The detachment of thought

Peter Gärdenfors

Cognitive Science
Lund University
Kungshuset
S-222 22 Lund, Sweden

E-mail: Peter.Gardenfors@lucs.lu.se

1. THE LADDER TO HUMAN THINKING

When comparing the human mind to that of other animals, several properties have been proposed as uniquely characterising human thinking. We are said to be alone in having a symbolic language, a free will, self-consciousness, and a sense of humor. I do not believe, however, that a sharp borderline can be drawn between human and animal cognition. Rather, human thinking has evolved away from that of our common ancestors among the primates in a series of small steps.

The main thesis in this article is that what characterises the evolution of human thinking is that it shows a successively greater detachment from the current environment. In contrast, animal cognition is more entrained with the external world as it is presented to the sensory organs. I will outline the detachment of thought in a series of steps. My objective is to analyse the *functions* of certain aspects of cognition rather than to study their neurophysiological foundations or their behavioral correlates.

My analysis builds on distinguishing between two kinds of mental representations: *cued* and *detached*. On the basis of this distinction, I will argue that as a result of an increasing detachment of representations, a ladder of cognitive functions can be perceived. The main steps, in the order they will be presented here, are planning, deception, self-awareness, free will, and linguistic communication.

2. CUED VS. DETACHED REPRESENTATIONS

I will argue that in order to understand the functions of most of the higher forms of cognition, one must rely on an analysis of how animals *represent* various things, in particular the surrounding world and its possibilities. There is an extensive debate in the literature on what is the appropriate meaning of ”representation” in this context (see e.g. Roitblat (1982), Vauclair (1990), Gärdenfors (1996) and Grush (1997)). Here I will not go into the intricacies of the debate, but only point out that there are different *kinds* of representations.

The central thesis of this paper is that in order to give an accurate analysis of many phenomena in animal and human cognition, it is necessary to distinguish between two kinds of representations: *cued* and *detached*.

A *cued* representation stands for something that is present in the current external situation of the representing organism. In general, the represented object need not be actually present in the actual situation, but it must have been triggered by something in a recent situation. Delayed responses, in the behaviorist's sense, are based on cued representations according to this characterization.

When, for example, a particular object is categorized as food, the animal will then act differently than if the same object had been categorized as a potential mate. I am not assuming that the animal is, in any sense, *aware* of the representation, only that there is some generalizing factor that determines its behavior.

In contrast, *detached* representations may stand for objects or events that are neither present in the current situation nor triggered by some recent situation. A memory of something, that can be evoked independently of the context where the memory was created, would be an example of a detached representation. Similarly, consider a chimpanzee, who performs the following sequence of actions: walks away from a termite hill, breaks a twig, peels its leaves off to make a stick, returns to the termite hill, and uses the stick to "fish" for termites. This behavior seems impossible to explain unless it is assumed that the chimp has a detached representation of a stick and its use.

A distinction that is similar to that between cued and detached representations has been made by Leslie (1987), who argues that "decoupled" representations are necessary for an organism to engage in *pretense*. When a child pretends that a banana is telephone, Leslie says (1987, p. 415) that the pretend representation must be "quarantined" from some of the sensory information. The perception of the banana must be complemented with information about telephones that is evoked from the child's memory. Leslie claims (1987, p. 416) that the emergence of pretense "is an early symptom of the human mind's ability to characterize and manipulate its own attitudes to information."

I am not claiming that it is possible to draw a sharp line between cued and detached representations. There are degrees of detachment, and, as will be seen below, there are different types of detachment. However, I still believe that the rough distinction between the two major kinds of representations is instrumental in that it directs our attention to key features of the representational forms.

Another caveat concerning my use of the notion of representation is that I am not making any ontological claims: I am not proposing that representations are

entities with some kind of reality status.[1] Rather, I view representations as *theoretical terms*, in the way standardly conceived of in philosophy of science (e. g., Sneed (1971)). Representations are theoretical idealizations, similar to "forces" in Newtonian mechanics, that are introduced to predict and explain empirical generalizations (cf. Lachman and Lachman 1982).

3. THE INNER WORLD

What is the main advantage of detached representations in comparison to cued ones? In order to answer this question, I will elaborate an idea introduced by Craik (1943, p. 61):

> "If the organism carries a "small-scale model" of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which are the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react on a much fuller, safer and more competent manner to the emergencies which face it."

Under the heading of the *inner environment* this kind of "small-scale model" has been made popular by Dennett: "the inner environment is simply any internal region that can affect and be affected by features of potential behavioral control systems" (1978, p. 79). Such an environment is necessary for representing objects (like food and predators), places (where food or shelter can be found), actions (and their consequences), etc., even when these things are not perceptually present. The evolution of this kind of representational power will clearly increase the survival values of the animal. As Dennett (1978, p. 77) puts it:[2]

> "Mutations equipped with such benign inner environments would have a distinct survival advantage over merely Skinnerian creatures in any exiguous environment, since they could learn faster and *more safely* (for trial and error learning is not only tedious; it can be dangerous). The advantage provided by such a benign inner environment has been elegantly expressed in a phrase of Karl Popper's: it 'permits our hypotheses to die in our stead.'"

As a tentative definition, the inner world of an animal will in this paper be identified with *the collection of all detached representations* of the animal and their

---

[1] Roitblat (1982, p. 395) seems to make this kind of claim in his response to the commentary: "Representations have exactly the same ontological status as genes […] . […] Genes, though not "observed," were proposed as *things*; their status was, and remains, that of a theoretical construct."

