Thinking about complex mental states:
language, symbolic activity and theories of mind

Emanuele Arielli

Abstract: One of the most important contributions in Roland Posner's work (1993) was the extension and development of the Gricean paradigm on *meaning* (1957) in a systematic framework, providing thus a general foundation of semiotic phenomena. According to this approach, communication consists in behaviors or artifacts based on reciprocal assumptions about the intentions and beliefs of the subjects involved in a semiotic exchange. Posner's model develops with clarity the hierarchical relationships of semiotic phenomena of different complexity, from simple pre-communicative behaviors (like indicating or signaling) to full communicative acts. Not only limited to linguistic communication, this framework can be successfully extended in the description of all kind of sign production, from gestures to artifacts.

This article discusses a key point of this paradigm, namely our faculty to attribute *mental states* to other individuals through the observation of their behavior, artifacts or texts. New insights from neuropsychology and developmental psychology are discussed in supporting the validity of this model of communication.

1. Two main criticisms against the *beliefs* and *intentions*-models of communication

The topic of mental state attribution, and its application to the study of communication, is plagued by two fundamental theoretical problems:[1]

1. The first is the propensity to *presuppose* mental states in circumstances where they are in fact absent. For example, we can attribute mental states to a story's character or to institutions. The problem is thus to determine when the attribution of mental states is realistic and when it is just a case of metaphorical or 'naïve' thinking.

2. Furthermore, the attribution processes in the communication models based on the recognition of beliefs and intentions usually reach high levels of complexity. Believing that „the speaker intends me to believe that she wants that I believe *x*" (or similar 'higher order' beliefs) sounds too complex to be realistic for some. Some authors (see Meggle

---

[1] The topics presented here have been initially developed in depth in Arielli 2006.

1983) even suggest that communication presupposes *infinite* levels of nested beliefs, which appears to be psychologically impossible.

To address the two problems concerning the *realism* and the *complexity* of attribution processes let consider following Aesopian fable (adapted from Ryan 1991, Betram 1979):

> The Fox and the Rooster
> Once a Dog and a Rooster went into the woods. Soon it grew dark. The Rooster said, 'Let us stay here all night. I will stay in this tree-top. You can sleep in the hollow trunk.' So the Dog and the Rooster went to sleep. In the morning the Rooster began to crow. A Fox heard him crow. He said, 'That is a rooster crowing. He must be lost in the woods. I will eat him for my breakfast.' Soon the Fox saw the Rooster in the tree-top. He said to himself: 'I must make him come down from the tree.' So he said to the Rooster, 'What a fine rooster you are! How well you sing! Will you come to my house for breakfast?' The Rooster said, 'Yes, thank you, I will come if my friend may come, too.' 'Oh yes', said the Fox, 'I will ask your friend. Where is he?' The Rooster said, 'My friend is in the hollow tree. He is asleep. You must wake him.' The Fox said to himself: 'Ha ha! I shall have two roosters for my breakfast!' So he puts his head into the hollow tree. Then he said, 'Will you come to my house for breakfast?'. Out jumped the Dog, and caught the Fox by the nose.

The story is straightforward and easy to understand, there is no intricate plot and it does not require any complex cultural reference or reasoning to see what has happened. We grasp the narrative of the story by building a representation or a model of the world in which the events take place. We build this model not only using a description of physical events, but more importantly we are able to describe the mental states of the story's characters, attributing beliefs and intentions to them.

Without these attributions we would not be able to understand the essence of the story. Let us consider, for example, the moment in which the Rooster hears the Fox's invitation to come to his house, but is aware of its true intentions. In this phase of the tale, the Rooster knows or believes that 1) the Fox intends to eat him and that 2) the Fox intends to make the Rooster believe that he is friendly and wants to invite him for breakfast (this is a complex knowledge: the Rooster *knows* that the Fox *wants* that the Rooster *believes* that the Fox *wants* to invite him for breakfast) and that 3) the Fox intends that the Rooster decides to come down from the top of the tree. Finally, of course, the Rooster knows that there is a dog in the trunk, while at the same times he knows that the Fox does not have this information. Subsequently, the plan worked up by the Rooster can be described as the intention to make the Fox believe that he

