abilities are based do not presuppose a 'theory of mind' in the infant. It is more parsimonious to take the data as evidence for the functioning of perceptual systems on information held in common between the participants in the interaction. In that important and original sense of 'mindreading', the minds of others are transparent to the baby and revealed in interpersonal behaviour (Coulter, 1979).

16

Precursors to a Theory of Mind: Understanding Attention in Others

SIMON BARON-COHEN

The Forked Path

A central problem in developmental psychology has been to explain how, during the course of normal development, children come to understand that they and others have such a thing as a mind, and that states of mind guide human action. Developmental psychologists addressing this problem have tended to go down one of two paths. The direct route entails testing young normal children's comprehension of different mental states, in the hope of tracing the origins of this ability (see, for example, Wellman, in press, and chapter 2). The indirect but, I would argue, equally fruitful path entails studying why some children *fail* to develop a concept of mind. I and others have been taking this 'abnormal' route by studying autistic children (Baron-Cohen, 1985; Leslie, 1987a; Frith, 1989; Hobson, 1989a). Such children reveal what may be necessary but insufficient for the development of a concept of mind. This chapter describes a journey a little way down this second path. I begin by summarizing the relevant research in autism.

The Autistic Child's Theory of Mind

In the past five years a number of experiments have suggested that autistic children are unable to attribute mental states to other people (Baron-Cohen, Leslie and Frith, 1985, 1986; Dawson and Fernald, 1987; Baron-Cohen, 1989a; Leslie and Frith, 1988; Perner, Frith, Leslie and Leekam, 1989; Harris and Muncer, 1988; Swettenham, 1990; Russell, Sharpe and Mauthner, forthcoming; Leekam and Perner, 1990; Sodian and Frith, 1990; Mitchell, 1989; Shaw, 1989). One study also suggests they may have a

purely behavioural notion of the function of the brain, and may even be completely unaware of the distinction between mental and physical entities (Baron-Cohen, 1989c). Together, such experiments point to a severe deficit in autistic children's 'theory of mind', as Premack and Woodruff (1978) called the ability to attribute mental states to oneself and others. By itself this could account for the striking abnormalities these children show in their social interaction and communication (Baron-Cohen, 1988), since both social and communicative skills depend to a large extent on the ability to take into account other people's beliefs, desires, thoughts, intentions, etc. (Dennett, 1978; Grice, 1957).

How are we to account for this impairment in the development of a theory of mind in autism? And what does such an impairment tell us about the factors underlying the normal development of a theory of mind? There have been two recent attempts to explain autistic children's deficits in this area, one computational-cognitive (Leslie, 1987a, 1988a; Leslie and Frith, 1990; and see Leslie, chapter 5), and one social-affective (Hobson, 1989a, b, 1990).[1] I shall first take a closer look at these two explanations. Then I shall consider some new evidence of joint-attention deficits in autism, and explore the possibility that a critical *precursor* in the development of a theory of mind lies in the infant's understanding of attention in others.

Leslie's Metarepresentation Theory of Autism

Leslie (see chapter 5) postulates a distinction in the development of the normally functioning cognitive system, between *primary representation* and *metarepresentation*. Primary representations function as ways of storing literal information about events in the world and are, as Leslie (1987a) poetically expressed it, 'sober' (p. 414), in that they include information that the system needs for its survival in the real world, e.g. There is a tiger out there. Metarepresentations, in contrast, function to allow the system to construct descriptions of hypothetical events, such as descriptions of pretend objects, thoughts, dreams, etc. Rather than referring to the outside world, they instead refer to other representations. Hence their name.

Leslie argues that metarepresentations have the special general form **Agent-Informational Relation-***'expression'*, where an **Agent** is a person or persons, and an '**expression**' is any *decoupled* representation. (A representation is said to be decoupled, in Leslie's theory, when a primary representation is copied and the normal reference links to the outside world then suspended.) An **Informational Relation** is any Intentional state (i.e. any mental state that is *about* something), such as thinking, knowing, believing, intending, pretending, etc. So, to give an example, a primary representation might have the form *There is a tiger*, whilst a metarepresentation derived from this (i.e. a decoupled copy of it) might have the form *John believes 'there is a tiger'*.

Notice that the Informational Relation, by preceding the expression, renders the expression 'opaque', i.e. suspends its normal truth implications. That is, the sentence is still true even if there is no tiger, so long as John believes there is one. Intentional states are thus said to possess the logical property of *referential opacity*, a term derived from Quine (see chapter 1). The form of metarepresentation that Leslie proposes is, of course, identical to the analysis of intentional propositions given by Bertrand Russell, and the logical properties he outlines for metarepresentation are identical to those documented by Gottlob Frege. (For a summary of this philosophical background, see Olson, 1988.)

Metarepresentation is the mechanism Leslie sees as underlying the ability to represent mental states. In its mature form this mechanism allows for a theory of mind, and in its earliest, most primitive form, it allows for the emergence of pretend play since, according to Leslie, the logical properties of pretence and mental state attribution are similar. This has significance when Leslie applies his model to autism: he predicts that damage to the capacity for metarepresentation should in one tragic blow give rise both to an inability to employ a theory of mind and an inability to engage in pretend play. In addition, this model predicts a specific profile of both *intact* and damaged cognitive skills in autism, (intact ones being based on primary representation, damaged ones requiring metarepresentation).