[2] For further arguments on the evolutionary value of the inner environment, cf. Sjölander (1993, pp. 4-6).

interrelations. Again, I am not assuming that the animal is *aware* of its inner world, nor of the processes utilizing this construct.

It seems that many species of animals have inner worlds. For example, the searching behavior of rats is best explained is if it is assumed that they have some form of "spatial maps" in their heads. Evidence for this, based on their abilities to find optimal paths in mazes, was collected by Tolman already in the 1930's. However, his results were swept under the carpet for many years since they were clear anomalies for the behaviorist paradigm. Vauclair (1987) provides a more recent analysis of the notion of a "cognitive mapping."

It is difficult to assess when detached representations first appeared in the animal kingdom, but a wild speculation is that it is coordinated with the development of the neo-cortex, i.e., roughly with the appearance of mammals. However, it is only with the development of *crossmodal* representations that we obtain advanced forms of an inner world (Davenport 1976, Murray 1990, Allott 1991).[3] It is interesting to note that the human language function does not reside in the same places in the brain as the call systems of the other apes (Deacon 1992). The call systems are *automatic* reactions that cannot be suppressed. The development of the areas in the frontal lobes allowed language to develop as a *voluntary*, i.e., detached, system (Barber and Peters 1992, p. 316).

In support of the general speculation concerning the correlation between detached representations and the neo-cortex, one can note that mammals *play*, but reptiles don't (Sjölander 1993). Playing is a way of building up a repertoire of behaviors that can be used at later occasions. However, this mechanism presupposes that the behaviors are represented in a detached way (see Gulz 1991). Only among the mammals ons finds evidence of *dreaming* in the form of rapid eye movements during sleep. Dreaming apparently presumes an inner world.

My aim in the remainder of the paper is to establish that existence of an inner world is a prerequisite for the evolution of many higher cognitive functions. The functions I will consider are planning, deception, self-awareness, free will, and linguistic communication.

4. IMMEDIATE AND ANTICIPATORY PLANNING

One of the main evolutionary advantages of an inner world is that it frees an animal who is seeking a solution to a problem from dangerous trial-and-error behavior. Jeannerod (1994) says that his general position is that "actions are driven by an internally represented goal rather than directly by the external world." By exploiting its inner world, the animal can *simulate* a number of

[3]Murray (1990) argues that the amygdala is crucial for crossmodal sensory–sensory associations.

different actions in order to "see" their consequences and evaluate them (also compare Grush (1997) and Barsoalou (to appear)). After these simulations, it can choose the most appropriate action to perform in the outer environment. Of course, the *success* of the simulations depends on how well the inner world is matched to the outer. Evolutionary selection pressures will, in the long run, result in a sufficient correspondence between the two environments. As the Norwegian poet Olav Haugen writes: "Reality is a hard shore against which the wave-borne dreamer strands."

The ability to envision various actions and their consequences is a necessary requirement for an animal to be capable of planning. Following Gulz (1991, p. 46), I will use the following criterion: An animal is planning its actions if it has a representation of a goal and a start situation and it is capable of generating a representation of partially ordered set of actions for itself for getting from start to goal. This criterion presupposes representations of (1) goal and start situations, (2) sequences of actions, and (3) the outcomes of actions. The representations of the actions must be detached, otherwise the animal has no *choice*. In brief, planning presupposes an inner world.

There are several clear cases of planning among primates and less clear cases in other species (see e. g. chapters 5, 7, 8 and 9 in Ellen and Thinus-Blanc, eds., 1987, and pp. 58-61 in Gulz 1991). The termite-fishing chimpanzee mentioned earlier is one such example. By the way, this is an example of planned *tool making*.

However, as Gulz (1991) emphasizes, all evidence for planning in non-human animals concerns planning for *present* needs. Apes and other animals plan because they are hungry or thirsty, tired or frightened. Oakley (1961 p. 187) notes that "Sultan, the chimpanzee observed by Kohler, was capable of improvising tools in certain situations. Tool-making occurred only in the presence of a visible reward, and never without it. In the chimpanzee the mental range seems to be limited to present situations, with little conception of past or future."

Man seems to be the only animal that can plan for *future* needs. Gulz (1991, p. 55) calls planning for present needs *immediate planning* while planning for the future is called *anticipatory planning*. Humans can predict that they will be hungry tomorrow and save some food, and we realize that the winter will be cold, so we start building a shelter already in the summer. The crucial distinction is that for an animal to be capable of anticipatory planning it must have a *detached* representation of its future needs. In contrast, immediate planning only requires a cued representation of the current need. There is nothing in the available evidence concerning animal planning, notwithstanding all its methodological problems, that suggest that any other species than *Homo*

has detached representations of their desires (cf. Gulz 1991, ch. 10).[4] The awareness of other animals concern here and now, while we are both here and in the future.

For example, apes and some other species manufacture tools for various purposes. But it seems to be only humans who *transport* tools over long distances. Bringing tools along is a clear indication of anticipatory planning the carrier realizes that there may be a *future need* for the tool. Based on the distribution of early stone tools found in Tanzania and Kenya, Toth (1985) argues that already *Homo habilis*, who lived between 2.5 and 1.7 million years ago, transported their tools several kilometers. In contrast, in studies of chimpanzees they have at most carried their tool a few hundred meters and the longest observed time lapse between the manufacturing of a tool and its use has been seventeen minutes. There are thus no signs that the chimpanzees have the cognitive capacities for anticipatory planning, while already *Homo habilis* seems to have planned for the future.

