Mind & Language, ISSN: 0268-1064
Vol. 13. No. 1 March 1998, pp 56–75.

The Norms of Cognitive Development

PHILIP GERRANS

Abstract: Once the notion of a precursive relationship between developmental stages is fully articulated in terms of the distinction between 'role' and 'realiser' states, it turns out that the 'Theory of Mind' literature operates with a notion of precursive relationships described at too high a level of abstraction to explain actual mechanisms of development. Furthermore, the tendency within that literature to explain precursive relationships in terms of role states with isomorphic linguistic/computational structures is misleading. Developmental relationships are more likely to exist between states which play a similar normative role in the agents's psychology than between states which can be described as sharing a similar computational architecture.

> A theory of normal development then, has to be more than a description of changes in surface behaviour. Developmental theory must reveal the structure of knowledge and ability that is reflected in surface behaviour. It must also show how brain growth and learning create such knowledge and ability. (Leslie and Roth, 1993)

1. Precursor States in the Development of Social Recognition and Response

The purpose of this paper is to suggest an alternative to the M-representation theory of the development of social cognition.[1] The M-representation hypothesis explains the child's development of the ability to interpret and predict behaviour as the result of acquisition of progressively more sophisticated sets of concepts, culminating in the M-representations necessary to pass the false-belief test. The essential defect of the theory is that it characterizes its domain using the conceptual repertoire of intentional (folk) psychology and attempts to 'reveal the structure of knowledge and ability' on

Address for correspondence: Department of General Philosophy, Sydney University, Camperdown NSW 2006, Australia.
Email: philip.gerrans@philosophy.su.edu.au.

[1] The M-representation theory is a specific version of the 'theory of mind hypothesis' which has dominated research in the area since 1985. See Baron-Cohen *et al*., 1985, or Gopnik and Wellman, 1995.

which social cognition depends via analysis of natural language. The children's behaviour in various experimental environments is then interpreted to show that particular types of behaviour, say pretend play, or passing the 'false-belief test', form part of the same developmental sequence in virtue of the logical structure of the concepts allegedly underlying these abilities.

The M-representation hypothesis builds on a philosophical tradition which places logical form at the centre of the project of explaining belief. The normative role of belief is to track the truth, which is why beliefs are individuated by their inferential role. And it is natural to explain this inferential role in terms of the logical form of the propositions which represent the content of the belief. As Frege said, 'truth is the aim of logic as beauty is that of aesthetics'. Thus, when we look at the role of belief in causing behaviour, it is no surprise that it should be explained as the state of a system which represents a proposition of a distinctive logical form. The functional, normative and logical/inferential roles of belief appear to coincide. Whether or not this is the correct way to understand the mental representations involved in the mental state of belief, as opposed to the sentences which attribute beliefs, is not a question to settle here. However, even if we assume that the functional and normative roles of belief are essentially connected with logical structure, we should not expect that hypothesis to apply to the other attitudes. The cognitive attitudes are the only ones whose norm places strong constraints on the logical form of the propositions they represent, just because of the connection between logical form and truth tracking. Thus my claim will be that the M-representation theory is misleading because it presents early behaviour which depends on noncognitive attitudes as less sophisticated precursors of belief attribution. In fact their roles in social interaction are quite different from that played by belief attribution, and their connection with the later abilities is not a matter of their being earlier stages of the maturation of the same modularized ability.

The alternative I propose looks for developmental relationships among abilities and attitudes which serve the same purpose, or as I shall put it have the same normative role, rather than among those which have the same imputed logical structure. Of course this approach raises a different, but related, problem for the theorist trying to understand early behaviour such as imitation, pointing and play and its links to later behaviour governed by cognitive norms. If early behaviours obey different norms, how is it that their sequential occurrence appears to be necessary (and in normal circumstances) sufficient for the development of the ability to attribute belief? The end of the paper is an attempt to answer this question, an attempt which has none of the unity and simplicity of the M-representational theory. However, the apparent fragmentation of social cognition into a set of disparate and loosely related abilities need not render it inexplicable, just complex and precarious. First, however, I need to show the difficulties with the M-representation approach which render the more inelegant normative approach attractive.

