorek and White 1993) show that the sophisticated behavior of the flock can emerge from the behaviors of single individuals who follow simple rules. In one model, birds flying in a flock just follow two basic rules: (a) try to position yourself as close as possible to the centre of the flock; (b) keep a certain minimal distance from your neighbors. These rules do not presume any cognitive capacities of the individuals beyond those of visual perception (other senses such as echolocation could in principle be used as well).

### **In-group versus Out-group**

One simple way to determine cooperation in an iterated PD or similar game is to divide the other individuals in an *in-group* and an *out-group*. The basic strategy is then to cooperate with everybody in the in-group and defect against everybody in the out-group. Cognitively, this strategy only demands that you can separate members of the in-group from the rest. In nature this is often accomplished via olfaction; for example, bees from a different hive smell differently and are treated with aggression. For a more advanced example, Dunbar (1996) speculates that dialects have evolved to serve as markers of the in-group among hominins. In general, in-group cooperation does not require any intentional communication and cognitively it only presumes a perceptual mechanism that can recognize the members of the in-group. Hence intersubjectivity does not play any role for this form of cooperation.

In all probability, the evolutionary origin of in-group formation is kinship selection. In a kin group, the outcomes of a PD game must be redefined due to the genetic relatedness of the individuals, and in many cases this makes cooperation the only evolutionarily stable strategy. In other words, a game that is a PD on the individual level, may be a perfectly cooperative game when the genes are considered to be the players. Kin groups can then form kernels from which larger cooperative groups can develop (Lindgren 1997: 351). In general, the in-group requires some form of marker, be it just physical proximity, that helps distinguish the in-group from the out-group. However, evolutionary biologists (e.g. Zahavi 1975) stress that such markers should be hard to fake in order to exclude free riders from the in-group.

Another way of generating an in-group is by identifying a common enemy. West et al. (2006) present empirical evidence that when a group playing an iterated PD game is competing with another group, cooperation within the group will increase. Bernhard et al. (2006) present an anthropological study of two groups in Papua New Guinea that support the same conclusion. The downside of this mechanism, as already Hamilton (1975) pointed out, is that what favors cooperation within groups, will also favor the evolution of hostility between groups. A parallel example from the world of apes concerns a flock of chimpanzees in the Gombe forest in Tanzania that became divided into a northern and a southern group. The chimpanzees in the two groups, who had earlier been playing and grooming together, soon started lethal fights against each other (de Waal 2005).

### **Reciprocal Altruism**

In an iterated PD, one player can *retaliate* against another's defection. Trivers (1971) argues that this possibility can make cooperation more attractive and lead to what he calls *reciprocal altruism*. This form of cooperation can be formulated as a slogan: "You scratch my back and I'll scratch yours." In

their seminal paper, Axelrod and Hamilton (1981) showed by computer simulations that cooperation can evolve in iterated PD situations. A Nash equilibrium strategy such as the well-known Tit-for-Tat can lead to reciprocal cooperative behavior among individuals that encounter each other frequently. The cognitive factors required for this strategy are, at least, the ability to *recognize* the individual you interact with and a *memory* of previous outcomes. *Trusting* another individual may be an emotional correlate that strengthens reciprocity. Following the lead of Frank (1988), such an emotion may have been selected for and thus become genetically grounded in a species that engages in reciprocal altruism. On this point, van Schaik and Kappelan (2006: 17) write:

[W]e must expect the presence of derived psychological mechanisms that provide the proximate control mechanisms for these cooperative tendencies [...]. [T]he evidence suggests a critical role for unique emotions as the main mechanisms. Extreme vigilance toward cheaters, a sense of gratitude upon receiving support, a sense of guilt upon being detected at free riding, willingness to engage in donation of help and zeal to dole out altruistic punishment at free riders – all these are examples of mechanisms and the underlying emotions that are less developed in even our closest relatives.

And Trivers (2006: 76) writes about his 1971 paper on reciprocity that its most important implication was that “it laid the foundation for understanding how a sense of justice evolved.”

Field studies of fish, vampire bats (Wilkinson 1984) and primates, among other species, have reported the presence of reciprocal altruism. Still, the evidence for reciprocal altruism in non-human species is debated and some laboratory experiments seem to speak against it (Stephens et al. 2002, Hauser et al. 2003). In line with this, Stephens and Hauser (2004) argue that the cognitive demands of reciprocal altruism have been underestimated (see also Hammerstein 2003). They claim that reciprocal altruism will be evolutionarily stable in a species only if (a) the temporal discounting of future rewards is not too steep; (b) they have sufficient discrimination of the value of the rewards to judge that what an altruist receives back is comparable to what it has given itself; and (c) there is memory capacity to keep track of interactions with several individuals.

Studies of temporal discounting reveal that the rates differ drastically between different species. Humans have by far, the lowest discount rate, which is a prerequisite for the prospective planning that will be presented below. An interesting problem, that should receive more attention within evolutionary game theory, is what factors (ecological, cognitive, neurological, etc) explain the discounting rate of a particular species.