5. THE HUMAN DILEMMA

Why is is cognitively more difficult to plan for future needs than for present ones? When you are planning to satisfy current needs it is necessary that you can represent possible actions and their consequences in your inner world and that the value of the consequences can be determined *in relation to your present need*. But no separate representation of this need is required. To plan for futuree needs, however, it is also necessary to represent these potential need (and realize that some of them will arise). Furthermore, the consequences of the actions you are planning to perform now must be evaluated in relation to these future needs.

From an evolutionary perspective, it appears obvious that the ability to plan for future needs is valuable. When the ability once has arisen, it will therefore spread in the population. The modern human *cannot but* plan for the future. The French poet Paul Valéry says in one of his aphorisms: "Man throws an arrow, tied to a rope, toward the future. It sticks in an *image* and he hauls himself toward this object."

I believe that this *foresight* results in a fundamental human predicament. The problem is that the actions that are appropriate for fulfilling future needs are often in conflict with those that satisfy present desires. For example, if I do not want to get cold later in the night, I should go out in the forest to look for more firewood, but right now I am warm and cosy and have no desire to leave the

[4]But the squirrel who is gathering and storing food for the winter, isn't it engaged in anticipatory planning? No, it is not planning at all. It has no *representation* of the winter, let alone its needs then. The gathering behavior is routine behavior of an instinctual nature that appears stereotypically without sensitivity to varying circumstances. For example, if one fills the squirrel's stores, it still continues gathering until the "urge" is gone (cf. Gulz 1991, p. 62).

camp fire. *We* must choose between acting for the present or for the future, while other animals, who presumably are not aware of their future needs, only chooses for the present moment. The conflict I am presenting here is similar to the conflict between the ego and the superego in psychoanalytic theory.

There are great individuals differences in how this dilemma is handled. The differences are well illustrated by the fable about the cricket and the ant. Some people have, like the ant in the fable, difficulties living in the present and they obtain their greatest satisfaction by planning for the future. These people set up retirement plans at the age of twenty-five.

Other people have hardly any foresight at all. Like the cricket in the fable they live from hand to mouth and do not worry about tomorrow. This disposition appears to be common among drug addicts. Using the stern terminology of economists, one could say this kind of people are discounting the future very rapidly.

The conflict between the present and the future self is closely related to what Kirkegaard calls ”despair” in his book *The Sickness unto Death*. He describes the unavoidable human dilemma in the following way:

> However much […] the despairer has succeeded in altogether losing his self, and in such a way that the loss is not in the least way noticeable, eternity will nevertheless make it evident that this condition is that of despair, and will nail him to his self so that the torment will still be that he cannot be rid of his self, and it will be evident that he cannot be rid of his self, and it will be evident that his success was an illusion. And this eternity must do, because having a self, being a self, is the greatest, the infinite, concession that has been made to man, but also eternity’s claim on him. (*The Sickness unto Death*, p. 51)

Kirkegaard also points out that the kind of ”sickness” that despair constitutes is unique for human beings:

> The possibility of this sickness is man’s advantage over the beast, and it is an advantage which characterizes him quite otherwise than the upright posture, for it bespeaks the infinite erectness or loftiness of his being spirit. (*The Sickness unto Death*, pp. 44-45)

6. DECEPTION

I want to analyse the evolution of self-consciousness as a series of comparatively small steps. A good planner must consider the actions of other individuals (in particular if the planner belongs to a social species). A special case of representations in the inner world concerns the minds of other individuals. In my opinion, the first step in the evolution of self-awareness is

you-awareness, i.e. when other agents are not only seen as acting things, but as having an inner world of their own, with beliefs, desires etc.

It is only when this representational capacity is accomplished that deliberate *deception* becomes possible. Deception, in its intentional sense, presumes a representation of other minds. To see this let us turn to the worthwhile survey of tactical deception in primates written by Whiten and Byrne (1988). After their initial attempt to define "tactical deception" was criticized in the commentary, they ended up with the following definition: "Acts from the normal repertoire of the AGENT, deployed such that another individual is likely to misinterpret what the acts signify, to the advantage of the AGENT" (1988, p. 271). The key word in this definition is "deployed." When this word refers to human behavior, it refers to an *intentional* act. I submit that this use of deception presupposes that the deceiver has some *representation* of how the individual to be deceived will *interpret* the deceiving act. In other words, deception presupposes that the inner world of the deceiver contains some form of *representation of the inner world of the target individual*. This is a special case of having what is called a "theory of mind". Note that deception presumes all the cognitive functions of (immediate) planning, and some more, i. e., an inner world containing a model of the inner world of other individuals. Thus, this analysis predicts that deception will occur later than planning in the evolution of cognitive functions.[5]

Most of examples in Whiten and Byrne's survey come from field observations of chimpanzees and baboons. For instance, Byrne (1995, p. 124) presents the following case concerning the foraging of baboons, where a juvenile uses a special tactic to obtain food from an adult animal:

> "The juvenile, named Paul, came across an adult female, Mel, just finishing the laborious process of digging up a corm. These were major sources of nutrition at that very dry, cold time of the year, but difficult to obtain from the hard ground; Paul was probably unable to dig his own. He looked around, seeing no other baboon, and screamed loudly. His mother, who was higher ranking than Mel, ran into view grunting aggressively and immediately pursued Mel. When they had both left the immediate area, Paul ate the corm."