is accepting his invitation and that there is a 'companion' down in the tree that still needs to be invited.

```
┌─────────────────────────────────────────────────────────────────┐
│ Reader's knowledge of the story's events                         │
│  ┌──────────────────────┐   ┌──────────────────────────────┐     │
│  │ The Fox's mind:      │   │ The Rooster's mind:          │     │
│  │                      │   │                              │     │
│  │  ┌────────────────┐  │   │  The Rooster knows or be-    │     │
│  │  │ The Fox believes│ │   │  lieves that his friend the  │     │
│  │  │ that there is   │  │   │  dog is in the trunk and that│     │
│  │  │ a Rooster lost  │  │   │   ┌────────────────────────┐ │     │
│  │  │ in the wood     │  │   │   │ the Fox intends to eat │ │     │
│  │  │ and that the    │  │   │   │ him and intends that   │ │     │
│  │  │ Rooster is alone│  │   │   │  ┌──────────────────┐  │ │     │
│  │  └────────────────┘  │   │   │  │ the Rooster believes│ │     │
│  │                      │   │   │  │ that the Fox likes │  │ │     │
│  │                      │   │   │  │ his crowing and that│ │     │
│  │                      │   │   │  │  ┌──────────────┐  │  │ │     │
│  │                      │   │   │  │  │ the Fox intends│ │  │ │     │
│  │                      │   │   │  │  │ to invite him │ │  │ │     │
│  │                      │   │   │  │  │ for breakfast.│ │  │ │     │
│  │                      │   │   │  │  └──────────────┘  │  │ │     │
│  │                      │   │   │  └──────────────────┘  │ │     │
│  │                      │   │   │  ┌──────────────────┐  │ │     │
│  │                      │   │   │  │ the Rooster intends│ │  │     │
│  │  ┌────────────────┐  │   │   │  │ to come down     │  │ │     │
│  │  │ The Fox intends │ │   │   │  └──────────────────┘  │ │     │
