

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Neuroscience and Biobehavioral Reviews xx (2005) 1–19

NEUROSCIENCE AND
BIOBEHAVIORAL
REVIEWSwww.elsevier.com/locate/neubiorev

Review

Theory of mind—evolution, ontogeny, brain mechanisms and psychopathology

Martin Brüne*, Ute Brüne-Cohrs

Center for Psychiatry and Psychotherapy, University of Bochum, Alexandrinenstr. 1-3, 44791 Bochum, Germany

Received 21 April 2005; revised 2 August 2005; accepted 8 August 2005

Abstract

The ability to infer other persons' mental states and emotions has been termed 'theory of mind'. It represents an evolved psychological capacity most highly developed in humans. The evolutionary origins of theory of mind can be traced back in extant non-human primates; theory of mind probably emerged as an adaptive response to increasingly complex primate social interaction. This sophisticated 'metacognitive' ability comes, however, at an evolutionary cost, reflected in a broad spectrum of psychopathological conditions. Extensive research into autistic spectrum disorders has revealed that theory of mind may be selectively impaired, leaving other cognitive faculties intact. Recent studies have shown that observed deficits in theory of mind task performance are part of a broad range of symptoms in schizophrenia, bipolar affective disorder, some forms of dementia, 'psychopathy' and in other psychiatric disorders. This article reviews the evolutionary psychology of theory of mind including its ontogeny and representation in the central nervous system, and studies of theory of mind in psychopathological conditions.

© 2005 Published by Elsevier Ltd.

Keywords: Theory of mind; Human evolution; Child development; Brain mechanisms of theory of mind; Psychopathology

Contents

1. Introduction	000
2. Theory of mind—adaptation to social complexity?	000
3. Ontogeny of theory of mind	000
4. CNS-representation of theory of mind	000
5. Testing theory of mind	000
6. Psychopathology of theory of mind	000
7. Developmental disorders	000
8. Personality disorders and other non-psychotic disorders	000
9. Schizophrenia and affective disorders	000
10. Brain damage and degenerative brain disorders	000
11. Discussion	000
References	000

1. Introduction

The term 'theory of mind' was originally proposed by primatologists Premack and Woodruff in a seminal article to suggest that chimpanzees may be capable of inferring mental states of their con-specifics (individuals of the same species) (Premack and Woodruff, 1978). Later on, the term was adopted by child psychologists to describe

* Corresponding author. Tel.: +49 234 5077155; fax: +49 234 5077235.
E-mail address: martin.bruene@ruhr-uni-bochum.de (M. Brüne).

the ontogenetic development of mental perspective taking in infants and young children (e.g. Leslie, 1987). In terms of psychopathology, the concept of a disturbed theory of mind has become increasingly influential to explain behavioral symptoms in children with autistic spectrum disorders (Baron-Cohen et al., 1985). It is now widely acknowledged and well buttressed by numerous empirical studies that autistic children and adults with Asperger's syndrome, a mild form of autism, have profound difficulties in appreciating the mental states of other individuals (e.g. Baron-Cohen, 1988, 1991; Baron-Cohen et al., 1997, 2001a; Buitelaar et al., 1999). Such deficits in mental state comprehension have been shown to be selective, that is, other cognitive capacities or 'non-social' intelligence may well be preserved in people with autism (Baron-Cohen et al., 1986; Baron-Cohen, 1991). The situation with other psychopathological conditions and psychiatric disorders is less clear. There is, however, growing evidence that impaired theory of mind may also lie at the core of certain psychotic symptoms in 'endogenous' psychoses and behavioral deviations found in heterogeneous disorders affecting frontal lobe functioning—from psychopathy to frontotemporal dementia.

Although somewhat speculative, it is conceivable that the strong desire inherent to human nature to attribute agency—we sometimes even ascribe intentions to inanimate objects—renders the cognitive faculty of theory of mind vulnerable to dysfunction. In other words, if theory of mind, as suggested here, is so central to human life, any functional impairment or structural disruption of the underlying neural substrates of this recently evolved cognitive capacity could be detrimental to social functioning.

In this article, we seek to review the evolutionary background, the ontogenetic development and the evidence for selective disorders of theory of mind in psychopathological conditions. Before doing so, we want to emphasize that theory of mind only represents one particular aspect of what has been labeled 'social cognition' (Brothers, 1990; Adolphs, 2001). The perception of social signals, motivation, emotion, attention, memory and decision-making equally contribute to the actual behavioral output in social interaction. As Adolphs (2001) has pointed out, the components and boundaries of social cognition are to a great deal ill defined. For the sake of clarity and space, we consider it necessary to narrow the view on theory of mind—acknowledging that in 'real-life' situations theory of mind is entrenched in a neural network that constitutes the 'social brain' of human and non-human primates (Dunbar, 2003).

2. Theory of mind—adaptation to social complexity?

In 1966 and 1976, respectively, Jolly and Humphrey argued independently of each other that primates have excess cognitive capacities beyond the needs for everyday

feeding and ranging. They suggested that it has been the social environment that primarily put evolutionary pressure on brain development in primates (Jolly, 1966; Humphrey, 1976). Indeed, primates are essentially gregarious animals, and group living certainly confers adaptive advantages on the individual such as better protection from predation and food sharing (Alexander, 1987). On the other hand, group living incurs the cost of directly competing for resources and sexual partners. This situation may have created specific selective pressures in primates to evolve 'social intelligence' (Whiten, 2000). The fact that primates (and not other taxa) have taken social intelligence so far may lie in the fact that their world is largely vision-dominated (Dunbar, 1998). Crucial in the context of primate group living with strong mutual dependency and complex interactions is the ability of individuals to identify others who cooperate and, even more importantly, who try to defect. That is, if an individual trusts that cooperation will be reciprocated, cheating could be an even more successful strategy for another subject. Thus, to counteract cheating one must be able to detect deception (Trivers, 1971). Dawkins and Krebs (1979) have reasoned that evolutionary arms races took place between as well as within species. This concept can well be expanded to include cognitive skills in primate species. Indeed, there is more than mere plausibility behind this logic. In an intriguing series of experiments, Cosmides (1989) was able to demonstrate that the performance of human test subjects on a reasoning task originally developed by Wason (1966) increased if the abstract conditional rules were replaced by a social scenario. For example, if subjects were shown four cards displaying 'A', 'B', '2', and '3' on one side, and were asked, which of the four cards had to be turned over to prove the conditional rule 'if there is a vowel on one side, then there is an even number on the other side', the majority failed to answer the question correctly. If, however, the cards displayed the words 'beer', 'coke', '16', and '25', and the rule was 'if a person drinks alcohol, then she must be over 18 years old', the solution was much easier (one needs to turn the 'beer' card and the '16' card to determine the potential violation of the rule). In other words, recasting the task into a social contract (Cosmides, 1989) improves participants' performance, perhaps because evolution favored the cognitive ability of cheating detection. In support of the evolutionary explanation, Sugiyama et al. (2002) found cross-cultural evidence for the existence of a universal cheater-detecting mechanism. Individuals of the Ecuadorian Shiwiar, non-literate hunter-horticulturalists, were as skillful at cheating detection in a modified Wason selection task as subjects from developed countries (Sugiyama et al., 2002). This and other studies based on evolutionary game theory suggest that social intelligence, including theory of mind could have evolved in order to facilitate cheating detection, and, perhaps more important for ancestral human societies, to reinforce cooperation. In the classical 'Prisoner's Dilemma' (Axelrod and Hamilton, 1981), two hypothetical suspects apprehended at the scene

of a crime, who are interrogated separately, have the option to cooperate ('it was neither of us'), to defect ('it was him'), or to confess ('it was me'). Depending on the expected punishment (e.g. 1 year in prison if both cooperate, 4 years if both defect, 5 years for the cooperator, if the other defects, who himself would escape punishment), both suspects face the dilemma whether to cooperate or to defect. If repeatedly played, the best strategy for the Prisoner's Dilemma is perhaps 'tit-for-tat', that is, to respond to defection by defecting, and to cooperate in response to cooperation (see Nowak and Sigmund, 1993). The problem of altruistic behavior and cooperation has led Trivers (1971) to suggest that in humans several psychological mechanisms evolved to protect against cheating and to reinforce cooperation, clearly including what was later called 'theory of mind'. Recent research has elaborated on how humans enforce social norms and cooperation by means that are largely not selfishly motivated (Fehr and Fischbacher, 2004). This review cannot do justice to the vast amount of literature accumulated on evolutionary game theory. For our purpose it may be sufficient to emphasize the intimate tie of human cooperation and deception with the cognitive capacity of inferring the mental states of putative allies or competitors. Interestingly, recent studies using functional brain imaging have confirmed that the brain areas activated during performance of a Prisoner's Dilemma Game and games involving reciprocal exchange remarkably overlap with the areas activated by theory of mind tasks (McCabe et al., 2001; Rilling et al., 2004).

Anatomical support of the social intelligence hypothesis comes from a comparison of brain size in primates. A simple benefit-cost-calculation suggests that there must be an explanation for the obvious disparity between a brain weight of about 2% of the body weight in adult humans and 20% utilization of the energy intake (Aiello and Wheeler, 1995). Such a costly organ in energetic and developmental terms must convey a certain advantage—otherwise, it would never evolve. Moreover, primates have big brains relative to body size, and in humans, the neocortex is three times greater and much more convoluted than expected for a primate of our brain size (Rilling and Insel, 1999). So the question is, does primate brain size primarily reflect their social intelligence? Dunbar has suggested the following equation: if, for example, the demand of information processing capacity (i.e. brain size) increases with the number of possible relationships, and if the average group size of different primate species serves as a measure of social complexity, brain size would be expected to correlate with group size in the respective species (Dunbar, 1998). In fact, statistical analysis revealed that there is a link of group size and the size of the neocortex in primates when the primary visual cortex V1 is excluded (because the size of the visual cortex is relatively stable in different primate species). With regard to humans (in ancestral conditions) an extrapolation would predict an average number of about 150 individuals with whom someone would have a personal

relationship. This strikingly matches the data of ethnological studies in hunter-gatherer societies and even in modern societies, where the average number of personal acquaintances is indeed about 150 people (Dunbar, 2003).

It is, on the other hand, at first sight perplexing that the neat correlation of group size with neocortex size does not fit for great apes. They usually live in much smaller groups than predicted. However, Byrne and colleagues could show a correlation of the neocortex ratio (defined as the ratio of the volume of the neocortex compared to the volume of the rest of the brain) with the number of complex social manipulations in different primate species—commonly referred to as 'tactical deception' (Byrne, 2003). Not only did tactical deception clearly occur more frequently in chimpanzees compared with any other primate species; equally important was the fact that the observation of tactical deception could only in chimpanzees be interpreted in a way that suggested evidence of mental perspective taking, i.e. a theory of mind (Byrne, 2003). It can therefore be concluded, whether we like the idea or not, that complex social interactions between individuals and the need to be capable of distinguishing between 'sincere' cooperation and defection has been a major driving force in the evolution of primate and human cognition.

