This paper appeared in Corballis, M, & Lea, S (eds) The descent of mind: psychological perspectives on hominid evolution. Oxford University Press 1999

Evolution of a theory of mind?

Simon Baron-Cohen

Departments of Experimental Psychology and Psychiatry

University of Cambridge

Downing St, Cambridge, CB2 3EB, UK

email: sb205@cus.cam.ac.uk

Acknowledgements: The author was supported by the Medical Research Council, the Wellcome Trust, and the Gatsby Foundation, during the period of this work. I am grateful to Steve Mithen and Andy Whiten for their comments on the first draft of this chapter.

Homo sapiens sapiens is arguably the only species which possesses a developed 'theory of mind'. By this I mean the ability to attribute the full range of mental states (both goal states and espistemic states) to ourselves and to others, and to use such attributions to make sense of and predict behaviour. 'Theory of mind' is the phrase coined by (Premack & Woodruff, 1978), and there are a set of synonyms for this ability: mind-reading (Whiten, 1991), mentalizing (Morton, Frith & Leslie, 1991), folk psychology (Wellman, 1990), and the Intentional stance (Dennett, 1987). In this chapter, I will use the term theory of mind, for convenience, whilst assuming that any of these alternative synonyms would be equally applicable.

There is considerable interest in tracing the evolution of a theory of mind, because of its central importance in modern human behaviour. Like language or bipedalism, a theory of mind can be taken as a major milestone in primate evolution. The importance of language and bipedalism is in some sense easier to see, or at least, the arguments are more familiar. Among other things, language enabled primates to manipulate the behaviour of conspecifics at a distance (a form of remote control), to obtain information about events that they had not directly witnessed, to inform others about events they had not directly witnessed, and to act co-operatively. Bipedalism enabled primates to use their forelimbs for other things than just locomotion, such as carrying, throwing, and transforming objects. It also enabled foraging into new niches by reducing water loss. For at least these reasons, language and bipedalism have been justifiably the subject of considerable research. But what is the equivalent enormous importance of a theory of mind? In what

way did the evolution of a theory of mind transform primate evolution? Is it really justifiable to equate the importance of a theory of mind with language or bipedalism?

In this chapter I try to do two things. First, I want to persuade you that that actually the evolution of a theory of mind is not only *as* important as these other developments, but in some respects, is *more* important than them. The main argument I will use here is that without a theory of mind, having the ability to speak or perceive speech would have been of little value. You can guess therefore that I will be arguing that a theory of mind must have *preceded* any ability to use language in the communicative way in which it is used today. Secondly, I want to question the claim by (Mithen, 1996) that a theory of mind evolved around 6 million years ago. This claim is made on the basis that existing species of ape have a full theory of mind, and our common ancestor with modern apes would have lived around 6 million years ago. My reason for questioning this date is that recent experimental evidence throws doubt on the idea that modern apes have a full theory of mind. Instead, from available evidence, I will argue that all we can conclude is that a theory of mind proper was certainly evident 40,000 years ago, but beyond then there is no substantive evidence for its existence.

1. The importance of a theory of mind

To grasp the importance of a theory of mind, consider the following list of 8 behaviours that depend on it: For each behaviour, I spell out the reason why each depends on a theory (or concept) of mind.

i. *Intentionally communicating with others*

We can define intentional communication as requiring a theory of mind, if we restrict it to those communicative acts that are produced in order to change the knowledge state of the listener. Thus, when a dog barks at a cat, this is not intentional communication because the dog is not intending to update the knowledge state of the cat. The effect of the bark may well be that the cat becomes aware (i.e. comes to know) that there is a dog nearby, but the dog's intention might have been far simpler: to make the cat run away. If there was an intention behind the bark, it was simply to change the cat's behaviour, without any necessary reference to changing the cat's knowledge state or mind.

In contrast, when I tell you that Liverpool won the football match, I am doing so in order to give you new information that I believe you do not have, and that you might be interested in or want. I am trying to change your knowledge state. Boring as it may be, this little utterance counts as intentional communication, and necessarily requires a theory of mind.

So, to intentionally inform others, one needs a concept that *others have minds* that can be informed or uninformed. Indeed, one needs a concept of information, which is itself intrinsically mentalistic. At least, as long as one defines the goal of intentional informing as being to change the other animal's knowledge state. Two more examples should serve to clarify why this definition is needed. If a person shouts "watch out!", to inform

the listener of an impending danger, the intention is to change the other person's knowledge about the current state of the environment. Equally, if a person produces a bus-timetable to inform another person about the times of future events, the intention is to change another person's knowledge about the future state of the environment.