│  │  │ that            │  │   │   └────────────────────────┘ │     │
│  │  │ ┌────────────┐ │  │   └──────────────────────────────┘     │
│  │  │ │ the Rooster │ │  │   ┌──────────────────────────────┐     │
│  │  │ │ believes    │ │  │   │ The Rooster intends to       │     │
│  │  │ │ that the Fox│ │  │   │ chase the Fox away and in-   │     │
│  │  │ │ likes his   │ │  │   │ tends that                   │     │
│  │  │ │ crowing and │ │  │   │  ┌────────────────────────┐ │     │
│  │  │ │ believes    │ │  │   │  │ the Fox believes that in│ │     │
│  │  │ │ that        │ │  │   │  │ the trunk there is an- │ │     │
│  │  │ │ ┌────────┐ │ │  │   │  │ other rooster and be-  │ │     │
│  │  │ │ │the Fox │ │ │  │   │  │ lieves                 │ │     │
│  │  │ │ │intends │ │ │  │   │  │  ┌──────────────────┐  │ │     │
│  │  │ │ │to invite│ │ │  │   │  │  │ that the Rooster be-│ │     │
│  │  │ │ │him for │ │ │  │   │  │  │ lieves the Fox to be│ │     │
│  │  │ │ │breakfast│ │ │  │   │  │  │ friendly          │  │ │     │
│  │  │ │ └────────┘ │ │  │   │  │  └──────────────────┘  │ │     │
│  │  │ └────────────┘ │  │   │  └────────────────────────┘ │     │
│  │  │ ┌────────────┐ │  │   │  ┌────────────────────────┐ │     │
│  │  │ │ the Rooster │ │  │   │  │ the Fox intends to     │ │     │
│  │  │ │ intends to  │ │  │   │  │ wake the companion in  │ │     │
│  │  │ │ come down   │ │  │   │  │ the trunk (meeting     │ │     │
│  │  │ │ from the    │ │  │   │  │ eventually the dog).   │ │     │
│  │  │ │ tree top.   │ │  │   │  └────────────────────────┘ │     │
│  │  │ └────────────┘ │  │   └──────────────────────────────┘     │
│  │  └────────────────┘  │                                        │
│  └──────────────────────┘                                        │
└─────────────────────────────────────────────────────────────────┘
```
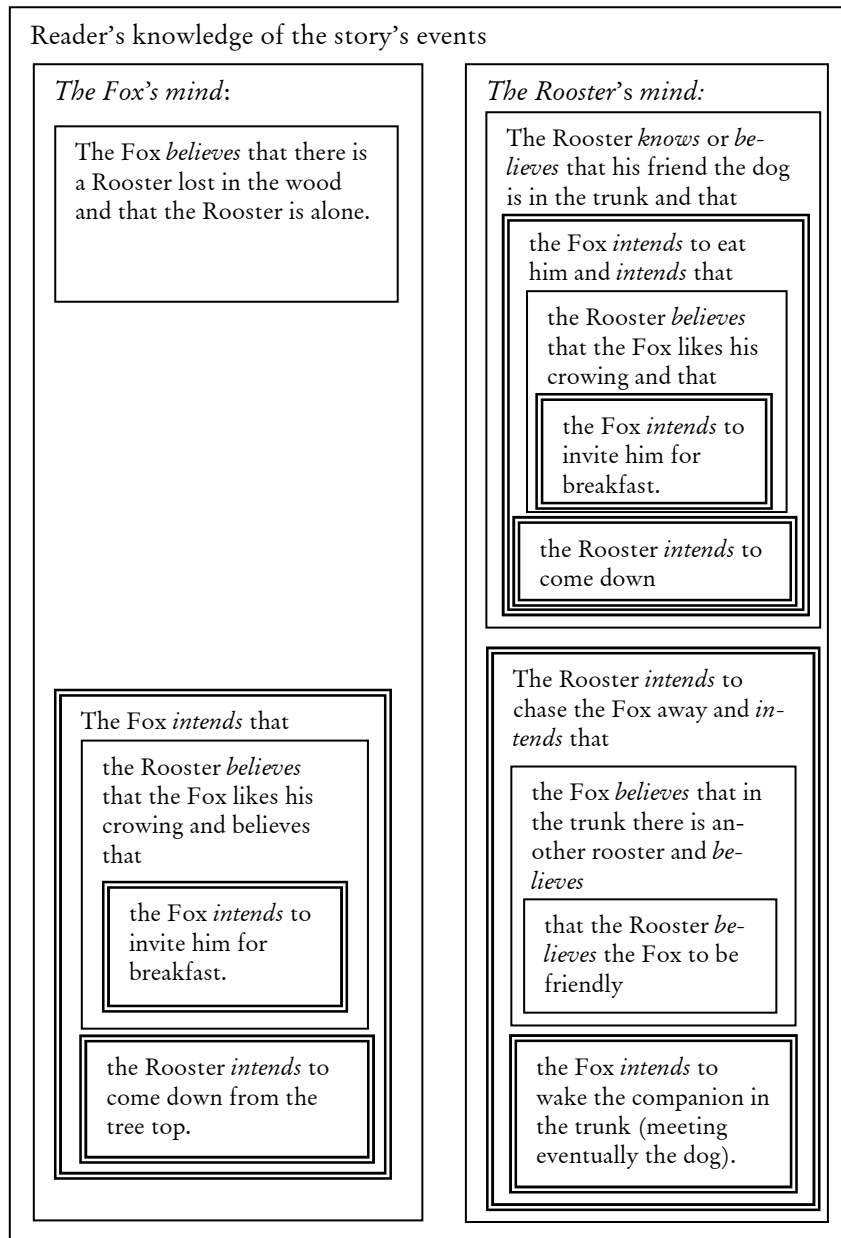
Figure 1