The cognitive demands Stephens and Hauser (2004) present explain, according to them, why reciprocal altruism is difficult to establish in other species than *Homo sapiens*. However, against their argument, one should present the three different types of reciprocity proposed by de Waal and Bosnan (2006: 103-104): (1) Individuals reciprocate on the basis of symmetrical features of dyadic relationships (kinship, similarities in age or sex). De Waal and Bosnan say that this is the cognitively least demanding form because “it requires no scorekeeping since reciprocation is based on pre-existing features of the relationship. [...] All that is required is an aversion to major, lasting imbalance in incoming and outgoing benefits.” (2) Attitudinal reciprocity where individual willingness to cooperate depends on the attitude the partner has recently shown. Cognitively, for this form of reciprocity, the “involvement of memory and scorekeeping seems rather minimal, as the critical variable is general social disposition rather than specific costs and benefits of exchanged behavior” (de Waal and Bosnan 2006: 104). (3) Calculated reciprocity, in which “individuals reciprocate on a behavioral one-on-one basis with a significant time interval.” This is the cognitively most demanding form since it “requires memory of previous events, some degree of score-keeping, partner-specific contingency between favors given and received” (de Waal and Bosnan 2006: 104). The reciprocity discussed by Stevens and Hauser (2004) is most similar to this third form and it is therefore not surprising that it would be the most difficult to achieve. When analyzing to what extent different species of animals exhibit reciprocal altruism, the distinctions made by de Waal and Bosnan (2006) must be kept in mind.

Cognitively, the attitudinal reciprocity defined by de Waal and Bosnan (2006) involve rather minimal capacities: Individuals must be recognized over time and the “attitude” of the partner must be remembered. Representing the attitude of the partner involves some form of intersubjectivity, at least representing the desires of the other.

Reacting to the desires of the partner is a general mechanism for improving cooperation beyond kin altruism. Charness and Dufvenberg (2006: 1580) introduce “guilt aversion” as a way to explain why individuals behave more cooperatively than predicted by game theory. According to them, “decision makers experience guilt if they believe they let others down.” This means that by considering the desires of your partners and having an emotional mechanism that makes you avoid letting them down (Frank 1988), the payoff matrix of, for example, a PD game will change and the player ends up being much cooperative than what is predicted in traditional game theory. The situation is analogous to kin selection where a PD is also turned into a cooperative game when the inclusive fitness of the kin is considered.

Charness and Dufvenberg (ibid.) claim that guilt aversion involves beliefs about the beliefs of others. However, this claim about the intersubjectivity of the decision maker seems to be too strong for the mechanism they are investigating. To wit, it is sufficient that the decision maker reacts to the desires of the other to realize that the other is let down. In spite of this reservation, the notion of guilt aversion seems central for analyzing many forms of human cooperative behavior and its merits further theoretical and experimental investigations.

An interesting question is to what extent guilt aversion can be shown to exist in other species. The so-called ultimatum game could be a benchmark for testing guilt aversion and thereby indirectly the capacity to represent the desires of others. In this game a proposer is offered a sum of money and is free to divide it in any proportion with a responder. The crucial feature is that the responder can accept or reject the offer. If she accepts, both players receive the proposed division, but if she rejects, both receive nothing. With human subjects the proposer typically offer the responder 40% to 50% of the amount given and responders in general reject offers under 20%. These results indicate that they want the proposers to take the desires of the responder into account and they punish a proposer who is too unfair.

By contrast, Jensen et al. (2007) showed that when chimpanzee were used as subjects and proposers was given a choice between an 8-2 and a 5-5 offer, 75% of the chimpanzee choose the 8-2 offer and it was rejected by only 5% of the chimpanzee responders. Even more drastically, when the chimpanzee proposers could choose between a 10-0 and an 8-2 offer, 46% choose the 10-0 offer (and it was rejected by only 44% of the chimpanzee responders). This experiment is at least an indication that chimpanzees do not exhibit the same guilt aversion as humans do (which in turn could mean that they have a limited capacity to represent the desires of others).

### **Cooperation on Future Goals**

Cooperation often occurs over an extended time span and then involves a form of planning. It is useful to make a distinction between *immediate* planning that involves current goals and *prospective* planning that concerns future non-present goals (Gulz 1991; Suddendorf and Corballis 1997; Gärdenfors 2003; Osvath and Gärdenfors 2005). Humans can predict that they will be hungry tomorrow and save some food, and we can imagine that the winter will be cold, so we start building a shelter already in the summer. The cognition of other animals concerns here and now, while humans are mentally both here and in the future. The central cognitive aspect of prospective planning is the capacity to *represent your future drives*.

The plans of apes and other animals depend in most cases on their current drive states: They plan because they are hungry or thirsty, tired or frightened. Bischof (1978) and Bischof-Köhler (1985) hypothesize that animals other than humans cannot anticipate future needs or drive states (see also Gulz 1991). Their cognition is therefore bound to their present motivational state (see also Suddendorf and Corballis 1997). This hypothesis, which is called the Bischof-Köhler hypothesis, has been supported by the previous evidence concerning planning in non-human animals. However, recent results indicate that, under experimental conditions, great apes and corvids have a rudimentary capacity to act in order to fulfill future needs (Mulcahy and Call 2006; Correia, Dickinson and Clayton 2007; Osvath and Osvath in press). They are not known, however, to use this capacity in the wild, which implies that prospective cognition plays a minor role, or no role at all, in the lives of great apes. More studies are required to determine the extent to which it can be said that the apes are on the cognitive brink of prospective planning. In spite of these surprising results, the human capacity for prospective planning seems to be far more advanced, in particular when it comes to cooperation.