Whereas intentionally informing others necessarily requires a theory of mind, unintentionally informing others does not. For example, if a trail of footprints in the sand was left unintentionally, the animal producing the footprints was probably not thinking about how another animal's knowledge state might be changed by this information. Indeed, the animal producing the footprints was probably not thinking about the footprints as information at all. In contrast, if a trail of footprints in the sand was left intentionally, the animal producing the footprints was probably thinking about another animal's thoughts - for example, wanting to make it possible for the other animal to *know* how to find him or her, or wanting to make the other animal *believe* the footprints lead to where the prey is.

ii. *Repairing failed communication with others*

Conversational repair is another good index of a speaker's theory of mind. If one animal is attempting to communicate to a listener, but is failing, the speaker can do one of two things: repeat the utterance in an identical fashion, or try to communicate the same message in a different way. The latter strategy is likely to indicate that the speaker

believes the listener has not *understood* the intended message, and that the speaker is trying a different method to get the listener to understand.

Thus, if I say "Have you seen it?", you are likely to either look at me quizzically, or ask me what on earth I'm talking about. If I simply repeat the utterance, you'll probably repeat your last response. If however I rephrase the utterance to "Have you seen my wallet?", you'll probably answer usefully. My rephrasing of the utterance depended on me assuming that my first attempt at communication failed because it was ambiguous in some way, and that by using less ambiguous words, you (the listener) would then understand it. All very mentalistic.

iii. *Teaching others*

Teaching others also necessarily requires a theory of mind. Again, this assumes we are restricting the definition of teaching to those behaviours produced by a more knowledgeable animal, with the intention of changing the knowledge state of a less knowledgeable listener. For example, a mother showing her juvenile daughter how to use a tool would count as an instance of teaching.

iv. *Intentionally persuading others*

Persuasion is an aspect of intentional communication, and as such it necessarily requires a theory of mind. But it is worth special mention here, because it is produced with the

specific intention of changing someone else's belief about the value of something. I might try to persuade you to buy A rather than B, or to go down route x rather than route y, or to choose me rather than him. Admittedly all of these also are produced with the intention of changing your behaviour, but the means to doing this is by changing your beliefs about the value of the different options.

In contrast, if a deer grows a huge pair of antlers, this may make a doe choose him over the deer with the smaller antlers, but this does not count as an act of intentional persuasion, according to the above definition. The deer's body did not set out to change the beliefs of the doe. Equally, a male gorilla beating his chest may have the effect of making another male gorilla turn and run away, but this again does not count as an act of intentional persuasion. It may be no more complex than the earlier example of the dog barking, causing the cat to run away. There is no evidence that the animal is considering the mental states of the audience.

v. *Intentionally deceiving others*

Intentional deception also requires a theory of mind. Indeed, we will define intentional deception as occurring when one animal attempts to place false information in the mind of another, or attempts to withhold true information from the mind of another. Thus, making a trail of footprints lead from locations A to B, and then swinging through the trees (thereby leaving no footprints) to hide in location C, would count as an instance of intentional deception. Rubbing out the trail of footprints from A to B would also count as

intentional deception. In both cases, the first animal is attempting to influence the knowledge state of another animal.

In contrast, a stick insect, whose appearance saves it from being eaten by a predator, is not engaging in a deception that requires any theory of mind. Indeed, the stick insect may not be thinking about anything, let alone the mind of its predator. The same applies to an animal with camouflage. True, by staying still, it may not be seen by its predator, but it was probably not aiming to make other animals *think* it was not there. It was not necessarily thinking about what other animals were thinking at all.

vi. *Building shared plans and goals*

Sharing a plan or goal with another animal requires a "meeting of minds". Both animals must recognise the intention of the other animal, and subsequently work out how to mesh their actions with those of the other animal to achieve the shared goal. Take this example:

A troop of chimpanzees are hunting a baby monkey, to eat it. The goal is to get the infant monkey away from its mother, scare off the adult monkeys, and kill the baby. They hunt as a team, and achieve the goal. This may not be evidence of building a shared plan in that each individual chimpanzee may simply be pursuing their own individual goals, which just happen to coincide. Thus, an infant monkey comes into sight, adult male chimpanzees love the taste of baby monkeys, so all the adult male chimps in the troop recognise the reward and aim for it. The adult male chimps' goals all coincide because

they all share the same taste or food preference. They mesh their actions with those of the other chimps, in the sense that as one chimp attempts to grab the infant monkey and gets beaten off, the next one jumps in and tries to do the same thing. Each chimp may even recognise the intentions of the others, representing for example that "He is trying to get the infant monkey". But this still falls short of being an example of building shared plans, in that a shared plan involves both animals recognising that they are both holding the same goal.