It is possible to describe in the same manner all beliefs and intentions from the Fox's standpoint. What we get is a structure in which beliefs about beliefs, intentions or beliefs about intentions - and so on - combine to form a complex system of mental states embedded in one another that we, the readers, attribute to the two characters in the fable. Figure 1 shows the breakdown of the characters' mental states when the Fox invites the Rooster to go down and the Rooster is already preparing his countermove (in this diagram, beliefs are represented as single bordered boxes, intentions as double bordered boxes).

In this analysis we form quite inelegant sentences ("believing that the other believes that ...", "intending that the other believes ...", and so on), revealing how surprisingly complex the representation of the characters' mental states is in a particular moment of this simple story. Still, the understanding of this complex hierarchical nesting of beliefs and intentions is a *necessary condition* in following the narrative. It may seem implausible to keep in mind these complex relationships, yet without these assumptions it would be impossible for the reader to get the story. Without grasping all the structure, we wouldn't be able to answer the simplest questions about what has happened, who outsmarted whom, and so on.

The fact that we attribute mental states to fictional (and non-human) characters in a fantasy tale gives us, in part, an answer to the first question posed at the beginning: the ability to represent and reason about mental states is independent from the question if we recognize those states as real or not. We know that we are dealing with a fictional story, still we don't have any problem pretending *as if* we were in front of 'real minds'.

Moreover, understanding a story as simple as this presupposes the grasping of complex nested mental states, and this gives us an answer to the second question: we are able to easily maintain complex representations of this kind. As in many types of cognitive processes, we are not *aware* of the complexity of these mechanisms, despite the spontaneity with which we master them. The cognitive mechanisms that are active during the reading of a text include the decoding of its linguistic signs, the construction of a model of the story and the mental states of its characters. Our mind is able to carry out those functions automatically, without our conscious awareness, sometimes even eluding our introspection. This is true in general processes such as perception, motor coordination, reasoning, but also in the production and comprehension of speech and texts.

It is to be noted that, according to the models of communication based on beliefs and intentions, the diagram in Figure 1 is quite complex, but not complex enough, since it does not describe the mental processes that are presupposed in every simple communicative act of the story's

characters. 'Make someone believing something' by means of a sentence ('Will you come to my house for breakfast?') should be further broken down in the well known triple-parted bundle of intentions and beliefs that are described by Grice's and Posner's model of communication as necessary conditions for a speech act.

2. The epistemological status of reasoning about mental states

Reasoning about mental states and attributing them to 'minds' (fictional or real), constitute the ground on which Grice's and Posner's models of communication are based. The epistemological and psychological status of mental states is however the object of traditional debate and has often been questioned. In the philosophical debate on the scientific validity of behavioral explanations in terms of mental states, the so-called *eliminativist* position considers intentional concepts (belief, knowledge, will, understanding, and so on) hopelessly incompatible with a scientific and naturalistic description. According to this position, we need to replace them with behavioral descriptions or, as suggested more recently, reduce them to the description of neuronal processes. According to this position, mental states are similar to pseudo-scientific notions like the medieval *phlogiston* or simply to everyday figure of speech like 'The sun rises', a sentence that has a meaning in the common language even if we know that this is scientifically unsound, since the sun does not evolve around the earth.

Still, notions such as belief, intention and other mental states are not only metaphorical labels, since they are an essential part of the way we understand the behavior of people and talk about it. A description in terms of physical and neuronal processes, although scientifically appropriate, would neither be realistic nor useful in everyday usage. That is, *folk psychological* explanations are not scientific, but this does not mean that they are wrong *models* of human behavior. Describing behaviors in mentalist terms has in fact strong explanatory and predictive validity. Although the psychology of common sense is not supported by the scrutiny of the scientific method, its validity is continually tested in everyday life, proving to be a good set of heuristics for the understanding of human action. If using mental states were entirely inadequate to explain and predict consistently a good part of our behavior, they would have never developed as an essential part of our language. With reference to this topic, Daniel Dennett (1987) proposed to consider the use of mental concepts as an attitude or *stance* specific to certain types of complex behaviors, for reasons of simplicity and efficacy. This happens also with respect to entities that do not have a real mind. For example I can use intentional concepts when I say that my bank's central computer 'knows' the status of

my account or the chess program I'm playing against 'intends' to block the progression of my pawn.

3. The neural basis of intention attribution

Recent research in areas such as developmental psychology, autism research and so-called *social-neuropsychology* have shown that reasoning and attributing mental states not only means using conventional labels, nor is a simple chosen *stance* to take for convenience. There are several indications that our brains are biologically wired to recognize and elaborate others' mental states and intentions. According to these findings, there is a neuronal basis of *mentalizing* (Frith and Frith 2006) or *mind reading*, that is of our ability to formulate hypotheses about others' mind beliefs and intentions, building representations of their mental states. This is a feature that evolution developed in humans and probably in apes.