For many forms of cooperation among animals, it seems that mental representations are not needed. If the common goal is *present* in the actual environment – for example as food to be eaten or an enemy to be fought – the collaborators need not have a joint representation before acting. If, on the other hand, the goal is distant in time or space, a *joint* representation of it must be produced before cooperative action can be taken. For example, building a common dwelling requires coordinated planning of how to obtain the building material and advanced collaboration in the construction. Among other things, this depends on forming joint intentions, which is an advanced form of intersubjectivity presumably unique to humans.

Another problem concerning collaboration on a future goal is that the *value* of the goal cannot be determined from the given environment. (Contrast a goal that is already present on the scene.) The value of the goal has to be estimated by each individual or communicated between individuals.

The hominin life style opened up many new forms of cooperation on future goals. Thus in relation to *Homo habilis* Plummer (2004: 139) mentions "cooperative behavior including group defense, diurnal foraging ... with both hunting and scavenging being practiced as the opportunities arose, and the ability (using stone tools) to rapidly dismember large carcasses so as to minimize time spent at death sites."

In general terms, cooperation on future goals requires *the inner worlds of the individuals to be coordinated*. For instance, the chief of a tribe can try to convince members of the tribe that they should cooperate in digging a common well that everybody will benefit from or in building a defensive wall that will improve the security of everybody. The goal requires efforts to be made by the members of the community, but it will, ideally, bring a net benefit to all involved.

The presence of mutual representations of a future goal will change a situation, which would be a PD without the presence of such representations, into a game where the cooperative strategy is the equilibrium solution. For example, if we live in an arid area, each individual (or family) will benefit by digging a well. However, if my neighbor digs a well, I may defect and take my water from his well, instead of digging my own. But if nobody digs a well, we are all worse off than if everybody does it. This is a typical example of a PD.

Now if somebody communicates the idea that we should cooperate in digging a communal well, then such a well, by being deeper, would yield much more water than all the individual wells taken together. Once such cooperation is established, the PD situation may disappear or at least be ameliorated, since everybody will benefit more from achieving the common goal. In game theoretical terms, digging a communal well will be a new equilibrium strategy. This example shows how the capacity of sharing future goals in a group can strongly enhance the value of co-operative strategies within the group. The upshot is that strategies based on future goals may introduce new equilibria that are beneficial for all participants.

The cognitive requirements for cooperating about future goals involve crucially the capacity to represent your drives or wishes at a future time, that is, prospective planning. Even if there is presently plenty of water you will foresee the dry season and that you will be thirsty then. This representation, and not your current drive state, forms the driving force behind a wish to cooperate in digging a well.

What are the communicative requirements then? Symbolic language is the primary tool by which agents can make their inner representation known to each other. In previous work (Brinck and Gärdenfors 2003; Gärdenfors 2003, 2004, in press; Osvath and Gärdenfors 2005), it has been proposed that there is a strong connection between the evolution of prospective cognition and the evolution of symbolic communication. In brief, the argument is that symbolic language makes it possible to cooperate about future goals in an efficient way. It is not necessary that the symbolic communication involves any syntactic structures – protolanguage (Bickerton 1990; Dessalles 2007) will do perfectly well (Gärdenfors in press).

Planning for future goals requires that the present goals can be suppressed, which hinges on the executive functions of the frontal brain lobes (Hughes, Russell and Robbins 1994). The future goals and the means to reach them are picked out and externally shared through symbolic communication. This kind of sharing gives humans an enormous advantage concerning cooperation in comparison to other species. I submit that there has been a co-evolution of cooperation on future goals and symbolic communication. However, without the presence of prospective cognition, the selective pressures that resulted in symbolic communication would not have emerged.

## Indirect Reciprocity

In a social species, an individual often faces a decision whether to cooperate or not. In the analyses of prisoners' dilemmas and similar games in standard game theory, it is taken for granted who the potential collaborators are. In practice, however, an important question is: How do you know *who* to cooperate with? (see e.g. Dessalles 2007: 360).

As we have noted, reciprocal altruism is found in some animal species and it is common among humans. However, in humans one often finds more extreme forms of altruism: "I help you and somebody else will help me." This form of cooperation has been called *indirect reciprocity* and it seems to be more or less unique to humans. A possible exception comes from a study by Warneken and Tomasello (2006). They let three human-raised juvenile chimpanzees watch when a human attempted, but failed to achieve different goals. The chimpanzees sometimes helped the humans, but mainly in situations where the human reached for but failed to grasp objects. However, it is not clear that these cases of helping are indirect reciprocity since the chimpanzees may expect reciprocation on part of the human. In contrast, Warneken and Tomasello (2006) also showed that 18-month-old human infants in a large majority of the cases helped an adult experimenter in various scenarios. Warneken et al. (2007) demonstrated in a follow-up experiment that the children helped irrespective of being rewarded for it. Thus the evidence that their behavior is true indirect reciprocity is quite strong. Summing up the experiments, Warneken (forthcoming) concludes that "both humans and chimpanzees act *for* others, but only humans act collaboratively *with* others." The difference seems to be that the children, but not the apes, excel in representing the intentions of others.

