The theory of mind hypothesis of autism: History and prospects of the idea

Spearman Medal Lecture,
Presented at the BPS London Conference,
December 1990.

Simon Baron-Cohen

Departments of Psychology and Child Psychiatry, Institute of Psychiatry, University of London.

In this paper I review some of the work revealing a cognitive deficit in autism, in the child's theory of mind. This deficit, which shows up reliably across a series of studies, would by itself wreak havoc with the child's social and communicative development, given its critical importance in these processes. Studying such cases of "mind-blindness" may help us not only to understand autism, but normal development too.

Requests for reprints should be addressed to:
Dr Simon Baron-Cohen
Psychology Department
Institute of Psychiatry
Denmark Hill
London SE5 8AF

The Psychologist: Bulletin of the British Psychological Society (1992), 5, 9-12.

AUTISM WAS first described in 1943 by Loe Kanner, a psychiatrist working in the John Hopkins Hospital in Baltimore. Early work by experimental psychologists focussed on the notion that autism was a developmental language disorder (ser Rutter, 1978, for a review). Two new insights led to a shift towards seeing autism as a disorder of cognition. First, studies revealed that language disorder alone was unable to account for the social abnormalities found in autism, since such social deficits were not commonly found in other language-impaired children (see Rutter, 1983). Secondly, a series of seminal studies by Beate Hermelin and Neil O'Connor (1970) revealed autism-specific cognitive deficits in the comprehension of meaning.

The cognitive approach to understanding autism seemed to fit with my own observations of these children. I had worked as a teacher of children with autism, and it seemed to me that the social and communicative abnormalities in autism could be conceptualised as the result of cognitive deficits, specifically in the child's understanding of the human world. Certainly, a motivational explanation seemed unlikely, as it was clear that these children were interested in people to some extent, (for example, they often asked questions about people's behaviour, appearance, and routines). But they appeared singularly oblivious to what others were *thinking*. For example, it did not seem to occur to them that others might think them odd, or funny, or that others might *think* anything at all. The other thing that struck me with much force was that their behaviour and speech seemed to be largely lacking in any *self-reflection* - they seemed to lack what we commonly think of as signs of self-consciousness. It seemed plausible that, unlike normal children, these children might be unable to understand themselves or others as *subjects*, with subjective states of mind, and instead might think of themselves and others purely as objects.

The "theory of mind" hypothesis

When I joined the MRC Cognitive Development Unit in 1982, I had the privilege of working with two outstanding researchers, Uta Frith and Alan Leslie. Together, we put forward the hypothesis that children with autism might be specifically impaired in the ability to represent mental states (such as beliefs, desires, intentions, etc), that is, in the development of a theory of mind. The phrase "*theory of mind*" came from Premack and Woodruff (1978) who defined it as follows:

> In saying that an individual has a theory of mind, we mean that the individual imputes mental states to himself and others ... A system of inferences of this kind is properly viewed as a theory, first because such states are not directly observable, and second, because the system can be used to make predictions, specifically about the behaviour of other organisms (p.515).

The development of a theory of mind in normal children seems to be a very early achievement, in progress during the first year of life (Baron-Cohen, 1989a,b, 1991a; Leslie & Happe, 1989) and increasing in complexity in early childhood (Astington, *et al.* 1988). The hypothesis that in autism there may be a specific impairment in the development of a theory of mind seemed particularly attractive because of the claim that a theory of mind is essential in order to understand and predict much of human behaviour (Dennett, 1978; Wellman, 1990).

Consider, for example, how to make sense of the following scenario: *A man comes out of a shop and walks off*

down the street. About half way down the street he sudddenly stops, turn around, runs back to the shop, and goes inside. (We instantly think to ourselves that the man must have **remembered** he left something in the shop, that he **wants** to retrieve it, and that he **believes** it will still be in the shop.) *The man then re-emerges from the shop, but this time he walks along slowly, scanning the ground.* (Now we make the assumption that whatever he **thought** was in the shop wasn't there, and that he now **believes** he may have dropped it on the pavement outside.) If we lacked the ability to refer to the man's beliefs, desires, etc, his actions would seem most peculiar.