An important question is, when exactly in primate evolution did theory of mind evolve. At this stage, we do not know the answer. However, as already mentioned, to some extent, we can trace back the evolution of theory of mind by studying our closest relatives. Behavioral observation of chimpanzees in the wild and in captivity suggests that they possess a capacity for deliberate coalition-forming and strategic deception (De Waal, 1982; Whiten and Byrne, 1997). Other behavioral patterns that possibly require theory of mind such as teaching, however, inconsistently occur in wild chimpanzees (Byrne, 1995). In experimental conditions, there is robust evidence for mirror self-recognition, symbolic representation and at least visual perspective taking in chimpanzees (Suddendorf and Whiten, 2001). Whether this is indicative of true theory of mind is still under dispute (Heyes, 1998; Tomasello et al., 2003; Povinelli and Vonk, 2003). While tactical deception and Premack and Woodruff's (1978) classic experiments indicate at least some evidence of theory of mind abilities in chimpanzees, a more parsimonious explanation could be that, although important precursors for theory of mind, such capacities could perhaps better be interpreted as species-specific means to predict the behavior of con-specifics without necessarily implying mental perspective taking (Tomasello et al., 2003).

In any event, there is a costly side of the coin of having the ability to mentalize. Advantageous as theory of mind may be for successful social interaction (thus, from an evolutionary perspective, increasing an individual's inclusive fitness), at the individual level it comes at the expense of slow maturation, and hence, late reproduction. Maximizing reproductive success should rather favor early sexual

maturation (and large litter size; Joffe, 1997). Primates, however, are extreme K-strategists, that is, their offspring grows slowly, multiple births are unusual, and birth intervals are long. Moreover, the already considerable extension of the juvenile period in primates reaches a maximum in humans. Interesting in this regard is the fact that the length of the juvenile period in primates is also positively correlated with the size of the non-visual cortex in the same way as group size is; it does not correlate with the length of gestation, lactation, and reproductive life span. This finding could be interpreted as supporting a relation of slow maturation to constraints of the social environment (Joffe, 1997). For example, the extension of the juvenile period in primates may have been crucial to acquire the vast amount of possible social behavioral ‘strategies’ (procedural rules) and when to employ these strategies (here, the term ‘strategy’ does not necessarily imply conscious awareness; Schmitt and Grammer, 1997). This process is not merely time-consuming. The real-life opportunities of testing possible consequences of such social strategies are limited in number. It is, therefore, conceivable that the need for mental simulation of social interaction might have speeded up the evolution of theory of mind. If mental simulation is involved (see below), then theory of mind not only comprises the representation of the mental states of other individuals, but also one’s own mental state (attachment theorists have termed this ability ‘reflective functioning’; Fonagy, 1997).

3. Ontogeny of theory of mind

At birth, human infants are essentially immature. The growth of the human brain extends well into the postnatal period. At the cellular level, synaptic pruning and myelination even takes place until after puberty and adolescence (De Bellis et al., 2001; Levitt, 2003).

In principle, the ontogeny of the theory of mind faculty does not so much differ from the maturation of other brain functions—just as an infant is not capable of jumping before sitting, standing and walking, the ability of appreciating one’s own and other’s mental states follows a distinct sequence of acquisition. Baron-Cohen has described a developmental model of a theory of mind mechanism (he usually prefers the notion of a ‘theory of mind’ module; Baron-Cohen, 1995). Here, we primarily refer to the modularity hypothesis of theory of mind, not only for didactical reasons, but also, because almost certainly there is an innate ‘hard-wired’ foundation of the theory of mind faculty (see below).

Very early in life, around six months of age, the human infant distinguishes between the motion of inanimate and animate objects. At about 12 months the infant develops the ability of what has been called joint attention. Joint attention refers to the cognitive capacity to form a triadic representation involving the infant’s own perception,

the perception of an agent, for example, her mother, and an object, at least if the object is within the infant’s sight. At the age of about 14–18 months, the human infant is able to turn its head into the direction the gaze of an agent suggests an object to be, and the infant begins to understand the mental states of desire, intention, and the causal relation between a person’s emotions and goals (Saxe et al., 2004). Between 18 and 24 months of age toddlers discover the difference between reality and pretense. This involves what Leslie has termed ‘decoupling’ (Leslie, 1987). The infant can distinguish between the representation of a real event and the representation of a hypothetical state (such as a thought) and starts to engage in ‘pretend play’. At about the same time the infant learns to recognize him- or her in a mirror (as per the notion above that chimpanzees are capable of mirror self-recognition, it may be worth mentioning that it would be an oversimplification to conclude that ontogeny merely recapitulates phylogeny; Povinelli, 1993). Not until the age of 3–4 years, however, is a child able to distinguish between his or her own and others’ beliefs and knowledge of the world, for example, that someone may hold false beliefs. Five to six year olds understand that someone can hold beliefs about another person’s beliefs (Wimmer and Perner, 1983; Perner and Wimmer, 1985). However, a recent study showed that there is still considerable instability of understanding false beliefs in 5-year-olds, especially when the false belief scenario is framed in relation to a person’s volitional (apparently less well predictable) action rather than a physical object as in the standard test described below (Rai and Mitchell, 2004).

Metaphor and irony comprehension entails the capacity to go beyond the literal meaning of an utterance, and children do not understand metaphor or irony before the age of six to seven (Ackerman, 1981). Likewise, they cannot reliably distinguish jokes from lies before age six to seven years (Sullivan et al., 1995). Even more complex is the comprehension of a ‘faux pas’ situation. A faux pas happens when a person says something she should not have said, not grasping her mistake. Understanding faux pas requires a developmentally advanced theory of mind ability because it requires simultaneous representation of two mental states: The perspective of the person who commits the faux pas, and the representation of the second-person involved who may feel hurt or irritated. ‘Faux pas’ may not be reliably understood before the age of 9 to 11 years (Baron-Cohen et al., 1999).

As pointed out above, the modular perspective of the ontogeny of theory of mind implies quite a static and inflexible development of this cognitive capacity. However, there is considerable impact of the social environment on the development of theory of mind skills in infants and children leading to recognizable individual differences. As Carpendale and Lewis (2004) have highlighted, young children apparently acquire theory of mind abilities at an earlier age if their parents frequently use expressions referring to mental states when talking to their infants

compared with children whose parents use such terms less often. In addition, the presence of older siblings speeds up young children's appreciation of other minds (overview in [Carpendale and Lewis, 2004](#)). Furthermore, it is noteworthy that, predictably from the evolutionary framework outlined above, these developmental steps of theory of mind constitute a human universal. Although cross-cultural evidence is still limited, [Avis and Harris \(1991\)](#) have clearly shown that understanding false belief emerges at a similar age in children of the Baka, preliterate hunter-gatherers in southeast Cameroon.

Finally, it is noteworthy that the development of theory of mind is clearly paralleled by language acquisition. In fact, understanding a speaker's intention is a precondition of learning new words. As [Frith and Frith](#) have pointed out, random associations of utterances with objects rarely occur when young children learn to speak ([Frith and Frith, 2003](#)) and a child begins to use words undoubtedly referring to mental states such as 'I think' at the age of four—the watershed of distinguishing between own and other's mental states.

In contrast to our growing understanding of children's acquisition of theory of mind abilities, fairly little is known about the development of theory of mind in adult humans. Due to the fundamental role of subjective experience and recall of past social interactions in theory of mind performance, we would expect a continuous refinement of mental state attribution throughout the adult human life span. On the other hand, selection pressure declines with age (particularly with respect to the post-reproductive life span). It is therefore conceivable that aging does not spare social cognitive abilities. Two studies of theory of mind abilities in older people have revealed conflicting results. [Happé et al. \(1998\)](#) found that people with a mean age of 73 years, although slower in performance, were superior on a theory of mind task compared to adolescents and young adults of about 14 years and 22 years of age, respectively. In contrast, a recent study revealed the opposite, namely a successive decline in theory of mind in adults aged between 60 and 74, and between 75 and 89, respectively, compared to younger adults ([Maylor et al., 2002](#)). Thus, at this stage there is still controversy whether and how theory of mind capacities change over the adult human life span.

4. CNS-representation of theory of mind

If primate brains, particularly neocortical structures, enlarged over evolutionary time due to selection pressures from the social environment, where exactly is theory of mind located in the human brain? Evidence comes from various sources. Comparative neuroanatomy and neurophysiology informs us which brain areas and corresponding functions came under selection pressure in non-human primates to evolve into the neural correlates of theory of mind in modern humans. In addition, functional brain

imaging studies and lesion studies in patients suffering from brain injuries or stroke may help localizing the brain circuits underlying theory of mind.

Before summarizing some of the most important empirical studies, it is necessary to point out that divergent theoretical conceptualizations of theory of mind exist. To some degree, this has considerable impact on how empirical findings are interpreted. (1) Drawing on [Fodor's \(1983\)](#) concept of a modular organization of the human mind, some theorists advocate the existence of a separate theory of mind module (e.g. [Scholl and Leslie, 1999](#)). Like other domain-specific cognitive capacities represented in the brain, which process only a certain class of information, the theory of mind mechanism is supposed to process information restricted to social inference. Cognitive mechanisms are assumed to work reliably, efficiently, and economically. According to the modular hypothesis, the development of theory of mind mainly depends on neurological maturation of the brain structures involved. Experience, on the contrary, may trigger the action of the theory of mind mechanism, but does not determine the makeup of the mechanism. (2) The 'metarepresentational' theory-theory (e.g. [Perner, 1991](#)) of theory of mind is somewhat distinct from the modular model. As [Flavell \(1999\)](#) has summarized, the theory-theory proposal holds, similar as the modular theory does, that the entities and the causal principles of theory of mind are specific (e.g. beliefs, desires, thoughts). Furthermore, the ingredients of the theory of mind faculty are interconnected, e.g. we are able to recognize that what we believe has impact on what we perceive ([Flavell, 1999](#)). In contrast to the modular theory, however, the theory-theory account ascribes a greater role of individual experience to the developmental formation of the theory of mind faculty by providing input for revising and reorganizing existing formations. (3) The simulation theory proposes that theory of mind relates to the ability to imaginatively 'put oneself into the shoes' of others (e.g. [Davies and Stone, 1995](#)). In contrast to the above-mentioned accounts of theory of mind, the simulation theory posits that appreciating other persons' mental states critically depends on introspection. Similar to the theory-theory model, the simulation hypothesis stresses the importance of experience for the modeling of theory of mind skills. As [Flavell \(1999\)](#) has emphasized, a concise theory of theory of mind has to converge on these models. Cross-cultural studies and the largely invariable developmental trajectories of theory of mind acquisition during infancy and childhood suggest that this cognitive faculty is, at least to some extent, innate; theory of mind also draws on introspection; it can be conceptualized as an 'informal' theory of mental states; memory, language and inhibitory control improve theory of mind; they are, however, not necessarily engaged in belief attribution and distinctly represented in the brain ([Saxe et al., 2004](#)); experience modifies theory of mind abilities ([Flavell, 1999](#)).

Now, what can we learn from primate research about theory of mind, in light of the fact that there is no unequivocal evidence of mental state attribution in non-human primates in general, and a virtual absence of theory of mind in monkeys? Single cell recordings in non-human primates convey important information about candidate cerebral representations of cognitive precursor capacities of what we call ‘true’ theory of mind in humans (the term ‘precursor capacities’ by no means ought to suggest a teleological interpretation, i.e. that something evolves in order to later suit a certain purpose).