Evaluation of Leslie's theory How far are these predictions supported? The evidence for autistic children's inability to employ a theory of mind has already been cited. And in line with the other main prediction from Leslie's theory, a number of studies have documented an absence or severe impoverishment in autistic children's pretend play (Riguet, Taylor, Benaroya and Klein, 1981; Ungerer and Sigman, 1981; Gould, 1986; Rutter, 1978; Wing, Gould, Yeats and Brierley, 1977; Baron-Cohen, 1987). In addition, an analysis of intact and damaged skills within autistic children's social cognition also seems to support Leslie's view. Thus, social cognitive skills which do not involve mental state attribution, such as visual self-recognition, face recognition, gender recognition, visual perspective-taking, and social schemata construction are *unimpaired* in autism (see Baron-Cohen, 1988, for a review), as are relationship recognition and the animate–inanimate distinction (Baron-Cohen, forthcoming a).

A critical test of whether autistic children have a specific deficit in representing mental representations would turn on how they perform on tests that require representation of non-mental representations. So far, the

only such tests that have been tried with autistic children are a false photograph test, and a false drawing test, both derived from Zaitchik (1990). Results from these studies show that autistic children have no difficulty with such tests (Leekam and Perner, 1990; Charman and Baron-Cohen, 1990) supporting Leslie's model. The only alternative psychological theory of the deficit in autistic children's theory of mind has been proposed by Hobson, to which we now turn.

Hobson's Social-Affective Theory of Autism

Hobson's (1989a, 1990) theory assumes that, from birth, normal infants are involved in 'reciprocal personal relations with others' (Hobson, 1990, p. 116), a claim for which there is considerable evidence (see Schaffer, 1984, for a review). Hobson argues that one source of such reciprocity is the infant's ability to perceive its caregiver's expressions of emotions. He further argues that infants do not *infer* emotions from bodily action, but rather that they directly perceive them. The ability to perceive emotions, Hobson maintains, leads to the phenomenon of 'social referencing' (Campos and Stenberg, 1980), that is, the infant monitoring the caregiver's facial expression to disambiguate novel events.

Hobson goes on to argue that both the production of gestural requests and social referencing lead to the ability to construe the world from other people's positions, and to construe the world as if it were other than it really is. He holds this latter ability responsible for the ability to pretend play. For Hobson then, the autistic child's deficits in understanding other people's mental states and in generating pretend play are consequences of an inability to understand and respond to emotions in others, a deficit postulated to be present in autistic infants.

Evaluation of Hobson's theory A number of controversies, both empirical and theoretical, surround Hobson's model of normal development. First, it is not clear why normal infants' understanding of emotions would by itself enable them to attribute non-emotional, opaque mental states (such as beliefs, thoughts, intentions, etc.). Hobson (1990) is vague on this point: he says that infants' understanding of emotions 'crystallize' (p. 117), allowing them to assume alternative standpoints, but the term 'crystallize' leaves it unclear as to exactly how this operates. Secondly, by what mechanism an understanding of alternative viewpoints necessarily produces the development of a symbolic capacity (as Hobson maintains) is also left unspecified: he writes that infants 'come to grasp' (Hobson, 1990, p. 117) how symbols function, but such phrases beg many questions.

From empirical studies of autistic children there are other reasons for taking issue with Hobson's theory. Hobson's key prediction for autism is that their understanding of emotions should be markedly impaired, and their affective 'personal relatedness' (Hobson, 1989a, p. 22) should also be severely disturbed. However, the idea that there is a global affective impairment in autism is not supported by the data, since autistic children do show some signs of attachment (Sigman and Ungerer, 1984; Hertzig, Snow, and Sherman, 1989; Shapiro, Sherman, Calamari and Koch, 1987; Dissanayake and Crossley, 1989; Sigman, Mundy, Ungerer and Sherman, 1987). Secondly, autistic children can also understand simple emotions, such as happy and sad, as outcomes of situations and desires (Baron-Cohen, forthcoming d; Harris, Coles and Tan – cited in Harris, chapter 19). Thirdly, autistic children do also use some simple emotion terms in their language in an appropriate way (Tager-Flüsberg, 1989). Fourthly, although Hobson's experimental data show that autistic children have difficulty in emotional expression recognition tests (Hobson, 1986a, b), recent studies suggest such deficits may be a function of verbal mental age (Ozonoff, Pennington and Rogers, 1990; Braverman, Fein, Lucci and Waterhouse, 1989; Hobson, Ouston and Lee, 1988a, b).

Nevertheless, an affective disturbance may exist in autism. For example, a number of studies agree that autistic children have some difficulty in *expressing* emotions facially (Snow, Hertzig and Shapiro, 1987; Yirmiya, Kasari, Sigman and Mundy, 1989; Dawson, Hill, Spencer and Galpert, 1988; Kasari et al., forthcoming), although at present it is unclear if this is simply a non-verbal sign of autistic children's inability to *communicate* with others, or if it represents a purely affective disturbance.