The M-representation hypothesis was developed in order to explain both the typical developmental trajectory and autistic deviations from that tra-

jectory. Thus the raw data for the theory are the sequence of behaviours exhibited in normal development up until the age of 5. A normal trajectory of development includes in approximate temporal order, allowing for some overlap (for example the most sophisticated forms of imaginative play and deception arrive after the ability to pass the false-belief test), the following types of behaviour which have attracted the interest of theorists of social cognition (Baron-Cohen, 1993; Wellman, 1993):

neonatal imitation
early affective exchange
gaze monitoring
protoimperative pointing
protodeclarative pointing
shared attention
pretend play in its less and more sophisticated forms, from object substitution to make believe
3-year-old performance on the false belief test
4-year-old performance on the false belief test

All of these behaviours are intersubjective, they involve an interchange of more or less sophistication between an infant and other agents. In autism, one of whose clinical indicators is a developmental deficit of social recognition and response, all, except protoimperative pointing, of these types of behaviour are delayed, deviant or absent. In particular, in a large-scale study conducted by Simon Baron-Cohen, the absence of shared attention proved a very good indicator of both a later failure to pass the false-belief test and a subsequent diagnosis of autism (Baron-Cohen and Swettenham, 1996). Failure to engage in pretend play was a slightly less reliable indicator of a later diagnosis of autism. Retrospective anecdotal evidence from parents of children ultimately diagnosed as autistic suggests that in many cases they are affectively disengaged, do not make eye contact or point protodeclaratively. Thus the hypothesis suggests itself that perhaps autism results from damage to a capacity whose normal maturation underlies the typical sequence of development which is delayed or absent in the autistic case (Baron-Cohen *et al.*, 1993).

To support this hypothesis requires a deeper understanding of the nature of precursive relationships between earlier and later stages of development. An earlier stage is a precursor of a later if in normal circumstances its occurrence is necessary and sufficient for the occurrence of a successor. So someone who claims, for example, that ability to pretend is a precursor of ability to attribute belief is claiming that in normal circumstances the ability to attribute belief will not develop unless the ability to pretend develops first. Precursive relationships can exist between both *role* and *realizer* states.[2] A role

[2] I do not intend this definition to be revisionary, merely a slightly more precise formulation of assumptions common to theorists on all sides of the debate, namely that the precursor relationship is one of counterfactual dependence sustained by a causal process

state, like pretending, is realized by the actual mental/brain processes on which that action depends. The distinction is familiar from functionalist accounts of mind which distinguish between the functional role of a mental state and the state of a physical system which enables it to instantiate that functional role. Functional roles are characterized according to their inputs/outputs, abstracting away from the actual physical structures on which they depend. A realizer state is the state which physically implements the functional specification, enabling the system to play its causal role intermediate between input and output.

Folk psychology is concerned with the interpretation and prediction of behaviour and thus classifies mental states by their functional role. However, the task of the cognitive psychologist is, ultimately, to show how information represented by realizer states enables them to occupy or play that distinctive functional role; for example, to show how the ability to pretend or imitate is the result of a state of the brain which instantiates a certain conceptual structure. Recall Leslie's remark that 'Developmental theory must reveal the structure of knowledge and ability that is reflected in surface behaviour. It must also show how brain growth and learning create such knowledge and ability'. In so doing the cognitive scientist will build a bridge between the mental states which are evidenced by typical behaviour and attributed in intentional vocabulary, such as pretence and imitation, and the concepts which are instantiated in the developing brain of the child. For example, a cognitive science account of development would show how the representations involved in imitation or pretence are linked ultimately to those involved in passing the false-belief test. Intuitively, it seems there are two ways such the representations at realizer level might be linked precursively. Either earlier representations are *transformed* into, or *trigger*, via an intermediate causal process, the development of successor representations. An example of a transformation of a representation is a game of object substitution in which the banana is treated as a telephone. A representation of a normal scene is transformed by the replacement of one of its typical elements. An example of triggering might be seen in the hypothesis that 'parameter setting' is involved in the acquisition of natural language, e.g. that parsing ability, after being 'switched' to head or tail first, then develops in a typical sequence (Pinker, 1994).