In contrast, consider another example: Two chimps carry a log, then lean it up against a high wall. One holds it still whilst the other scrambles up it. When the climber reaches the top, he then turns and holds the log whilst the other chimp scrambles up it. This counts as a convincing example of building a shared plan in that the goal is not achievable without the help of the other animal, and both animals cannot help eachother without realising what both are aiming at. With this joint plan in mind, they can recognise why the other is taking the different role that they are (e.g.: the holder of the log, versus the climber up the log).

vii. *Intentionally sharing a focus or topic of attention*

The same argument applies to the sharing of a focus on attention. Two animals can coincidentally look at the same target. This is not shared attention, if each animal is simply aware only of his or her own viewpoint. Shared attention is necessarily mentalistic in that both animals must be aware of the other animal being aware of looking

at the same target at they are. Thus, I see you turn to look out of the window. If I then look out of the window, this is not shared attention. If I see you have looked back and have seen me looking out of the window, then this probably is. More convincing is when you point out the window, and keep pointing until I turn to look out the window. Chances are that if this is genuine shared attention, then I will acknowledge in some way that I have seen what you were trying to get me to see. I will turn back to look at you, and smile, or nod, etc.,

viii. *Pretending*

Last on this list is pretending. This is different from intentional deception in that the intention is not to mislead or plant a false belief in an audience, but simply to pretend. The intention is to temporarily treat one object as if it is another, or as if it had attributes that it clearly does not have. Pretending necessarily requires a theory of mind in that one has to be able to switch between thinking about one's *knowledge* of the real identity of the object, and its current *pretend* identity. Pretending only exists in the mind of the pretender. It is not an intrinsic part of the object.

The empirical contribution of studying autism

Let's take stock. We have surveyed 8 behaviours which are claimed to all require a theory of mind. The reason for this brief survey was to illustrate quite how important a theory of mind is. Without a theory of mind, none of these behaviours would be seen.

This is true by definition, if the analysis of the above 8 behaviours is correct. But it is also true empirically: children with autism are a natural test of this in that many of these children fail the standard test of understanding false beliefs, suggesting they have difficulties in the development of a theory of mind, and they fail to show the above behaviours in the normal way (see (Baron-Cohen, 1995), for a review of the evidence; this is summarized in Table 1). Indeed, autism is a clear illustration of what human life would be like if one lacked a theory of mind. The most devastating effect is on the ability to socialize, communicate, and use imagination. It is hard to think of aspects of our psychology that are more central or important than these. Certainly, I hope you agree, at least as important as language (syntax), or bipedalism.

insert Table 1 here

Language without a theory of mind

These children also show us quite how useless a language capacity is without a theory of mind. Strip out a theory of mind from language use and you have an individual who may have some syntax, the ability to build a vocabulary, and a semantic system. Crucially, what would be missing from their language use and comprehension is 'pragmatics' - being able to decipher the speaker's communicative intentions, decipher non-literal language, read "between the lines", understand jokes, and tailor one's speech to fit the

listener's background mental states (their knowledge, interest, expectations, etc.,). This is the aspect of language that is missing from the language of most children with autism (Baron-Cohen, 1988a; Baron-Cohen, 1988b; Paul & Cohen, 1985; Surian, Baron-Cohen & Van der Lely, 1996; Tager-Flusberg, 1992; Tager-Flusberg, 1993).

The relationship between language and theory of mind is likely to be a very complex one, for several reasons. First, understanding that words *refer* presumes a concept of intention or goal. Second, mapping reference correctly, in language acquisition, is massively facilitated by joint attention (Baldwin, 1991), itself an early form of mind-reading (Baron-Cohen, 1995). These two points imply that normal 12-18 months old language learners benefit by *first* having the mental state concepts of intention and attention. Without this, the infant would be left with the puzzle of what people are doing when they are talking. Third, language serves as a virtual "print-out" of a speaker's mind, for the listener, giving the listener access to a description of the speaker's thoughts. Fourth, syntax can serve to disambiguate a speaker's intended meaning; that is, syntax is used for the informing function (Cheyney, personal communication). Given this set of connections between language and theory of mind it may be no surprise that children with autism (who are impaired in theory of mind) invariably show language delay.