With reference to the understanding of stories, a study by Fletcher et al. (1995) has shown that certain brain regions are activated during the reading of events requiring the attribution of mental states (like Aesop's fable at the beginning), but they are not active if we are reading about events that are purely physical.

A widespread view about the faculty of understanding others' mental states is offered by the so-called *simulation theory* (see Goldman 1993; Gallese and Goldman 1998), a philosophical explanation that has been supported by recent neurological findings. According to it, we are able to attribute mental states to others because we 'put ourselves in their shoes', we represent their motives, beliefs and intentions as if we were in their situation. Simulating others' cognitive states using our own cognitive resources allow us to understand and predict the behavior of other people. A possible neural source of this simulation skill has been given by the discovery of so-called *mirror neurons* (Rizzolatti and Arbib 1998, Rizzolatti et al. 2001), first in the brains of monkeys and then of humans. These are a specialized set of neurons that are activated only when a subject carries an action (e.g. grasping an object) but also when the subject observes someone else performing the same action. This would suggest the presence of a common mechanism for both the production and the perception of an action. According to many researchers, mirror neurons form the basis of our ability to simulate and understand the mental states of other subjects. The ability to attribute mental states would have allowed the evolution of complex social interactions, favoring the effective coordination between group members and giving rise to communication: it has been in fact suggested (Rizzolatti and Arbib 1998) that mirror neurons can offer a neurobiological explanation for the evolution of language, and

the fact they are concentrated in Broca's area - one of the centers of production and comprehension of language - seems to confirm this fact. This hypothesis would be a fascinating and controversial biological evidence of the close relationship between the understanding of mental states and communication.

Research on brain mechanisms about mind reading is an area in full development. In particular the findings on the neural basis of our ability to coordinate and interact with others are building a new disciplinary area, called *social neuroscience* (Singer 2006, Ciaramitano et al. 2007) that offers new scientific support to the classic models of communication based on intentions and beliefs recognition.

4. Theories of mind

Parallel to the neuropsychological research, another view supporting the centrality and reality of mental states attribution suggests that during our development we build *theories of mind* in order to explain the behavior of other subjects (Gopnik 1996, Gopnik and Meltzoff 1997). All children develop different models of the mind's working during their growth, with which they try to explain and predict the behavior of other people. The assumptions and the *rules* we formulate about others' minds are progressively tested with our experience, and then refined, revised or rejected during our development, in a similar fashion as a scientist testing hypothesis and developing theories.

Developmental psychologists since Jean Piaget have tried to determine when and how the ability to 'read' others' minds emerges. Recently it has been argued that the ability to understand others' mental states is based on a *modular* brain function, similar to those that govern perception and language (Leslie 1994, Baron-Cohen 1995). Leslie (1992) suggests that a module (called Theory of Mind Mechanism) allow us to naturally develop 'naïve' psychological explanation of others' behavior and to attribute mental states.

Determining how and when the ability of a child to understand the mind of others appears is a difficult task. First experimental studies have shown that around the age of four there is a remarkable change in the ability to attribute beliefs to others. In a classic experiment on *false beliefs* (Wimmer and Perner 1983), the researcher presented to a child a fictional character (Maxi) hiding a toy in a box and then leaving the room. After Maxi had left, the object was then moved by the experimenter to a different place before the child's eyes. Afterward Maxi went back into the room and the experimenter asked the child to predict where Maxi would probably look for the toy. Adults and children above age four attributed to

Maxi a false belief, and expected that he would search for the toy in the box where it was initially hidden. Children under the age of four years usually answered that Maxi would look in the new hiding place, according to what *they* knew.

It would be a mistake, however, to conclude that these studies show that under the age of four we don't have an understanding of others' mental states. The inability to attribute a (false) belief does not exclude the ability to attribute mental states otherwise. In fact, it has been shown that the understanding of desire and emotional states of others is reached well before the age of three, and also that children, even before they are able to pass the false belief test, are good at forecasting other people's behavior and preferences (Repacholi and Gopnik 1997; Meltzoff 1995; Bellagamba and Tomasello, 1999).