Some Philosophical Issues

Before leaving Hobson's and Leslie's accounts, it is worth examining some of the philosophical issues underlying them. Hobson proposes that perceptual experience plays an important role in the development of knowledge of other minds. He does this on the basis of philosophical objections (by, for example, Wittgenstein) to the view that such knowledge is simply *inferred* from one's own experience: neither observation of other people's behaviour nor experience of one's own mental states could by themselves lead one to the inference that mental states underlie behaviour in others. Hobson's alternative is to assume that certain basic mental states must therefore be *directly* perceivable in behaviour and in facial configuration (cf. chapter 17).

Leslie's theory, outlined earlier, provides an alternative solution to the philosophical 'problem of other minds'. He agrees with Hobson that mental

states cannot be inferred from behaviour, but disagrees with him that mental states are therefore directly observable. Instead, Leslie suggests that human beings are hard-wired (biologically pre-programmed) to develop a capacity for metarepresentation towards the end of the first year of life, a capacity which by itself allows human beings to represent that they and other systems can represent (think about) things. Such a position is similar to Fodor's (1987, pp. 132–3) nativist theory of mind hypothesis:

> Here's what I would have done if I had been faced with this problem in designing *Homo sapiens*. I would have made a knowledge of commonsense *Homo sapiens* psychology *innate*; that way nobody would have to spend time learning it . . . The empirical evidence that God did it the way I would have isn't, in fact, unimpressive . . . Suffice it that (1) Acceptance of some form of intentional explanation appears to be a cultural universal. There is, so far as I know, no human group that doesn't explain behaviour by imputing beliefs and desires to the behaviour. (And if an anthropologist claimed to have found such a group, I wouldn't believe him.) (2) At least in our culture, much of the apparatus of mentalistic explanation is apparently operative quite early.[2]

One critical philosophical difference, then, between Leslie (and Fodor) on the one hand, and Hobson on the other, is in the role they attribute to experience in the development of a normal concept of mind. Hobson sees perceptual experience as a necessary factor, whilst Leslie (and Fodor) do not. This leads Hobson (1990) to predict that *congenitally blind* children, who obviously lack perceptual experience of their caregiver's facial expressions, should also be severely impaired in the development of a normal theory of mind. In contrast, Leslie's theory suggests that the development of a metarepresentational capacity will be largely, if not completely, independent of perceptual experience. Indeed, this is why Leslie calls a theory of mind a 'module' (see Leslie, chapter 5).

From this, it follows that Leslie's theory predicts blind children should develop a theory of mind at the same mental age as other non-autistic children. (Blindness may well slow down 'mental age' development across the board, but according to Leslie's theory there is no reason to predict that their theory of mind will be *specifically* delayed.) Testing blind children on an appropriate version of the Sally-Anne marble test (Baron-Cohen et al., 1985) may be a critical next step in order to evaluate Leslie's and Hobson's theories further.[3]

This brief excursion into two major psychological theories of autism has, I hope, illustrated some of the value of exploring the 'abnormal' route as a way to understand the normal development of a theory of mind. In the next part of this chapter I move on to consider some recent evidence of *joint-attention* deficits in autism, and to examine their implications.

Joint-Attention Behaviours in Autism

From observational studies, it appears that joint-attention behaviours occur far less frequently in autistic children than in non-autistic control groups (Loveland and Landry, 1986; Landry and Loveland, 1989; Mundy, Sigman, Ungerer and Sherman, 1986; Sigman, Mundy, Ungerer and Sherman, 1986). These behaviours include 'referential looking' (as occurs when I look at what you are looking at, and attempt to get you to look at something by using the direction of my eye gaze), and gestures such as giving, showing and pointing. In normal children, referential looking is present in the majority of eight-month-olds (Scaife and Bruner, 1975), and gestures such as giving, showing and pointing emerge between nine to 12 months old (Bruner, 1983; Stern, 1985). Joint-attention deficits are likely to be the earliest *social* deficits in autism yet identified.

Do Joint-Attention Behaviours Require Metarepresentation?

Mundy and Sigman (1989) consider that the data indicating joint-attention deficits in autism refute the metarepresentation theory of autism. They argue this on the grounds that metarepresentation only emerges at 12 to 18 months (with pretend play – see Leslie, 1987a), whilst joint-attention behaviours emerge in normal development *prior* to 12 months. However, as several commentators have pointed out (Harris, 1989; Baron-Cohen, 1989b) this argument is rather tenuous, since pretend play can also be present in normal development from as early as nine months of age (Bates et al., 1979). In other words, the two abilities may not be as distant in their chronology of emergence as Mundy and Sigman suggest.

Hobson (1989b) suggests that joint-attention deficits indicate an incapacity to share or participate in other people's experiences, and suggests this incapacity stems from autistic children being unable to understand patterns of 'bodily expressiveness' (p. 199) as essentially linked with the observed person's feelings. His emphasis on joint-attention behaviours as entailing sharing experiences seems plausible. Hobson (1989b) goes on to suggest that such behaviours are likely to involve metarepresentation (p. 201) since the child is 'psychologically engaged with someone else's psychological engagement with the world and with oneself' (p. 201).