2. When did a theory of mind evolve?

Existing primatological evidence

Let's turn to the question of most relevance to this book: the evolutionary question. Here things are necessarily speculative, as we attempt to peer into the mists of time, but there are two strategies available for us to answer this question. First, do existing monkey and ape species have a full theory of mind? If so, we can assume a theory of mind evolved as early as the common ancestor between us and these existing primate species. Secondly, what clues does the palaeo-archeological record give us? Mithen (1996) calls this second strategy "cognitive archeology" - inferring their behaviour (and thence their cognitive abilities) from fossil records of early hominids, and from their tool use, cave painting, etc.,

The 6 million years hypothesis

Using the first strategy, of looking at extant species of monkey and ape, has led field observers to conclude that monkey species show little if any theory of mind, but that modern apes do. That is, they show signs of deception, in their natural behaviour, which is one hallmark of a theory of mind (Byrne & Whiten, 1991). This leads Byrne and Whiten to conclude that the common ancestor of modern humans and apes, who lived around 6 million years ago, had elements of a theory of mind. This is not "very very old" in evolutionary terms (for example, the common ancestor between modern humans and monkeys lived around 35 million years ago - see Figure 1), but it is still very old compared to the alternative hypothesis, reviewed next.

insert Figure 1 here

The 40,000 years hypothesis

There is reason to doubt the 6 million year old hypothesis is correct. This is because experimental tests of a theory of mind in modern apes have found it hard to obtain convincing evidence of the ability to attribute beliefs to others. Even Premack and Woodruff, who first asked the question about whether chimpanzees could attribute false beliefs (Premack & Woodruff, 1978) finally concluded they could not, when their chimps failed a better controlled test of the same ability (Premack, 1988). In fact, (Povinelli & Eddy, 1996) have found it hard to obtain evidence that chimpanzees even understand gaze as an indicator of the mental state of attention, in the way human children do.

There is a second reason to doubt the 6 million year old hypothesis. If modern apes (chimpanzees, oran utans, and gorillas) have a theory of mind, why don't we see signs of the 8 behaviours listed above in their natural behaviour? (Cheney & Seyfarth, 1990) put forward some interesting data suggesting monkeys may have 'words' but they conclude that monkeys use such 'words' to change the behaviour (rather than the mental state) of the listener. Equally, to take the last item on the earlier list, why don't we see evidence of pretend play in monkeys or apes - a behavior which in human children emerges very early, around 18 months of age (Leslie, 1987)?

Palaeo-archeological evidence

We can then turn to the alternative strategy for dating the evolution of a theory of mind, namely, the palaeo-archeological evidence. This tells us that *tool use* was evident from 2-3 million years ago, and became more sophisticated over time. By itself, this is not evidence that these hominid ancestors had a theory of mind.

Around 30,000 years ago one sees the earliest examples of *cave paintings*, but again, this is not evidence that the artist had a theory of mind. Some people might be intially drawn to conclude that any animal capable of art must have a theory of mind. But recall that many children with autism, who fail tests of a theory of mind, and who show none of the 8 behaviours listed earlier, are able and even gifted artists (Charman & Baron-Cohen, 1992; Charman & Baron-Cohen, 1993; Selfe, 1977). Art clearly requires an ability to represent representations, but not necessarily an ability to represent mental states like beliefs. (For more on this distinction, see (Charman & Baron-Cohen, 1995; Leekam & Perner, 1991; Leslie & Thaiss, 1992).

The first fiction

Mithen (1996) has performed an invaluable service to cognitive neuroscience in reviewing some much more relevant evidence from palaeo-archaeology. Around the same time, 30,000 years ago, we see the first evidence of special forms of art: statues of *impossible entities*, such as the half-man-half lion ivory statuette from Hohlenstein-Stadel, southern Germany, dated around 30-33,000 years ago (Figure 2), and the painting

of the half-man-half-reindeer, from Trois-Freres, Ariege, in France, dated around 30,000 years ago (Figure 3).

insert Figures 2 and 3 here

These are of interest because they are representations of *fictions*. They are necessarily representations of the artist's mind, of the artist thinking about his or her own thoughts. They are also, incidently, direct evidence of the capacity for pretend play. Animals that are half-man-half-lion have never existed, except in the world of the imagination, of pretence. (I am happy to be corrected on this point). So, here we can say with some confidence that a full theory of mind must be at least as old as 30-40,000 years.

