Mind Blindness and the Brain in Autism

Review

Uta Frith
UCL Institute of Cognitive Neuroscience
17 Queen Square
London WC1N 3AR
United Kingdom

Experimental evidence shows that the inability to attribute mental states, such as desires and beliefs, to self and others (mentalizing) explains the social and communication impairments of individuals with autism. Brain imaging studies in normal volunteers highlight a circumscribed network that is active during mentalizing and links medial prefrontal regions with posterior superior temporal sulcus and temporal poles. The brain abnormality that results in mentalizing failure in autism may involve weak connections between components of this system.

In this review, I will discuss a productive and successful, though still controversial, theory of autism. This theory attempts to explain the social and communication failure that is the very core of autistic disorder. The cognitive cause for this failure is assumed to be "mind blindness." This concept presupposes that normal individuals have the capacity to "mind read," that is, to attribute mental states to self and other. This is referred to as the "theory of mind" or "mentalizing." The theory assumes that this capacity, far from being the product of complex logical inference, rests on a dedicated neurocognitive mechanism. I will review the evidence that this mechanism is impaired in both severe and mild forms of autism. Its putative neural basis can give clues to the underlying brain abnormalities in autism.

The Autism Spectrum
It is now widely agreed that autism is a neurodevelopmental disorder (for reviews see Bailey et al., 1996; Happé and Frith, 1996; Lord et al., 2000). Autism persists throughout life. It varies in degree of severity and can occur at all levels of ability, so that it is now generally assumed that there is a spectrum of autistic disorders. Asperger syndrome, a milder variant, and currently distinguished from other forms of autism by the lack of linguistic or cognitive delay, is often not diagnosed until late childhood or even adulthood. The diagnosis of autistic disorder is based on behavioral criteria set out in diagnostic handbooks such as the ICD-10 (World Health Organization, 1992) and DSM-IV (American Psychiatric Association, 2000). Autism was first identified and labeled by Kanner (1943) and Asperger (1944).

The causes of autism are largely genetic (see Maestrini et al., 2000, for a review of susceptibility genes). There is no known medical treatment, but well-structured behavioral treatments have beneficial effects, and high levels of compensatory learning can occur. The

Correspondence: u.frith@ucl.ac.uk

prevalence of autism spectrum disorders is now estimated at between 0.3% and 0.7%. The increase of diagnosed cases in recent years can be accounted for by increased awareness of the disorder in all its variants and the use of wider diagnostic criteria (Fombonne, 1999). The male to female ratio is approximately 3 to 1, becoming more extreme with higher levels of ability.

Individuals with autistic spectrum disorder have striking limitations in social relatedness and in the ability to communicate verbally and nonverbally. They are often aloof in childhood and remain egocentric even after having learned the basic rules of social interaction. They may have no speech, or very delayed speech, and even those who become verbally fluent still have problems in comprehension. Individuals with autism also have other characteristic features, such as restricted interests, motor stereotypies, and obsessive tendencies. They can have excellent rote memory and may possess savant skills.

Autism is a disorder that affects many cognitive functions; however, it does not imply a global information processing deficiency (Scheuffgen et al., 2000). While the hallmark of the disorder is a failure of social communication, this does not imply a global lack of social ability. Rather, autism appears to be caused by one or more specific, i.e., circumscribed, cognitive deficits. At the same time, such modular deficits would have developmental repercussions on general adaptive functioning (Frith and Happé, 1998). This is in line with current ideas about innate domain-specific mechanisms with a circumscribed basis in the brain (for a discussion of current theories see Black, 1998). Arguably the most relevant of these deficits in the origin of autism is a subtle but devastating deficit in human social insight, on which this review will focus. This can be referred to as the mind blindness hypothesis (e.g., Baron-Cohen, 1995).

Mind Reading and Mind Blindness
Individuals with autistic disorder have occasionally commented on what they perceive as an unfathomable yet ubiquitous ability of other people to "mind read" during ordinary social interactions. Normal people indeed behave as if they have an implicit theory of mind, and this allows them to explain and predict others' behavior in terms of their presumed thoughts and feelings. To give an example: you might observe me in my office bent over a filing cabinet drawer pulling out and putting back folders. You would make sense of this behavior by mentalizing, that is, automatically recognizing that I am looking for a paper that I *believe* is in one of the folders and that I *wish* to retrieve. You would think this even if you knew that the paper was not there. To explain my behavior, it is immaterial whether the missing file is in the cabinet or really somewhere else. Suppose that you say to me "Try Debbie's desk," and I respond with "I might have known." Without mentalizing, this everyday exchange would seem like complete non sequiturs. Further, without mentalizing, you might come up with an outlandish interpretation of what I was doing—perhaps

practicing back bending and finger moving? The important point of the example is that for an instantaneous interpretation of ordinary behavior, we automatically take account of the mental state of people, their desires, and their beliefs.

The Cognitive Basis of Mind Reading

Leslie (1987) proposed that the ability to represent mental states is based on a dedicated cognitive mechanism. This mechanism includes a "decoupler" and an "expression raiser" and transforms primary representations (impressions of the physical world) into secondary representations. These are "decoupled" from reality and raised into expressions "in quotes." They can thus be attached to an agent's intentional stance; for example, agent A *believes*, *desires*, etc., that "*x is the case.*" Mentalizing can thus be conceived of as representing an agent's propositional attitude to states in the world, thus keeping apart someone's attitude to states in the world and actual states in the world. This is why children are not confused when their mother holds a banana to her face and pretends it is a telephone.

According to Leslie (1987), the first florid manifestation of the ability to mentalize is seen in the young child's enjoyment of pretence, from around 18 months. Here the child acts as if realizing that when mother is using a banana as a telephone, she is taking a propositional attitude to a particular object, which does not interfere with the child's learning about real telephones and real bananas. The implications of this proposal are radical: a neural system is required that supports the processing of specific information in relation to agents and is not tied to a particular modality. If there is such a system in the normal case, then we can envisage this system being dysfunctional from birth, resulting in a difficulty with the intentional stance. This difficulty would result in mind blindness.

The development of this radical proposal as a neurocognitive theory owed much to the timely coincidence of some highly novel ideas and experiments in the late 1970s and early 1980s. They concerned the need to explain understanding of mental states, such as *beliefs,* in the chimpanzee (Premack and Woodruff, 1978) and in young children (Wimmer and Perner, 1983). Likewise, there was the need to explain the spontaneous enjoyment of *make believe* in infancy (Leslie, 1987). At the same time, it had been documented that young children with autism lacked spontaneous make-believe play (Wing and Gould, 1979). New questions could now be asked: how did understanding of mental states such as belief and pretense evolve? How does it develop in the normal child? What is different in the brain of individuals with autism that impairs this development?

The Development of Mentalizing
in the Normal Child

If there is a dedicated mechanism for mentalizing, incorporating such functions as a "decoupler" and an "expression raiser," when does it come into play and how does it enable learning? Clearly, a newborn child does not possess fully functioning mentalizing ability. Nevertheless, the assumption is that the brain comes equipped with a start-up kit, and that a normal species-specific social environment will tune it up and get it into action. The main purpose of an innate start-up mechanism is that it should lead to fast learning about the properties of its domain, with culture shaping the content of the knowledge that is acquired. The development of the social brain involves many other processes as well, such as the perception of faces, voices, and movements of conspecifics, and these may well be prerequisites for the development of mentalizing.