A number of candidate structures have been identified in non-human primate brains that have undergone adaptive modifications to constitute in humans a neural network of theory of mind. Single cell recordings in macaque monkeys have revealed that neurons in the middle portion of the temporal lobe, particularly in the superior temporal sulcus (STS), selectively fire when monkeys observe the gaze direction of other monkeys. These neurons are also active when the animals observe goal-directed behavior (Gallese and Goldman, 1998). In humans, functional brain imaging studies have revealed that a homologous area of the temporal lobe is activated by observation of seemingly purposeful movements of inanimate objects (as opposed to random movements), and even when still photographs depict ‘implied’ motion (Kourtzi and Kanwisher, 2000). For example, such activity could be elicited by showing human subjects pictures of a discus thrower in action, whereas no such activity could be measured when the discus thrower was at rest. Activity in parts of the STS, therefore, is linked to the observation of intentional movements. Although this does not imply conscious awareness, the representation of ‘intentions’ is certainly a critical aspect of theory of mind. In a variety of functional imaging studies during theory of mind task performance the blood flow increased in an area of the STS adjacent to the part that was activated by monitoring biological motion (Grossman and Blake, 2002).

The temporal lobes of non-human primates also contain a specific type of cells called ‘mirror neurons’ due to their unique quality to discharge during both the execution of a certain hand or mouth action or by the mere observation of the same behavior carried out by another individual. These neurons have also been found in greater density in the ventral premotor cortex of macaque monkeys, an area that is possibly homologous to the Broca area in humans (Gallese and Goldman, 1998). In an ingenious series of experiments, the group of Rizzolatti has demonstrated that mirror neurons selectively fire when monkeys observe a hand movement of which the terminal part is hidden from their view. In other words, a subset of mirror neurons is active when the monkey can only ‘infer’ or predict the result of the incompletely visible action (Umiltà et al., 2001). Mirror neurons may therefore be crucially involved in understanding action-goal states. In humans, Fadiga et al. (1995) have shown in an experiment using transcranial magnetic stimulation (TMS) that the observation of a goal-directed hand movement

elicited enhanced motor evoked potentials (MEP). Notably, these enhanced MEPs could be measured precisely in those muscles the observer would use when carrying out the action himself.

The discovery of mirror neurons in humans offers an explanation of how the ability to imitate the actions of others could have evolved into the capacity to simulate the mental states of other individuals (i.e. theory of mind) (Williams et al., 2001). However, as Frith and Frith (1999, 2001) have pointed out, for theory of mind it is not sufficient to represent goal-directed actions. It is also necessary to be able to distinguish between behavior generated by self or others. And indeed, there are at least two other important brain regions involved in theory of mind. We believe that simulating other people’s mental states does not necessarily involve conscious reflection, but is readily available to conscious awareness. For example, transference and counter-transference in dyadic psychotherapeutic settings always implicate the mutual, largely unconscious attribution of mental states such as intentions, desires and beliefs, and it is the goal of psychodynamic approaches to unveil these unconscious processes and to make them accessible to the conscious mind. For conscious reflection on one’s own and other’s mental states an individual needs computational resources beyond the capacity for imitation and action simulation, and a candidate structure involved in this task is the inferior parietal cortex. Recent research using functional brain imaging has revealed that the left and right hemisphere are differentially involved in first versus third-person perspective. First-person perspective was shown to activate the left inferior parietal cortex, whereas third-person perspective activated the corresponding region on the right side of the human brain (Ruby and Decety, 2001). Interestingly, when a subject imitates the action of another person, more activation is found in the left inferior parietal cortex, but more activation is found on the opposite side when subjects view their actions being imitated. These experimental results support the assumption that the right inferior parietal cortex may be critical for consciously representing others’ minds, whereas the left inferior parietal cortex may be involved in representing one’s own mental states (Decety and Chaminade, 2005).

The other brain area that has consistently been shown to be engaged in theory of mind is the anterior cingulate cortex (ACC). The ACC receives input from the motor cortex and the spinal cord, from the ipsilateral prefrontal cortex, and from the thalamus and brainstem nuclei (Paus, 2001). It is highly heterogeneous in terms of its cytoarchitecture and functional organization. The ACC is now conceived of as an important mediator of motor control, cognition, and arousal regulation (Paus, 2001). In monkeys, for example, the most rostral part of the ACC is active prior to the execution of self-initiated movements (Frith and Frith, 1999). Most interesting from an evolutionary viewpoint and with respect to theory of mind is that the anterior ACC inconsistently forms a paracingulate sulcus

Table 1
Overview of brain imaging studies of theory of mind in chronological order

Author(s); published	Sample (<i>n</i>)	Mean age	Sex m/f	Brain imaging technique	ToM method/tasks	Activated brain areas in ToM tasks
Goel et al., 1995	9 healthy subjects	24.7	5/5	PET [¹⁵ O]H ₂ O	Presentation of familiar and unfamiliar objects requiring inference of others' attribution of their function (i.e. ToM). One non-ToM condition involving inference of function of unfamiliar objects from their form. Two control conditions: visual and semantic attributes of known objects.	The left medial frontal lobe (approx. BA 9) and the left temporal lobe (approx. BAs 21, 39/19 and 38) were specifically activated by the theory of mind condition.
Fletcher et al., 1995	6 healthy subjects	38	6/0	PET [¹⁵ O]H ₂ O	Story comprehension tasks necessitating attribution of mental states. Two control tasks: 'physical' stories not requiring mental state attribution and passages of unlinked sentences.	The left medial frontal gyrus (approx. BA 8) and the posterior cingulate cortex were activated only in the mental state attribution condition.
Happé et al., 1996	5 patients with Asperger syndrome and normal intellectual functioning	24	5/0	PET	Story comprehension tasks necessitating attribution of mental states. Two control tasks: 'physical' stories not requiring mental attribution and passages of unlinked sentences.	Patients with Asperger syndrome activated an area of the medial prefrontal cortex, which was more ventrally located compared to the area activated in control subjects; no further group difference of activation pattern.
Gallagher et al., 2000	6 healthy subjects	30	5/1	fMRI	Comparing a story task and a cartoon tasks, both requiring theory of mind.	Both conditions activated the medial prefrontal cortex (specifically the paracingulate gyrus).
Brunet et al., 2000	8 healthy subjects	23.3	8/0	PET [¹⁵ O]H ₂ O	Nonverbal comic strips in three different conditions involving attribution of intentions to story characters, physical logic and knowledge about objects' properties.	During mental state attribution the right middle and medial prefrontal cortex (incl. BA 9), the right inferior prefrontal cortex (BA 47), the right inferior temporal gyrus (BA 20), the left superior temporal gyrus (BA 38), the left cerebellum, the bilateral anterior cingulate, and the middle temporal gyri (BA 21) were activated.
Russell et al., 2000	5 schizophrenic patients	36	5/0	fMRI	Reading the mind in the eyes test.	Increased signal response in the left inferior frontal gyrus (BA 44/45/46) and medial frontal lobes (BA 45/9), left middle and superior temporal gyrus (BA 21/22) in healthy subjects. Reduced activation in the middle/inferior frontal cortex (BA 9/44/45) in schizophrenia patients.
McCabe et al., 2001	7 controls 12 healthy subjects	40 n.m.	7/0 n.m.	fMRI	Two-person 'trust and reciprocity' games with human or computer counterparts for cash reward.	Within the group of cooperators prefrontal regions were more active when they were playing a human, rather than when playing a computer. No significant differences in activation within the group of non-cooperators.
Vogele et al., 2001	8 healthy subjects	25–36	8/0	fMRI	ToM stories, physical stories and unlinked sentences, and 'self and other ascription' stories and 'self ascription' stories.	ToM tasks led to increased neural activity in the anterior cingulate gyrus whereas 'self' led to increased activity in the temporoparietal junction and anterior cingulate cortex. There was a significant interaction of self perspective and ToM in the right lateral prefrontal cortex.
Calder et al., 2002	9 healthy subjects	58.3	9/0	PET [¹⁵ O]H ₂ O	Three conditions of eye gaze direction: 100% direct, 50% direct-50% averted, and 100% horizontally averted at or away from participant.	There was a linear relationship between increasing proportions of horizontally averted eye gaze and increased rCBF in the medial prefrontal cortex (approx. BA 8/9). Increasing proportions of direct eye gaze were associated with increased rCBF particularly in the right middle and superior temporal gyri.

Table 1 (continued)

Author(s); published	Sample (<i>n</i>)	Mean age	Sex m/f	Brain imaging technique	ToM method/tasks	Activated brain areas in ToM tasks
Ferstl and von Cramon, 2002	9 healthy subjects	24	5/4	fMRI	Presentation of related and unrelated sentence pairs requiring logical explanation or ToM processing.	The frontomedian cortex was activated in coherent and non-coherent trials when ToM instructions were given, and in coherent but not in non-coherent trials when logical/non-ToM instructions were given.
Brunet et al., 2003	7 patients with schizophrenia	31	7/0	PET [¹⁵ O]H ₂ O	Picture stories involving attribution of intentions by selecting one of three options depicting the logical ending of the story.	In contrast to controls, the schizophrenic patients did not show activation of the right prefrontal cortex during attribution of intentions.
Calarge et al., 2003	8 controls 13 healthy volunteers	23.3 26.5	8/0 6/7	PET [¹⁵ O]H ₂ O	Making up a ToM story about a given scenario.	A complex activation pattern was found comprising the medial frontal cortex, superior and inferior frontal regions, the paracingulate gyrus, the cingulate gyrus, the angular gyrus, the anterior pole of the temporal lobe, and the right cerebellum, predominantly on the left.
Nieminen-von Wendt et al., 2003	8 subjects with Asperger's syndrome	28.1	8/0	PET [¹⁵ O]H ₂ O	Auditorily given ToM stories and physical (control) stories.	During ToM tasks both groups showed increased activation in the occipitotemporal area bilaterally, the right temporal lobe, the thalamus, and the midbrain. The activation in the medial prefrontal area was more intensive and extensive in the control group.
Saxe and Kanwisher, 2003	8 controls 25 healthy subjects	31.5	8/0 13/12	fMRI	Visually presented stories of false belief, mechanical inference, human action and nonhuman objects.	There was activation of the temporoparietal junction bilaterally only during tasks requiring reasoning about the content of mental states. The left temporoparietal junction was activated during presentation of photographs as well as objects, whereas the right temporoparietal junction showed a trend towards greater activity during presentation of people.
Walter et al., 2004	subgroup of 14 healthy subjects 13 healthy subjects		7/7 6/7	fMRI	Whole body photographs in a range of postures, and inanimate objects. Comic strips requiring understanding of a person's intention in a social interaction or a person's intention in a non-social action.	The anterior paracingulate gyrus was activated by intentional social interactions and in the prospective social intention condition, but not by intentional physical action, whereas in all conditions the anterior cingulate and the superior temporal sulcus were activated.
Grezes et al., 2004	12 healthy subjects 6 healthy subjects	24.75 25–39	6/6 4/2	fMRI	A second condition with future intentional social interaction. Presentation of videotapes of the subjects themselves and other actors/participants lifting and carrying a box of different weights. Then subjects had to judge whether the actor had the correct or false expectation of the weight by their nonverbal behaviour.	Contrasting perception of one's own action with actions of others showed activation in the dorsal premotor cortex, the left frontal operculum, the left intraparietal sulcus and the left cerebellum, which occurred earlier for perception of one's own action compared with the actions of others.

Rilling et al., 2004	19 healthy subjects	28.1	8/11	fMRI	Examination of the main effect of receiving feedback from a human partner or a computer in the Ultimatum Game and Prisoner's Dilemma Game (PDG).	The dorsomedial prefrontal cortex and the rostral anterior cingulate gyrus were activated only in competition with human partners but not computers in the Ultimatum Game. In both tasks only human partners activated the right mid STS, a region spanning the hypothalamus, ventral thalamus and midbrain, and central regions of the hippocampus. In the PDG computer partners also activated the anterior paracingulate cortex, the right STS, thalamus and the left lingual gyrus.
German et al., 2004	16 healthy subjects	18–29	8/8	fMRI	Video clips with actors performing simple everyday actions or pretending to perform a similar set of actions under covert conditions.	There was activation of the medial prefrontal area (approx. BA 9/6/32, 9, 10), the inferior frontal gyrus bilaterally (approx. BA 44, 47), the temporo-parietal regions (approx. BA 21, 22), and of the parahippocampal areas including the amygdala when subjects viewed pretended actions.