Supporting evidence: adornment

There is further evidence that at this time, early Homo sapiens sapiens had a theory of mind. The archeological record shows the existence of burial at this time, which implies that our ancestors then were concerned about death. By itself, this does not tell us that they could think about the mind. But Mithen points out that burials around 28,000 years ago also include the dead person being *adorned with jewelry*. For example, at Sungir, in Russia, a 60 year old man was buried with an adolescent male and female. All three individuals were decorated with thousands of ivory beads, necklaces, and bracelets (see Figure 4).

insert Figure 4 here

Now why would someone adorn themselves, or adorn their dead relative? This behaviour can be taken as evidence that the decorator cared about how other people perceived the adorned person - they wanted an audience to *think* the person was beautiful, or of high status, or worthy of an after-life, or whatever. If this is not an excessively rich interpretation of jewelry-use, then this is an additional strand of evidence that around 30-40,000 years ago, our ancestors had a theory of mind.

Is early religion relevant?

If the archeological evidence from these burials also indicates the existence of religion, then this might also be supportive evidence for the existence of a theory of mind 40,000 years ago. This is because anthropological evidence (Boyer, 1990) suggests that the common feature of all current religions is that a supernatural agency is postulated - a god, a spirit - who can *communicate* with you, possibly judge (i.e. *think about*) you, and who can be appeased by ritual acts. The idea of a supernatural agency of this kind would be impossible without a theory of mind. Indeed, the idea that ritual actions might cause good outcomes or ward off bad ones is itself a belief in intentional causation rather than purely physical causation.

But let's leave religion out of it, since adorned skeletons in graves are not clear evidence of religion. Let's stick to our two strong clues: art of a purely fictional kind (Figures 2

and 3) and adornment (Figure 4). Here we can jump back to ask about these behaviours in autism, since children with autism who lack a theory of mind should also not produce art of a purely fictional kind, or bother with adornment. What does the evidence show?

Back to autism

Consistent with this idea, children with autism do draw, but tend to draw objects they have seen (buildings, cars, electricity pylons, train-stations, etc). When challenged to draw purely fictional entities, like a "man that could never exist", whilst normal 4 year old children produce sketches of men with two heads or three arms (Karmiloff-Smith, 1990), children with autism do not (Scott & Baron-Cohen, 1996). Those children with autism who are more able, and who can pass first-order theory of mind tests, can draw such fictional entities (Craig, Baron-Cohen & Scott, submitted). This is empirical evidence that the kind of art we think of as involving pretend play, and is a good indicator of whether the artist has a theory of mind. There is no systematic evidence about adornment in autism, but it is widely noted that such children pay little attention to how they appear to others, for example, showing little if any signs of embarrassment (Baron-Cohen, Spitz & Cross, 1993) or interest in fashion (Baron-Cohen, 1993).

3. Conclusions

Mithen (1996) clearly supports the 6 million year old hypothesis:

> " A specialized domain of social intelligence first appeared in the course of human evolution after 55 million years ago. This gradually increased in complexity with the addition of further mental modules, such as that for a theory of mind between 35 and 6 million years ago" (p.94)

He bases this conclusion on the evident "social intelligence" of monkeys and apes today. However, there is a danger of confounding social intelligence with theory of mind. It is clear that many monkey species and the apes show social intelligence in that they form alliances, keep track of social status, and behave tactically in grooming those allies they depend on (De Waal, 1989; Whiten, 1991). Whilst this is fascinating, and may be evidence of social intelligence evolving independently of general intelligence, it is not necessarily evidence of the possession of a theory of mind. For the latter, one needs signs of one or more of the 8 behaviours listed in Section 1, above. For that reason, in this chapter I remain more cautious in concluding that a theory of mind had in all likelihood evolved by 40,000 years ago - but that before this, there is as yet no clear evidence for it.

Mithen's conclusions are also based on his claim that "both monkeys and apes also engage in intentional communication" (p. 161). Here again one sees a potential confound. Clearly monkeys and apes vocalize or gesture intentionally, but this is not the same as "intending to communicate", as defined in Section 1 above. Monkeys and apes may be vocalizing or gesturing with the intention to alter the behaviour of the listener or audience, but there is no compelling evidence yet that they are vocalizing or gesturing

with the intention to alter the mental states of their listener or audience (Cheney and Seyfarth 1993).