Sensitivity and learning about the inner states of agents starts early and proceeds rapidly. Early signs of such sensitivity are seen in the phenomenon of shared attention (Carpenter et al., 1998). Children in the first year of life automatically follow another person's gaze, seemingly attending to the other person's focus of interest. Shared attention is accompanied by other signs of mentalizing. For instance, referential looking, where children check the mother's expressive attitude toward a novel object before approaching or avoiding it (Repacholi, 1998). The ability to imitate complex and arbitrary but intentional actions of others—as opposed to their accidental actions—is another sign of the inexorable progress of mentalizing ability and is achieved in the middle of the second year of life (Meltzoff, 1995).

Young children aged 2–3 years learn to understand and use mental state verbs (want, know, pretend) before they learn color names (Bretherton, 1992). Mentalizing ability is also important as a facilitator of learning in other domains. For instance, according to Bloom (2000), mentalizing has a critical function in enabling children to learn the meanings of words. Thus, children don't learn words by mere association of word sound and object in view. Such association is inherently ambiguous and error prone, as speaker and listener may look at different objects. Instead, children learn by tracking the speaker's referential intention, for example, by taking into account the speaker's gaze (Baldwin et al. 1996). The effortless ease with which children as young as 5 (and usually before 8 years of age) acquire advanced concepts such as false belief, deception, white lie, and double bluff is remarkable.

Experimental Studies of Mentalizing Failure
in Autism

The mind blindness theory predicts that the milestones of the normal development of mentalizing should be absent at the appropriate age in young children with autism. In particular, they should fail to follow another person's gaze, fail to point at or show objects of interest—both signs of shared attention—and fail to understand make-believe play. Baron-Cohen et al. (1996) looked for these signs in a large prospective population-based study of infants aged 18 months. At age 3 years, when a firm diagnosis of autism can be made, the co-occurrence of these early signs was found to predict the diagnosis remarkably well. Taken together, these three signs of impaired mentalizing in early life proved reliable enough to serve as a first infant screening test for autism (Baird et al., 2000). It is possible that even some preconditions for the development of mentalizing may be absent. Attentional preferences for human agents, their faces, their voices, and their movements, which are probably important triggers for the mentaliz-
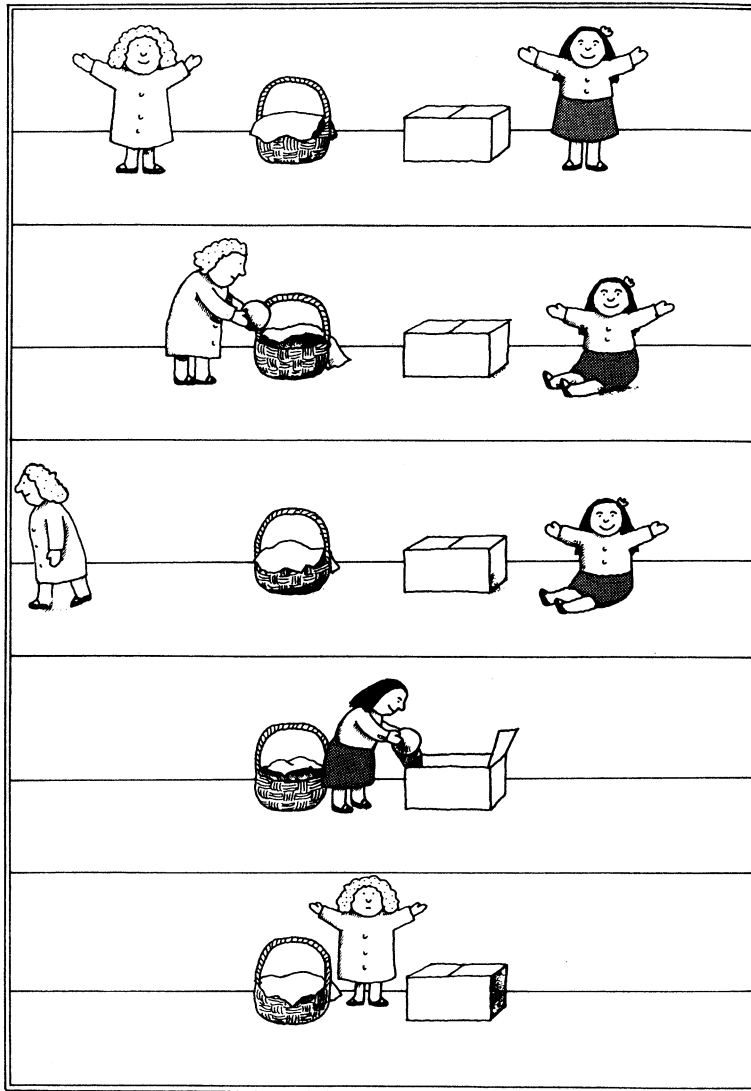
Figure 1. Scenario of Sally-Ann Task
With kind permission from the artist, Axel Scheffler.

ing mechanism, may be lacking in autism. For instance, preschool children with autism did not show a preference for speech over nonspeech stimuli as do other children (Klin, 1991). Nor do older children show a spontaneous preference for facial expressions over other salient stimuli, such as hats (Hobson, 1993). Face recognition difficulties are common throughout the autism spectrum, perhaps because of a lack of social interest early in life. In a neuroimaging study, Schultz et al. (2000) found that brain activation patterns in adults with autism did not distinguish between faces and objects, in contrast to normal adults.

The mind blindness hypothesis was originally proposed and tested by Baron-Cohen et al. (1985, 1986). The argument was that if the social impairment in autism arises from a failure of the mentalizing mechanism as conceptualized by Leslie (1987), then children with autism should be unable to represent mental states such as beliefs. They should be unable to understand and predict behavior in terms of someone's belief even when having achieved the appropriate level of verbal and cognitive development. The test was a false belief task origi-

nally devised by Wimmer and Perner (1983), who showed that normally developing children aged 4 and above passed this test.

In the Sally-Ann task, shown in Figure 1, the following scenario is enacted either with two dolls or two real people: Sally has a basket and Anne has a box. Sally puts a marble into her basket, and then she goes out for a walk. While she is outside, naughty Anne takes the marble from the basket and puts it into her own box. Now Sally comes back from her walk and wants to play with her marble. Where will she look for the marble? The answer seems obvious to a 4 year old child: Sally will look inside her basket. Why? Because that is where Sally *thinks* it is. The marble is really in Anne's box, but Sally doesn't know this. She was not there when Anne transferred the marble. Children with autism, with a mental age of 4 years and above, had difficulty with this task. Unlike normally developing children, and unlike children with Down syndrome, they indicated that Sally would look in Anne's box.