Nowak and Sigmund (2005) show that, under certain conditions, indirect reciprocity can function as an evolutionarily stable strategy. However, as we shall see, their explanation depends on strong assumptions about the communication of the interactors. In their definition of indirect reciprocity, any two players are supposed to interact at most once with each other. This is an idealizing assumption, but it has the effect that the recipient of a defect (cheat) act in a PD cannot retaliate. Thus all strategies that have been considered for the iterated PD are excluded. So the question is under what conditions indirect reciprocity can evolve as an evolutionarily stable strategy.

The key concept in Nowak and Sigmund's evolutionary model is that of individual *reputation* (see also e.g. Leimar and Hammerstein 2001; Panchanathan and Boyd 2004; Fehr 2004). Typically, the reputation of individual *x* is built up when some members of the society observe *x*'s behavior towards third parties and convey what they have observed to other members of the society. The level of *x*'s reputation can then be used by other individuals when they are deciding whether to help *x* in a situation of need. It should be noted that the reputation is not something that is visible to all others, unlike status markers such as a raised tail among wolves. Rather each individual must keep a private account of the reputation of all others. Semmann et al. (2005) have demonstrated experimentally that building a reputation through cooperation is valuable for future social interactions, not only within but also *outside* one's own social group.

In these interactions it is important to distinguish between *justified* and *unjustified* defections. If a potential receiver has defected repeatedly in the past, the presumptive donor can be justified in defecting, as a way of punishing the receiver. However, the donor then runs the risk that his own reputation will drop. To prevent this, the donor should communicate that the reason for defecting is that the receiver has a bad reputation. This will make it possible for other individuals to represent the intentions of the donor.

Nowak and Sigmund (2005) stipulate that a strategy is first order if the assessment of an individual *x* in the group depends only on *x*'s actions. More sophisticated strategies distinguish between justified and unjustified defections. A strategy is categorized as second order if it depends on the reputation of the receiver and third order if it additionally depends on the reputation of the donor. Nowak and Sigmund then show that only eight of the possible strategies are evolutionarily stable and that all these strategies depend on the distinction between justified and unjustified defection.

Nowak and Sigmund also note, without spelling out any details, that indirect reciprocity seems to require some form of "theory of mind." An individual watching a second individual (donor) helping or not helping a third (a receiver) in need must judge that the donor does something, respectively, "good" or "bad" to the receiver. The intersubjectivity required for such comparisons includes empathy and representing the desires of the receiver, but also, more importantly, the ability to represent the *intention* of the donor.

The success of indirect reciprocity thus heavily depends on the mechanism of reputation. This means that indirect reciprocity, in the model presented by Novak and Sigmund (2005), requires several cognitive and communicative mechanisms. The cognitive requirements are (at least): (a) recognizing the relevant individuals over time, (b) remembering and updating the reputation scores of these individuals and (c) representing the desires of others to judge whether a particular donor action is “good” or “bad”. Apart from this, the communication system of the individuals in the group must be able to (d) identify individuals in their absence, e.g. by names, (e) expressions to the effect that “x was good to y” and “y was bad to x”. There seems to be no animal communication system that can handle these forms. Thus it is no surprise that *Homo sapiens* is the only species exhibiting indirect reciprocity.<sup>6</sup> The communication system need not be a language with full syntax, but a form of *protolanguage* along the lines discussed by Bickerton (1990) is sufficient. The communication can be symbolic, but it is possible that a system based on *miming* (Donald 1991; Zlatev, Persson and Gärdenfors 2005) would be sufficient.

The trust that is built up in reciprocal altruism is *dyadic*, that is, a relation between two individuals. In contrast, reputation is an emergent *social* notion involving everybody in the group. On this point, Nowak and Sigmund (2005: 1296) write: “Indirect reciprocity is situated somewhere between direct reciprocity and public goods. On the one hand it is a game between two players, but it has to be played within a larger group.”

As noted above, an individual in a donor situation should signal that the potential recipient has a “bad reputation” before defecting, in order not to lose in reputation because of misinterpretation from the onlookers. This is another form of communication that presumes (f) expressions of the type “y has bad reputation”.

Of course, the need for communication is dependent on the size of the group facing the PD situations. In a tightly connected group where everybody sees everybody else most of the time, there is no need for a reputation mechanism. One can compare with how the ranking within such a group is established. If I observe that x dominates y and I know that y dominates me, there is no need for (aggressive) interaction or communication to establish that x dominates me. However, in large and loosely connected groups, these mechanisms are not sufficient, but some form of communication about non-present individuals is required. This form of “displacement” of a message is one of the criteria that Hockett (1960) uses to identify symbolic language.

There may be still other mechanisms that influence the reputation of an individual. In many situations people are not only willing to cooperate but ready, also, to punish free riders, i.e. individuals who benefit from the cooperation of others without cooperating themselves. The punishing behavior is difficult to explain, both because the punishment is costly and because the cooperative non-punishers benefit (Fehr and Gächter 2002). Thus punishment should decrease in frequency.<sup>7</sup> Barclay (2005) presents some evidence that the reputation of a punisher rises, and that people are more willing to cooperate with punishers. This suggests that punishing behavior, in combination with the mechanisms presented above, is rewarded in the long run (see also Sigmund, Hauert and Nowak 2001) and can stabilize cooperation in iterated, prisoners’ dilemma-style interactions (Lindgren 1997). Unlike human conduct, natural animal behavior appears not to involve altruistic punishment (van Schaik and Kappeler 2006), although de Waal and Brosnan (2006) report an experimentally induced refusal to cooperate that is costly for the non-cooperative individual.