There is some agreement that a nine months old child first reaches a level of behavior's complexity in which we would say that she is able to act intentionally and to take account of the intentionality of others. According to Tomasello (2003), this faculty is linked to the ability to recognize causal relationships in the world. Every child rapidly learns to build a sort of 'map' of the causal regularities of the environment (cf. Gopnik et al. 2004), whose complexity grows from the simple physical causality of directly observable events to hidden causal mechanisms concerning the behavior of others: mental states are in this sense invisible causes that move people.

5. The case of autism

Important clues about the development of the ability to attribute mental states come from autism studies. A peculiar characteristic of autistic patients seems to be their inability to understand and reason about the mental states of other people, making it impossible to develop a proper communicative competence, independently of the development of other cognitive abilities. Autistic children, for example, are unable to pass the false belief test, even at an age at which other children are able to make correct predictions. Moreover, they are not able to understand simple narratives, such as the one presented at the beginning, that require an understanding of mental states, whereas they do not have problems with the description of purely mechanical events (Baron-Cohen, Leslie and Frith 1986).

The clinical research of Leslie (1992, 1994), Baron-Cohen et al. (Baron-Cohen, Leslie and Frith 1985, Baron-Cohen, 1995) have led to the hypothesis that one cause of autism is a defect in those areas of the brain responsible for 'mind reading', like Leslies's Theory of Mind module .

This defect explains the difficulty of the autistic subjects to develop a proper communicative competence, even though they do not have deficiencies in the mastery of the syntactic and semantic rules of a language. What is lacking, is the ability to use a language as a tool to influence and understand the mental states of other people (Leslie 1994). Depending on the severity of the condition, autistic individuals often lack the understanding of indirect communication, like irony, or figurative speech, that require inferences about what a speaker really means beyond what he literally said.

6. Conclusion

Communication requires the ability to reason about the mind of other people. The models of communication that take origin from Grice's suggestions and were developed in its full formal clarity by Posner (1993) are based on the assumption that we are able to keep track of very complex relationships between mental states such as intentions and beliefs. This approach has been put into criticism from two sides, concerning firstly the epistemological reality of mental concepts, and secondly showing how these models seem to be psychologically unrealistic, since they presuppose complex cognitive processes that seem at odds with the ease of human semiotic behavior.

In this short essay I tried to argue that the 'stance' we tend to take in attributing mental states even to fictional or non-human entities is revealing of a faculty of our minds in building complex representations of nested relationships of beliefs and intentions, as the opening example of Aesop's tale showed. The fact that even a simple story requires a complex structure of mental concepts indicates that most part of our reasoning about mental processes is implicit and automatic.

Moreover, recent research on the neurological basis of mind reading and on the development of the ability to make assumption about others' mental states has given growing support to the hypothesis of a biological and evolutionary origin of this faculty, offering an empirical and not only theoretical evidence of Posner's contribution to the foundation of a general theory of semiotics.

References

Arielli, Emanuele 2006: *Cognizione e comunicazione*, Bologna: Il Mulino

Baron-Cohen, Simon 1995: *Mindblindness: An essay on autism and theory of mind*, Cambridge, Mass.: MIT Press

Baron-Cohen, Simon, Alan Leslie & Uta Frith 1985: „Does the Autistic Child Have a Theory of Mind?", in: *Cognition*, 21: 37–46

Baron-Cohen, Simon, Alan Leslie & Uta Frith 1986: "Mechanical, Behavioural and Intentional Understanding of Picture Stories in Autistic Children", in: *British Journal of Developmental Psychology*, 4: 113–125

Bellagamba, Francesca & Michael Tomasello 1999: "Re-enacting Intended Acts: Comparing 12- and 18-Month-Olds", in: *Infant Behavior and Development*, 22: 277–282

Bertram, Bruce 1979: "Analysis of Interacting Plans as a Guide to the Understanding of Story Structure", in: *Poetics*, 9, 1–3: 295–311

Carruthers, Peter & Peter Smith (eds.) 1996: *Theories of Theories of Mind*, Cambridge: Cambridge University Press