BA, Brodmann area; fMRI, functional magnetic resonance imaging; n.m., not mentioned; PET, positron emission tomography; rCBF, regional cerebral blood flow; STS, superior temporal sulcus; ToM, theory of mind.

that is present in only 30–50% of individuals and possibly still under selection pressure (Paus, 2001). This area consistently ‘lights up’ in functional brain imaging studies during theory of mind task performance (Gallagher and Frith, 2003). Moreover, the ACC of apes and humans contains a spindle shaped cell type (thus termed ‘spindle cells’) unique to apes and humans. Spindle cells have not been found in monkeys, and the density of spindle cells in the ACC of apes correlates inversely with the species’ genetic distance from humans. That is, the density is lowest in orang utans, intermediate in gorillas, higher in chimpanzees and highest in humans (Nimchinsky et al., 1999). The exact function of spindle cells is as yet not known. However, in light of what has been said about social complexity and social intelligence in primates we would speculate, in line with Frith and Frith (2003) that spindle cells evolved to gain inhibitory control. ‘Voluntary’ suppression of any immediate response in social interaction and reward delay may also relate to the execution of tactical decision.

As to our present knowledge, it is most likely that theory of mind involves a neural network including the temporal lobes, the inferior parietal cortex, and the frontal lobes (Vogeley et al., 2001; overview in Gallagher and Frith, 2003; Frith and Frith, 2003; Saxe and Kanwisher, 2003).

The functional brain imaging studies of theory of mind are summarized in Table 1.

5. Testing theory of mind

The ‘gold standard test’ of comprehending other persons’ minds is to grasp that others can hold false beliefs that are different from one’s own (correct) knowledge (Dennett, 1978). The classic ‘Sally-and-Anne-Test’ (Wimmer and Perner, 1983) experimentally creates a situation in which a test person has to distinguish his or her own knowledge that an object has been hidden by one character (Anne) in the absence of another person (Sally) from the knowledge of the other characters involved. The crucial question is where Sally would look for the object when she returned: The location it was before she left the scene, or the place where Anne had moved it. Children under the age of four usually perform quite poorly on this test. The cognitive capacity to pass the test requires the ability to ‘metarepresent’ Sally’s mental state, i.e. ‘I know that she does not know where the object really is’. The Sally-and-Anne-Test, therefore, encompasses what is called understanding a ‘first order’ false belief.

More sophisticated cognitive capacities involving a theory of mind include the understanding of higher order false belief tasks (e.g. Perner and Wimmer, 1985), metaphor, irony, and faux pas. It has been argued that understanding metaphor requires at least first order theory of mind comprehension, whereas irony involves second order theory of mind, because these processes relate to the ability

to go beyond the literal meaning of utterances by inferring what the speaker actually might have intended (Happé, 1994; Langdon et al., 2002b).

In adults with psychopathological conditions, short stories involving double bluff, mistakes, persuasions or white lies (Happé, 1994), cartoons or other visually presented material has been used to assess theory of mind abilities. In theory of mind research in schizophrenia, for instance, short stories with or without use of props and picture sequencing tasks have been given to patients, as well as, tests of comprehension of hints behind indirect speech, metaphor and irony. Over the years, the pictorial theory of mind material has been modified in order to better control for interference with attention, memory, 'general' intelligence, and verbalization. One problem in early studies in schizophrenia was that patients not only performed poorly on theory of mind tasks, but also often failed to correctly respond to the control or 'reality' questions (e.g. Frith and Corcoran, 1996; Drury et al., 1998). Therefore, 'physical' control tasks of similar complexity but without the requirement to refer to mental states have been introduced, in particular to address the question whether theory of mind deficits are specific (Frith and Corcoran, 1996; Sarfati et al., 1997; Langdon et al., 1997; Drury et al., 1998; Brunet et al., 2003). Moreover, Baron-Cohen et al. (1997, 2001a) developed a more realistic test perhaps tapping into theory of mind, where subjects are asked to infer mental states of persons of whom the eye region is depicted only. However, this and other theory of mind tests aim at dissecting 'cognitive' and 'affective' mental state attribution (Shamay-Tsoory et al., 2005), where the concept of theory of mind overlaps with empathy (in German called 'Einfühlung'), such that the validity of the 'Eyes Test' as a theory of mind task has been criticized (Jarrold et al., 2000).

In any event, the problem of 'real-life' task presentation cannot satisfactorily be resolved in experimental laboratory 'off-line' test conditions. People with psychiatric disorders who may be biased in their belief formation, for instance, may in 'neutral' test situations be relatively unaffected in abstract reasoning tasks (Simpson et al., 1998). Also, comparability of theory of mind studies in normals or populations suffering from psychopathologies may to some extent be limited due to subtle differences of task presentation, e.g. using picture stories, which depict the outcome of the scene versus 'open ended' picture sequences (Sarfati et al., 1997a).

6. Psychopathology of theory of mind

The concept of theory of mind is appealing to clinical psychopathologists, because theory of mind in its most sophisticated expression is unique to humans and because its absence or impaired functioning may account for quite a broad spectrum of behavioral abnormalities in both children and adults. In the following section, we therefore briefly

summarize some of the empirical findings regarding theory of mind in psychopathological conditions.

7. Developmental disorders

In this paragraph, we put emphasis on those studies addressing the issue of impaired theory of mind acquisition during child development. The starting point of empirical research into this matter was Baron-Cohen (1988) intriguing question whether the autistic child has a 'theory of mind'. Since, Kanner's (1943) and Asperger's (1944) groundbreaking publications clinicians have sought to explain why autistic children would behave so socially withdrawn and unempathically the way they actually do. Autistic children actively avoid eye contact or close body contact. They frequently engage in stereotyped behaviors and fail to establish emotional relationships. A wealth of research has now demonstrated that autistic children are extremely impaired in appreciating the mental states of other individuals. The theory of mind deficit found in autistic children correlates with their social behavioral abnormalities and impaired pragmatic use of language (Baron-Cohen, 1991, 1995). Even individuals with high functioning autism or Asperger's syndrome who pass relatively simple false belief tasks have difficulties in theory of mind tasks that also address empathic abilities such as appreciating the mental states when an individual's eye region is depicted only (Baron-Cohen et al., 2001a; but see Roeyers et al., 2001 for divergent results, and Brent et al., 2004 for a comparison of different theory of mind tasks in children with autistic spectrum disorders). Most important is the fact that impaired theory of mind in autism is independent of general intelligence and that other cognitive capacities are left intact. In fact, many children with autism have an even superior technical understanding (sometimes termed 'folk physics' as opposed to 'folk psychology') compared to their age-related peers. On the other hand, it has been shown that children with other developmental disorders such as specific language impairments (Perner et al., 1989), Down's syndrome (Baron-Cohen et al., 1986; Russell et al., 1991) or Williams syndrome (Karmiloff-Smith et al., 1995) are remarkably unimpaired in their ability to mentalize despite lower intelligence in the latter two groups compared to normal children. Likewise, children with ADHD who have marked executive functioning and attention problems are apparently unimpaired in their theory of mind abilities, except perhaps those with the most severe attention deficits (Buitelaar et al., 1999; overview in Kain and Perner, 2003). Thus, impaired theory of mind in autistic children is not merely a problem of attention or general intelligence. Taking the evidence together, individuals with autistic spectrum disorders are specifically impaired in their capacity to mentalize and to empathize relative to other mental faculties and, although speculative, this deficit of

social intelligence could be balanced by a more advanced technical intelligence (Baron-Cohen et al., 2001b).

8. Personality disorders and other non-psychotic disorders

Empirical research on theory of mind in personality disorders has largely focused on what has been conceptualized as ‘psychopathy’. Psychopaths have been characterized as superficially charming, but otherwise unreliable, ‘cold-hearted’ and unresponsive individuals. These emotional deficits have been found to be present in psychopaths from childhood on. It has therefore been argued that psychopathic individuals who are impaired in empathizing with others could have impaired theory of mind abilities, too. Contrary to intuition, however, current evidence suggests that psychopathic individuals are unimpaired in their ability to appreciate the mental states of others, at least in experimental conditions (Blair et al., 1996; Richell et al., 2003). Even in the more demanding ‘Reading the Mind in the Eyes Test’ (see above), which clearly comprises an empathic element, psychopaths perform equally well compared to non-psychopathic controls (Richell et al., 2003). Interestingly, Mealey and Kinner (2003) have reasoned from an evolutionary point of view that psychopathic individuals develop theory of mind abilities enabling them to understand others in purely instrumental terms devoid of empathic feelings. Psychopathy can accordingly be understood as evolved non-cooperative ‘cheater-morph’, a personality type that may persist in populations at a low prevalence rate.

At least two other lines of research into theory of mind in personality disorders are worth mentioning. One originally related to the question as to whether theory of mind deficits in schizophrenia were trait or state-dependent. Langdon and Coltheart (1999) tested non-clinical subjects who scored highly on a schizotypy questionnaire and compared their theory of mind abilities to low-level schizotypy subjects. Interestingly, in support of the former assumption they found that individuals with high schizotypy scores performed more poorly on theory of mind tasks than low-level schizotypics. So, if there were a continuum from normalcy to psychosis, with schizoid personality and schizotypy lying somewhere in-between, these results would indicate that theory of mind deficits were trait rather than state dependent.

Secondly, it has been argued from a psychoanalytical perspective that patients with severe personality pathologies could be impaired in their theory of mind abilities for at least two reasons: (1) As already mentioned, favorable early experiences within the family milieu have the potential of speeding up the theory of mind abilities of young children. On the contrary, it is thus conceivable that an abusive or depriving family environment can impede the development of theory of mind skills. (2) Fonagy (1989, 1991) suggested

from a psychoanalytical point of view that a patient’s theory of mind abilities (or ‘reflective functioning’) might be functionally inhibited as a defense mechanism, which could be crucial to take into account in analyzing transference and countertransference during therapy.

In relating ‘reflective functioning’ to attachment, a recent study has tested theory of mind abilities in patients with anorexia nervosa (Tchanturia et al., 2004). Patients with anorexia nervosa performed more poorly on a variety of theory of mind measures, but also on control tasks that did not involve theory of mind. Thus, this study failed to demonstrate a selective theory of mind deficit in anorexia nervosa, but more subtle tests tapping into theory of mind may be useful to refine this issue in patients with eating disorders or severe personality disorders.

9. Schizophrenia and affective disorders

Schizophrenia and affective disorders, commonly subsumed under the obsolete term ‘functional psychoses’ usually manifest in adolescence or adulthood. In contrast to theory of mind, deficits that emerge due to gross brain pathology (see below) or primarily due to developmental or personality disorders, schizophrenia and affective disorders have an intermediate position—we simply do not know whether theory of mind has developed normally in individuals suffering from ‘functional psychoses’.