### **Commitments, Contracts and Conventions**

Commitments and contracts are special cases of cooperation on the future. They rely on an advanced form of intersubjectivity that allows for joint beliefs. In general, joint beliefs form the basis for much of human culture. They make many new forms of cooperation possible. For example, to promise something only means that you intend to do it. On the other hand, when you *commit* yourself to a second person to do an action, you intend to perform the action in the future, the other person wants you to do it and intends to check that you do it, and there is joint belief concerning these intentions and desires (Dunin-Keplicz and Verbrugge 2001). Unlike promises, commitments can thus not arise unless the agents achieve joint beliefs and have prospective cognition. It has been argued (Tomasello 1999; Gärdenfors 2003, 2007) that the capacity for joint beliefs is only found in humans.

For similar reasons, *contracts* cannot be established without joint beliefs and prospective cognition. For example, Deacon (1997) argues that marriage is the first example of cooperation where

symbolic communication is needed. Because marriage is a form of contract, it is an advanced form of cooperation on the future. If the theory presented in this paper is correct, it could, however, not be first form of cooperation for which symbols are needed.

In human societies, many forms of cooperation are based on *conventions*. Conventions also presume *common beliefs* (often called common knowledge). For example, if two cars meet on a gravel road in Australia, then both drivers know that this co-ordination problem has been solved by driving on the left hand side numerous times before, both know that both know this, both know that both know that both know this, etc, and they then both drive on their left without any hesitation (Lewis, 1969). Many conventions are established without explicit communication, but, of course, communication makes the presence of a convention clearer.

Conventions function as virtual governors in a society. Successful conventions create equilibrium points, which, once established, tend to be stable. The convention of driving on the left hand side of the road will force me to “get into step” and drive on the left. The same applies to *language* itself: A new member of a society will have to adjust to the language adopted by the community. The meaning of the linguistic utterances emerges from the individuals' meanings.<sup>8</sup> There are, of course, an infinite number of possible “equilibrium” languages in a society, but the point is that once a language has developed it will have strong explanatory effects regarding the behavior of the individuals in the society.

When linguistic communication develops individuals will come to share a relatively stable system of *meanings*, i.e. components in their inner worlds that communicators can exchange between each other. In this way, language fosters a *common structure* of the inner worlds of the individuals in a society (see Gärdenfors and Warglien 2006).

### **The Cooperation of Homo Oeconomicus**

*Homo oeconomicus*, rational man, is assumed to have perfect reasoning powers. He is logically omniscient, a perfect Bayesian when it comes to probability judgments and a devoted utility maximizer. He also has a complete theory of mind in the sense that he can put himself totally in the shoes of his opponents (and vice versa) and see them as perfect Bayesians when reasoning about what would be the possible equilibrium strategies in the game he is facing. When he communicates, his messages have a logically crisp meaning. He cooperates, if and only if this maximizes his utility according to the description of the game.

The problem with *Homo oeconomicus* is that he seldom faces the complexities of reality, where the game to be played is not well defined and where its outcomes and the strategies available may fluctuate from minute to minute, where the utilities of the outcomes are uncertain, where reasoning time and memory resources are limited, where communication is full of vagueness and misunderstandings and where his opponents have all kinds of idiosyncracies and cognitive shortcomings, partly depending on their ecological conditions. As Trivers (2006: 79) writes: “I know of no species, humans included, that leaves any part of its biology at the laboratory door; not prior expectations, nor natural proclivities, nor ongoing physiology, nor arms or legs, nor whatever.”

In game theory, it is almost exclusively strategies within a fixed game that have been studied. In nature, strategies that can be used in a variety of different but related games are more realistic. How should the rational man act when meeting an opponent in flesh and blood, with all kinds of biological constraints, when he knows that his opponent is not fully rational, but does not know what the exact limitations of the opponent are? Game theory without biology is esoteric; with biology it is as complicated as life itself (see e.g. Eriksson and Lindgren 2002).

## **Synthesis**

The somewhat eclectic list of different forms of cooperation and their cognitive and communicative prerequisites that have been presented above is summarized in table 1. The “delayed gratification” mentioned as a requirement for cooperation on future goals refers to the capacity to abstain from current gains in order to obtain a larger one later.

\*\*\* Insert Table 1 about here \*\*\*

The different kinds of cooperation presented here are ordered according to increasing cognitive demands. They do not exhaust the field of possibilities and there is most certainly a need to make finer divisions. Nevertheless, the table clearly shows that the cognitive and communicative constraints are important factors when analyzing evolutionary aspects of cooperation. In particular, the table exhibits that the more advanced forms of cooperation also require advanced forms of intersubjectivity.

As far as is known, it is only humans who clearly exhibit calculated reciprocal altruism, indirect reciprocity, cooperation on future goals and cooperation based on common beliefs (conventions and contracts). This correlates well with the levels of intersubjectivity that are required for these forms of cooperation.