Leslie and Happé (1989) also consider joint-attention behaviours require metarepresentation. Their reasons are, however, different to Hobson's. They argue that joint-attention behaviours such as pointing and showing are examples of 'ostensive communication', and that such communication

requires a theory of mind, if a Gricean analysis of 'speech acts' is accepted (Grice, 1957). And because a theory of mind requires metarepresentation, it follows that joint-attention behaviours also require this. This argument mirrors that put forward for communicative deficits in autism (Baron-Cohen, 1988).

An Experimental Investigation of Pointing

In order to tease out the role of a theory of mind in autistic children's joint-attention deficits, I carried out an experiment (Baron-Cohen, 1989b). I focused on just one of these behaviours (pointing) in more detail, testing comprehension and production of two different functional uses of pointing. The first type was called *proto-imperative pointing*, and this was defined as pointing 'in order to use another person to **obtain** an object' (ibid., p. 117). The second was called *proto-declarative pointing*, and this was defined as pointing 'in order to **comment** or **remark** on the world to another person' (ibid., p. 118).

I reasoned that proto-imperative pointing need not require taking into account the other person's mental states, in that it might merely comprise acting to cause another person to get an object or give an object – physical, not mental, interactions. Proto-declarative pointing, in contrast, is likely to involve taking into account the other person's mental states, in that it appears to comprise acting to cause another person to 'take notice of' and 'comment on' (rather than obtain) an object. Commenting, arguably, entails mental – not physical – interaction. I predicted therefore that if autistic children's joint-attention deficits were due to their inability to attribute mental states to others, then proto-imperative pointing would be unimpaired in autism, whilst protodeclarative pointing would be impaired.

Boxes 16.1 and 16.2 show the scoring categories used in this study, and tables 16.1 and 16.2 show the results. The results of this study supported the predictions: both in production and comprehension of proto-imperative pointing autistic children were unimpaired, whilst in production and comprehension of proto-declarative pointing autistic children showed severe difficulties. Such difficulties were not seen in non-autistic normal and developmentally delayed control children of equivalent mental age.[4]

If, as we assumed earlier, proto-declarative pointing requires a theory of mind, then its impairment in this study is consistent with Leslie's metarepresentation theory. What about the other assumption made earlier, that proto-imperative pointing does not require a theory of mind? It has been suggested that proto-imperative pointing can be analysed purely in terms of physical–causal understanding (Harding and Golinkoff, 1979), and

Box 16.1 Scoring comprehension of pointing

1 *Understanding proto-imperatives*. The act of labelling a pointing gesture as a protoimperative depends on the viewer interpreting it as such. We therefore recorded how the subject interpreted a pointing gesture. The experimenter said 'I am going to use my finger to say something. What am I saying?' He then faced the subject and pointed to one of four toys in turn, positioned in a semicircle close to the subject but at some distance from the experimenter. The experimenter then waited to see how the subject responded.

Scoring. Four response categories (two pass and two fail) were rated:

a If the child picked up the object and handed it to the experimenter, this was rated as a pass, the subject interpreting the gesture as a protoimperative (request for object).
b If the subject said 'you're saying "Give me that toy"', or some such phrase, this was rated as a pass, the subject articulating the meaning of the gesture as a protoimperative.
c If the subject did nothing, or made some inappropriate response (e.g. touching the index finger or imitating the gesture), this was rated as a fail.
d If the subject named or commented on the toy, this was rated as the child interpreting the gesture as a protodeclarative, in which case the experimenter repeated the gesture insistently, in order to emphasize that a different response was expected. If the child did not then produce a response described in (a) or (b) above, this was scored as a fail, although understanding protodeclarative pointing was noted.

2 *Understanding proto-declaratives*. The experimenter said 'Now I am going to use my finger to say something else. What am I saying?' He then walked over to the window, looked up to sky and pointed, and then looked across to the subject, whilst still pointing up at the sky. The experimenter then waited to see how the subject would respond. He then repeated this with three locations: through the open door, into his briefcase and into his jacket pocket. Each of these locations was selected because the experimenter could see an object which the subject could not see (e.g. a plane or bird in the sky, a person in the corridor or toy in the briefcase or pocket).

Scoring. Five response categories (three pass and two fail) were rated:

a If the child looked in the direction of his point (often involving the child moving nearer the experimenter), this was rated as a pass, interpreting the gesture as a protodeclarative.
b If the child said 'What is it?' or 'What are you looking at?' or named the

object, etc. (i.e. expressed interest in the gesture as a proto-declarative.)

c If the child said 'You want me to look at something you can see', or an equivalent phrase (e.g. 'You're saying "There's a plane in the sky"'), this was scored as a pass, the subject articulating the meaning of the gesture as a proto-declarative.

d If, however, the subject attempted to pick something up and give it to the experimenter, this was rated as a fail, the subject (inappropriately) interpreting the gesture as a proto-imperative. Understanding proto-imperative pointing was however noted.

e Finally, if the subject made no response, or an inappropriate response (such as touching the index finger, etc.), this was rated as a fail.