One interpretation of this behavior is that Paul imagines that if he screams, *his mother will think* that Mel has hurt Paul and then chase away Mel so that Paul can eat the corm. If this interpretation is correct, Paul's behavior will be a case of intentional deception involving a model of the inner world of another animal. However, there is another weaker interpretation where Paul's behavior is seen as a case of trial-and-error, and thus not intentional deception. If this kind of behavior is successful the first time it is tried, Paul will be more likely to

[5]This thesis is most naturally interpreted as a statement about phylogeny, but can also be given an ontogenetical meaning.

repeat it in the future. If this interpretation is the correct one, his behavior is an example of conditioning that does not involve any representation of the inner world of another individual.

Unfortunately, almost all evidence concerning potential deception is based on more or less anecdotical material. Lacking controlled experiments, it is therefore strongly debatable whether the evidence can establish that deception in the intentional sense occurs among other animals than humans.[6] Heyes (1998) has criticized the strong interpretations and she claims that there is no clearcut evidence that apes or monkeys have any "theory of mind". But the debate is still going on. Primatologists have countered Heyes' analysis. Byrne (1995, pp. 132-134) claims, for example, that there are cases of *counterdeception* that cannot be explained unless intentional deception is involved.

7. SELF-AWARENESS

Deception, in the full intentional sense presupposes that the deceiver has a representation of the dupe's inner world. On this level, an animal can have goals concerning the intentions of other individuals, e.g., *want* somebody to *believe* that an attack would fail. This is an example of a second-order intention.

But a smart agent will not be duped: he will realize that somebody is trying to deceive him and counteract. Hence the really smart deceiver will foresee the reasoning of such a smart agent (cf. Dennett 1988). The important aspect of this escalation in smartness is that it can only work if the deceiver-to-be realizes that the agent he wants to deceive not only has her own representations of the external world, but that her inner world *contains a representation of the deceiver himself*.

Do animals other than humans have self-awareness? Gallup's (1977) experiments show that chimpanzees and orangutans, but no other primates, can recognize themselves in mirrors.[7] And when it comes to recognize oneself in a photograph, only chimpanzees have been successful.

But recognizing oneself in a mirror or on a photograph only requires awareness of one's own *body*, not of one's own mind. The final step in the evolution of higher-level inner representation is small but crucial for self-awareness in its proper sense: I must realize that the inner world of my opponent does not only

---

[6]However, there are cases when it is clear that deception is *not* taking place: The partridge feigning a broken wing to lure away the fox from her chickens is not *fooling* the fox. "Fooling" presumes an intention to make somebody else misinterpret the fooling act. There is no evidence that the partridge has any *representation* of what the fox thinks. She merely acts instinctively when the fox approaches, and can hence not have any intention to fool.

[7]Epstein, Lanza and Skinner (1980) performed a similar experiment intending to show that also pigeons can learn the same kind of behavior. Davis (1989) argues, in my opinion convincingly, that their experiment does not show that pigeons have any form of self-awareness.

contain a representation of myself as a bodily agent, but as *an agent with inner representations as well*. I believe that it is only after this insight that the agent can become self-conscious in the sense that it can form representations of its own representations.

In other words, *self-awareness* can then develop as a shortcut in the representations involved in the deception game: I can in my inner world have a detached representation of my own inner world. The most important aspect of this form of detachment is that I can *attend* to different aspects of my inner world and use this attention in my planning. Other animals can only attend to the external world as it is represented to them. In particular, I can attend to my own feelings and desires and from a representation of how these feelings could be different.

However, I submit that this kind of self-awareness could never develop without the previous establishment of a representation of the inner world of other individuals. In other words, I claim that an "I"-experience must be preceded by a "you"-experience.

8. FREE WILL

There is a strong connection between the possibility of a *free will* and the capability to plan for future needs. Harry Frankfurt (1971) proposes in a classic paper with the title "Freedom of the will and the concept of a person" that a necessary condition for an individual to be a person is that he or she does not only wants something but also wants to want it. Frankfurt calls such wishes "second order wants".

When we reflect upon our choices we are often, but not always, satisfied with them. For example, a smoker frequently wishes that he or she did not want to smoke. In such a case a (first order) wish to smoke is in conflict with a second order wish not to smoke. What is called *akrasia* or weakness of the will is precisely a conflict where the first order desire wins. Many animals have a will, but it is presumably only humans who can ponder on their wishes and want them to be different. It is essential to note that such reflections presume that you have detached representations of your own wishes.

Animals who plan can *choose* in the sense that they can image different actions and their consequences in their inner worlds and then perform the action that is judged to lead to the best consequences. But this does no entail that they have a *free* will. In his analysis, Frankfurt demands that to have a free will it is not sufficient that one can choose, but also be capable of choosing one's will, that is, one should be able to fulfil one's second order desires. In other words, one is free to have the will one wants. A drug addict who repeatedly succumbs to his cravings does not have a free will in this sense.

Having a free will presupposes a form of self-awareness in the sense that one must be aware of one's desires to be capable of wanting to have another desire.