So, a theory of mind gives one a ready device for understanding social behaviour. We might predict that if one lacked a theory of mind, if one were blind to the existence of mental states, the social world would seem chaotic, confusing, even frightening. At the worst this might lead one to withdraw from it completely, but at the very least it might lead to very odd attempts at interaction with people, treating them as lacking "minds", and therefore behaving towards them in a similar manner to the way one approaches inanimate objects. Since the behaviour of children with autism is often described in these terms (Kanner, 1943), it seemed worthwhile considering if there were abnormalities in their theory of mind.

But this hypothesis was attractive for a second reason. Speech Act Theorists such as Grice (1967/1975), Austin (1962), and Searle (1965) had argued that a theory of mind is also essential for normal communication, both verbal and non-verbal. Put simply, the argument is that all communication requires both participants to take into account the background knowledge and presuppositions of the other person in the dialogue, as well as their intentions in communicating. Such mental state attribution is necessary, it is argued, if a dialogue is to respect the conversational rules of *pragmatics* - if it is to be appropriate and relevant to the social context (Sperber & Wilson, 1986).

The relationship between a theory of mind and pragmatics is explained more fully elsewhere (Baron-Cohen, 1988). But, again, it leads to the prediction that if one was unable to appreciate other people's mental states, communication would go seriously awry. The idea that in autism there might be an impairment in the development of a theory of mind therefore seemed to be a parsimonious candidate hypothesis to explain two key symptoms, namely the social and communicative abnormalities.

Testing the "theory of mind" hypothesis

A simple but stringent experimental test of normal children's understanding of the mental state "belief" was developed by Heinz Wimmer and Josef Perner, in 1983. They selected belief as the mental state to test because this is arguably the clearest case of a mental state that is about something in the world (Dennett, 1978). That is, it is a mental state that possesses Intentionality (Searle, 1965, 1979). Their test was based on a puppet story in which a character holds a false, and therefore different, belief to that held by the child. Children are scored as passing this test if they demonstrate that they take into account the story character's different belief, and can predict the story character's action, given her false belief.

An adaptation of Wimmer and Perner's (1983) test is shown in Figure 1. We gave this test to children with autism, as well as to a group of children with mental handicap (all with the diagnosis of Down's Syndrome), and a group of normal children. The beauty of this paradigm is that although it requires a verbal mental age of

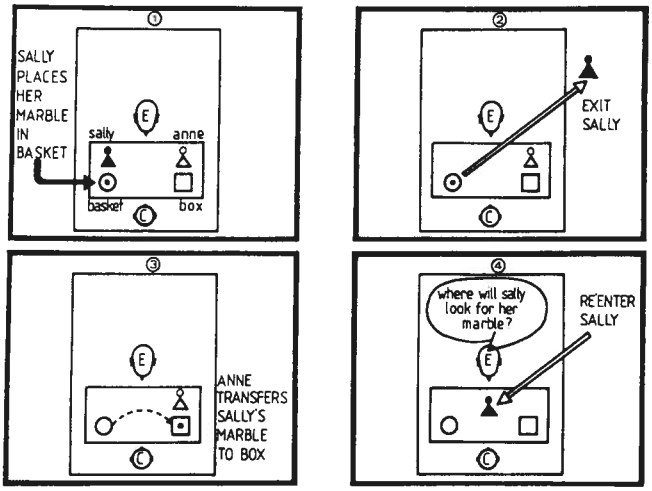

Figure 1: Scenario of a simple test of a theory of mind (reproduced from* Cognition, *21, 37-46)

about four years old to comprehend the narrative, the test does not require any expressive language abilities. The subject is only required to point to one location or the other, in response to key questions. Furthermore, control questions establish if the child can understand reality as distinct from belief. Indeed, this is why testing understanding of false belief is such a good test of children's concept of belief, in that in such cases belief does not match reality.

The critical event in the story is that Sally is not present when Anne moves the object from A to B. If asked where on her return Sally will look for the object, the correct answer is location A, since this is where she originally put it, and where she therefore believes it still is. We found that whilst 86 per cent of the Down's Syndrome subjects, and 85 per cent of the normal children passed this test, only 20 per cent of the subjects with autism did so, and this was so despite this group having a higher mental and chronological age than the two comparison groups (Baron-Cohen, *et al.* 1985). Instead, 80 per cent of the subjects with autism indicated that Sally would look for the object at location B, where the object really was.