Box 16.2 Scoring production of pointing

Scoring. The videotapes were rated by two independent judges, using the following criteria:

a Pointing was scored as proto-imperative if it was rated as unmistakably part of an attempt to obtain an object (e.g. (i) Sam points at toy held by Anna, and repeats the point until Anna hands the toy over; or (ii) Sam points to a toy held by Anna, and then abandons the gesture through Anna's lack of response, and finally takes the toy himself).

b Pointing was scored as proto-declarative if it was rated as unmistakably part of an attempt to comment on an object or event, that is, to draw another person's attention to the object or event, as an end in itself (e.g. (i) Sam points at a boy who is crying, then looks over to an adult, and continues pointing until the adult turns around and looks; or (ii) Sam points at an object or event which could not be an object of possession, such as an airplane out of the window, or a car across the road, and then looks over to Anna, and waits until she also looks at the same event).

Pointing rated as ambiguous by both judges was ignored.

as an 'instrumental' gesture (Zinober and Martlew, 1985). Since physical–causal understanding itself only requires primary representation of objects in a particular temporal and spatial relationship to each other (Leslie, 1988b), and indeed since physical causality is well understood in autism (Baron-Cohen, Leslie and Frith, 1986; Curcio, 1978), intact functioning of proto-imperative pointing in autism is also in line with predictions from the metarepresentation theory.

Is the Gricean account that Leslie and Happé (1989) proposed sufficient for all joint-attention behaviours? (Reminder of their account: joint-attention

Table 16.1 Average number of children (and percentages in each group) passing comprehension test of proto-imperative and proto-declarative pointing

Groups	*n*	Proto-imp		Proto-dec		Dissociation	
		Pass	%	Pass	%	Rate[a]	%
Autistic	20	14	70.0	2[b]	10.0	12[c]	60
Down's syndrome	14	11	78.5	12	85.7	1	7.2
Normal	27	25.25	93.5	26	96.2	0.75	2.3

[a] Dissociation rate = average number of children passing one and failing other.
[b] $P < 0.001$.
[c] $P < 0.05$.

Table 16.2 Number of children (and percentages in each group) producing proto-imperative and proto-declarative pointing[a]

Groups	*n*	Proto-imp		Proto-dec		Dissociation	
		Pass	%	Pass	%	Rate[b]	%
Autistic	10	4	40	0[c]	0	4	40
Mental handicap	10	8	80	7	70	1	10
Normal	10	7	70	9	90	2	20

[a] First judge only.
[b] Dissociation rate = the number of children passing one and failing the other.
[c] $P < 0.005$.

behaviours are communicative, communication requires a theory of mind, a theory of mind requires metarepresentation, therefore joint-attention behaviours require metarepresentation.) I think a Gricean account may be valid for proto-declarative pointing, and perhaps for showing, because these are, arguably, communicative, but there are some joint-attention behaviours which may not be forms of ostensive communication (e.g. referential looking and, arguably, giving). I have found it useful instead to develop a cognitive account of joint-attention behaviours, be they forms of ostensive communication or not (Baron-Cohen, 1989d). This account is sketched below.

Understanding Attention in Others

I begin with an analysis of proto-declarative pointing, although I hope to show that the level of analysis adopted is sufficiently general to be relevant for *all* forms of joint-attention behaviours. This analysis is framed in terms of the infant developing a concept of *attention*. Using this concept, what the nine- to 12-month-old infant understands by someone else's proto-declarative point is that the person is focusing his or her attention on a specific object or event. However, understanding attention cannot simply entail understanding that another person can see, since a number of experiments show that the latter is within autistic children's comprehension (Hobson, 1984; Leslie and Frith, 1988; Baron-Cohen, 1985, 1989b). An example of a test of what another person can see is illustrated in figure 16.1. The experimenter simply directs his or her eyes at one object, chosen at random, and asks the child 'What am I looking at?'.

Primary representation is sufficient for the infant to represent a person, and to represent if that person is seeing or not seeing (i.e. has his/her eyes open and unobstructed, etc.). Primary representation and the application of rules of 'geometry' and 'line-of-sight' (Lempers, Flavell and Flavell, 1977) are therefore likely to be sufficient to pass a test like that shown in figure 16.1. In contrast, understanding attention, I argue, must require understanding that vision (or audition) can be directed selectively, and that its direction depends on the person finding the object or event *of interest*. But is primary representation sufficient in order for the infant to represent a person attending, that is, to represent a person as seeing something of interest or not of interest?

The simplest computational way of achieving this that I can imagine is by the infant representing a person's representation of the object as marked by a simple **+ive or −ive valence tag**. This mechanism could be elaborated with a larger number of such valence tags (e.g. for representing an object as dangerous versus safe, etc.), but even the simple version of this mechanism suggests that metarepresentation may be required for joint-attention behaviours. On this model, then, autistic children's joint-attention deficits would be consistent with Leslie's metarepresentation theory.