The inability of children with autism to understand false belief tasks at the appropriate age has been con-
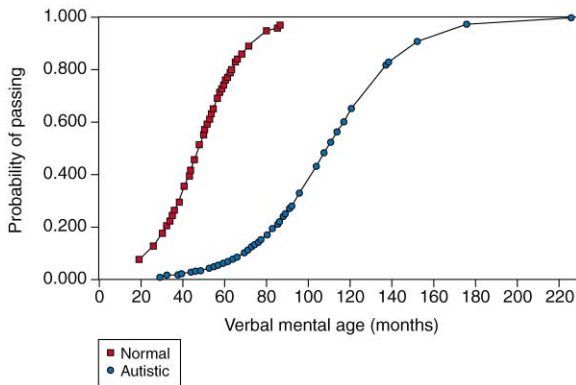
Figure 2. Relationships Between Verbal Ability and Success on False Belief Tasks

Redrawn from Happé's meta-analysis of data from 70 children with autism, 34 children with mental retardation, and 70 normally developing children, showing the cumulative likelihood of passing false belief tasks with increasing verbal mental age.

With kind permission from the author and the publishers of *Child Development* (Blackwell).

firmed subsequently in a number of studies (see chapters in Baron-Cohen et al, 1999). Happé (1995) showed that the verbal mental age of children with autism who are on the cusp of understanding false belief exceeds that of normally developing children by 4 or more years. Thus, as shown in Figure 2, a verbal mental age of 8 and above appears to be necessary to pass standard false belief tests in the case of autism, but only about 4 in the case of normal development.

Implications of Success and Failure on False Belief Tasks

False belief tasks are deceptively simple, but they tap many different abilities and can be solved in different ways. The mind blindness hypothesis is often misunderstood as meaning that people with autism do not possess an explicit theory of mind and never can possess such a theory. Instead, the hypothesis is about the failure of the mentalizing start-up mechanism, not about a "theory." Despite a dysfunctional start-up mechanism, able individuals with autism, and especially those with Asperger syndrome, can come to understand mental states through compensatory learning. However, not only do they acquire this understanding late, but they are slow and error prone on more advanced mentalizing tasks.

If success on false belief tasks is not always easy to interpret, neither is failure. The mentalizing deficit hypothesis predicts failure on the Sally-Anne and similar tasks, but there are many other reasons for failure. For instance, the Sally-Anne test requires working memory and the ability to inhibit reality-oriented responses, i.e., pointing to the place where the object really is. For a convincing demonstration, it is necessary to show success on a task that is in every respect the same but that does not involve thinking about mental states.

An example of such a task uses a scenario in which the critical question concerned the pictorial content of

a photograph. As illustrated in Figure 3, children were shown a teddy bear sitting on a chair. A Polaroid photograph was taken of the scene. The photo was put aside, and the teddy bear was moved to a bed. The critical question was whether the invisible photo showed the teddy bear on the bed or on the chair. The answer is obviously, "on the chair." Compare this to the scenario in the Sally-Anne task, where the question was whether the invisible belief in Sally's mind was that the marble was in the basket or in the box. A belief can become false, while still being held as true in the person's mind. Just so, a photograph can become out of date, still depicting an old scene. The results of a comparison of the two experiments by Leslie and Thaiss (1992) are shown in Figure 4. As predicted by the theory, understanding of false photographs, but not understanding of false beliefs, was well within the comprehension of children with autism. In the case of the normally developing children, the situation was if anything the reverse: they found it easier to answer questions about Sally's belief than questions about the photograph. This suggests that the Sally-Ann task requires a measure of inhibitory capacity, and that normally developing children under the age of 4 struggle with the task because of domain-general limitations (e.g., inhibitory failure, salience of reality), while children with autism fail due to specific difficulties in "reading minds." This experiment strongly supports the idea that mentalizing rests on a separable, or modular, cognitive mechanism.

Mind Blindness Explains the Social Communication Impairments in Autism

The mind blindness hypothesis is an example of a particular model of developmental disorders. The claim of this model is that a single circumscribed cognitive deficit can result in a variety of symptoms that may superficially look unrelated and that span a wide range of severity (Frith et al., 1991). Thus, a deficit in mentalizing can account parsimoniously for the core impairments in socialization, communication, and imagination that characterize the autism spectrum. At the same time, it is specific enough to predict unimpaired function in other domains, assuming there were no additional cognitive deficits. In fact, there *are* other deficits.

The mind blindness hypothesis has never claimed to account for the presence of repetitive behavior and narrow obsessively pursued interests in autism. It cannot account for motor problems, perceptual processing anomalies, or the commonly found superior rote memory skills. Other theories address these features (Russell, 1998; Happé, 1999). However, mind blindness may be able to explain some of the language abnormalities. In autism, muteness, language delay, echoing of speech, and idiosyncratic use of language are highly typical features. Even in cases of age-appropriate or precocious appearance of language, a defining feature of Asperger disorder, parental observations suggest that the first words were often unusual and that vocabulary acquisition was different from that in normally developing children. To investigate the apparently odd pattern of word learning Baron-Cohen et al. (1997) used the ingenious discrepant looking paradigm, where speaker and listener attend to different objects while the speaker utters
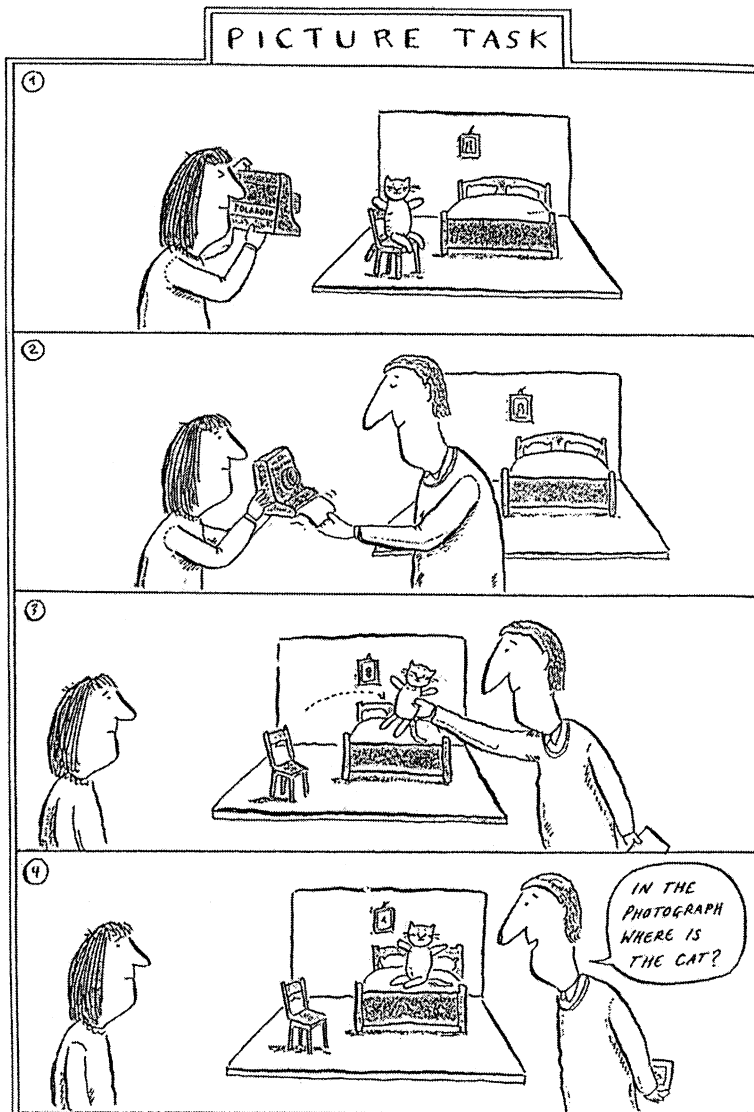
a new word. They demonstrated that children with autism made errors in mapping the word to the object that they happened to be looking at at the time, showing mere association learning. Control children matched for mental age did not make such errors, but instead mapped the word to the object that the speaker was looking at. To be guided by the speaker's referential intention is a sign of mentalizing, and its absence in autism goes some way toward explaining the unusual development of language in autism.