The analysis, or a more detailed version of it, can be used to make predictions about the possible forms of cooperation that can be found in different animal species. For example, if a monkey species can be shown to recognize members of its own troop and to remember and evaluate the results of some previous encounters with individual members, but if two monkeys cannot communicate about a third individual, then the prediction is that reciprocal altruism between the members of the species might be expected, but not indirect reciprocity. In this way, a causal link from cognitive capacities to cooperative behaviors can be established. Conversely, but perhaps less reliably, if a species does not exhibit a particular form of cooperative behavior, then it can be expected that they do not have the required cognitive or communicative capacities.

## Acknowledgements

This article builds on some material from Gärdenfors (2007) and (in press). I would like to thank members of the seminars in Cognitive Science and Philosophy at Lund University, in particular Ingar Brinck and Mathias Osvath, and of the project group on Games, Action and Social Software at the Netherlands Institute for Advanced Studies for very helpful comments. I also want to thank the Swedish Research Council for general support that made this research possible.

## References

- Axelrod R, Hamilton WD (1981) The evolution of cooperation. *Science* 214: 1390-1396.
- Barclay P (2005) Reputational benefits for altruistic punishment. *Evolution and Human Behavior* 27: 325-344.
- Bernhard H, Fischbacher U, Fehr E (2006) Parochial altruism in humans. *Nature* 442: 912-915.
- Bickerton D (1990) *Language and Species*. Chicago, IL: The University of Chicago Press.
- Bischof N (1978) On the phylogeny of human morality. In: *Morality as a Biological Phenomenon* (Stent G, ed), 53-74. Berlin: Abakon.
- Bischof-Köhler D (1985) Zur Phylogenese menschlicher Motivation. In: *Emotion und Reflexivität* (Eckensberger LH, Lantermann ED, eds), 3-47. Vienna: Urban Schwarzenberg.
- Boesch C, Boesch-Achermann H (2000) *The Chimpanzees of the Tai Forest: Behavioural Ecology and Evolution*. Oxford: Oxford University Press.
- Brinck I, Gärdenfors P (2003) Co-operation and communication in apes and humans. *Mind and Language* 18, 484-501.
- Brownell CA, Carriger MS (1990) Changes in cooperation and self-other differentiation during the second year of life. *Child Development* 61: 1164-1174.
- Bowles S, Gintis H (2003) The origins of human cooperation. In: *The Genetic and Cultural Origins of Cooperation* (Hammerstein P, ed), 429-443. Cambridge, MA: MIT Press.
- Call J, Tomasello M (1999) A nonverbal false belief task The performance of children and great apes. *Child Development*, 70(2): 381-395.
- Charness G, Dufwenberg M (2006) Promises and partnership. *Econometrica* 74: 1579-1601.
- Correia S P C, Dickinson A, Clayton N S (2007) Western scrub-jays anticipate future needs independently of their current motivational state. *Current Biology* 17: 856-61.
- Csibra G (2003) Teleological and referential understanding of action in infancy. *Philosophical Transactions of the Royal Society in London* 358: 447-458.

- Dawkins R (1976) *The Selfish Gene*. Oxford: Oxford University Press.
- Deacon T W (1997) *The Symbolic Species*. London: Penguin Books.
- Dessalles J-L (2007) *Why We Talk*. Oxford: Oxford University Press.
- De Waal F B M, Brosnan S F (2006) Simple and complex reciprocity in primates. In: *Cooperation in Primates and Humans* (Kappeler PM, Van Schaik CP, eds), 85-105. Berlin: Springer.
- Donald M (1991) *Origins of the Modern Mind*. Cambridge, MA: Harvard University Press.
- Dunbar R (1996) *Grooming, Gossip and the Evolution of Language*, Cambridge, MA: Harvard University Press.
- Dunin-Kepliz B, Verbrugge R (2001) A tuning machine for cooperative problem solving. *Fundamenta Informatica* 21: 1001-1025.
- Eriksson A, Lindgren K (2002) Cooperation in an unpredictable environment. In: *Proceedings of Artificial Life VIII* (Standish RK, Abbass MA, Bedau HA, eds), 394-399. Cambridge, MA: MIT Press.
- Emery N J (2000) The eyes have it: the neuroethology, function and evolution of social gaz., *Neuroscience and Biobehavioral Reviews* 24: 581-604.
- Fehr E (2004) Don't lose your reputation. *Nature* 432: 449-450.
- Fehr E, Gächter S (2002) Altruistic punishment in humans. *Nature* 415: 137 - 140.
- Frank R (1988) *Passions within Reason: The Strategic Role of the Emotions*. New York, NY: Norton.
- Gärdenfors P (1993) The emergence of meaning. *Linguistics and Philosophy* 16: 285-309.
- Gärdenfors P (2001) Slicing the theory of mind. *Danish Yearbook for Philosophy* 36: 7-34.
- Gärdenfors P (2003) *How Homo Became Sapiens*. Oxford: Oxford University Press.
- Gärdenfors P (2004) Cooperation and the evolution of symbolic communication. In: *The Evolution of Communication Systems* (Oller K, Griebel U, eds), 237-256. Cambridge, MA: MIT Press.
- Gärdenfors P (2007) *The cognitive and communicative demands of cooperation*. In: *Hommage à Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz*. <http://www.fil.lu.se/hommageawlodek/>
- Gärdenfors P (2008) Evolutionary and developmental aspect of intersubjectivity. In: *Consciousness Transitions – Phylogenetic, Ontogenetic and Physiological Aspects* (Liljenström H, Århem P, eds), 281-385. Amsterdam: Elsevier.
- Gärdenfors P (in press) The role of cooperation in the evolution of protolanguage and language. In: *The Evolution of Mind and Culture* (Hatfield G, ed) Philadelphia, PA: Pennsylvania State Museum.
- Gärdenfors P, Warglien M (2006) Cooperation, conceptual spaces and the evolution of semantics. In: *EELC 2006* (Vogt P et al eds), 16 – 30. LNAI 4211, Berlin: Springer Verlag.
- Gopnik A, Astington J W (1988) Children's understanding of representational change, and its relation to the understanding of false belief and the appearance-reality distinction, *Child Development* 59: 26-37.
- Gulz A (1991) *The Planning of Action as a Cognitive and Biological Phenomenon* Lund: Lund University Cognitive Studies 2.
- Hamilton WD (1975) Innate social aptitudes of man, an approach from evolutionary genetics. In: *Biosocial Anthropology* (Fox R, ed), 133-157. London: Malaby Press.
- Hammerstein P (2003) Why is reciprocity so rare in social animals? A protestant appeal. In: *The Genetic and Cultural Origins of Cooperation* (Hammerstein P, ed), 83-93. Cambridge, MA: MIT Press.
- Hare B, Call J, Agnetta B, Tomasello M (2000) Chimpanzees know what conspecifics do and do not see, *Animal Behaviour* 59: 771-85.
- Hare B, Call J, Tomasello M (2001) Do chimpanzees know what conspecifics know?, *Animal Behaviour* 61: 139-51.
- Hauser MD, Chen MK, Chen F, Chuang E (2003) Give unto others: genetically unrelated cotton-top tamarin monkeys preferentially give food to those who altruistically give food back. *Proceedings of the Royal Society of London B* 270: 2363-2370.
- Hockett CF (1960) The origin of speech. *Scientific American* 203(3): 88-96.
- Hughes C, Russel J, Robbins TW (1994) Evidence for executive dysfunction in autism. *Neuropsychologia* 32: 477-492.
- Jensen K, Call J, Tomasello M (2007) Chimpanzees are rational maximizers in an ultimatum game. *Science* 318: 107-109.