Frith (1992) was the first to suggest that psychotic symptoms found in schizophrenia could indicate impaired theory of mind. Formal thought disorders such as incoherence, knight’s move etc. could arise from a patient’s inability to take into account an interlocutor’s mental state. That is, a thought-disordered patient might falsely infer that the interlocutor shares a common knowledge with the patient, i.e. knows what the patient knows. Likewise, schizophrenic patients who have difficulties in experiencing their behavior as the result of their own intentions may interpret their actions as being under alien control. Frith (1992) has therefore argued that impaired theory of mind in schizophrenia may account for (1) disorders of ‘willed action’, e.g. negative and disorganized symptoms, (2) disorders of self-monitoring, e.g. delusions of alien control and voice-commenting hallucinations or other ‘passivity’ symptoms, and (3) disorders of monitoring other persons’ thoughts and intentions, including delusions of reference and persecution. By contrast, Hardy-Baylé (1994) has argued that an executive planning deficit would lie at the core of schizophrenic patients’ impaired capacity to mentalize and that such problems may therefore largely occur in patients with thought and language disorganization. Both models have gained empirical support (Corcoran et al., 1995; Sarfati et al., 1999; summarized in Brüne, 2005b). Although not as severe as those seen in autism, patients with schizophrenia have specific theory of mind deficits that deteriorate with acuity or chronicity of the disorder

Table 2
Overview of brain lesion studies of theory of mind in chronological order

Author(s); published	Sample (<i>n</i>)	Mean age (years)	Sex m/f	Brain imaging technique	ToM method/tasks	Main results
Siegal et al., 1996	17 right-hemisphere-damaged patients (RHD)	69.2	7/10	CT	False-belief task (modified Sally-Anne-Test).	RHD patients but not LHD patients were impaired in their ability to appreciate false beliefs.
	11 left-hemisphere-damaged patients (LHD)	70.3	8/3			
Winner et al., 1998	13 patients with right hemisphere brain damage following stroke	59.5	6/7	CT or MRI	Short stories involving lies and jokes.	Patients with RHD had difficulties in making second-order mental state attributions, which was correlated with the ability to distinguish jokes from lies.
Stone et al., 1998	5 patients with bilateral damage to the orbitofrontal cortex (OFC)	34–51	n.m.	CT and/or MRI	First and second-order false-belief tasks and faux pas tasks	In contrast to patients with DFC, patients with damage to the OFC were impaired in detecting faux pas, but neither group was impaired in first and second-order ToM tests.
	5 patients with unilateral damage in the left dorsolateral prefrontal cortex (DFC)	64–80				
Happé et al., 1999	14 patients with right hemisphere damage following stroke (RHD)	64	5/9	CT or MRI	Short stories involving double bluff, mistakes, persuasions and white lies, humorous cartoons and non-ToM control stories and cartoons, comprehension of funny cartoons versus altered versions.	Patients with RHD, but not LHD, showed specific impairments in tasks requiring ToM compared with healthy controls.
	5 patients with left hemisphere damage following stroke (LHD)	67	4/1			
Rowe et al., 2001	19 controls	73	9/10		First and second-order false belief stories.	RF and LF patients were impaired in ToM compared with controls. ToM impairment was independent of executive and intellectual functioning.
	15 patients with right frontal lobe lesions (RF)	40.2	6/9	CT or MRI		
Fine et al., 2001	16 patients with left frontal lobe lesions (LF) following neurosurgery	44.2	8/8		False belief tests (stories), cartoons and metaphor comprehension.	The patient showed selective impairment in ToM in the absence of executive functioning deficits.
	31 controls	42.9	13/18			
	Case-study with congenital lesion in the lateral part of the basal nuclei of the left amygdala	32	1/0	MRI		
Stuss et al., 2001	13 controls	30	13/0		Visual perspective taking and deception task.	Right frontal lesions were associated with impaired visual perspective taking, and particularly right ventromedial lesions with impaired detection of deception.
	32 patients with focal lesions of different etiologies in frontal brain areas		n.m.	CT or MRI		
	Right frontal <i>n</i> =4	54.25				

Bird et al., 2004	left frontal <i>n</i> =8	57.38				
	bifrontal <i>n</i> =7	49.57				
	non-frontal right <i>n</i> =5	46.2				
	left <i>n</i> =8	48.88				
	14 controls	52.0				
	Case study with bilateral anterior cerebral artery infarction	62	0/1	CT and MRI angiography	Picture sequencing task, short stories involving double bluff, mistakes, persuasions, white lies, and violation of social norms, faux pas, and moving objects test.	The patient had difficulties in understanding unintentional violations of social norms and in detecting faux pas, but was unimpaired in picture sequencing and short stories tasks.
Shamay-Tsoory et al., 2005	12 controls	64.1	n.m.			
	26 patients with focal lesions in prefrontal cortex (PFC)***, (,)	34.12		CT or MRI	Short stories (second-order false belief, irony, and social faux pas). Assessment of empathic abilities using the Interpersonal Reactivity Index (IRI). Assessment of recognition of facial expression and recognition of affective prosody.	Patients with ventromedial (and particularly right ventromedial) lesions were significantly impaired in understanding faux pas as well as irony, but not in second-order false belief tasks as compared with patients with posterior lesions and normal controls. Neither facial expression recognition, nor prosody contributed to group differences in faux pas and irony understanding. Empathic ability was correlated with ToM performance in PFC patients.
Samson et al., 2005	Left PFC, <i>n</i> =6		20/6			
	Right PFC <i>n</i> =7		8/5			
	Bilateral PFC <i>n</i> =13		10/3			
	13 patients with lesions of posterior cortex (PC) (left PC <i>n</i> =9, right PC <i>n</i> =4) of various etiologies	40.46				
	13 controls	34.2				
	Case study of a stroke patient with right prefrontal and temporal damage (inferior and middle frontal gyri, superior temporal gyrus)	56	1/0	MRI	Short non-verbal videos as false-belief tasks requiring low and high levels of self-perspective inhibition when attributing beliefs to someone else. Simulation play of attributing visual experiences and emotions to someone else.	The patient had a selective deficit in inhibiting his own perspective, showing a high proportion of egocentric errors in both social and non-social conditions.

CT, computed tomography; MRI, magnetic resonance imaging; n.m., not mentioned; ToM, theory of mind.

(Langdon et al., 2001; Pickup and Frith, 2001; overview in Frith, 2004). That is, these deficits are probably independent of other cognitive dysfunctions such as attention, set-shifting capacity, general intelligence and so forth (Lee et al., 2004). Patients with negative or disorganized symptoms seem to perform most poorly on theory of mind tasks, whereas studies of patients with predominantly paranoid symptoms have revealed mixed results (e.g. Langdon et al., 1997). It is unclear, however, as to what extent impaired theory of mind in schizophrenia contributes to social behavioral deviations (overviews in Corcoran, 2000; Brüne, 2005b). In conversational interactions, relatively stable outpatients with schizophrenia have been found to appreciate the mental states of their interlocutors (McCabe et al., 2004). This finding could illustrate the difference between 'online' and 'offline' mentalizing, a problem that warrants further investigation (Frith, 2004).

In contrast to schizophrenia, theory of mind in affective disorders remains an under-explored area of research. Clinically, thought disorders with respect to both form and content are characteristic of affective disorders, suggesting possible theory of mind difficulties. In an early study, patients with affective disorders who served as clinical control group were found unimpaired relative to healthy persons in their ability to appreciate mental states (e.g. Doody et al., 1998). A study in patients with bipolar affective disorder, however, revealed impaired theory of mind in both acutely depressed and manic patients, whereas remitted patients were unimpaired relative to healthy controls (Kerr et al., 2003). In contrast, Inoue et al. (2004) found remitted patients with unipolar or bipolar depression impaired in their ability to appreciate second order false belief tasks. This finding was independent of age, sex, and duration of illness or general intelligence. Similarly, Bora et al. (2005) discovered theory of mind deficits in euthymic patients with bipolar affective disorder. However, a link of theory of mind difficulties in affective disorders with other cognitive functions such as working memory, attention or with other specific psychopathological symptoms, e.g. thought disorder, warrants further investigation.

10. Brain damage and degenerative brain disorders

Assessing theory of mind in patients suffering from brain lesions following stroke, brain tumor operation or degenerative disorders differs from the studies outlined above, because it can be taken for granted that theory of mind developed normally in these individuals.

A number of studies on theory of mind in patients with brain damage to the frontal lobes following cerebral artery infarction or following tumor excision have demonstrated that patients with right frontal lesions are impaired in a variety of tasks involving theory of mind. These tests include, for instance, the appreciation of second order mental states, distinguishing jokes from lies or recognizing deception (e.g.

Siegal et al., 1996; Winner et al., 1998; Happé et al., 1999; Stuss et al., 2001; Rinaldi et al., 2002). These deficits are largely independent of other cognitive dysfunction, and overall less pronounced in patients with left frontal lesions, although some studies have revealed mixed results (Rowe et al., 2001). Bilateral damage to the orbitofrontal cortex has been shown to be associated with difficulties in understanding faux pas (Stone et al., 1998). An interesting case study in a patient with congenital damage to the left amygdala, who was diagnosed with schizophrenia and Asperger's syndrome, underscored the assumption that theory of mind deficits may occur independently of executive dysfunction. This patient had profound difficulties in a variety of tests involving theory of mind but was virtually unimpaired on several executive functioning tests (Fine et al., 2001). Most disturbing for our current understanding of the cerebral representation of theory of mind is the fact that a patient with bilateral anterior cerebral infarction and extensive damage to the medial prefrontal cortex was at best very mildly impaired in her ability to appreciate mental states, whereas profound executive deficits were present (Bird et al., 2004). This case challenges the common notion that normal functioning of the medial prefrontal cortex is indispensable for the execution of theory of mind. However, Shamay-Tsoory et al. (2005) have recently found that damage to the ventromedial prefrontal cortex is specifically associated with difficulties in inferring other persons' emotions, rather than false beliefs, suggesting that the attribution of affective states is differently represented compared with intention and belief representation. The studies of theory of mind in patients with focal brain damages are summarized in Table 2.

With respect to normal aging, the existing studies have produced mixed results (see above). However, in a mixed population of 'frail' older nursing home residents Washburn et al. (2003) found theory of mind impairments to be associated with poor social functioning, even after controlling for other cognitive function.

Saltzman et al. (2000) examined theory of mind abilities and executive functions in non-demented patients suffering from Parkinson's disease (PD). Compared with normal older control subjects and young students, patients with PD were impaired on theory of mind tasks but also on measures of executive functioning, suggesting little evidence for a specific theory of mind deficit in PD.

Only a few studies have looked at theory of mind capacities in other neurodegenerative diseases such as Alzheimer's disease (AD) and frontotemporal dementia (FTD). Cuerva et al. (2001) found patients with mild to moderate AD to be impaired only on the more complex second order false belief tasks compared to healthy age-matched control subjects. The patients who performed poorly on these tasks were also more severely impaired on tasks of verbal memory, abstract thinking, verbal comprehension, and naming. As the presented material (read-out short stories) posed high demands on cognitive abilities,

the results could largely be explained by this confound rather than by a specific theory of mind deficit in AD.