Using a model of the mindreading system shown in Figure 5, there is better evidence for the ability to attribute goal states (ID, or the Intentionality Detector) being as old as Mithen suggests, in that chimpanzees can clearly recognize *goal* states (Premack & Woodruff, 1978). They are also acutely aware of gaze direction (EDD, or the Eye Direction Detector), suggesting they are monitoring when they might be the target of another's perception (Chance, 1967). Less clear cut is whether they show shared attention (SAM, or the Shared Attention Mechanism) (Povinelli & Eddy, 1996). This means that elements of mindreading may be as old as 6-35 million years, and evolution may have "tinkered with old parts" under selection pressure, to produce a theory of mind mechanism (ToMM) more recently.

insert Figure 5 here

Finally, we might consider that the presence of a Broca's area in the brain 200,000 years ago (as inferred from cranial evidence - (Mithen, 1996) implies a theory of mind may be at least this old, in that language without a theory of mind would be functionally very limited. But this is only indirect evidence for the existence of a theory of mind.

Conclusions

A theory of mind is a powerful means of making sense of the social world. It enables explanations and prediction of the behaviour of agents, and communication (beyond animal signalling). Given its centrality to what makes the human mind essentially human, its evolution needs investigation. Palae-oarchaelogical evidence shows it was in place at least 40,000 years ago, and comparative data from studies of existing primates shows that aspects of a theory of mind may be as old as 6 million years. Specifically, recognising *volitional* states and a sensitivity to *eye-direction* may be a skill we share with the apes, and therefore with our common ancestor 6 million years ago. In contrast, *shared attention* and recognising *epistemic* states may be unique to Homo Sapiens and may therefore have evolved more recently. In terms of the model of the "mindreading system" shown in Figure 5, ID and EDD may be phylogenetically older (at least 6 million years) than the more recent SAM and ToMM. This leads to the idea that a theory of mind did not necessarily evolve all at once, but by degrees.

Figure Legends

Figures 1-4 are reproduced from Mithen (1997) with kind permission.

Figure 1: Simplified model of primate evolution

Figure 2: Statuette from Hohlenstein-Stadel, 30-33,000 years old

Figure 3: Painting from Trois-Fieres, 30,000 years old

Figure 4: Skeleton from Sringir, 30,000 years old

Figure 5: The Mindreading System (Baron-Cohen, 1995)

Table 1: Evidence for theory of mind impairments in autism

First-order mind-reading tests

(i) the mental-physical distinction (Baron-Cohen, 1989a; Wellman & Estes, 1986)

(ii) The functions of the mind test (*ibid*).

(iii) The appearance-reality distinction (*ibid*); (Flavell, Green & Flavell, 1986)

(iv) First-order false belief tasks (Baron-Cohen, Leslie & Frith, 1985; Baron-Cohen, Leslie & Frith, 1986; Leekam & Perner, 1991; Perner, Frith, Leslie & Leekam, 1989; Reed & Peterson, 1990; Swettenham, Baron-Cohen, Gomez & Walsh, 1996; Wimmer & Perner, 1983).

(v) The "seeing leads to knowing" test (Baron-Cohen & Goodhart, 1994; Leslie & Frith, 1988; Pratt & Bryant, 1990)

(vi) Recognising mental state words test (Baron-Cohen et al., 1994).

(vii) simple causes of emotion (such as situations and desires) vs complex causes of emotion (such as beliefs) (Baron-Cohen, 1991; Baron-Cohen et al., 1993); (Harris, Johnson, Hutton, Andrews & Cooke, 1989)

(x) Recognizing the eye-region of the face as indicating when a person is thinking and what a person might want (Baron-Cohen, Campbell, Karmiloff-Smith, Grant & Walker, 1995; Baron-Cohen & Cross, 1992).

(xi) The accidental-intentional distinction (Phillips, 1993).

(xii) Deception (Baron-Cohen, 1992; Sodian & Frith, 1992; Yirmiya, Solomonica-Levi & Shulman, 1996), premised on understanding that people's beliefs can differ and therefore can be manipulated.

(xiii) Tests of understanding metaphor, sarcasm, and irony - these all being intentionally non-literal statements (Happe, 1993).

(ix) Pragmatics (Baron-Cohen, 1988b); see also (Tager-Flusberg, 1993), e.g. recognizing violations of pragmatic rules, such as the Gricean Maxims of conversational cooperation (Surian et al., 1996). Since many pragmatic rules involve tailoring one's speech to what the listener needs to know, or might be interested in, this can be seen as intrinsically linked to a theory of mind.

Second-order mind-reading tests

(i) second-order false belief tests (Baron-Cohen, 1989b); (Perner & Wimmer, 1985), that is, tests of understanding what one character thinks another character thinks. Such second-order reasoning is usually understood by normal children of 5-6 years of age (Sullivan, Zaitchik & Tager-Flusberg, 1994),

(ii) bluff and double bluff (Happe, 1994).