Ciaramidaro, Angela, Mauro Adenzato, Ivan Enrici, Susanne Erk, Lorenzo Pia, Bruno G. Bara, Henrik Walter 2007: „The Intentional Network: How the Brain Reads Varieties of Intentions", in: *Neuropsychologia*, 45, 13: 3105–3113

Dennett, Daniel 1987: *The Intentional Stance*, Cambridge, Mass: MIT Press

Fletcher, Paul C., Francesca Happé, Uta Frith, Susan C. Baker, Raymond J. Dolan, Richard S. Frackowiak & Chris D. Frith 1995: "Other Minds in the Brain: A Functional Imaging Study of Theory of Mind in Story Comprehension", in: *Cognition*, 57: 109–128

Frith, Chris D., Uta Frith 2006: „The Neural Basis of Mentalizing", in: *Neuron*, 50, 4: 531–534

Gallese, Vittorio & Alvin I. Goldman 1998: „Mirror Neurons and the Simulation Theory of Mind-Reading", in: *Trends in Cognitive Sciences*, 12:493–501

Goldman, Alvin I. 1993: "The psychology of folk psychology", in: *Behavioral and Brain Sciences*, 16: 15–28

Gopnik, Alison & Andrew N. Meltzoff 1997: *Words, Thoughts and Theories*, Cambridge, Mass.: MIT Press

Gopnik, Alison 1996: "Theories and modules: Creation Myths, Developmental Realities and Neurath's Boat", in Carruthers et al. (eds.) 1996: 169–183

Gopnik, Alison, Clark Glymour, David Sobel, Laura Schulz & Tamara Kushnir 2004: „A Theory of Causal Learning in Children: Causal Maps and Bayes Nets", in: *Psychological Review*, 111: 1–31

Grice, Herbert Paul 1957: "Meaning", in: *Philosophical Review*, 66: 377–388

Jorna, René, Barend van Heusden & Roland Posner (eds.) 1993: *Signs, Search and Communication*, Berlin: de Gruyter

Leslie, Alan M. 1992: „Pretense, Autism, and the Theory-of-Mind Module", in: *Current Directions in Psychological Science*, 1: 18–21

Leslie, Alan M. 1994: „Pretending and Believing: Issues in the Theory of Theory-of-Mind-Module", in: *Cognition*, 50: 211–238

Meggle, Georg 1981: *Grundbegriffe der Kommunikation*, Berlin: de Gruyter

Meltzoff, Andrew N. 1995: „Understanding the Intentions of Others: Re-enactment of Intended Acts by 18-Month-Old Children", in: *Developmental Psychology*, 31: 1–16

Posner, Roland 1993: „Believing, Causing, Intending: The Basis for a Hierarchy of Sign Concepts in the Reconstruction of Communication", in: Jorna, van Heusden & Posner (eds.) 1993: 215–270

Repacholi, Betty M. & Alison Gopnik 1997: „Early Reasoning About Desires: Evidence from 14- and 18-Month-Olds", in: *Developmental Psychology*, 33: 12–21

Rizzolatti, Giacomo & Michael A. Arbib 1998:" Language Within Our Grasp", in: *Trends in the Neurosciences*, 21: 188–194

Rizzolatti, Giacomo., Leonardo Fogassi, Vittorio Gallese 2001: „Neurophysiological Mechanisms Underlying the Understanding and Imitation of Action", in: *Nature Reviews Neuroscience*, 2: 661–670

Ryan, Marie-Laure 1991: *Possible Worlds, Artificial Intelligence and Narrative Theory*, Bloomington: Indiana University Press

Singer, Tania 2006: „The Neuronal Basis and Ontogeny of Empathy and Mind Reading: Review of Literature and Implications for Future Research", in: *Neuroscience & Biobehavioral Reviews*, 30, 6: 855–863

Tomasello, Michael 2003: *Constructing a Language: A Usage-Based Theory of Language Acquisition*, Cambridge, Mass: Harvard University Press

Wimmer, Heinz & Joseph Perner 1983: „Beliefs About Beliefs: Representations and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception", in: *Cognition*, 13: 103–128