By contrast, the frontal variant of frontotemporal dementia (fvFTD) is characterized by changes in personality and social behavior while most cognitive domains are relatively preserved, at least in the early stages of the disorder. From a clinical perspective this could be indicative of a selective theory of mind deficit in FTD. In a study, comparing patients with fvFTD with mild AD and healthy control subjects [Gregory et al. \(2002\)](#) found fvFTD patients to perform significantly worse on all theory of mind tasks with increasing impairment relative to task complexity. AD patients again failed only on the more cognitively demanding second order false belief tasks indicating an interference with cognitive performance rather than impaired theory of mind per se. Interestingly, theory of mind impairment correlated with measures of behavioral disturbance in FTD patients as well as with the degree of frontal atrophy. A further study addressed theory of mind abilities in FTD compared with Huntington's disease (HD) and normal controls. HD is a degenerative disease with predominant involvement of subcortical structures, especially the striatum, leading to frontal cortical atrophy by deafferentiation. HD is associated with involuntary limb movements and at the behavioral level with altered social conduct. Both patient groups performed worse on theory of mind tasks compared to normal controls. Overall, however, HD patients were better at appreciating mental states than FTD patients ([Snowden et al., 2003](#)). FTD patients gave more literal interpretations without being able to grasp what was funny about a story, whereas HD patients formulated hypotheses indicative of theory of mind, although sometimes deviating from conventional interpretation.

In conclusion, the few available data seem to indicate that patients with frontotemporal dementia may have a specific theory of mind deficit. By contrast, the evidence for a primary impairment of theory of mind in AD, HD or PD is fairly weak.

11. Discussion

In this article, we have sought to examine a specific aspect of social cognition in an evolutionary perspective. The ability to infer mental states of other individuals, referred to as 'theory of mind' probably emerged in primates due to selection pressures from the social environment ([Brothers, 1990](#)). Tracing back the evolutionary history of this cognitive faculty, we found evidence that theory of mind most likely evolved from the capacity to monitor biological motion and from imitation behavior, and now involves a neural network including the frontal lobes, the STS, the ACC, and the inferior parietal cortex ([Abu-Akel, 2003](#); [Decety and Chaminade, 2005](#)).

Theory of mind is certainly most highly developed in humans. But it comes at a certain cost. The evolution of big brains is energetically expensive and the ontogenetic

acquisition of theory of mind is extremely time-consuming. Theory of mind comprises an innate cognitive capacity represented in a dedicated neural network. This, by no means, excludes the possibility that the actual development of the ability to infer mental states is highly dependent on environmental input, i.e. social interaction with other humans ([Carpendale and Lewis, 2004](#)). On the contrary, in line with [Fonagy \(1991\)](#) we argue that unfavorable social conditions during early childhood may seriously obstruct normal development of theory of mind. In addition, theory of mind evolved in humans to cope with a social environment that is in many ways fundamentally different from our present social environment. Thus, there are countless possible impediments for theory of mind to develop properly—be they genetic in origin, environmental or both.

Psychopathology, we propose, almost always involves disturbances of social reasoning or theory of mind. The actual manifestations of impaired theory of mind, however, can be highly diverse. We have tried to review the most important findings in this area of research accumulated over the past 10 years or so. Due to space limitations and conciseness, this overview is unavoidably incomplete. In brief, in developmental disorders such as autism, the acquisition of theory of mind can be fundamentally retarded ([Baron-Cohen et al., 2001b](#)). Studies in patients with acquired brain lesions inform us that theory of mind can also secondarily be impaired—in individuals who had normal theory of mind abilities prior to the event ([Stuss et al., 2001](#)). In a great deal of psychopathological conditions such as personality disorders or 'functional' psychoses, our knowledge is limited as to whether theory of mind developed normally during ontogeny (e.g. [Brüne, 2005b](#)).

Recent research developments aim at linking theory of mind abilities with linguistic skills of patients and healthy control subjects. There is some evidence for [Sperber and Wilson's \(2002\)](#) hypothesis that theory of mind is a prerequisite for the pragmatic use of human language. A few studies into schizophrenia have shown, for example, that a violation of the rules of pragmatic use of language is linked to patients' impaired theory of mind ([Greig et al., 2004](#); [Corcoran and Frith, 2005](#); [Brüne and Bodenstein, 2005](#)), and little is known whether analogous links between pragmatics and theory of mind can be found in other neuropsychiatric disorders.

If the emergence of a theory of mind was advantageous in terms of survival and reproduction in hominids, as put forward in the introductory paragraphs, this cognitive capacity should somehow relate to an individual's actual social behavioral skills. While studies on the association of theory of mind with social behavioral competence in normal populations are lacking, some clues stem from studies of theory of mind and behavioral abnormalities in psychopathological conditions. For instance, theory of mind predicts the level of social expertise in

schizophrenic patients (Roncone et al., 2002; Brüne, 2005a), and Abu-Akel and Abushua'leh (2004) found a significant interaction between theory of mind skills in patients with paranoid schizophrenia and their history of violent behavior. The violent patients performed superior on task involving second order theory of mind compared with a non-violent sample with schizophrenia; by contrast, the violent patients performed more poorly on a test requiring basic empathy skills, which suggests that empathy and theory of mind represent different domains within the social module and affect the actual social behavior (at least in patients) to a different extent (Abu-Akel and Abushua'leh, 2004; Brüne, 2005a). These issues are critical, not only for understanding patients' behavior, but also for developing cognitive training devices in this domain. Gambini et al. (2004), for instance, were able to demonstrate that patients' delusional beliefs could be modified when the patients were encouraged to shift their perspective from first-person to third-person, acknowledging the viewpoint of the interviewer. This illustrates that patients with schizophrenia could probably benefit from cognitive training in the social domain. Manuals for such trainings are underway (e.g. Moritz et al., in press).

The theoretical importance of theory of mind for everyday life requires further studies in both healthy and mentally ill populations. For example, research into theory of mind in healthy and psychiatrically ill persons could ideally be combined with studies of individual differences in game-theoretical scenarios. Based on our evolutionary outline, we predict that individuals with theory of mind deficits would perhaps be more likely to cooperate when defecting would be the better strategy, if they had difficulties in understanding deception or detecting cheating. In contrast, paranoia-prone patients would perhaps be more likely to defect based on their false assumption of being cheated by others (Brüne and Bodenstein, 2005). Another line of research worth pursuing relates to the possible impact of sexual selection on the evolution of theory of mind. To our knowledge, sex differences in theory of mind have not been systematically examined in adult populations. We propose that women would be better at theory of mind tasks, because due to their higher parental investment in potential offspring (Trivers, 1972), women would face a higher cost if cheated by males. In partial support of this hypothesis, one study has found individual differences between male and female preadolescents, where young females outperformed preadolescent boys in their theory of mind abilities and social skills (Bosacki et al., 1999). Finally, it would be interesting to see whether theory of mind deficits in psychopathological conditions have predictive value in relation to relapse or response to treatment, and more empirical work is needed to examine what diagnostic groups may improve upon theory of mind training devices.

References

- Abu-Akel, A., 2003. A neurobiological mapping of theory of mind. *Brain Res. Rev.* 43, 29–40.
- Abu-Akel, A., Abushua'leh, K., 2004. 'Theory of mind' in violent and nonviolent patients with paranoid schizophrenia. *Schizophr. Res.* 69, 45–53.
- Ackerman, B., 1981. Young children's understanding of a false utterance. *Dev. Psychol.* 31, 472–480.
- Adolphs, R., 2001. The neurobiology of social cognition. *Curr. Opin. Neurobiol.* 11, 231–239.
- Aiello, L.C., Wheeler, P., 1995. The expensive tissue hypothesis. *Curr. Anthropol.* 36, 184–193.
- Alexander, R.D., 1987. *The Biology of Moral Systems*. De Gruyter, New York.
- Asperger, H., 1944. Die "autistischen Psychopathen" im Kindesalter. *Arch. Psychiat. Nervenkrankh.* 117, 76–136.
- Avis, J., Harris, P.L., 1991. Belief-desire reasoning among Baka children: Evidence for a universal conception of mind. *Child. Dev.* 62, 460–467.
- Axelrod, R., Hamilton, W.D., 1981. The evolution of cooperation. *Science* 211, 1390–1396.
- Baron-Cohen, S., 1988. Social and pragmatic deficits in autism: Cognitive or affective? *J. Autism Dev. Disord.* 18, 379–402.
- Baron-Cohen, S., 1991. The theory of mind deficit in autism: How specific is it? *Br J. Dev. Psychol.* 9, 301–314.
- Baron-Cohen, S., 1995. *Mindblindness: An Essay on Autism and Theory of Mind*. Bradford/MIT Press, Cambridge, MA.
- Baron-Cohen, S., Leslie, A., Frith, U., 1985. Does the autistic child have a 'theory of mind'? *Cognition* 21, 37–46.
- Baron-Cohen, S., Leslie, A., Frith, U., 1986. Mechanical, behavioural and intentional understanding of picture stories in autistic children. *Br. J. Dev. Psychol.* 4, 113–125.
- Baron-Cohen, S., Jolliffe, T., Mortimore, C., Robertson, M., 1997. Another advanced test of theory of mind: evidence from very high functioning adults with autism or Asperger syndrome. *J. Child. Psychol. Psychiatr.* 38, 813–822.
- Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., Plaisted, K., 1999. Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *J. Autism Dev. Disord.* 29, 407–418.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., Plumb, I., 2001a. The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger Syndrome or high-functioning autism. *J. Child Psychol. Psychiatr.* 42, 241–251.
- Baron-Cohen, S., Wheelwright, S., Spong, A., Schill, V., Lawson, J., 2001b. Studies of theory of mind: Are intuitive physics and intuitive psychology independent? *J. Dev. Learn. Dis.* 5, 47–78.
- Bird, C.M., Castelli, F., Malik, O., Frith, U., Husain, M., 2004. The impact of extensive medial frontal lobe damage on 'theory of mind' and cognition. *Brain* 127, 914–928.
- Blair, R.J.R., Sellars, C., Strickland, I., Clark, F., Williams, A., Smith, M., Jones, L., 1996. Theory of mind in the psychopath. *J. Forens. Psychiatr.* 7, 15–25.
- Bora, E., Vahib, S., Gonul, A.S., Akdeniz, F., Alkan, M., Ogut, M., Eryavuz, A., 2005. Evidence for theory of mind deficits in euthymic patients with bipolar disorder. *Acta Psychiatr. Scand.*
- Bosacki, S., Astington, J.W., 1999. Theory of mind in preadolescence: Relations between social understanding and social competence. *Soc. Dev.* 8, 237–255.
- Brent, E., Rios, P., Happé, F., Charman, T., 2004. Performance of children with autism spectrum disorder on advanced theory of mind tasks. *Autism* 8, 283–299.
- Brothers, L., 1990. The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Concepts Neurosci.* 1, 27–51.