With this apparatus in place we can see that the most elegant hypothesis about the development of social cognition is that there is a discrete, modularized capacity for the development of folk psychology, whose normal maturation is manifested by the presence of the behaviour listed above and whose impairment at some early or prenatal stage results in the typically deviant developmental profile of autism. Substantiating this hypothesis

---

in a physical system, the human brain. The issue then becomes characterizing that relationship in a way which generates perspicuous hypotheses as to the nature of that causal process.

requires an account of the transitions between realizer states which bring about the observable behavioural sequence which produces the role state classifications which are the data of folk psychology. No one has yet attempted such an account for all aspects of the normal sequence from birth to the age of 4, but there are elegant attempts to create precursive sequences between some of the key landmarks of development whose absences are also indicators of autism: for example, pretence, shared attention and belief attribution.

The M-representation hypotheses as developed by Leslie, and in a slightly different version by Simon Baron-Cohen, are two such attempts. A similar idea is found in the work of Perner, who initially coined the expression 'metarepresentation'. All three have a similar explanation of the conceptual leap the normal child makes between the ages of 3 and 4 which enable it to pass the false-belief test. The false-belief test tests the ability of the child to predict behaviour in situations where others will act on false, out of date, information which differs from that perceptually available to the child. For example, a child watches while another person sits at a table which has a teddy bear and a basket on it. The other person leaves the room and the teddy bear is placed in the basket. The child is then asked to say or point where the other person will look for the bear on her return. Three-year-olds say the basket, four-year-olds correctly predict that the other will look where the bear was (Wellman, 1993). In effect the child is being tested for her ability to understand that others may act on representations which have gone out of date, a form of false belief. There is no asymmetry between first and third person on this sort of test (Gopnik and Meltzoff, 1993). The 3-year-old child is just as bad at overriding present perceptual information in order to self ascribe out-of-date beliefs.

How should we explain the difference between the 3- and 4-year-old performance at this sort of task? A related question is 'how should we explain typical autistic performance on the false belief task?' Most autistics never pass the false-belief task, but even those who do are severely delayed and the suspicion remains that they do not do so using the same representational device as the normal person. Autistics are not just stuck at the 3-year-old stage of social cognition. An autistic of 3-year-old mental age will tend to exhibit the typical deficits in social recognition and response, stereotypy, language deficit and affective disengagement which differentiate them from their peers. Someone who holds to a version of the M-representational theory needs to show how damage to a precursive stage of the metarepresentational module could produce, not only failure to pass the false-belief test, but the typically deviant developmental profile of autism. Leslie's reformulation of the metarepresentational hypothesis, together with his explanation of the link between the concepts involved in pretence and belief attribution, is the beginning of an attempt to meet this requirement.

2. M-Representation

Declarative sentences, paintings and beliefs all stand in representational relations to their objects. To represent that relation is to metarepresent, to form a higher-order representation of a first-order representational relation. Examples would be 'that sentence says that x is F' or 'that painting is about the relation between x and y' or 'she believes/hopes/expects etc. that x is F'. The initial formulations of the metarepresentational hypothesis proposed that autism might be the result of a generalized inability with metarepresentation which manifests itself in social contexts as an inability to form beliefs about others' mental states, hence the failure to pass the false-belief test. However, this hypothesis, as it stood, could not account for the difference between normal and autistic children, because autistic children performed better than normal children on analogues of the false-belief test conducted with photographs and maps (Zaitchek, 1990). That is to say, they were better than normal children at detecting where maps or photos had gone out of date because of a change in location of relevant items: an analogue of the false-belief test where a subject's belief is falsified by the change of location of a salient object.

Leslie's explanation is that the false-belief test tests a specialized capacity for metarepresentation of *psychological states* which is insulated from a generalized metarepresentational capacity. This form of metarepresentation, restricted to psychological states, he calls M-representation. The essential feature of M-representation, as he defines it, is a logical structure borrowed from some philosophical analyses of *de re* belief attribution.[3] The proposed analysis is that Belief is a three place relational predicate ranging over agents, objects and properties. Thus the logical form of 'S believes (*de re*) that x is F' is: Belief(S, x, F).

Leslie suggests that this conceptual structure is first mentally represented at about the age of 2, as the result of maturation of a modular device devoted specifically to the understanding of agents. It coexists with another metarepresentational device devoted to nonintentional representative media. Leslie now has an apparently elegant explanation of the difference between 3- and 4-year-old and normal and autistic performance on the false-belief and false-photo tests. The false-photo task is solved by the nonintentional metarepresentational device (NID), the false-belief task by the M-representation device (Theory of Mind Module, ToMM). The difference between normal and autistic performance on false-beliefs and false-photo tests is explained by the fact that the NID is undamaged in autism whereas the ToMM never comes on line in autistics, evidenced by their difficulty with pretence.