Applied to pointing, this theory assumes that the autistic child develops the 'index-finger outstretched' hand configuration that is hard-wired in our species alone (see Butterworth, chapter 15), but then uses it functionally only for proto-imperatives because a physical–causal concept is all that is available to the child. Proto-declarative pointing does not develop, according to this model, because the child lacks the ability to represent that another

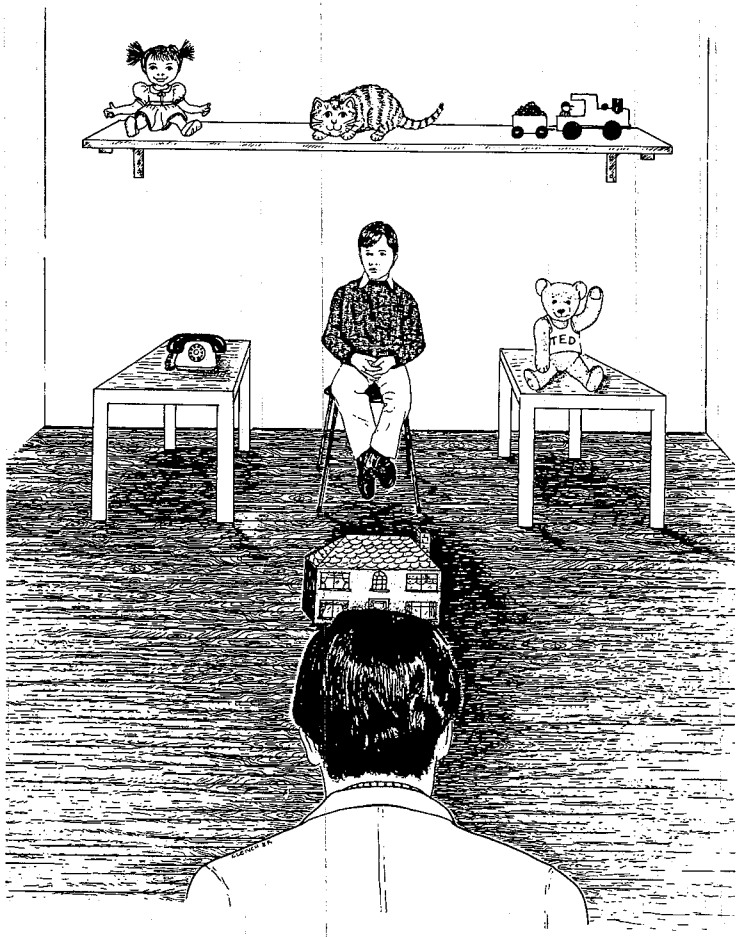


Figure 16.1 *Experimental test of perceptual role-taking (from Baron-Cohen, 1989b).*

person might represent an object as having a different valence (interesting versus uninteresting).

I would like to suggest that this analysis in terms of understanding attention in others applies not only to comprehension of proto-declarative pointing but also to its production, and to comprehension and production of the range of joint-attention behaviours mentioned earlier (referential looking, showing, giving, offering), all of which appear absent or severely impoverished in autism. For example, in production of all of these behaviours, normal children often look back at the other person's face and continue to point, show, give, etc., until the adult 'takes notice' of the target of the child's action. The child does not simply point, show, etc., but monitors the effect of this on the other person's direction of vision. This suggests that in producing the behaviour the child is trying to influence the other person's attention to an object (positively or negatively). Similarly, comprehension of any of these joint-attention behaviours is likely to entail understanding that another person represents an object as interesting or uninteresting, etc.

The claim that understanding attention as a mental state in others occurs during the period from seven to nine months and constitutes the origins of a theory of mind[5] is consistent with the analysis developed by Stern (1985), and corresponds to the age at which Trevarthen and Hubley (1978) have argued '*secondary* intersubjectivity' develops. Stern (1985, p. 134) has argued that Trevarthen and Hubley's claim of *primary* intersubjectivity being present at three to four months is incorrect, and instead suggests Trevarthen should replace his term 'secondary intersubjectivity' simply with 'intersubjectivity', since seven to nine months is the earliest point at which intersubjectivity or a theory of mind can be convincingly demonstrated.[6]

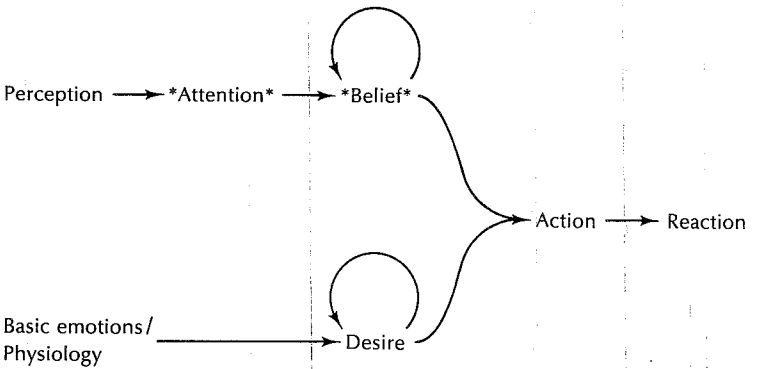
My formulation of proto-declarative pointing deficits in autism in terms of an inability to represent attention in others may also help to make sense of the dozens of reports in the clinical literature of autistic children taking an adult by the wrist and placing the adult's hand on a doorknob, etc., rather than using eye-gaze strategies to communicate the same intent, strategies which Gómez (chapter 13) argues rely on understanding shared attention, and which seem to be beyond many non-human primates.[7] Before placing this in the context of a broader theory of autism I am developing, I shall briefly consider the relevance of this to the child's developing theory of other mental states.