This result lent preliminary support to the hypothesis that in autism there is a failure to develop a theory of mind. The data from the control groups further revealed that such a deficit must be autism-specific, rather than the result of general developmental delay. Finally, the data from control questions revealed that children with autism had no difficulty answering questions involving memory, or questions which did not involve mental state attribution. During the subsequent years this pattern of results has been replicated several times, using a number of different paradigms (eg: Leslie & Frith, 1988; Perner, *et al.* 1989; Baron-Cohen, *et al.* 1986; Baron-Cohen, 1989c, d, 1991b; Leekam & Perner, in press; Reed & Petersen, 1990). These replications suggests that this deficit is fairly robust.

How specific is the deficit?

The question arises as to the specificity of this deficit. First, is it specific to autism? Other childhood clinical populations that have so far been tested appear unimpaired on such tests, and these populations include children with the following diagnoses: Down's Syndrome, mental handicap of unknown aetiology, language-impairment, emotional disturbance, and deafness. (For references to these, see the studies mentioned in the previous paragraph, plus Siddens, *et al.* 1990, and Leslie & Sellers, 1990). Further clinical populations remain to be tested, but the deficit does seem

to be autism-specific.

Secondly, is the deficit specific to understanding the mental state of belief? Some of the studies mentioned earlier have tested understanding of other mental states such as *know, think, and pretend* by children with autism, and these too appear very difficult for them. (For a discussion of pretence and autism, see Baron-Cohen, 1987, and Harris, in press). This is also borne out in studies of the spontaneous speech of children with autism, which rarely contains any mental state terms that refer to thinking, knowing, pretending, or believing (Tager-Flusberg, 1989). This last study found that they do talk about simple emotions and desires, and more experimental approaches confirm that their understanding of these is in line with their mental age (Baron-Cohen, 1991d).

Naturally, before concluding that the cognitive deficit is specific to their theory of mind, it would be necessary to show that other aspects of *social cognition* are unimpaired in autism. Social cognition can be defined as that part of cognition which is used in everyday interaction with the social world. Studies which have examined this question show that children with autism are unimpaired in a range of social cognitive tasks. These include visual self-recognition, peer recognition, distinguishing self from other people, person-permanence (that is, understanding that people continue to exist even when you can no longer see them), distinguishing animate and inanimate objects, and perceiving social relationships. More importantly, children with autism are also unimpaired on tests of perceptual role-taking, that is, judging what another person can see. This array of findings is reviewed in more detail elsewhere (Baron-Cohen, 1988, 1989a, 1991c).

The deficit in their theory of mind thus contrasts with this growing list of unimpaired social cognitive skills, and in particular contrasts strongly with their good performance on tests of perceptual role-taking, suggesting that the impairment is in conceptual (rather than perceptual) role-taking. Finally, measures of communication and social functioning in autism appear to correlate with performance on tests of a theory of mind (Perner *et al.* 1989; Siddens *et al.* 1990), providing important evidence of the validity of these tests.

A case of specific developmental delay?

On all tests of theory of mind in autism, an identifiable subgroup of children consistently pass. These comprise about 30 per cent of children with autism with a mental age (verbal or non-verbal) above 4 years old. From this it becomes clear that a theory of mind cannot be impaired in a blanket sense in all children with autism. One hypothesis that has been advanced is that there may be a *delay* in the development of a theory of mind in autism, such that all children with autism are late in developing this ability, but that after a severe delay it may emerge in some (Baron-Cohen, 1989c).

Some evidence for this notion comes from a study of those children with autism who pass *first-order* belief attribution tests (that is, who are able to make attributions of the form **Sally thinks x**), but who are nevertheless impaired in more complex tests of a theory of mind, at the level of *second-order* belief attribution (Baron-Cohen, 1989c). (An attribution at this level takes the form **Mary thinks that John thinks x**). First-order belief attribution is easily within normal four-year-olds' understanding, whilst second-order belief attribution is within normal six-year-olds' understanding (Perner & Wimmer, 1985). Such studies therefore suggest that a minority of children with autism do reach the stage of a normal four-year-old in the development of a theory of mind, but are nevertheless delayed in reaching the stage even of a normal seven-year-old.