The ToMM and the NID operate in concert with a Selection Processor (SP)

[3] The belief industry in philosophy is far too large to summarize, and constantly developing. An overview which explains the reasons for the relational formation initially made popular by Kaplan (Kaplan, 1975) is contained in Salmon and Soames (1988).

whose function is to override current perceptual evidence and select the appropriate earlier representation (the scene in the previous photo or as previously seen by the agent in the false-belief test) and feed it to the appropriate module. It is the coming on line of SP at around the age of 4 which explains differential autistic and normal performance on the false-belief test. Autistics and normal 3-year-olds fail that test for different reasons: the 3-year-old for lack of an SP, the autistic for lack of a ToMM. Autistics have intact SP and NID and can thus do extremely well on the false-photo and map tests.

Why then does the autistic reach 3-year-old standard on the false-belief test if she has no functioning ToMM? The answer is that she presses into service other inferential devices, language, general cognition, or perhaps information gleaned from the function of NID is employed by analogy. These possibilities are tasks for future research.

Leslie suggests that ToMM may in fact consist of a number of sub modules. He suggests at least two. The first, ToMM 1, attributes perceptual states and goals to agents, and figures in protoimperative pointing, shared attention and the understanding of desire. The second, ToMM 2, involving M-representation, develops later and includes abilities like protodeclarative pointing, pretence and belief. Pretence plays a crucial role for Leslie since it allows him to construct a precursive chain based on the development of M-representations.

There are at least three problems with this way of explaining things.

The first is that it makes it look as if autism is the result of the failure of a mechanism which comes on line at the age of 2, evidenced by pretence, as a result of the maturation of a mechanism whose function is to generate M-representations in psychological contexts. However, there is good anecdotal evidence that autism is manifest earlier than this in lack of affective response and exchange, possibly in failure of early imitation, and almost certainly in an absence of normal gaze monitoring and shared attention. In the case of shared attention, the evidence is more than anecdotal. Baron-Cohen's survey suggests that absence of shared attention is the most reliable indication of a later diagnosis of autism as well as of failure to pass the false-belief test. Thus shared attention has at least as good credentials as pretence to be a precursor of belief attribution, yet Leslie allocates them to different modules.

The second reason for being suspicious of this solution is that it is equally easy to analyse protodeclarative pointing and shared attention as depending on M-representation. Protodeclarative pointing is the indication or demonstration of an object in order to make it an object of interest or pleasure. It is thus very plausibly grouped with shared attention, which is distinguished by being accomplished by eye movement and the monitoring of others' eye movement, yet Leslie groups protodeclarative pointing with pretence and assigns shared attention to the functioning of a different module.

It is worth noting that Baron-Cohen, who treats shared attention as the crucial precursor in the development of social cognition, does so because of

the importance of the monitoring of gaze as a medium of communication and intersubjective engagement (Baron-Cohen, 1994). Further support for this emphasis on the role of gaze monitoring comes from work on primates, particularly chimpanzees, which suggests that attention monitoring is at the heart of social interchange which sustains the complex and volatile chimpanzee phallocracy. So there is evidence that gaze and attention monitoring is a basic mechanism of social interchange in primate species. Humans have complex verbal and cultural behaviour as well, but it seems plausible that the earliest human interactions would be similar to those of other primate species (De Waal 1982, Humphreys 1983). This thesis gains force when we consider that the autistic child, who does not monitor another's gaze or direct attention by eye movement, never develops the more sophisticated forms of recognition and response which make up social cognition. Once again, however, these types of consideration would tend to link shared attention with gaze monitoring and protodeclarative pointing, since all these activities involve looking into the eyes of the other as a way of gaining or sharing information.