(iii) decoding complex mental states from the expression in the eye-region of the face (Baron-Cohen, Jolliffe, Mortimore & Robertson, in press; Baron-Cohen, Wheelwright & Jolliffe, in press).

References

Baldwin, D. (1991). Infants' contribution to the achievement of joint reference. Child Development, 62, 875-890.

Baron-Cohen, S. (1988a). Without a theory of mind one cannot participate in a conversation. Cognition, 29, 83-84.

Baron-Cohen, S. (1988b). Social and pragmatic deficits in autism: cognitive or affective? Journal of Autism and Developmental Disorders, 18, 379-402.

Baron-Cohen, S. (1989a). The autistic child's theory of mind: a case of specific developmental delay. Journal of Child Psychology and Psychiatry, 30, 285-298.

Baron-Cohen, S. (1989b). Perceptual role-taking and protodeclarative pointing in autism. British Journal of Developmental Psychology., 7, 113-127.

Baron-Cohen, S. (1991). Do people with autism understand what causes emotion? Child Development, 62, 385-395.

Baron-Cohen, S. (1992). Out of sight or out of mind: another look at deception in autism. Journal of Child Psychology and Psychiatry, 33, 1141-1155.

Baron-Cohen, S. (1993). Are children with autism acultural? Commentary on Tomasello et al. Behavioral and Brain Sciences, 16, 512-513.

Baron-Cohen, S. (1995). Mindblindness: an essay on autism and theory of mind.: MIT Press/Bradford Books.

Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., Grant, J., & Walker, J. (1995). Are children with autism blind to the mentalistic significance of the eyes? British Journal of Developmental Psychology, 13, 379-398.

Baron-Cohen, S., & Cross, P. (1992). Reading the eyes: evidence for the role of perception in the development of a theory of mind. Mind and Language, 6, 173-186.

Baron-Cohen, S., & Goodhart, F. (1994). The "seeing leads to knowing" deficit in autism: the Pratt and Bryant probe. British Journal of Developmental Psychology, 12, 397-402.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (in press). An even more advanced test of theory of mind: evidence from very high functioning adults with autism or Asperger Syndrome. Journal of Child Psychology and Psychiatry.

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

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and Intentional understanding of picture stories in autistic children. British Journal of Developmental Psychology, 4, 113-125.

Baron-Cohen, S., Ring, H., Moriarty, J., Shmitz, P., Costa, D., & Ell, P. (1994). Recognition of mental state terms: a clinical study of autism, and a functional neuroimaging study of normal adults. British Journal of Psychiatry, 165, 640-649.

Baron-Cohen, S., Spitz, A., & Cross, P. (1993). Can children with autism recognize surprise? Cognition and Emotion, 7, 507-516.

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (in press). Is there a "language of the eyes"? Evidence from normal adults and adults with autism/Asperger's Syndrome. Visual Cognition.

Boyer, P. (1990). Tradition as truth and communication: Cognitive description of traditional discourse. New York: Cambridge University Press.

Byrne, R., & Whiten, A. (1991). Computation and mindreading in primate tactical deception. In A. Whiten (Ed.), Natural theories of mind, . Oxford: Basil Blackwell.

Chance, M. (1967). The interpretation of some agonistic postures: the role of "cut-off"acts and postures. Symposium of the Zoological Society of London, 8, 71-89.

Charman, T., & Baron-Cohen, S. (1992). Understanding beliefs and drawings: a further test of the metarepresentation theory of autism. Journal of Child Psychology and Psychiatry, 33, 1105-1112.

Charman, T., & Baron-Cohen, S. (1993). Drawing development in autism: the intellectual to visual realism shift. British Journal of Developmental Psychology, 11, 171-185.

Charman, T., & Baron-Cohen, S. (1995). Understanding models, photos, and beliefs: a test of the modularity thesis of metarepresentation. Cognitive Development, 10, 287-298.

Cheney, D., & Seyfarth, R. (1990). How monkeys see the world: University of Chicago Press.

Craig, J., Baron-Cohen, S., & Scott, F. (submitted). An investigation of genuine creativity in autism. .

De Waal, F. (1989). Peacemaking among primates: Harvard University Press.

Dennett, D. (1987). The Intentional Stance. Cambridge, Mass: MIT Press/Bradford Books.