Possible origins of an impaired theory of mind

It cannot be the case that development is entirely normal in children with autism until the age at which children usually pass tests of false belief, namely at three or four years old. We can be sure of this for two reasons. First, social and communicative abnormalities in autism begin before 36 months of age (DSM-IIIR, 1987). Secondly, normally developing 1 and 2 year olds also fail tests of false belief, yet nevertheless show social and communicative abilities not seen in autism. It is therefore possible that in infancy there are *precursors* to the development of a theory of mind.

One likely precursor to a theory of mind is the ability to understand another person's *attentional state*, that is, the ability to appreciate what another person is attending to, or what they find of interest (Baron-Cohen, 1989b; 1991a). This develops very early, and is certainly present by 10-14 months of age, as manifested in young children's production and comprehension of "joint-attention" behaviours (Bruner, 1983). These include giving and showing objects, pointing to objects, and monitoring what other people are looking at. Giving and showing objects are, I take it, self-explanatory, but pointing merits a little elaboration. Pointing occurs when a person extends their index finger towards an object, and it seems to occur for at least two different functions. The first is in order to obtain an object (so-called "protoimperative pointing"), whilst the second is in order to comment or show interest in an object (so-called "protodeclarative pointing").

Interestingly, children with autism are impoverished in the amount of giving and showing behaviours they produce (Sigman, *et al.* 1986), and while they do use and understand protoimperative pointing, they almost never use or understand protodeclarative pointing (Baron-Cohen, 1989a). They also monitor other people's direction of gaze less than other children (Sigman *et al*, 1986). Since such gaze-monitoring starts as early as three months of age in normal infants (Scaife & Bruner, 1975), these deficits in understanding attention may hold important clues about the normal developmental origins of a theory of mind, and its impairment in autism. It also suggests that early diagnosis of autism might in part be possible on the basis of the absence of joint-attention behaviours. We are currently testing this possibility.

Future prospects of the "theory of mind" hypothesis

Some questions for future research include the following. First, is such a deficit amenable to any form of psychological intervention? It is clear that broad-based social skills training packages have little effect on deficits in the child's theory of mind (Howlin, 1989), but whether specific training which focuses precisely on this cognitive deficit will have any benefits is an open question, and one which Pat Howlin, Julie Hadwin and I are currently exploring. Secondly, where in the brain might a dysfunction exist, to disrupt the development of a theory of mind in autism? The new scanning techniques that allow imaging of the brain *during* cognitive tasks may provide an answer to this question. One possibility is that the brain system responsible for the production of joint-attentional behaviours in normal infants may be dysfunctional in autism. Thirdly,

what are the lessons for understanding normal development in this work? Many of us continue to study autism not only to understand the disorder, but also becuse we hold the belief that the study of psychopathology allows a "carving of cognition at its seams" (McCarthy & Warrington, 1990: 370). Pathology illustrates what is necessary and sufficient for normal development. It is my hope that the research I have described here may not only allow our basic understanding to inch forward, but may lead to useful clinical applications.

Acknowledgements

I would like to express deep gratitude to my colleagues at the MRC Cognitive Development Unit for their support of and contributions to the work I carried out whilst there. I have also benefitted from support from other departments: University College, London, and the Institute of Psychiatry. Portions of this paper first appeared in the *International Review of Psychiatry* (vol.2, 1990: 77-88), under the title *Autism: a specific cognitive disorder of "mind-blindness"*. This also formed the substance of the Boyd McCandless Award Lecture at the APA San Francisco Conference, August 1991. Correspondence should be addressed to the author at the Department of Psychology.

References

Astington, J., Harris, P. & Olson, D. (1988) *Developing theories of mind.* New York: Cambridge University Press.

Austin, J. (1962) *How to do Things with Words.* Oxford: Basil Blackwell.

Baron-Cohen, S. (1987) Autism and symbolic play. *British Journal of Developmental Psychology*, 5, 139-148.

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

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

Baron-Cohen, S. (1989b) Joint attention deficits in autism: towards a cognitive analysis. *Development and Psychopathology*, 1, 185-189.