What of those individuals without the benefit of a start-up mechanism, who learn about mental states through conscious effort? Slow learning based on forming associations between behavior and outcomes will allow the gradual acquisition of mental state concepts. In everyday life, many individuals with autistic disorder show that they have learned the rules of social convention, but they lack the intuition to discern situations where these rules become inappropriate and are thrown by playfulness and irony. However, even without the intuitive ability to mentalize, social interaction with others

who can make the appropriate allowances for mind blindness can still be a rich source of experience and learning.

Alternative views of the social impairment in autism have often focused on emotional dysfunction. Studies by Hobson (1993) and Sigman and colleagues (see Sigman and Capps, 1997) suggest that children with autism are less responsive to the emotions displayed by others. For instance, they show little concern when an adult cries out in pain, pretending to be hurt, except when their attention was strongly engaged. On the other hand, contrary to popular belief, failure of bonding or attachment does not appear to be a distinguishing characteristic of autism in early childhood. Attachment would appear to be one of those components of social cognition that are dissociable from mentalizing. It is possible that responsiveness to specific emotions is another dissociable social component.

Impaired emotional processing may be secondary to mind blindness. Experimental studies (Baron-Cohen et al., 1997; Adolphs et al., 2001) suggest that individuals
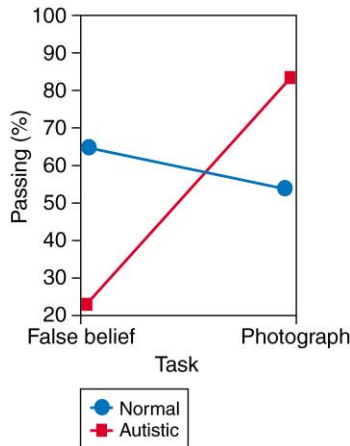
Figure 4. Success and Failure of False Belief and False Photograph Tasks

Redrawn from the results of an experiment by Leslie and Thaiss of 20 normally developing children with a mean age of 4 years and 15 children with autism with a mean age of 12 years. Here passing two false belief tasks, one of them illustrated in Figure 1, was compared with passing two false photograph tasks, one of them illustrated in Figure 3.

With kind permission from the authors and the publishers of *Cognition* (Elsevier).
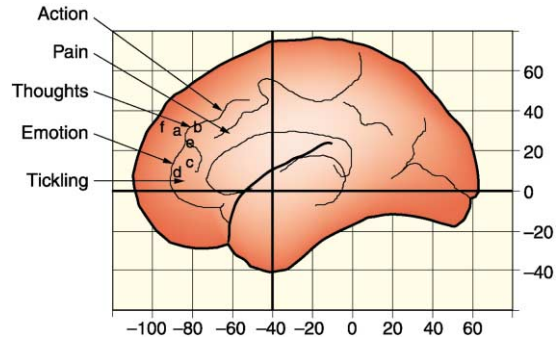

Figure 5. Schematic Representation of Areas in the Medial Prefrontal Cortex Activated by Mentalizing and Related Tasks

Average peak activations during mentalizing are represented by letters for the following studies: a, Goel et al., (1995); b, Fletcher et al., (1995); c, Brunet et al., (2000); d, Vogeley et al., (2001); e, Gallagher et al., (2000); and f, Castelli et al., (2000). These peaks, which are very medial, cluster in the most anterior part of the of paracingulate cortex, a region on the border between anterior cingulate and medial prefrontal cortex, with a range of x values from -6 to $+2$ on average (for y and z values, see figure). The brain volume indicated by the peak activations is circumscribed by a space of $8 \times 18 \times 24$ mm. The same region was also activated in tasks that required subjects to report on their own mental states as indicated by the tips of arrows. The self-reported inner states have included actions (Carter et al., 1998), pain (Rainville et al., 1997), thoughts (McGuire et al., 1996), emotions aroused by pictures (Lane et al., 1997), and sensations experienced during tickling (Blakemore et al., 1998).

with autism are impaired when having to interpret complex social emotions from faces rather than simple basic emotions. Individuals on the severe extreme of the autism spectrum may never make deliberate eye contact and perhaps may not distinguish between biological agents and mechanical objects. This severe form of the condition is characterized by a degree of social detachment that exists over and above mind blindness. However, global asocial behavior is not the rule in autism spectrum disorders.

Functional Brain Imaging and the Neural Substrate of Mentalizing

As yet, only a few studies have investigated the neurophysiological substrate of mentalizing, This is partly due to the difficulties in designing suitable mentalizing tasks with closely matched control tasks (i.e., tasks that differ only in the requirement to mentalize). Existing studies of normal volunteers have used contrasting conditions with stories (Fletcher et al. 1995; Gallagher et al. 2000; Vogeley et al. 2001), cartoons (Gallagher et al. 2000), picture sequences (Brunet et al., 2000), and animated geometric shapes (Castelli et al. 2000). In all these studies, a network of brain regions was identified that was consistently active during mentalizing over and above the specific demands of the respective tasks. This (essentially bilateral) network appears to be the distinctive signature of mentalizing. The peaks of activation are in (1) the medial prefrontal cortex, in particular, the most anterior part of paracingulate cortex, a region on the border between anterior cingulate and medial prefrontal cortex (very medial); (2) the temporal-parietal junction at the top of the superior temporal gyrus (stronger on the right); and (3) the temporal poles adjacent to the amygdala (somewhat stronger on the left). Figures 5–7

show the location of the average peak activations in the six studies quoted above. The figures also show peak activations in these same regions obtained in other highly relevant imaging studies, which can inform us about the function of these regions and how they might contribute to the ability to mentalize.