Understanding Attention and Other Mental States

Wellman (chapter 2) and Wellman and Bartsch (1988) have provided data which suggest that in the normal development of a theory of mind, understanding of **desire** precedes understanding of **belief**. On this basis Wellman suggests that two-year-olds can be construed as *desire psychologists* only, whilst three-year-olds can be construed as *belief–desire psychologists*. In addition, Wellman and Bartsch suggest that at both ages children's understanding of these mental states causally link with other internal states, such as emotions, and physiological drives (e.g. hunger). Wellman and Bartsch's model is shown in figure 2.1, p. 22.

As can be seen from that figure, Wellman and Bartsch posit that as part of their theory of mind young children know that perceptions can directly cause beliefs. The earlier discussion of the data from autism suggested that at a certain stage (estimated to be prior to seven months of age) children might understand perception without understanding attention. We therefore propose the following revision to Wellman and Bartsch's model: in their theory of mind, young children must first understand that perception can be directed selectively as attention, and then that attention can then cause beliefs. This refinement of Wellman and Bartsch's model is shown in figure 16.2.

Notice we postulate that understanding attention is specifically located in the development of understanding beliefs. This last claim is made on the



* — * internal states that are not understood in autism

Figure 16.2 *Proposed revision to Wellman and Bartsch's model (see figure 2.1 for the original model).*

basis of data from autism suggesting that autistic children are not pure behaviourists, but rather are more like desire psychologists. Thus, as cited earlier, Harris, Coles and Tan (1989) and Harris and Muncer (1988) have reported that autistic children can attribute desires to other people, given a minimum mental age, and this has been replicated by Baron-Cohen (forthcoming c). Finally, in autistic children's protocols describing picture stories of human action, of all the possible mental state terms they could have produced, only desire terms were used (Baron-Cohen, Leslie and Frith, 1986). Longitudinal language data analysed for spontaneous production of internal state terms in autism confirms this picture (Tager-Flüsberg, 1989).

Such data suggest that joint-attention deficits do not affect understanding of simple desires (or the emotions these cause), but may well affect understanding of even simple beliefs. This may be because beliefs are always opaque, whilst desires and emotions need not be: beliefs therefore require metarepresentation, whilst desires and emotions need not always (see Wellman, chapter 2). Such an analysis raises the idea that understanding attention may be critical not only as a precursor to understanding beliefs, but to all mental states that require metarepresentation. In the final section of this chapter, I will briefly consider the notion of a deficit in the autistic child's understanding of attention within the context of a broader theory of autism.

The Specific Developmental Delay Theory of Autism

Our research in autism started from the discovery that the majority of autistic children were impaired in their ability to pass tests of *first-order* belief attribution (of the form **Sally thinks the marble is in the basket** – Baron-Cohen et al., 1985, 1986). I later found that the minority of autistic children who were unimpaired at this level were nevertheless impaired at the next level up, in their ability to make *second-order* belief attributions (of the form **John thinks that Mary thinks the icecream van is in the park** – Baron-Cohen, 1989a). This raised the hypothesis that autism may be a case of *specific developmental delay* in the acquisition of a theory of mind, with different autistic children delayed at different points in this developmental sequence. The more recent research suggests there may be yet *earlier* delays in arriving at the stage of joint-attention behaviours, and we hypothesize that this may be a necessary first step in the normal construction of a theory of mind.

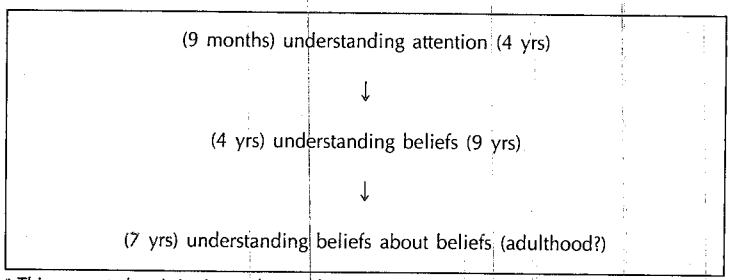
Some predictions follow from this specific developmental delay theory. First, autistic children who fail tests of proto-declarative pointing will also fail tests of first-order belief attribution. Secondly, those autistic children who do pass tests of proto-declarative pointing may succeed on tests of first-order belief attribution, but even these are likely to fail to understand second-order beliefs. This specific developmental delay model is shown in box 16.3.