Kaminski J, Riedel J, Call J, Tomasello M (2005) Domestic goats, *Capra hircus*, follow gaze direction and use social cues in an object task. *Animal Behaviour* 69: 11–18.

Lehmann L, Keller L (2006) The evolution of cooperation, altruism - A general framework and a classification of models. *Journal of Evolutionary Biology* 19: 1365-1376.

Leimar O, Hammerstein P (2001) Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society London B* 268: 745–53.

Lewis D (1969) *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Lindgren K (1997) Evolutionary dynamics in game-theoretic models. In: *The Economy as an Evolving Complex System II* (Arthur B, Durlauf S, Lane D, eds), 337-367. Reading, MA: Addison-Wesley.

Lorek H, White M (1993) Parallel bird flocking simulation, available online via <http://citeseer.ist.psu.edu/lorek93parallel.html>.

Maynard Smith J (1982) *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Mitchell P (1997) *Introduction to Theory of Mind: Children, Autism and Apes*. London: Arnold.

Mulcahy N J, Call J (2006) Apes save tools for future use, *Science* 312: 1038-1040.

Neill DB (2001) Optimality under noise: higher memory strategies for the alternating prisoner's dilemma. *Journal of Theoretical Biology* 211: 159-180.

Nowak MA, Sigmund K (2005) Evolution of indirect reciprocity. *Nature* 437: 1291-1298.

Osvath M, Gärdenfors P (2005) Oldowan culture and the evolution of anticipatory cognition. *Lund: Lund University Cognitive Studies* 122.

Osvath M, Osvath H (in press) Ape forethought: Self-control and envisioning in the face of future events. *Animal Cognition*.

Panchanathan K, Boyd R (2004) Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* 432: 499-502.

Perner J, Leekam S, Wimmer H (1987) Three-year-old's difficulty with false belief: the case for a conceptual deficit. *British Journal of Developmental Psychology* 5: 125-37.

Plummer T (2004) Flaked stones and old bones: Biological and cultural evolution at the dawn of the dawn of technology. *Yearbook of Physical Anthropology* 47: 118–64.

Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* 4: 515–26.

Preston SD, de Waal F (2003) Empathy: Its ultimate and proximal bases. *Behavioral and Brain Sciences* 25: 1-72.

Reynolds CW (1987) Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics* 21(4): 25-34.

Richerson PJ, Boyd RT, Heinrich J (2003) Cultural evolution of human cooperation. In: *The Genetic and Cultural Origins of Cooperation* (Hammerstein P, ed), 357-388. Cambridge, MA: MIT Press.

Semmann D, Krambeck H-J, Milinski M (2005) Reputation is valuable within and outside one's own social group. *Behavioral Ecology and Sociobiology* 57: 611-616.

Sigmund K, Hauert C, Nowak M (2001) Reward and punishment. *PNAS* 98: 10757-10762.

Stephens DW, McLinn, Stevens JR (2002) Discounting and reciprocity in an iterated prisoner's dilemma. *Science* 298: 2216-2218.

Sterelny K (2003) *Thought in a Hostile World*. Oxford: Blackwell Publishing.