Why these particular regions and what do they have in common? Clearly, the system identified is tailor-made
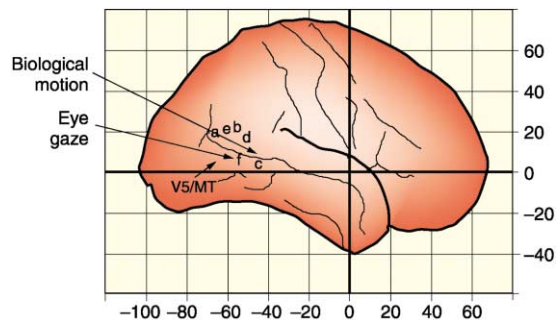

Figure 6. Schematic Representation of Areas in the Temporo-Parietal Region Activated by Mentalizing and Related Tasks

The average peaks of activation in mentalizing studies are indicated by letters a–f (for key see legend to Figure 5). The activations are bilateral, with x values ranging from 41 to 59, but most studies find greater activations in the right hemisphere, as illustrated in the diagram. Arrows point to the average peak activation of two studies involving eye gaze (Wicker et al., 1998; Hoffman and Haxby, 2000) and to the average peak activation obtained in four studies of biological motion (Puce et al., 1998; Bonda et al., 1996; Calvert et al., 1997; Grezes et al., 1999). A third thicker arrow refers to the area V5/MT which processes pure motion. This region is activated by all kinds of motion, whether or not it is biological.
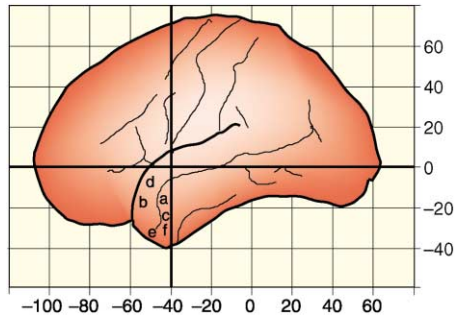
Figure 7. Schematic Representation of Areas in the Peri-Amygdaloid Cortex at the Temporal Poles Activated by Mentalizing Tasks

The average peaks of activation in mentalizing studies are indicated by letters a–f (for key see legend to Figure 5). The activations are bilateral, with x values ranging from 36 to 58, but most studies find greater activations in the left hemisphere, as illustrated in the diagram. The peak activation for mentalizing obtained by Baron Cohen et al. (1999), using an eye gaze interpretation paradigm, is in the amygdala, which is not represented in this figure as it lies in a region some 20 mm more medial.

for processing the intentions of biological agents. As Figure 5 shows, the same space of the medial frontal region is also activated by tasks that imply awareness of the self. As Figure 6 shows, the superior temporal sulcus, mainly on the right, is also activated by tasks that require detection of biological agents. Further studies suggest that this is not confined to biological motion in the visual modality. Activations are shown with such diverse stimuli as faces (Kesler-West et al., 2001), speech (Belin et al., 2000), multimodal cues (Kawashima et al., 1999), and contextual cues to intention (Toni et al., 2001). Other studies suggest that the temporal poles, stronger on the left, are also activated when facts about other agents and the self are remembered, e.g., familiar faces and scenes (Nakamura et al., 2000), and familiar emotionally laden stimuli in different modalities (Dolan et al., 2000).

Unfortunately, none of the studies to date throws light on how a link between all three regions might result in mentalizing. This ability is clearly more than the sum of its parts. If mentalizing crucially involves "decoupling" to keep apart real states of affairs and mental states (Leslie, 1987), then what neurophysiological process might underpin it? One key approach to this problem is the comparative study of brain physiology in autism. If mentalizing is dysfunctional in autism, then the associated brain abnormality should point us in the right direction.

**Evidence from Functional Brain Imaging
of Mentalizing in Autism**
Only three studies to date have explicitly studied individuals with autism on mentalizing tasks. Happé et al. used a story paradigm in a PET study, comparing six normal adults with five able adults with Asperger syndrome. Subjects were scanned while reading stories and answering questions about complex mental states or nonmental inferences, against a baseline of reading and remembering unconnected sentences. While both types of subjects answered the questions satisfactorily, differ-

ences were shown in brain activation. The Asperger group showed less activation in the critical medial prefrontal region, while their peak activation was in a more ventral region of frontal cortex.

In an fMRI study, Baron-Cohen et al. (1999) compared a group of six able individuals with autism with a group of twelve controls. Subjects were asked to judge inner states of people from photographs of the eye region, deciding which of two words best described their mental/emotional state. The contrast was to judge whether the photo was that of a male or a female. Compared to the control group, people with autism demonstrated less extensive activation in frontal regions and no activation in the amygdala.

Castelli et al. (unpublished data) showed silent animations to ten able adults of normal intelligence with autistic disorder and to ten normal adults. The animations featured two triangles moving about on a screen [for examples see http://www.icn.ucl.ac.uk/groups/UF/Research/animations.html]. In one condition they were scripted to elicit attribution of mental states (e.g., coaxing, mocking). In another condition, the triangles moved randomly. This was the contrast that was used to highlight the mentalizing system. During mentalizing, the autism group showed less activation than the controls in the three previously identified brain regions. However, they showed identical activation during mentalizing in one additional region, the occipital gyrus. The activation of this region suggests that both groups devoted more intensive visual analysis to the critical animations. However, connectivity between the occipital and temporoparietal regions was weaker in the autism group than in the controls. This finding provides a clue to a possible reason for mind blindness. The underactivation of the system may be due to a bottleneck for interactive influences between lower and higher order perceptual processing areas. These findings are still preliminary but support the notion of a dysfunction in the specific neural substrate for mentalizing in autism.

**Evidence from Anatomical Studies
of the Brain in Autism**
Can mentalizing failure in autism be linked to some structural abnormality in one or more of the regions of the mentalizing system? Some preliminary evidence for such a possibility exists. Abell et al. (1999) reported structural magnetic resonance imaging (MRI) data on 15 high-functioning individuals with autistic disorder. A voxel-based whole brain analysis identified gray matter differences relative to 15 age- and IQ-matched controls in a distributed system possibly centered on the amygdala. *Decreases* of gray matter were found in *anterior* parts of this system, in particular the paracingulate sulcus and inferior frontal gyrus. The paracingulate region was extremely close to the region that was found to be less active in individuals with autistic disorder in the Happé et al. (1996) and Castelli et al. (unpublished data) imaging studies. *Increases* in gray matter were also found in the *posterior* parts, that is, the peri-amygdaloid cortex and the middle temporal and inferior temporal gyrus. Increases in cerebellar structures were also found. Another structural MRI study (Howard et al. 2000), using volumetric measures, also found an enlargement

in the amygdaloid region in able individuals with autistic disorder. While there are theories of amygdala dysfunction in autism (Baron-Cohen et al., 2000; Howard et al., 2000; Adolphs et al., 2001), the evidence so far suggests that this region is only one component among several that might play a causal role in the origin of mind blindness.

There is also evidence from the few existing histoanatomical studies of autistic brains for abnormalities in these particular brain regions. For example, Bauman and Kemper (1994), in an important series of studies, reported cellular abnormalities in post mortem brains of individuals with autistic disorder, in particular, reduced neuronal cell size and increased cell packing density in regions of the limbic system comprising the hippocampal complex, subiculum, entorhinal cortex, amygdala, mamillary body, medial septal nucleus, and anterior cingulate. Outside the limbic system, reduced numbers of Purkinje cells were found in the posterior and inferior regions of the cerebellum.

Evidence from Acquired Brain Lesions

Given that the anterior part of paracingulate cortex, the superior temporal sulcus at the temporo-parietal junction, and the temporal poles, have been reliably activated in neuroimaging studies of mentalizing, what can we learn from acquired lesions of these areas? We do not expect to find patients suffering the equivalent of autism. For one thing, the effects of developmental brain abnormalities would be different from those of accidentally acquired lesions; for another, there is more to autism than social communication impairment. However, we can gain information on whether intact functioning of these regions is necessary for mentalizing success.