The general assumption in the specific developmental delay theory is that autistic children's physical–causal knowledge is mental age appropriate and the only delayed aspect of their development that is specific to autism is in their theory of mind. Currently, we could sketch in some age/stages in the theory as follows: in terms of their theory of mind, most autistic children of less than four-years-old (in producing no joint-attention behaviours) may be similar to normal six-month-olds. After this, some autistic children may develop proto-declarative pointing and pretend play, although its late emergence means it remains impoverished. And years later still, some autistic children may progress beyond this level and succeed at understanding beliefs, although they remain very odd socially as a result of such a severe specific delay. We are currently collecting longitudinal data to test these notions further.[8]

Conclusions

In this chapter I have begun to sketch a model of the *emergence* of a theory of mind which attempts to account for the data from both normal and

Box 16.3 Developmental stages in the understanding of beliefs (normal timescale: on left; autistic[a] timescale; on right)



[a] This suggested autistic timescale may be a generous one: nine years is the earliest age at which we have found autistic children who can make first-order belief attributions, but for many autistic children the delay may be more severe than this. For yet others, there may be no development in their theory of mind at all (Humphries and Baron-Cohen, 1990).

autistic children. This model is in one sense a radical one, in that it places the onset of a capacity for metarepresentation in human development earlier than has hitherto been argued. On the other hand, it is consistent with analyses which do their utmost to avoid unnecessarily rich attributions in infancy (Stern, 1985).

The central distinction in my theory between understanding seeing versus understanding attending hinges on the latter requiring representation of the object of something perceived as 'interesting' (etc.) to the viewer. It is noteworthy that, in Bates et al.'s (1979) view, 'protocommunications' are all about sharing *interest* in objects. In this sense, appreciating that objects may or may not be of interest to others may be the drive behind all communication, and may help to explain why, even in those autistic children whose syntax, lexicon, phonology and even semantics are all intact, spontaneous communication in terms of two-way sharing of interests is seldom seen (Baron-Cohen, 1988).

In closing, we should keep in the mind that caregivers of autistic children are no different from caregivers of other children in the social input they provide (Cox, Rutter, Newman and Bartak, 1975), and yet autistic children are severely delayed in the development of a theory of mind. Such children therefore allow us to begin to answer the question 'Which cognitive structures and processes must be present for an infant to be able to take advantage of the "tutorial environment" (Butterworth and Grover, 1988) that their caregiver provides as part of social experience?' We know that all caregivers, the world across, spend many enjoyable hours during the infant's first year of life playing the 'peekaboo' game (Bruner, 1983). We can surmise that, apart from its possible role in developing turn-taking skills (as Bruner has discussed), this game functions to focus the infant on the seeing–attending distinction. Quite whether children who lack the concept of attentional states, such as our model postulates autistic children do, could enjoy such games, becomes an important question for future research.

ACKNOWLEDGEMENTS

I am grateful to Alan Leslie, Peter Hobson, Vasu Reddy and Sue Leekam for comments on an earlier draft of this chapter. Parts of this work were presented in talks at the MRC Cognitive Development Unit, London (May 1989), at the Oxford Workshop on Children's Theories of Mind (Chair: Paul Harris, June 1989), and in the Symposium on Pointing at the BPS Developmental Psychology Conference, University of Surrey (Chair: George Butterworth, September 1989). Figure 16.1 was drawn by Cathy Clench, St Bartholomew's Hospital, Medical Illustration Department, London.

NOTES

1 A summary of these two theories can also be found in Baron-Cohen (1988).

2 Fodor obviously didn't know about autistic children when he wrote this. Such children may be the exception to his statement about there being 'no human group that doesn't explain behaviour by imputing beliefs and desires to the behaviour'. Note that it was from anthropology that he expected any challenge to this statement to come, but as it has turned out, the challenge has come from abnormal psychology.

3 Such an experiment is currently being tried by Maggie Minter and Peter Hobson of the Institute of Psychiatry, London (personal communication).

4 Butterworth (chapter 15) cites the aphorism of 'When the finger points at the moon, the fool looks at the finger' to highlight the view that comprehension of proto-declarative pointing is linked to cognitive development. The results from the study reported here (Baron-Cohen, 1989b) can clarify this further: although a minimum MA of ten to 14 months may be necessary in order to understand proto-declarative pointing, low IQ by itself is *not* a cause of failure to comprehend proto-declarative pointing, as the non-autistic developmentally delayed control subjects demonstrate.

5 Leslie's (1987a) theory pinpointed the origins of metarepresentation at about 14 to 18 months of age, with the emergence of pretend play. His amended theory (Leslie and Happé, 1989) suggests that the origins of metarepresentation predate this, in being based in 12-month-olds' ostensive communication. The theory I am developing here suggests that the origins of metarepresentation may predate even this: that in understanding that others have attentional states, infants of seven to nine months reveal their capacity for metarepresentation.

Naturally, this does not imply that infants understand that other people have *beliefs (that x)*, but understanding attention may imply they understand that other people can *think (of x)*, to use Perner's (in press b) distinction.

6 Trevarthen and Hubley's (1978) notion of primary intersubjectivity is based on reciprocity seen between the infant and his or her caregiver. Stern (1985, p. 135) argues that there is no evidence that this is inter-mental, and moreover, unlike the behaviours comprising Trevarthen and Hubley's notion of secondary intersubjectivity, reciprocity in the neonatal period is not unique to humans but is seen in other social animals.

7 In parallel studies between London and Madrid, Juan-Carlos Gómez, Wendy Phillips and I are currently exploring how autistic children use eye gaze in situations where they desire an object.

8 Ages in this paragraph are relative indications of some predictions this theory makes. Until good longitudinal data is available, these ages will remain as heuristic estimates.