Stevens JR, Hauser M (2004) Why be nice? Psychological constraints on the evolution of cooperation. *Trends in Cognitive Science* 8: 60-65.

Suddendorf T, Corballis MC (1997) Mental time travel and the evolution of human mind. *Genetic, Social and General Psychology Monographs* 123: 133-167.

Suddendorf T, Whiten A (2001) Mental evolution an development: Evidence for secondary representation in children, great apes, and other animals. *Psychological Bulletin* 127(5): 629-650.

Tomasello M (1999) *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

Tomasello M, Call J (2006) Do chimpanzees know what others see – or only what they are looking at? In: *Rational Animals* (Hurley S, Nudds M, eds), 371-384. Oxford: Oxford University Press.

Tomasello M, Carpenter M, Call J, Behne M, Moll H (2005) Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* 28: 675–735.

- Trivers RL (1971) The evolution of reciprocal altruism. *Quarterly Review of Biology* 46: 35–57.
- Trivers RL (2006) Reciprocal altruism: 30 years later. In: *Cooperation in Primates and Humans*, (van Schaik CP, Kappeler PM, Oxford eds), 67-103. Berlin: Springer.
- van Schaik CP, Kappeler PM (2006) Cooperation in primates and humans: closing the gap. In: *Cooperation in Primates and Humans*, (van Schaik CP, Kappeler PM, Oxford eds), 85–105. Berlin: Springer.
- Warneken F (in press) The Origins of Human Cooperation from a Developmental and Comparative Perspective. In: *The Evolution of Mind and Culture* (Hatfield G, ed) Philadelphia, PA: Pennsylvania State Museum.
- Warneken F, Hare B, Melis A P, Hanus D, Tomasello M (2007) Spontaneous altruism by chimpanzees and young children. *PloS Biology* 5: 1414–20.
- Warneken F, Tomasello M (2006) Altruistic helping in human infants and young chimpanzees. *Science* 311: 1301-3.
- Wellman H M, Liu D (2004) Scaling of theory-of-mind tasks, *Child Development* 75: 523–41.
- West S A, Gardner A, Shuker D M, Reynolds T, Burton-Chellow M, Sykes E M, Guinness M A, Griffin A S (2006) Cooperation and the scale of competition in humans, *Current Biology* 16, 1103-1106.
- Wilkinson G S (1984) Reciprocal food sharing in the vampire bat, *Nature* 308, 181–184.
- Zahavi A (1975) Mate selection - a selection for a handicap, *Journal of Theoretical Biology* 53, 205-214.
- Zlatev J, Persson T, Gärdenfors P (2005) Bodily mimesis as the ‘missing link’ in human cognitive evolution, *Lund University Cognitive Studies* 121, Lund.

<i>Type of cooperation</i>	<i>Cognitive demands</i>	<i>Communicative demands</i>
Flocking behavior	Perception	None
In-group-out-group	Recognition of group member	None
Reciprocal altruism (attitudinal reciprocity)	Individual recognition, minimal memory, empathy, reacting to the desires of others	Expressing desires
Reciprocal altruism (calculated reciprocity)	Individual recognition, memory, slow temporal discounting, empathy, reacting to the desires of others	Expressing desires
Cooperation on future goals	Individual recognition, memory, prospective planning, delayed gratification, reacting to desires, joint intentions	Symbolic communication in the form of protolanguage
Indirect reciprocity	Individual recognition, (episodic) memory, slow temporal discounting, reacting to the emotions and intentions	Symbolic communication in the form of language with names and syntax for roles
Commitment and contract	Individual recognition, memory, prospective planning, joint beliefs	Symbolic communication Protolanguage
Cooperation based on conventions	Joint beliefs	None, but enhanced by symbolic communication
The cooperation of <i>Homo oeconomicus</i>	Full intersubjectivity, perfect rationality	Perfect symbolic communication

Table 1: The cognitive and communicative demands of different forms of cooperation.

---

<sup>1</sup> In Gärdenfors (2007, in press) I also consider the communicative requirements for cooperation and tie them to the evolution of language.

<sup>2</sup> A stronger criterion, focused on human cooperation, is formulated by Bowles and Gintis (2003): "An individual behavior that incurs personal costs in order to engage in a joint activity that confers benefits exceeding these costs to other members of one's group."

<sup>3</sup> An evolutionarily stable strategy is a strategy that an individual cannot improve on, given that everybody else adopt the same strategy.

<sup>4</sup> Detached representations are called secondary representations in Suddendorf and Whiten (2001) and decoupled representations by Sterelny (2003)

<sup>5</sup> Tomasello et al. (2005: 675) write that "the crucial difference between human cognition and that of other species is the ability to participate with others in collaborative activities with shared

---

goals and intentions: shared intentionality." However, their notion rather concerns representing the behavior of others as goal-directed.

<sup>6</sup> In contrast, van Schaik and Kappelan (2006, p. 15) write: "The three basic conditions for reputation are individual recognition, variation in personality traits, and curiosity about the outcome of interactions involving third parties." In my opinion, however, these criteria are too weak since they do not include that reputation is communicated.

<sup>7</sup> In line with Frank (1988), Fehr and Gächter (2002) say that negative emotions towards defectors are the proximate causes of altruistic punishment.

<sup>8</sup> This form of emergence of a social meaning of language is analyzed in detail in Gärdenfors (1993).