Some studies exist where typical theory of mind tasks have been used with the appropriate control tasks in patients with brain lesions. Happé et al. (2001) showed that a patient who had undergone stereotactic anterior capsulotomy (which severs fronto-thalamic fibers) for intractable depression was specifically impaired on mentalizing tasks following surgery. He was reported to show deterioration in his everyday social behavior. He also failed cartoon tests and story tests of theory of mind. Group studies of patients with prefrontal lesions, which most probably included the critical medial prefrontal region identified in brain imaging studies, also show theory of mind deficits on a variety of tasks (Stone et al., 1998; Channon and Crawford, 2000; Stuss et al., 2001). Importantly, the evidence from the patients who suffer mentalizing failure suggests independence from performance on executive function tasks, which is also thought to be dependent on frontal lobe function (Rowe et al. 2001; Blair and Cipolotti, 2000).

Reports on patients with damage to the superior temporal sulcus at the temporo-parietal junction, mainly on the right, have not so far included mentalizing tasks. However, the right hemisphere stroke patients studied by Happé et al. (1999) with verbal and nonverbal mentalizing tasks could well have included such lesions. These authors found impairments and communication failure as typically seen in some cases of autism, but only in their right hemisphere patients, not in their left hemisphere patients. A study of a patient with congenital

left amygdaloid lesion and a diagnosis of Asperger syndrome showed severe impairment on large variety of mentalizing tasks (Fine et al., 2001). It would be interesting to study mentalizing performance in patients with semantic dementia who suffer from lesions in the temporal pole.

The neuropsychological studies to date suggest that the medial prefrontal cortex may be necessary for mentalizing, but it seems unlikely that it is also sufficient. For lesions in other regions identified as part of the mentalizing system in brain imaging studies, data are as yet too sparse. Other lesion cases too could be informative, in particular in the cerebellum, which has been found to be active during mentalizing in at least some of the few extant studies.

In summary, the results from neuropsychological, structural, and functional imaging studies to date, together with findings on cellular abnormalities in autistic brains, provide some converging evidence for the critical brain abnormalities leading to mind blindness.

Preliminary Thoughts on the Evolution of Mind Reading

The social brain is complex (Brothers, 1997), and very old, but the mentalizing system appears to be of more recent origin. Monkeys, who are known for their complex social lives, are unable to mentalize (Cheney and Seyfarth, 1990), in contrast to chimpanzees and bonobos, who appear to have only incipient mentalizing skills but can engage in deception (de Waal, 1992). Mentalizing adds a new dimension to the repertoire of social interactions. It allows the manipulation of others in particularly subtle ways and reaches far beyond the ability to manipulate their behavior by direct instrumental action. Frith and Frith (2000) speculated that the brain system dedicated to the representation of mental states evolved from the dorsal action system rather than from the ventral object identification system. They argued that much of the social intelligence already so well developed in the monkey could be seen as deriving from the ventral system. It depends upon complex and sophisticated object recognition: recognition of subtle differences in emotional expression, recognition of other individuals, and recognition of their status and relationships. Mentalizing, in contrast, required the development of the capacity to represent actions, and the goals and intentions of agents implicit in actions performed by agents.

Both goal directed movement and eye gaze of other agents provide clues to their desires, and the ability to detect such clues may be a first step in the evolution necessary for mentalizing. The ability to detect goal directed movements is already found in animals without even the incipient ability to mentalize. Neuroimaging studies have pinpointed the temporal-parietal junction at the top of the superior temporal sulcus during both the detection of eye gaze and of mentalizing (see Figure 6). What is known about cells in this part of the cortex? In their work with monkeys, Perrett et al. (1989) have identified cells in the superior temporal sulcus (STS) that respond to moving hands and faces but not to the movement of inanimate objects. Moreover, cells in STS, just as the "mirror neurons" (Gallese et al. 1996) in lateral inferior frontal regions of the macaque brain (F5), re-

spond to the observation of specific actions (e.g., a precision grip). Intriguingly, a neuron in anterior cingulate cortex (close to the area with peak activations in mentalizing studies [Figure 5]) in a patient undergoing neurosurgery was found to respond when the patient received a pinprick and also when he watched pinpricks to the examiner's fingers (Hutchison et al. 1999). It is plausible that mirror mechanisms form an early evolutionary link to mentalizing. Speculatively, their function underpins not only the automatic computation of an agent's goal directed actions, but of an agent's intention toward the self (prey or predator; friend or foe).

However, the detection of agency still does not get us anywhere near the ability to mentalize. How and where might this task be accomplished by neurons? The medial frontal cortex, in particular the most anterior part of the paracingulate cortex, is a promising candidate for the critical next step toward the evolution of mentalizing. First, it is active during the attribution of mental states to others and during the monitoring of inner states of the self (see Figure 5). Second, lesions in this area have been associated with mentalizing failure. Third, abnormal function as well as abnormal structure has been shown in autistic individuals in this region.

Very little is known about cells in anterior cingulate and adjacent medial prefrontal areas. However, an unusual type of projection neuron, spindle cells, has been identified in the anterior cingulate cortex (layer Vb) of bonobo, chimpanzee and man, but not in any other primate species or other mammals (Nimchinsky et al. 1999). The authors suggest that spindle cells in the anterior cingulate might represent a population of specialized neurons that could integrate inputs with emotional overtones and project to motor centers controlling vocalization or facial expression. While the function of the spindle cells is as yet unknown, it is notable that their appearance coincides with observations of incipient mentalizing in chimpanzees and bonobo but lack of mentalizing in monkeys.

Concluding Remarks

Mind blindness makes sense of the core social and communication impairments of individuals with autism. The hypothesis rests on robust experimental evidence and has the unique advantage of unifying the core symptoms that define the spectrum of autistic disorders by a single explanation and is able to account for the heterogeneity that is associated with autistic spectrum disorders.

Mentalizing rests on a separable brain system and can be selectively damaged by acquired brain lesions. The physiological basis of mentalizing remains unknown and appears to involve a complex, essentially bilateral network of cortical regions. In imaging studies, the most consistently activated regions are paracingulate sulcus in medial prefrontal regions, superior temporal sulcus at the temporo-parietal junction (more strongly on the right), and peri-amygdaloid cortex at the temporal poles (more strongly on the left). Preliminary findings suggest that the brain abnormality that results in autism compromises the functional connectivity of this network and leads to reduced activations in all three regions. Converging evidence from autism and acquired brain lesions suggests that an intact medial prefrontal region is necessary for mentalizing.

Experimental evidence shows that the typical social communication impairment of autism can be well explained by impairment in the mentalizing mechanism. Able individuals with autism spectrum disorders can with time and practice achieve awareness of mental states by compensatory learning. In normally developing children, the mentalizing mechanism allows fast learning of socially and culturally transmitted knowledge, including the meaning of words. Since children with autism spectrum disorders can be very intelligent and can learn by other means, the underlying brain abnormality must be sufficiently specific and circumscribed so as not to compromise general information processing ability. This has implications for a modular view of the development of cognitive functions.

References

Abell, F., Krams, M., Ashburner, J., Passingham, R., Friston, K., Frackowiak, R., Happé, F., Frith, C., and Frith, W. (1999). The neuroanatomy of autism: a voxel-based whole brain analysis of structural scans. Neuroreport 10, 1647–1651.