Kirkegaard expresses the connection in the following way:

> In general, what is decisive with regard to the self is consciousness, that is to say self-consciousness. The more consciousness, the more will; the more will, the more self. Someone who has no will at all is no self. But the more will he has, the more self-consciousness he has too. (*The Sickness unto Death*, p. 59)

The question is now: What is the evolutionary value of being able to choose one's own will, of being able to abstain from what you currently want? The answer, as always, is that a free will increases our fitness. Our desires have been moulded during an extended evolutionary process where the ecological conditions have been different from what they are nowadays. If we can choose our will, we can better adjust to the new circumstances. To take an everyday example, food rich in sugar content has been scarce during most of the time humans have existed. Since such food is rich in energy, we have developed a craving for ripe fruits and other things rich in sugar. In our modern world, we are rather confronted with a profusion of sugar and in the long run overconsumption is detrimental to our health. Who can let a second order want control and choose not to consume excessive amounts of sugar will presumably have a slightly higher evolutionary fitness than those who succumb to the temptations. In this way, one can, for good and for bad, let one's wisdom take power over one's desires.

Free will is necessary for *responsibility* Being responsible means that one can realise the potential consequences of the actions one chooses and realise that one can behave differently. Small children and animals are not responible since they do not have this capability (this does not mean that they cannot learn to avoid undesirable behaviour). Similarly, *morality* presupposes a free will. Formulating and obeying a moral rule requires that one has detached representations of one's goals and desires and that one can let the detached goal of the moral rule override one's current desires.

9. LINGUISTIC COMMUNICATION

Thinking does not presume a language. Humans, as well as animals, can simulate sequences of actions in their inner worlds. Such simulations are, among other things, necessary for planning. Language is, in my opinion, a very late phenomenon on the evolutionary scene. As I have tried to show in the previous sections, an individual can have a great deal of cognitive functions, including self-awareness and free will, without having a symbolic language.[8]

[8]A similar point is made by Donald (1991).

Against this position, it can be argued that we all have the experience of something like an omnipresent inner monologue (or dialogue) while we are engaged in thinking. I believe, however, that this experience is deceptive. Firstly, we can "think" without language. Consider, for example, the mental simulation of a slalom skier or a high jumper. Secondly, and more importantly, the inner speech is best interpreted as just parts of the *simulations* in the inner environment. The inner soliloquy is part of what we *perceive* in the inner environment (see Gärdenfors (1995)).

In contrast, I believe that language presumes the existence of an intricate inner world. In order to make this clear, I will introduce a distinction between *signals* and *symbols*. Both signals and symbols are tools of communication. The fundamental difference between them is that *the reference of a symbol is a detached representation, while a signal refers to a cued representation*. In other words, a signal refers to something in the outer environment, while a symbol refers to the inner world. Language consists of symbols — it can be used to talk about things not present in the current situation. This idea can be traced back to Hockett's (1960) notion of "displacement." Glasersfeld expresses the point as follows:

> "we can talk not only about things that are spatially or temporally remote, but also about things that have no location in space and never happen at all [...] in order to become a symbol, the sign must be detached from input. What the sign signifies, i.e., its meaning, has to be available, regardless of the contextual situation" (von Glasersfeld, 1977, p. 64).

With few exceptions, linguistic communication is achieved with the aid of symbols. Sjölander (1993, pp. 5-6) explains elegantly what is missing in animal communication:

> The predominant function of language is to communicate about that which is not here and not now. A dog can 'say': I am angry, I want water, I want to go out, I like you, etc. But it has no communicative means enabling it to 'say': I was angry yesterday, nor can it 'say': I will be angry if you lock me up tonight again, and I will chew up the carpet. Likewise, the dog can 'say': There is a rat here! but it cannot 'say': There is a rat in the next room.
>
> […] Clearly, if you live in the present, communicating mainly about how you feel and what you want to do in the moment, the biological signals inherent in each species are sufficient. A language is needed only to communicate your internal representation of what could be, what has been, and of those things and happenings that are not present in the vicinity.

A similar characterization can be found in von Glasersfeld (1977, pp. 63-65), who traces the idea back to Langer (1948). She clearly distinguishes symbols from signals:

> A term which is used symbolically and not signally does *not* evoke action appropriate to the presence of its object. [...] Symbols are not proxy for their objects, but are *vehicles for the conception of objects*. To conceive a thing or a situation is not the same as to "react toward it" overtly, or to be aware of its presence. In talking about things we have conceptions of them, not the things themselves; and *it is the conceptions, not the things, that symbols directly "mean."* Behavior toward conceptions is what words normally evoke: this is the typical process of thinking. (Langer 1948, p. 61)

Symbols refering to something in one person's inner world can be used to communicate as soon as the listeners have, or are prepared to add, the corresponding references in their inner worlds.[9] The actual conditions of the outer situation need not play any role for the communication to take place: two prisoners can talk fervently about life on a sunny Pacific island in the pitch dark of their cell.

Following Peirce's (1932) trichotomy of signals (which he calls indices), icons, and symbols, the role of *icons* can be characterized as follows. Like symbols, icons refer to detached representations, but unlike symbols, the choice of representation is not arbitrary. On the contrary, an icon in some aspects *resembles* the thing it represents.

Many animals have intricate systems of signals, for example, the dances of bees. However, even if their dances seem to have a kind of grammar, it still consists only of *signals*. The bees categorize, in a sophisticated way, places where nectar can be found. The crucial point is that they only use their dances in a situated manner, and thus the dances are not symbols according to my criterion. The same point is made by von Glasersfeld (1976, p. 222): "In my terms, the bees do not qualify for symbolicity, because they have never been observed to communicate about distances, directions, food sources, etc., without actually coming from, or going to, a specific location."