Flavell, J. H., Green, E. R., & Flavell, E. R. (1986). Development of knowledge about the appearance-reality distinction. Monographs of the. Society for Research in Child Development, 51.

Happe, F. (1993). Communicative competence and theory of mind in autism: A test of Relevance Theory. Cognition, 48, 101-119.

Happe, F. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. Journal of Autism and Development Disorders, 24, 129-154.

Harris, P., Johnson, C. N., Hutton, D., Andrews, G., & Cooke, T. (1989). Young children's theory of mind and emotion. Cognition and Emotion, 3, 379-400.

Karmiloff-Smith, A. (1990). Constraints on representational change: Evidence from children's drawing. Cognition, 34, 57-83.

Leekam, S., & Perner, J. (1991). Does the autistic child have a metarepresentational deficit? Cognition, 40, 203-218.

Leslie, A. M. (1987). Pretence and representation: the origins of "theory of mind". Psychological Review, 94, 412-426.

Leslie, A. M., & Frith, U. (1988). Autistic children's understanding of seeing, knowing, and believing. British Journal of Developmental Psychology, 6, 315-324.

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: evidence from autism. Cognition, 43, 225-251.

Mithen, S. (1996). The Prehistory of the Mind: Penguin.

Morton, J., Frith, U., & Leslie, A. (1991). The cognitive basis of a biological disorder: autism. Trends in Neurosciences, 14, 434-438.

Paul, R., & Cohen, D. J. (1985). Comprehension of indirect requests in adults with autistic disorders and mental retardation. Journal of Speech and Hearing Research, 28, 475-479.

Perner, J., Frith, U., Leslie, A. M., & Leekam, S. (1989). Exploration of the autistic child's theory of mind: knowledge, belief, and communication. Child Development, 60, 689-700.

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that..." Attribution of second-order beliefs by 5-10 year old children. Journal of Experimental Child Psychology, 39, 437-471.

Phillips, W. (1993). Understanding intention and desire by children with autism: University of London.

Povinelli, D., & Eddy, T. (1996). Chimpanzees: Joint visual attention. Psychological Science, 7, 129-135.

Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). Child Development, 61, 973-983.

Premack, D. (1988). `Does the chimpanzee have a theory of mind?' revisited. In R. Byrne & A. Whiten (Eds.), Machiavellian Intelligence: social expertise and the evolution of intellect in monkeys, apes, and humans, . Oxford: Oxford University Press .

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a `theory of mind'? Behaviour and Brain Sciences, 4, 515-526.

Reed, T., & Peterson, C. (1990). A comparative study of autistic subjects' performance at two levels of visual and cognitive perspective taking. Journal of Autism and Developmental Disorders, 20, 555-568.

Scott, F., & Baron-Cohen, S. (1996). Imagining real and unreal objects: an investigation of imagination in autism. Journal of cognitive Neuroscience, 8, 400-411.

Selfe, L. (1977). Nadia: A Case of Extraordinary Drawing Ability in an Autistic Child. London: Academic Press.

Sodian, B., & Frith, U. (1992). Deception and sabotage in autistic, retarded, and normal children. Journal of Child Psychology and Psychiatry, 33, 591-606.

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers Can Attribute Second-Order Beliefs. Developmental Psychology, 30, 395-402.

Surian, L., Baron-Cohen, S., & Van der Lely, H. (1996). Are children with autism deaf to Gricean Maxims? Cognitive Neuropsychiatry, 1, 55-72.

Swettenham, J., Baron-Cohen, S., Gomez, J.-C., & Walsh, S. (1996). What's inside a person's head? Conceiving of the mind as a camera helps children with autism develop an alternative theory of mind. Cognitive Neuropsychiatry, 1, 73-88.

Tager-Flusberg, H. (1992). Autistic children's talk about psychological states: deficits in the early acquisition of a theory of mind. Child Development, 63, 161-172.

Tager-Flusberg, H. (1993). What language reveals about the understanding of minds in children with autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), Understanding other minds: perspectives from autism, : Oxford University Press.

Wellman, H. (1990). Children's theories of mind. Bradford: MIT Press.

Wellman, H., & Estes, D. (1986). Early understanding of mental entities: a reexamination of childhood realism. Child Development, 57, 910-923.

Whiten, A. (1991). Natural theories of mind. Oxford: Basil Blackwell.

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.

Yirmiya, N., Solomonica-Levi, D., & Shulman, C. (1996). The ability to manipulate behaviour and to understand manupulation of beliefs: A comparison of individuals with autism, mental retardation, and normal development. Developmental Psychology, 32, 62-69.