Adolphs, R., Sears, L., and Piven, J. (2001). Abnormal processing of social information from faces in autism. J. Cog. Neurosci. 13, 232–240.

American Psychiatric Association. (2000). Diagnostic and Statistical Manual of Mental Disorders, 4th Edition DSM-IV-TR (Text Revision) (Washington, DC: American Psychiatric Association).

Asperger, H. (1944). "Autistic psychopathy" in childhood. In Autism and Asperger Syndrome, U. Frith, ed. and trans., 1991 (Cambridge: Cambridge University Press), pp. 37–92.

Bailey, A.J., Phillips, W., and Rutter, M. (1996). Autism: integrating clinical, genetic, neuropsychological, and neurobiological perspectives. J. Child Psychol. Psychiatry. 37, 89–126.

Baird, G., Charman, T., Baron-Cohen, S., Cox, A., Swettenham, J., Wheelwright, S., and Drew, A. (2000). A screening instrument for autism at 18 months of age: a 6-year follow-up study. J. Am. Acad. Child Adolesc. Psychiatry. 39, 694–702.

Baldwin, D.A., Markman, E.M., Bill, B., Desjardins, R.N., Irwin, J.M., and Tidball, G. (1996). Infants' reliance on a social criterion for establishing word-objects relations. Child Dev. 67, 3135–3153.

Baron-Cohen, S. (1995). Mindblindness: An Essay on Autism and Theory of Mind (Cambridge, MA: MIT Press).

Baron-Cohen, S., Leslie, A., and Frith, U. (1985). Does the autistic child have a "theory of mind"? Cognition 21, 37–46.

Baron-Cohen, S., Leslie, A., and Frith, U. (1986). Mechanical, behavioural and intentional understanding of picture stories in autistic children. Brit. J. Dev. Psych. 4, 113–125.

Baron-Cohen, S., Cox, A., Baird, G., Swettenham, J., Nightingale, N., Morgan. K., Drew, A., and Charman, T. (1996). Psychological markers in the detection of autism in infancy in a large population. Brit. J. Psychiatry 168, 158–163.

Baron-Cohen, S., Wheelwright, S., and Joliffe, T. (1997). Is there a "language of the eyes"? Evidence from normal adults and adults with autism or Asperger syndrome. Visual Cognition 68, 48–57.

Baron-Cohen, S., Tager-Flusberg, H., and Cohen, D. (1999). Understanding Other Minds II: Perspectives from Autism and Cognitive Neuroscience (Oxford: Oxford University Press).

Baron-Cohen, S., Ring, H.A., Bullmore, E.T., Wheelwright, S., Ash-

win, C., and Williams, S.C.R. (2000). The amygdala theory of autism. Neurosci. Biobehav. Rev. *24*, 355–364.

Bauman, M., and Kemper, T.L. (1994). Neuroanatomic observations of the brain in autism. In The Neurobiology of Autism, Bauman, M., and Kemper, T.L., ed. (Baltimore: Johns Hopkins University Press), pp. 119–145.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. Nature *403*, 309–312.

Black, I. (1998). Genes, brain, and mind: the evolution of cognition. Neuron *20*, 1073–1080.

Blair, R.J.R., and Cipolotti, L. (2000). Impaired social response reversal: A case of "acquired sociopathy." Brain *123*, 1122–1141.

Blakemore, S.J., Wolpert, D., and Frith, C.D. (1998). Central cancellation of self-produced tickle sensation. Nat. Neurosci. *1*, 635–640.

Bloom, P. (2000). How Children Learn the Meaning of Words (Cambridge, MA: MIT Press).

Bonda, E., Petrides, M., Ostry, D., and Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. J. Neurosci. *16*, 3737–3744.

Bretherton, I. (1992). Social Referencing, Intentional Communication and the Interfacing of Minds in Infancy (New York: Plenum Press).

Brothers, L. (1997). Friday's Footprint: How Society Shapes the Human Mind (Oxford University Press).

Brunet, E., Sarfate, Y., Hardy-Bayle, M.C., and Decety, J. (2000). A PET investigation of the attribution of intentions with a nonverbal task. Neuroimage *11*, 157–166.

Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P.K., Woodruff, P.W., Iverson, S.D., and David, A.S. (1997). Activation of auditory cortex during silent lip reading. Science *276*, 593–596.

Carpenter, M., Nagell, K., and Tomasello, M. (1998). Social cognition, joint attention and communicative competence from 9 to 15 months of age. Monogr. Soc. Res. Child Dev. *63*, 1–143.

Carter, C., Braver, T., Barch, D., Botvinick, M., Noll, D., and Cohen, J. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. Science *280*, 747–749.

Castelli, F., Happé, F., Frith, U., and Frith, C.D. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. Neuroimage *12*, 314–325.

Channon, S., and Crawford, S. (2000). The effects of anterior lesions on performance of a story comprehension test: left anterior impairment on a theory of mind-type task. Neuropsychologia *38*, 1006–1017.

Cheney, D.L., and Seyfarth, R.M. (1990). How monkeys see the world: Inside the mind of another species (Chicago: Chicago University Press).

Dolan, R.J., Lane, R., Chua, P., and Fletcher, F. (2000). Dissociable temporal lobe activations during emotional episodic memory retrieval. Neuroimage *11*, 203–209.

Fine, C., Lumsden, J., and Blair, J. (2001). Dissociation between "theory of mind" and executive functions in a patient with early left amygdala damage. Brain *124*, 287–298.

Fletcher, P.C., Happé, F., Frith, U., Baker, S.C., Dolan, R.J., Frackowiak, R.S.J., and Frith, C.D. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. Cognition *57*, 109–128.

Fombonne, E. (1999). The epidemiology of autism: a review. Psychol. Med. *29*, 769–786.

Frith, U., and Happé, F. (1998). Why specific developmental disorders are not specific: On-line and developmental effect in autism and dyslexia. Dev. Sci. *1*, 267–272.

Frith, C.D., and Frith, U. (2000). The physiological basis of theory of mind. In Understanding Other Minds, Perspectives from Developmental Cognitive Neuroscience, Second Edition, S. Baron-Cohen, H. Tager-Flusberg, and D.J. Cohen, eds. (Oxford: Oxford University Press), pp. 335–356.

Frith, U., Leslie, A., and Morton, J. (1991). The cognitive basis of a biological disorder. Trends Neurosci. *14*, 433–438.

Gallagher, H., Happé, F., Brunswick, N., Fletcher, P.C., Frith, U., and Frith, C.D. (2000). Reading the mind in cartoons and stories: An fMRI study of "theory of mind" in verbal and non-verbal tasks. Neuropsychologia *38*, 11–21.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the prefrontal cortex. Brain *119*, 593–609.

Grezes, J., Costes, N., and Decety, J. (1999). The effects of learning and intention on the neural network involved in the perception of meaningless actions. Brain *122*, 1875–1887.

Goel, V., Grafman, J., Sadato, N., and Hallett, M. (1995). Modelling other minds. Neuroreport *6*, 1741–1746.

Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. Child Dev. *66*, 643–855.