In spite of all attempts to teach apes various forms of symbolic codes (see e.g. Savage-Rumbaugh, Shanker, and Taylor (1998)), humans seem to be the only animals that use language in a fully detached way. Even though the pygmy chimpanzee Kanzi's performance is quite impressive, his use of symbols is dependent on the context: they mainly express requests to "direct teacher's attention to places, things and activities" (Savage-Rumbaugh et al., (1985, p. 658)). Human children, in contrast, very early use language outside the context of request. Vauclair (1990, p. 319) notes that "the use of symbols by apes is closely tied to the achievement of immediate goals, because the referents occur in the context of behavior on their objects." This is congenial with Gulz' (1991) conclusion that only humans are anticipatory planners. My conjecture is that

[9]For a model theoretic account of how such communication can be established, see Gärdenfors (1993).

this capability is required for the complete detachment of language. We are still waiting for Kanzi to tell us a story by the camp fire.

The fact that a language consists of symbols refering to detached representations is a necessary, but far from sufficient, condition to separate language from other forms of communication. I next turn to what needs to be added to this condition.

The first thing to notice is that human linguistic communication presumes an advanced kind of inner world. To see this, let us turn to Grice's (1957, 1969) theory of meaning. His initial definition in the second paper reads as follows (1969, p. 151):[10]

> "*U* meant something by uttering x" is true iff, for some audience *A*, *U* uttered *x* intending
>
> (1) *A* to produce a particular response *r*.
> (2) *A* to think (recognize) that *U* intends (1).
> (3) *A* to fulfill (1) on the basis of his fulfillment of (2).

Although he defines "meaning," I am more interested in applying the definition to linguistic communication in general. The feature I want to focus on here is that condition (2) expresses a *third-order intention* (Dennett 1978, p. 277-278): U intends A to think that U intends something. Gomez (1994, p. 68) claims that a truly requestive situation like "May I have some salt, please?" even involves a *fifth-order* level of intentionality: U *wants* A to *understand* that she *wants* him to *understand* the she *wants* the salt.[11] The upshot is, if I am correct, that a full-blown linguistic communication presumes a mind that is capable of you-awareness as well as self-awareness. A consequence of this is that language, in the normal sense, is most likely a *very* recent phenomenon in the evolution of thinking.

10. THE ROLE OF GRAMMAR

There are forms of communication that do not involve a grammatical structure. Bickerton (1990, p. 122) calls such a simplified mode of language a *protolanguage* and argues that

> "there is a mode of linguistic expression that is quite separate from normal human language and is shared by four classes of speakers: trained apes, children under two, adults who have been deprived of language in their early years, and speakers of pidgin".

[10]This definition is revised several times in the second paper, but the more complicated versions have the same general structure as the definition given here.

[11]However, he also claims that the *mutuality* of intentional communication can be achieved by "attention contact" without metarepresentations of the inner environment of the other (Gomez 1994, p. 73).

However, he never defines what constitutes protolanguage, but only characterizes it negatively by comparing it to ordinary language. He presents five types of differences: Protolanguage is less ordered than ordinary language, it contains no null elements,[12] it does not always respect the valence relations of verbs, it does not allow expansion of utterances, and it hardly contains any grammatical items (Bickerton 1990, pp. 122-126).

As regards the evolutionary timing of the transition from protolanguage to language, several authors (Bickerton 1990, Donald 1991, Lieberman 1992) speculate that this is essentially concurrent with the transition from *Homo erectus* to *Homo sapiens* (on the order of 200.000 years ago). One of the anatomical changes that occurs in this transition between the two species is the lowering of the larynx, which clearly is connected with the development of a spoken language (Lieberman 1992).

So whence grammar? As the human societies grew more complex, speed and efficiency in communication was rewarded. Barber and Peters (1992, p. 311) argue that[13]

> the need for fast and efficient processing is thus a major force that drives language away from iconicity and toward systematicity – and this in turn drives language toward arbitrariness […], for the following reason. In the long run it is less effort to deal with a tightly patterned system with a small number of reusable parts (both items and rules) than to deal with a sprawling system with many, many unique parts. But the reduction to reusable parts and patterns destroys most of the iconicity […], and at the same time compresses a great deal more information into a small number of rules: it radically increases the "depth" of the system while decreasing the algorithmic complexity […]."

A similar point is made by Savage-Rumbaugh and Rumbaugh (1993, pp. 86-87), who note the need for communication that is *independent of context* as one of the evolutionary forces behind the development of grammar:

> It will also be argued that syntax, rather than being biologically predetermined, is a skill which arises naturally from the need to process sequences of words rapidly. As overall intelligence increased, spurred by the ever-increasing use of language for planning future activities,

---

[12]"Null element" is a syntactic notion refering to places in a sentence where one can infer (using government and binding theory) that some constituent should be present, but where there is no explicit constituent.

[13]Ellegård (1977, p. 142) speaks about the "double articulation" of language and remarks concerning the evolution of grammar: "My hypothesis is thus that the double articulation of human speech emerged as a necessary consequence of the increasing number of signs, and the increasing demands for fast and more or less automatized production. The reaction of the brain toward these demands was the double articulation in phonemes and morphemes." (My translation).