- Brüne, M., 2005a. Emotion recognition, 'theory of mind' and social behavior in schizophrenia. *Psychiatr. Res.* 133, 135–147.
- Brüne, M., 2005b. 'Theory of mind' in schizophrenia: A review of the literature. *Schizophr. Bull.* 31, 21–42.
- Brüne, M., Bodenstein, L., 2005. Proverb comprehension reconsidered - 'theory of mind' and the pragmatic use of language in schizophrenia. *Schizophr. Res.* 75, 233–239.
- Brunet, E., Sarfati, Y., Hardy-Baylé, M.C., Decety, J., 2000. A PET investigation of the attribution of intentions with a nonverbal task. *NeuroImage* 11, 157–166.
- Brunet, E., Sarfati, Y., Hardy-Baylé, M.C., Decety, J., 2003. Abnormalities of brain function during a nonverbal theory of mind task in schizophrenia. *Neuropsychologia* 41, 1574–1582.
- Buitelaar, J.K., van der Wees, M., Swaab-Barneveld, H., van der Gaag, R.J., 1999. Theory of mind and emotion-recognition functioning in autistic spectrum disorders and in psychiatric control and normal children. *Dev. Psychopath.* 11, 39–58.
- Byrne, R.W., 1995. *The Thinking Ape*. Oxford University Press, Oxford.
- Byrne, R.W., 2003. Tracing the evolutionary path of cognition. In: Brüne, M., Ribbert, H., Schiefelhövel, W. (Eds.), *The Social Brain. Evolution and Pathology* pp. 43–60.
- Calarge, C., Andreasen, N.C., O'Leary, D.S., 2003. Visualizing how one brain understands another: A PET study of theory of mind. *Am. J. Psychiatry* 160, 1954–1964.
- Calder, A.J., Lawrence, A.D., Keane, J., Scott, S.K., Owen, A.M., Christoffels, I., Young, A.W., 2002. Reading the mind from eye gaze. *Neuropsychologia* 40, 1129–1138.
- Carpendale, J.I.M., Lewis, C., 2004. Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behav. Brain Sci.* 27, 79–96.
- Corcoran, R., 2000. Theory of mind in other clinical conditions: is a selective 'theory of mind' deficit exclusive to autism?. In: Baron-Cohen, S., Tager-Flusberg, H., Cohen, D.J. (Eds.), *Understanding Other Minds*, second ed. Oxford University Press, Oxford.
- Corcoran, R., Frith, C.D., 2005. Thematic reasoning and theory of mind. Accounting for social inference difficulties in schizophrenia. *Evol. Psychol.* 3, 1–19.
- Corcoran, R., Mercer, G., Frith, C.D., 1995. Schizophrenia, symptomatology and social inference: Investigating 'theory of mind' in people with schizophrenia. *Schizophr. Res.* 17, 5–13.
- Cosmides, L., 1989. The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* 31, 187–276.
- Cuerva, A., Sabe, L., Kuzis, G., Tiberti, C., Dorrego, F., Starkstein, S., 2001. Theory of mind and pragmatic abilities in dementia. *Neuropsychiatr. Neuropsychol. Behav. Neurol.* 14, 153–158.
- Davies, M., Stone, T., 1995. *Mental Simulations: Evaluations and Applications* 1995.
- Dawkins, R., Krebs, J.R., 1979. Arms races between and within species. *Proc. R. Soc. Lond. B* 205, 489–511.
- De Bellis, M.D., Keshavan, M.S., Beers, S.R., Hall, J., Frustaci, K., Masalehdan, A., Noll, J., Boring, A.M., 2001. Sex differences in brain maturation during childhood and adolescence. *Cereb. Cortex* 11, 552–557.
- De Waal, F.B.M., 1982. *Chimpanzee Politics*. Cape, London.
- Decety, J., Chaminade, T., 2005. The neurophysiology of imitation and intersubjectivity. In: Hurley, S., Chater, N. (Eds.), *Perspectives on Imitation. Vol. 1. Mechanisms of Imitation and Imitation in Animals*. MIT Press, Cambridge, MA.
- Dennett, D.C., 1978. *Brainstorm: Philosophical Essays on Mind and Psychology*. MIT Press, Cambridge, MA.
- Doody, G.A., Gotz, M., Johnstone, E.C., Frith, C.D., Owens, D.G., 1998. Theory of mind and psychoses. *Psychol. Med.* 28, 397–405.
- Dunbar, R.I.M., 1998. The social brain hypothesis. *Evol. Anthropol.* 6, 178–190.
- Dunbar, R.I.M., 2003. The social brain: Mind, language, and society in evolutionary perspective. *Annu. Rev. Anthropol.* 32, 163–181.
- Drury, V.M., Robinson, E.J., Birchwood, M., 1998. 'Theory of mind' skills during an acute episode of psychosis and following recovery. *Psychol. Med.* 28, 1101–1112.
- Fadiga, L., Fogassi, L., Pavesi, G., Rizzolatti, G., 1995. Motor facilitation during action observation: A magnetic stimulation study. *J. Neurophysiol.* 73, 2608–2611.
- Fehr, E., Fischbacher, U., 2004. Social norms and human cooperation. *Trends Cogn. Sci.* 8, 185–190.
- Ferstl, E.C., von Cramon, D.Y., 2002. What does the frontomedian cortex contribute to language processing: Coherence or theory of mind? *NeuroImage* 17, 1599–1612.
- Fine, C., Lumsden, J., Blair, R.J.R., 2001. Dissociation between 'theory of mind' and executive functions in a patient with early left amygdala damage. *Brain* 124, 287–298.
- Flavell, J.H., 1999. Cognitive development: Children's knowledge about the mind. *Annu. Rev. Psychol.* 50, 21–45.
- Fletcher, P.C., Happé, F., Frith, U., Baker, S.C., Dolan, R.J., Frackowiak, R.S., Frith, C.D., 1995. Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition* 57, 109–128.
- Fodor, J., 1983. *The Modularity of Mind*. MIT Press, Cambridge, MA.
- Fonagy, P., 1989. On tolerating mental states: Theory of mind in borderline personality. *Bull. Anna Freud Centre* 12, 91–115.
- Fonagy, P., 1991. Thinking about thinking: Some clinical and theoretical considerations in the treatment of a borderline patient. *Int. J. Psychoanal.* 72, 639–656.
- Fonagy, P., 1997. Attachment and reflective function: Their role in self-organization. *Dev. Psychopathol.* 9, 679–700.
- Frith, C.D., 1992. *The Cognitive Neuropsychology of Schizophrenia*. Lawrence Erlbaum, Hove, UK.
- Frith, C.D., 2004. Schizophrenia and theory of mind. *Psychol. Med.* 34, 385–389.
- Frith, C.D., Corcoran, R., 1996. Exploring 'theory of mind' in people with schizophrenia. *Psychol. Med.* 26, 521–530.
- Frith, C.D., Frith, U., 1999. Interacting minds - a biological basis. *Science* 286, 1692–1695.
- Frith, U., Frith, C.D., 2001. The biological basis of social interaction. *Curr. Dir. Psychol. Sci.* 10, 151–155.
- Frith, U., Frith, C.D., 2003. Development and neurophysiology of mentalizing. *Philos. Trans. R. Soc. Lond. B* 358, 459–473.
- Gallagher, H.L., Happé, F., Brunswick, N., Fletcher, P.C., Frith, U., Frith, C.D., 2000. Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia* 38, 11–21.
- Gallagher, H.L., Frith, C.D., 2003. Functional imaging of 'theory of mind'. *TICS* 7, 77–83.
- Gallese, V., Goldman, A., 1998. Mirror neurons and the simulation theory of mind-reading. *TICS* 2, 493–501.
- Gambini, O., Barbieri, V., Scarone, S., 2004. Theory of mind in schizophrenia: First person vs third person perspective. *Conscious. Cogn.* 13, 39–46.
- German, T.P., Niehaus, J.L., Roarty, M.P., Giesbrecht, B., Miller, M.B., 2004. Neural correlates of detecting pretense: Automatic engagement of the intentional stance under covert conditions. *J. Cogn. Neurosci.* 16, 1805–1817.
- Goel, V., Grafman, J., Sadato, N., Hallett, M., 1995. Modeling other minds. *Neuroreport* 6, 1741–1746.
- Gregory, C., Lough, S., Stone, V., Erzinclioğlu, S., Martin, L., Baron-Cohen, S., Hodges, J., 2002. Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: Theoretical and practical implications. *Brain* 125, 752–764.
- Greig, T.C., Bryson, G.J., Bell, M.D., 2004. Theory of mind performance in schizophrenia: Diagnostic, symptom, and neuropsychological correlates. *J. Nerv. Ment. Dis.* 192, 12–18.
- Grezes, J., Frith, C.D., Passingham, R.E., 2004. Inferring false beliefs from the actions of oneself and others: An fMRI study. *NeuroImage* 21, 744–750.