Happé, F. (1999). Autism: Cognitive deficit or cognitive style. Trends Cog. Sci. *3*, 216–222.

Happé, F., and Frith, U. (1996). The neuropsychology of autism. Brain *119*, 1377–1400.

Happé, F., Ehlers, S., Fletcher, S., Frith, U., Johannsson, M., Gillberg, C., Dolan, R., Frackowiak, R., and Frith, C. (1996). "Theory of mind" in the brain. Evidence from a PET scan study of Asperger Syndrome. Neuroreport *8*, 197–201.

Happé, F., Brownell, H., and Winner, E. (1999). Acquired "theory of mind" following stroke. Cognition *70*, 211–240.

Happé, F., Malhi, G., and Checkley, S. (2001). Acquired mind-blindness following frontal lobe surgery? A single case study of impaired "theory of mind" in a patient treated with stereotactic anterior capsulotomy. Neuropsychologia *39*, 83–90.

Hobson, P. (1993). Autism and the Development of Mind. (Hove, UK: Erlbaum Assoc.).

Hoffman, E.A., and Haxby, J.V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nat. Neurosci. *3*, 80–84.

Howard, M., Cowell, P., Boucher, P., Broks, P., Mayes, A., Farrant, A., and Roberts, N. (2000). Convergent neuroanatomical and behavioural evidence of an amygdala hypothesis of autism. Neuroreport *11*, 2931–2935.

Hutchison, W.D., Davis, K.D., Lozano, A.M., Tasker, R.R., and Dostrovsky, J.O. (1999). Pain-related neurons in the human cingulate cortex. Nat. Neurosci. *2*, 403–405.

Kanner, L. (1943). Autistic disturbances of affective contact. Nervous Child *2*, 217–250.

Kawashima, R., Imaizumi, S., Mori, K., Okada, K., Goto, R., Kiritani, S., Ogawa, A., and Fukuda, H. (1999). Selective visual and auditory attention toward utterances – a PET study. Neuroimage *10*, 209–215.

Kesler-West, M.L., Andersen, A.H., Smith, C.D., Avison, M.J., Davis, C.E., Kryscio, R.J., and Blonder, L.X. (2001). Neural substrates of facial emotion processing using fMRI. Brain Res. Cogn. Brain Res. *11*, 213–226.

Klin, A. (1991). Young autistic children's listening preferences in regard to speech: A possible characterization of the symptom of social withdrawal. J. Autism Dev. Disord. *21*, 29–42.

Lane, R.D., Fink, G.R., Chua, P.M., and Dolan, R.J. (1997). Neural activation during selective attention to subjective emotional responses. Neuroreport *8*, 3969–3972.

Leslie, A. (1987). Pretence and representation: The origins of "theory of mind." Psychol. Rev. *94*, 412–426.

Leslie, A., and Thaiss, L. (1992). Domain specificity in conceptual development: Evidence from autism. Cognition *43*, 467–479.

Lord, C., Cook, E.H., Leventhal, B.L., and Amaral, D.G. (2000). Autism spectrum disorders. Neuron *28*, 355–363.

Maestrini, E., Paul, A., Monaco, A.P., and Bailey, A. (2000). Identifying autism susceptibility genes. Neuron *28*, 19–24.

McGuire, P.K., Silbersweig, D.A., and Frith, C.D. (1996). Functional neuroanatomy of verbal self-monitoring. Brain *119*, 907–917.

Meltzoff, A. (1995). Understanding the intentions of others: Re-

enactment of intended acts by 18-month-old children. Dev. Psychol. *31*, 838–850.

Nakamura, K., Kawashima, R., Sato, N., Nakamura, A., Sugiura, M., Kato, T., Hatano, K., Ito, K., Fukuda, H., et al. (2000). Functional delineation of the human occipito-temporal areas related to face and scene processing. Brain *123*, 1903–1912.

Nimchinsky, E.A., Gilissen, E., Allman, J.M., Perl, D.P., Erwin, J.M., and Hof, P.R. (1999). A neuronal morphologic type unique to humans and great apes. Proc. Natl. Acad. Sci. USA *96*, 5268–5273.

Perrett, D.I., Harries, M.H., Bevan, R., Thomas, S., Benson, P., Mistlin, A., Chitty, A., Hietanen, J., and Ortega, J. (1989). Frameworks of analysis for the neural representation of animate objects and actions. J. Exp. Biol. *146*, 87–113.

Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? Behav. Brain Sci. *1*, 515–526.

Puce, A., Allison, T., Bentin, S., Gore, J.C., and McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. J. Neurosci. *18*, 2188–2199.

Rainville, P., Duncan, G.H., Price, D.D., Carrier, B., and Bushnell, M.C. (1997). Pain effect encoded in human anterior cingulate, but not somatosensory cortex. Science *277*, 968–971.

Repacholi, B.M. (1998). Infants' use of attentional cues to identify the referent of another person's emotional expression. Dev. Psychol. *34*, 1017–1025.

Rowe, A.D., Bullock, P.R., Polkey, C.E., and Morris, R.G. (2001). "Theory of mind" impairments and their relationship to executive functioning following frontal lobe excisions. Brain *124*, 600–616.

Russell, J., ed. (1998). Autism as an Executive Disorder. (Oxford: Oxford University Press).

Scheuffgen, K., Happé, F., Anderson, M., and Frith, U. (2000). High "intelligence," low "IQ?" Speed of processing and measured IQ in children with autism. Dev. Psychopathol. *12*, 83–90.

Schultz, R.T., Gauthier, I., Klin, A., Fulbright, R.K., Anderson, A.W., Volkmar, F., Skudlarski, P., Lacadie, C., Cohen, D.J., and Gore, J.C. (2000). Abnormal ventral temporal cortical activity during face discrimination among individuals with autism and Asperger's syndrome. Arch. Gen. Psychiatry *57*, 331–340.

Sigman, M., and Capps, L. (1997). Children with Autism. A Developmental Perspective (Cambridge, MA: Harvard University Press).

Stone, V.E., Baron-Cohen, S., and Knight, R.T. (1998). Frontal lobe contributions to theory of mind. J. Cog. Neurosci. *10*, 640–656.

Stuss, D., Gallup, G., Jr., and Alexander, M. (2001). The frontal lobes are necessary for "theory of mind." Brain *124*, 279–289.

Toni, T., Thoenissen, D., and Zilles, K. (2001). Movement preparation and motor intention. Neuroimage *14*, S110–117.

De Waal, F.B.M. (1992). Intentional deception in primates. Evol. Anthropol. *1*, 86–92.

Vogely, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., Maier, W., Shah, N.J., Fink, G.R., and Zilles, K. (2001). Mind reading: Neural mechanisms of theory of mind and self-perspective. Neuroimage *14*, 170–181.

Wicker, B., Michel, F., Henaff, M.A., and Decety, J. (1998). Brain regions involved in the perception of gaze: a PET study. Neuroimage *8*, 221–227.

Wimmer, H., and Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition *13*, 103–128.

Wing, L., and Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. J. Autism Dev. Disord. *9*, 11–29.

World Health Organization (1992). The ICD-10 classification of mental and behavioural disorders: Clinical Descriptions and Diagnostic Guidelines (Geneva: World Health Organization).