> communications became increasingly complex and increasingly independent of context. When complex ideas began to require groups of words for their expression, it became essential to devise a means to specify which of the words in a group modified (or were related to) which other words. Syntactical rules were developed to solve this dilemma. Such rules were the inevitable outgrowth of complex symbolic communication involving multiple symbols.

Adding grammar to a communication system thus increases its efficiency. For a neuroscientist, the question is in what way the brain must change in order to achieve this capacity. It seems that the grammaticality of language has probably not evolved as an independent cognitive ability (in contrast to Chomsky's claims concerning a "language acquisition device"). Rather, it could build on already existing structures since it seems to be tied to a more general capability of combining actions into sequences. Neurologically, sequencing is typically lateralized to the left hemisphere in humans. Corballis (1989) argues that in the course of evolution, sequencing emerged in the left hemisphere and was essential for *tool-making* and other practical skills.[14] In order to reproduce or create a tool, a sequence of actions had to be performed, and old elements of action sequences must be recombined to produce something new. The practical, mainly manual, ability forms the basis for all kinds of sequencing and was extended to sequencing of symbols, which then resulted in a grammatical language. Also *playing* involves sequences of motor actions performed on symbolically used objects (see Vauclair and Vidal 1994). And remember that Piaget has always emphasized that play and imitation are cardinal for the development of symbolic capacities. Apes seem to lack the sequencing capability that could explain why they are bad at imitating action sequences, why they never invent new forms of play, as well as why they cannot learn more than a protolanguage.[15]

11. CONCLUSION

In this article, I have outlined the evolution of human thinking as a series of steps, where each step has been characterised in terms of the representational capacities of the organism. My presentation has been based on the distinction between cued and situated representations.

The inner world of an animal has been definedas the collection of all its detached representations. I have argued that the notion of an inner world must serve as a basis for all higher cognitive functions like planning, deception, self-awareness, fre will and language. My main thesis is that the general trend in the

---

[14]See also Allott (1991) and Tomasello (1991).

[15]See also Barber and Peters (1992, p. 344) and Donald (1991, pp. 70-75). On the other hand, Bickerton (1990, p. 139) argues that "it is tool-making and *protolanguage* that share the same processes."

evolution of cognition is that more and more representations become detached. This has led to increasing cognitive flexibility, but also to the human dilemma involved in weighing one's present desires to the needs that are predicted for the future. In other words, I view the evolution of cognition as the story of the detachment of thought.

REFERENCES

Allott, R. (1991): "The motor theory of language," in *Studies in Language Origins, Vol 2*, Raffler-Engel, W. von, Wind, J. and Jonker, A., eds., John Benjamins, Amsterdam, pp. 123-157.

Barber, E. J. W. and Peters, A. M. W. (1992): "Ontogeny and phylogeny: What child language and archeology have to say to each other," in *The Evolution of Human Languages*, J. A. Hawkins and M. Gell-Mann, eds., Addison Wesley, Redwood City, CA, pp. 305-351.

Barsalou, L. W. (to appear): "Perceptual symbol systems", to be published in *Behavioral and Brain Sciences*.

Bickerton, D. (1990): *Language and Species*, The University of Chicago Press, Chicago.

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

Craik, K. (1943): *The Nature of Explanation*, Cambridge University Press, Cambridge.

Davenport, R. K. (1976): "Cross-modal perception in apes," in S. R. Harnad, H. D. Steklis, and J. Lancaster, eds., *Origins and Evolution of Language and Speech*, Annals of the New York Academy of Science 280, pp. 143-149.

Davis, L. H. (1989): "Selfconsciousness in chimps and pigeons," *Philosophical Psychology 2*, pp. 249-259.

Deacon, T. W. (1992): "Brain-language coevolution," in *The Evolution of Human Languages*, J. A. Hawkins and M. Gell-Mann, eds., Addison Wesley, Redwood City, CA, pp. 49-83.

Dennett, D. (1978): *Brainstorms: Philosophical Essays on Mind and Psychology*, The MIT Press, Cambridge, MA.

Dennett, D. (1988): "Why creative intelligence is hard to find," *Behavioral and Brain Sciences 11,* pp. 253.

Donald, M. (1991): *Origins of the Modern Mind*, Harvard University Press, Cambridge, MA.

Ellegård, A. (1979): "Om det mänskliga språkets ursprung," in *Kungliga Vitterhets-, Historie- och Antikvitetsakademins Årsbok*, Stockholm, pp. 131-147.

Ellen, P. and Thinus-Blanc, C., eds. (1987): *Cognitive Processes and Spatial Orientation in Animal and Man: Volume I Experimental Animal Psychology and Ethology,* Martinus Nijhoff Publishers, Dordrecht.

Epstein, R., Lanza, R.P. and Skinner, B.F. (1980): "Self-awareness in the pigeon," *Science 212*, pp. 695-696.

Frankfurt, H. (1971): "Freedom of the will and the concept of a person", *Journal of Philosophy 68*, pp. 5-20.

Gallup, G.G. (1977): "Self-recognition in primates," *American Psychologist 32,* pp. 329-338.

Gärdenfors, P. (1992): "Medvetandets evolution" (The evolution of consciousness), Chapter 5 in *Blotta tanken*, Doxa, Nora.

Gärdenfors, P. (1993): "The emergence of meaning," *Linguistics and Philosophy 16*, pp. 285-309.

Gärdenfors, P. (1995): "Speaking about the inner environment," paper prepared for the Nobel symposium on *Mind and language*, Stockholm, August 1994.