- Grossman, E.D., Blake, R., 2002. Brain areas active during visual perception of biological motion. *Neuron* 35, 1167–1175.
- Happé, F.G.E., 1994. An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistics, mentally handicapped and normal children and adults. *J. Autism. Dev. Disord.* 24, 129–154.
- Happé, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., Dolan, R., Frackowiak, R., Frith, C., 1996. 'Theory of mind' in the brain. Evidence from a PET scan study of Asperger syndrome. *Neuroreport* 8, 197–201.
- Happé, F.G., Winner, E., Brownell, H., 1998. The getting of wisdom: Theory of mind in old age. *Dev. Psychol.* 34, 358–362.
- Happé, F.G., Brownell, H., Winner, E., 1999. Acquired 'theory of mind' impairments following stroke. *Cognition* 70, 211–240.
- Hardy-Baylé, M.C., 1994. Organisation de l'action, phénomènes de conscience et représentation mentale de l'action chez des schizophrènes. *Actual. Psychiat.* 20, 393–400.
- Heyes, C.M., 1998. Theory of mind in nonhuman primates. *Behav. Brain Sci.* 21, 101–148.
- Humphrey, N.K., 1976. The social function of intellect. In: Bateson, P.P.G., Hinde, R.A. (Eds.), *Growing Points in Ethology*. Cambridge University Press, Cambridge pp. 303–317.
- Inoue, Y., Tonooka, Y., Yamada, K., Kanba, S., 2004. Deficiency of theory of mind in patients with remitted mood disorder. *J. Affect. Disord.* 82, 403–409.
- Jarrod, C., Butler, D.W., Cottingham, E.M., Jimenez, F., 2000. Linking theory of mind and central coherence bias in autism and in the general population. *Dev. Psychol.* 36, 126–138.
- Joffe, T.H., 1997. Social pressures have selected for an extended juvenile period in primates. *J. Hum. Evol.* 32, 593–605.
- Jolly, A., 1966. Lemur social behaviour and primate intelligence. *Science* 153, 501–506.
- Kain, W., Perner, J., 2003. Do children with ADHD not need their frontal lobes for theory of mind? A review of brain imaging and neuropsychological studies. In: Brüne, M., Ribbert, H., Schiefelhövel, W. (Eds.), *The Social Brain. Evolution and Pathology*, pp. 197–230.
- Kanner, L., 1943. Autistic disturbance of affective contact. *Nerv. Child* 2, 217–250.
- Karmiloff-Smith, A., Grant, J., Bellugi, U., Baron-Cohen, S., 1995. Is there a social module? Language, face processing and theory of mind in William's syndrome and autism. *J. Cogn. Neurosci.* 7, 196–208.
- Kerr, N., Dunbar, R.I.M., Bental, R.P., 2003. Theory of mind in bipolar affective disorder. *J. Affect. Disord.* 73, 253–259.
- Kourtzi, Z., Kanwisher, N., 2000. Activation in human MT/MST by static images with implied motion. *J. Cogn. Neurosci.* 12, 48–55.
- Langdon, R., Coltheart, M., 1999. Mentalising, schizotypy, and schizophrenia. *Cognition* 71, 43–71.
- Langdon, R., Michie, P.T., Ward, P.B., McConaghy, N., Catts, S., Coltheart, M., 1997. Defective self and/or other mentalising in schizophrenia: A cognitive neuropsychological approach. *Cogn. Neuropsychiatr.* 2, 167–193.
- Langdon, R., Coltheart, M., Ward, P.B., Catts, S.V., 2001. Mentalising, executive planning and disengagement in schizophrenia. *Cognit. Neuropsychiatr.* 6, 81–108.
- Langdon, R., Davies, M., Coltheart, M., 2002. Understanding minds and understanding communicated meanings in schizophrenia. *Mind Lang.* 17, 68–104.
- Lee, K.H., Farrow, F.D., Spence, S.A., Woodruff, P.W.R., 2004. Social cognition, brain networks and schizophrenia. *Psychol. Med.* 34, 391–400.
- Leslie, A., 1987. Pretence and representation: The origins of 'theory of mind'. *Psychol. Rev.* 94, 412–426.
- Levitt, P., 2003. Structural and functional maturation of the developing primate brain. *J. Pediatr.* 143 (4), S35–S45.
- Maylor, E.A., Moulson, J.M., Muncer, A.M., Taylor, L.A., 2002. Does performance on theory of mind tasks decline in old age? *Br. J. Psychol.* 93, 465–485.
- McCabe, K., Houser, D., Ryan, L., Smith, V., Trouard, T., 2001. A functional imaging study of cooperation in two-person reciprocal exchange. *Proc. Nat'l. Acad. Sci. USA* 98, 11832–11835.
- McCabe, R., Leudar, I., Antaki, C., 2004. Do people with schizophrenia display theory of mind deficits in clinical interactions? *Psychol. Med.* 34, 401–412.
- Mealey, L., Kinner, S., 2003. Psychopathy, Machiavellianism and theory of mind. In: Brüne, M., Ribbert, H., Schiefelhövel, W. (Eds.), *The Social Brain. Evolution and Pathology*, pp. 355–372.
- Moritz, S., Burlon, M., Woodward, T.S., 2005. *Metacognitive Skill Training for Schizophrenia (MCT) Manual*. VanHam Campus Verlag, Hamburg.
- Nieminen-von Wendt, T., Metsahonkala, L., Kulomaki, T., Aalto, S., Autti, T., Vanhala, R., von Wendt, L., 2003. Changes in cerebral blood flow in Asperger syndrome during theory of mind tasks presented by the auditory route. *Eur. Child Adolesc. Psychiatry* 12, 178–189.
- Nimchinsky, E.A., Gilissen, E., Allman, J.M., Perl, D.P., Erwin, J.M., Hof, P.R., 1999. A neuronal morphologic type unique to human and great apes. *Proc. Natl. Acad. Sci. U.S.A.* 96, 5268–5273.
- Nowak, M., Sigmund, K., 1993. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma Game. *Nature* 364, 56–58.
- Paus, T., 2001. Primate anterior cingulate cortex: Where motor control, drive and cognition interface. *Nature Rev. Neurosci.* 2, 417–424.
- Perner, J., 1991. *Understanding the Representational Mind*. MIT Press, Cambridge, MA.
- Perner, J., Wimmer, H., 1985. 'John thinks that Mary thinks...'. Attribution of second-order beliefs by 5-10 year old children. *J. Exp. Child Psychol.* 39, 437–471.
- Perner, J., Frith, U., Leslie, A., Leekam, S., 1989. Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. *Child. Dev.* 60, 689–700.
- Pickup, G.J., Frith, C.D., 2001. Theory of mind impairments in schizophrenia: Symptomatology, severity and specificity. *Psychol. Med.* 31, 207–220.
- Povinelli, D.J., 1993. Reconstructing the evolution of mind. *Am. Psychol.* 48, 493–509.
- Povinelli, D.J., Vonk, J., 2003. Chimpanzee minds: Suspiciously human? *TICS* 7, 157–160.
- Premack, D., Woodruff, G., 1978. Does the chimpanzee have a 'theory of mind'? *Behav. Brain Sci.* 4, 515–526.
- Rai, R., Mitchell, P., 2004. Five-year-old children's difficulty with false belief when the sought entity is a person. *J. Exp. Child Psychol.* 89, 112–126.
- Richell, R.A., Mitchell, D.G.V., Newman, C., Leonard, A., Baron-Cohen, S., Blair, R.J.R., 2003. Theory of mind and psychopathy: Can psychopathic individuals read the 'language of the eyes'? *Neuropsychologia* 41, 523–526.
- Rilling, J.K., Insel, T.R., 1999. The primate neocortex in comparative perspective using magnetic resonance imaging. *J. Hum. Evol.* 37, 191–223.
- Rilling, J.K., Sanfey, A.G., Aronson, J.A., Nystrom, L.E., Cohen, J.D., 2004. The neural correlates of theory of mind within interpersonal interactions. *NeuroImage* 22, 1694–1703.
- Rinaldi, M.C., Marangolo, P., Baldassarri, F., 2002. Metaphor comprehension in right brain-damaged subjects with visuo-verbal and verbal material: a dissociation (re)considered. *Cortex* 38, 903–907.
- Roeyers, H., Buysse, A., Ponnet, K., Pichal, B., 2001. Advancing advanced mind-reading tests: Empathic accuracy in adults with a pervasive developmental disorder. *J. Child Psychiat. Psychol.* 42, 271–278.
- Roncione, R., Falloon, R.H., Mazza, M., DeRisio, A., Pollice, R., Necozone, S., Morosini, P., Casacchia, M., 2002. Is theory of mind in schizophrenia more strongly associated with clinical and social functioning than with neurocognitive deficits? *Psychopathology* 35, 280–288.

- Rowe, A.D., Bullock, P.R., Polkey, C.E., Morris, R.G., 2001. 'Theory of mind' impairments and their relationship to executive functioning following frontal lobe excisions. *Brain* 124, 600–616.
- Ruby, P., Decety, J., 2001. Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nature Neurosci.* 4, 546–550.
- Russell, J., Mauthner, N., Sharpe, S., Tidswell, T., 1991. The 'windows task' as a measure of strategic deception in pre-schoolers and autistic subjects. *Br. J. Dev. Psychol.* 9, 331–349.
- Russell, T.A., Rubia, K., Bullmore, E.T., Soni, W., Suckling, J., Brammer, M.J., Simmons, A., Williams, S.C., Sharma, T., 2000. Exploring the social brain in schizophrenia: Left prefrontal underactivation during mental state attribution. *Am. J. Psychiatry* 157, 2040–2042.
- Saltzman, J., Strauss, E., Hunter, M., Archibald, S., 2000. Theory of mind and executive functions in normal human aging and Parkinson's disease. *J. Int. Neuropsychol. Soc.* 6, 781–788.
- Samson, D., Apperly, I. A., Kathirgamanathan, U., Humphreys, G.W., 2005. Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain* 128, 1102–1111.
- Sarfati, Y., Hardy-Baylé, M.C., Besche, C., Widlöcher, D., 1997. Attribution of intentions to others in people with schizophrenia: a non-verbal exploration with comic strip. *Schizophr. Res.* 25, 199–209.
- Sarfati, Y., Hardy-Baylé, M.C., Brunet, E., Widlöcher, D., 1999. Investigating theory of mind in schizophrenia: Influence of verbalization in disorganized and non-disorganized patients. *Schizophr. Res.* 37, 183–190.
- Saxe, R., Kanwisher, N., 2003. People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *NeuroImage* 19, 1835–1842.
- Saxe, R., Carey, S., Kanwisher, N., 2004. Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annu. Rev. Psychol.* 55, 87–124.
- Schmitt, A., Grammer, K., 1997. Social intelligence and success: Don't be too clever in order to be smart. In: Whiten, A., Byrne, R.W. (Eds.), *Machiavellian Intelligence II. Extensions and Evaluations*. Cambridge University Press, Cambridge, pp. 86–111.
- Scholl, B.J., Leslie, A., 1999. Modularity, development and 'theory of mind'. *Mind Lang* 14, 131–153.
- Shamay-Tsoory, S.G., Tomer, R., Berger, B.D., Goldsher, D., Aharon-Peretz, J., 2005. Impaired "affective theory of mind" is associated with right ventromedial prefrontal damage. *Cogn. Behav. Neurol.* 18, 55–67.
- Siegal, M., Carrington, J., Radel, M., 1996. Theory of mind and pragmatic understanding following right hemisphere damage. *Brain Lang.* 53, 40–50.
- Simpson, J., Done, J., Vallée-Tourangeau, F., 1998. An unreasoned approach: a critique of research on reasoning and delusions. *Con. Neuropsychiat.* 3, 1–20.
- Snowden, J., Gibbons, Z., Blackshaw, A., Doubleday, E., Thompson, J., Craufurd, D., Foster, J., Happé, F., Neary, D., 2003. Social cognition in frontotemporal dementia and Huntington's disease. *Neuropsychologia* 41, 688–701.
- Sperber, D., Wilson, D., 2002. Pragmatics, modularity and mind-reading. *Mind Lang.* 17, 3–23.
- Stone, V.E., Baron-Cohen, S., Knight, R.T., 1998. Frontal lobe contributions to theory of mind. *J. Cogn. Neurosci.* 10, 640–656.
- Stuss, D.T., Gallup, G.G., Alexander, M.P., 2001. The frontal lobes are necessary for 'theory of mind'. *Brain* 124, 279–286.
- Suddendorf, T., Whiten, A., 2001. Mental evolution and development: Evidence for secondary representation in children, great apes, and other animals. *Psychol. Bull.* 127, 629–650.
- Sugiyama, L.S., Tooby, J., Cosmides, L., 2002. Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia. *Proc. Nat'l. Acad. Sci.* 99, 11537–11542.
- Sullivan, K., Winner, E., Hopfield, N., 1995. How children tell lie from joke: The role of second order mental state attribution. *Br. J. Dev. Psychol.* 13, 191–204.
- Tchanturia, K., Happé, F., Godley, J., Treasure, J., Bara-Carril, N., Schmidt, U., 2004. 'Theory of mind' in anorexia nervosa. *Eur. Eat. Dis. Rev.* 12, 361–366.
- Tomasello, M., Call, J., Hare, B., 2003. Chimpanzees understand psychological states - the question is which ones and to what extent. *TICS* 7, 153–156.
- Trivers, R., 1971. The evolution of reciprocal altruism. *Quart. Rev. Biol.* 46, 35–57.
- Trivers, R., 1972. Parental investment and sexual selection. In: Campbell, B. (Ed.), *Sexual Selection and the Descent of Man*. Aldine-Atherton, Chicago.
- Umiltà, M.A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., Rizzolatti, G., 2001. I know what you are doing: A neurophysiological study. *Neuron* 31, 155–165.
- Vogele, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., Maier, W., Shah, N.J., Fink, G.R., Zilles, K., 2001. Mind reading: Neural mechanisms of theory of mind and self-perspective. *NeuroImage* 14, 170–181.
- Walter, H., Adenzato, M., Ciaramidaro, A., Endici, I., Pia, L., Bara, B.G., 2004. Understanding intentions in social interaction: the role of the anterior paracingulate cortex. *J. Cogn. Neurosci.* 16, 1854–1863.
- Washburn, A., Sands, L., Walton, P., 2003. Assessment of social cognition in frail older adults and its association with social functioning in the nursing home. *Gerontologist* 43, 203–212.
- Wason, P.C., 1966. Reasoning. In: Foss, B.M. (Ed.), *New Horizons in Psychology*.
- Whiten, A., 2000. Social complexity and social intelligence. *Novartis Found. Symp.* 233, 185–196.
- Whiten, A., Byrne, R.W., 1997. *Machiavellian intelligence II. Extensions and Evaluations*. Cambridge University Press, Cambridge.
- Williams, J.H.G., Whiten, A., Suddendorf, T., Perrett, D.I., 2001. Imitation, mirror neurons and autism. *Neurosci. Biobehav. Rev.* 25, 287–295.
- Wimmer, H., Perner, J., 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128.
- Winner, E., Brownell, H., Happé, F., Blum, A., Pincus, D., 1998. Distinguishing lies from jokes: theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain Lang.* 62, 89–106.