Baron-Cohen, S. (1989c) 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. (1989d) Are autistic children behaviourists? An examination of their mental-physical and appearance-reality distinctions. *Journal of Autism and Developmental Disorders*, 19, 579-600.

Baron-Cohen, S. (1991a) Precursors to a theory of mind: Understanding attention in others. In A. Whiten (ed) *Natural Theories of Mind.* Oxford: Basil Blackwell.

Baron-Cohen, S. (1991b) The development of a theory of mind in autism: deviance and delay? *Psychiatric Clinics of North America*, 14, 33-51.

Baron-Cohen, S. (1991c) The theory of mind deficit in autism: how specific is it? *British Journal of Developmental Psychology*, 9, 301-314.

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

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.

Bruner, J. (1983) *Child's Talk: Learning to use Language.* Oxford: Oxford University Press.

Dennett, D. (1978) *Brainstorms: Philosophical Essays on Mind and Psychology.* USA, Harvester Press.

DSM-III-R (1987) *Diagnostic and Statistical Manual of Mental Disorders, revised 3rd Edition,* Washington DC: American Psychiatric Association.

Grice, H.P. (1975) Logic and conversation. In R. Cole & J. Morgan (eds) *Syntax and Semantics: Speech Acts.* New York: Academic Press. (Original work published in 1967).

Hermelin, B. & O'Connor, N. (1970) *Psychological Experiments with Autistic Children.* London: Pergamon Press.

Harris, P. (in press) Pretence and planning. In S. Baron-Cohen, H. Tager-Flusberg, & D.J. Cohen, (eds) *Understanding Other Minds: Perspectives from Autism.* Oxford University Press.

Howlin, P., (1989) Changing approaches to communication training with autistic children. *British Journal of Disorders of Communication*, 24, 151-168.

Kanner, L. (1943) Autistic disturbance of affective contact. Nervous Child. 2, 217-250. Reprinted in L. Kanner, (1973) *Childhood Psychosis: Initial Studies and New Insights.* New York: John Wiley and Sons.

Leekam, S. & Perner, J. (in press) Does the autistic child have a theory of representation? *Cognition*

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. & Happe, F. (1989) Autism and ostensive communication: the relevance of metarepresentation. *Development and Psychopathology*, 1, 205-212.

Leslie, A.M. & Sellars, C. (1990) The deaf child's theory of mind. Unpublished ms, MRC Cognitive Development Unit, 17 Gordon St, London, WC1.

McCarthy, R. & Warrington, E. (1990) *Cognitive Neuropsychology.* Academic Press.

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.

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

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

Rutter, M. (1978) Language disorder and infantile autism. In M. Rutter and E. Schopler (eds) *Autism: a Reappraisal of Concepts and Treatment.* New York: Plenum.

Rutter, M. (1983) Cognitive deficits in the pathogenesis of autism. *Journal of Child Psychology and Psychiatry*, 24, 513-531.

Scaife, M. & Bruner, J. (1975) The capacity for joint visual attention in the human infant. *Nature*, 253, 265.

Searle, J. (1965) What is a speech act? In M. Black (ed) *Philosophy in America* London: Allen and Unwin.

Searle, J. (1979) What is an intentional state? *Mind* 88, 74-92.

Siddens, F., Happe, F., Whyte, R. & Frith, U. (1990) Theory of mind in everyday life: an interview-based study with autistic, retarded, and disturbed children. Poster presented at the European Conference of Developmental Psychology, Stirling University, August 1990.

Sigman, M., Mundy, P., Ungerer, J. & Sherman, T., (1986) Social interactions of autistic, mentally retarded, and normal children and their caregivers. *Journal of Child Psychology and Psychiatry*, 27, 647-656.

Sperber, D. & Wilson, D. (1986) *Relevance: Communication and Cognition.* Oxford: Basil Blackwell.

Tager-Flusberg, H. (1989) An analysis of discourse ability and internal state lexicons in a longitudinal study of autistic children. Paper presented at the SRCD Conference, Kansas City, April 1989.

Wellman, H. (1990) *The Child's Theory of Mind.* Bradford Books/MIT Press.

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.