Gärdenfors, P. (1996): "Cued and detached representations in animal cognition", *Behavioural Processes 36*, pp. 263-273.

Glasersfeld, E. von (1976): "The development of language as purposive behavior," in S. R. Harnad, H. D. Steklis, and J. Lancaster, eds., *Origins and Evolution of Language and Speech*, Annals of the New York Academy of Science 280, pp. 212-226.

Glasersfeld, E. von (1977): "Linguistic communication: theory and definition," in D. M. Rumbaugh, ed., *Language Learning by a Chimpanzee: The LANA Project*, Academic Press, New York, pp. 55-71.

Gomez, J. C. (1994): "Mutual awareness in primate communication: A Gricean approach," in *Self-awareness in Animals and Humans*, S. T. Parker, R. W. Mitchell, and M. L. Boccia, eds., Cambridge University Press, Cambridge, pp. 61-80.

Grice, H. P. (1957): "Meaning," *The Philosophical Review 66*, pp. 377-388.

Grice, H. P. (1969): "Utterer's meaning and intentions," *The Philosophical Review 78*, pp. 147-177.

Grush, R. (1997): "The architecture of representation", *Philosophical Psychology 10*, pp. 5-23.

Gulz, A. (1991): *The Planning of Action as a Cognitive and Biological Phenomenon*, Lund University Cognitive Studies 2, Lund.

Heyes, C. (1998): "Theory of mind in nonhuman primates", *Behavioral and Brain Sciences 21*, pp. 101-115.

Hockett, C. F. (1960): "The origin of speech," *Scientific American 203(3)*, pp. 88-96.

Jeannerod, M. (1994): "The representing brain, neural correlates of motor intention and imagery," *Behavioral and Brain Sciences*, pp. 187-202.

Kirkegaard, S. (1849/1989): *The Sickness unto Death*, Penguin Classics, London.

Lachman, R. and Lachman, J. L. (1982): "Memory representations in animals: Some metatheoretical issues," *Behavioral and Brain Sciences 5*, pp. 380-381.

Leslie, A. M. (1987): "Pretense and representation: The origins of "theory of mind"," *Psychological Review 94*, pp. 412-426.

Lieberman, P. (1992): "On the evolution of human language," in *The Evolution of Human Languages*, J. A. Hawkins and M. Gell-Mann, eds., Addison Wesley, Redwood City, CA, pp. 21-47.

Murray, E. A (1990): "Representational memory in nonhuman primates," in *Neurobiology of Comparative Cognition*, R. P. Kesner and D. S. Olton, eds., Lawrence Erlbaum Associates, Hilldale, NJ, pp. 127-155.

Oakley, K. P. (1961): "On man's use of fire, with comments on tool-making and hunting," in S. L. Washburn, ed., *Social Life of Early Man*, Aldine Publishing Co., Chicago, pp. 176-193.

Peirce, C. S. (1932): *Collected Papers of Charles S. Peirce*, ed. Hartshorne, Weiss and Burks, Harvard University Press, Cambridge, MA.

Roitblat, H. L. (1982): "The meaning of representation in animal memory," *Behavioral and Brain Sciences 5*, pp. 353-372.

Savage-Rumbaugh, E. S., Rumbaugh, D. M, and McDonald, K. (1985): "Language learning in two species of apes," *Neuroscience and Biobehavioral Review 9*, pp. 653-665.

Savage-Rumbaugh, E. S., and Rumbaugh, D. M. (1993): "The emergence of language", in Gibson, K. R. and Ingold, T., ed., *Tools, Language and Cognition in Human Evolution*, Cambridge University Press, Cambridge, pp. 86-108.

Savage-Rumbaugh, E. S., Shanker, S. G. and Taylor, T. J. (1998): *Apes, Language and the Human Mind*, Oxford University Press, Oxford.

Sjölander, S. (1993): "Some cognitive breakthroughs in the evolution of cognition and consciousness, and their impact on the biology of language," *Evolution and Cognition 3*, pp. 1-10.

Sneed, J. (1971): *The Logical Structure of Mathematical Physics*, Reidel, Dordrecht.

Stadler, M. (1989): "On learning complex procedural knowledge," *Journal of Experimental Psychology: Learning, Memory and Cognition 15*, pp. 1061-1069.

Tomasello, M. (1991): "Processes of communication in the origins of language," in *Studies in Language Origins, Vol 2*, Raffler-Engel, W. von, Wind, J. and Jonker, A., eds., John Benjamins, Amsterdam, pp. 85-97.

Vauclair, J. (1987): "A comparative approach to cognitive mapping," in P. Ellen and C. Thinus-Blanc, eds., *Cognitive Processes and Spatial Orientation in Animal and Man: Volume I Experimental Animal Psychology and Ethology,* Martinus Nijhoff Publishers, Dordrecht, pp. 89-96.

Vauclair, J. (1990): "Primate cognition: From representation to language," in S. T. Parker and K. R. Gibson, eds., *"Language" and intelligence in monkeys and apes*, Cambridge Univesity Press, Cambridge, pp. 312-329.

Vauclair, J. and Vidal, J.-M. (to appear): "Discontinuities in the mind between animals and humans," paper presented at the conference on *Cognition and Evolution*, Berder, March 1994.

Whiten, A. and Byrne, R. W. (1988): "Tactical deception in primates," *Behavioral and Brain Sciences 11,* pp. 233-73.