In fact Baron-Cohen has explained shared attention and linked it with belief attribution in a way which is almost identical to Leslie's M-representation theory, except that it credits shared attention with M-representational structure. At the level of natural language analysis there seems no reason to say that the child who points or directs attention is not 'indicating that X is F' or 'indicates, of X, that it is F' where 'indication' can be a matter of gesture or eye movement. And presumably, shared attention is evidence of the fact that the child is able to attribute the same sort of ability to others, perhaps by substituting 'sees' into the slot occupied in her own case by 'indicates'. If this is correct, and we accept the hypothesis that argument structure is grounds for a hypothesis that an earlier ability is precursive of another, then perhaps protodeclarative pointing is the first time such a structure is evidenced. Later, the child can substitute more sophisticated attitudes of 'pretence' or 'belief' into the argument slot initialized by 'indicates'. As I said earlier, none of this should be taken too seriously because it is all conducted at the level of natural language and, no doubt, there are other ways to regiment the structure of the propositions used to attribute these sorts of attitudes.

At present, note only that Leslie allocates shared attention and belief to different modules when, if argument structure is the crucial factor, it seems that protodeclarative pointing, shared attention, pretence and belief really ought to fit together. Of course Leslie has his reasons for his focus on pretence, which brings me to a third objection raised by Currie's paper in this issue. Crucial to Leslie's focus on pretence is the fact that belief is an indirect attitude. That is, it is one conceived of as having the possibility of being true or false in virtue of its representative structure. Once again, Leslie is heir to a philosophical tradition which analyses this property in virtue of logical form. In the case of belief, current fashion treats it as the three-place relation

between agents, objects and properties mentioned earlier, the structure Leslie describes as M-representation.

If we follow Baron-Cohen in assigning metarepresentational structure to concepts other than belief, then perhaps the false-belief test marks a significant conceptual division *within* the set of concepts distinguished by their argument structure, a shift to the ability to attribute indirect attitudes. If this is the case, then M-representation structure might be a necessary but not a sufficient condition for the attribution of belief and other indirect attitudes.

However, for Leslie indirect attitudes and M-representation go together. In other words, he explains the functional role of belief in virtue of its logical form. Pretence plays the central role in his theory because he treats it as a quasi-indirect attitude. According to his analysis, in order to pretend, the child has to override the direct evidence of perception for the sake of a make-believe or a game. In effect, pretence becomes the entertaining of false belief within a limited context. So the child who holds a banana up to her ear and then talks into it is 'pretending, of the banana, that it is a telephone'. From here it is a short conceptual leap, once this structure is in place, to conceiving of all mental states as M-representations, signalled by passing the false-belief test.

It is hard to know if this is an empirical claim or established by philosophical analysis alone. In either case there are reasons to doubt it. The empirical evidence, while not conclusive, indicates that most make-believe play seems to be the result of variation on scripted behavioural routines which can be equally well explained as requiring the transformation of the represented routine rather than its metarepresentation (Fivush 1987, Lillard 1993). If this is correct, it would fit just as well if not better with the idea that the conceptual ability involved is the transformation of remembered representations rather than their metarepresentation.[4] For example, a child who puts a tennis ball in a bowl and then pretends to eat it need not have this argument structure in place; pretence (agent, tennis ball, orangehood). The pretence could equally well be accomplished by the substitution of one element of a represented behavioural sequence for another: i.e. the tennis ball for an orange. In either case the resultant role would still be described as 'pretending the tennis ball is an orange'.

However, Leslie's most important argument for the M-representational theory of pretence is philosophical, based on the ability, essential to make-believe, to distinguish pretence from belief in the act of make-believe. This tracking problem is important because make-believe situations are not entirely imaginary, they usually involve a background of veridical perceptual belief. The pretend food is in a real bowl for example. Leslie's point is that, in order to distinguish the attitudes, the child must be able to metarepresent them, in other words to represent to herself 'I *see/believe/know* the bowl is a bowl and I am *pretending* the tennis ball is an orange'.

---

[4] The point is developed in Sterelny forthcoming.

However, as Currie points out, if the child cannot distinguish between pretence and belief at first-order level, semantic ascent will not help. The child would need to be able to distinguish 'I pretend that I pretend that p' from 'I believe that I pretend that p', 'I pretend that I desire that p' etc., exactly the distinction the resort to metarepresentation was supposed to make. The child needs some way to distinguish the attitudes before she can metarepresent them *qua* distinctive type of attitude. In other words, the pretence and belief cannot be distinguished by the slot they fill in a logical structure because they need to be recognized as the attitude they are before they can be assigned to the appropriate slot.

3. Distinguishing the Attitudes by Normative Role

One obvious answer is to say that the attitudes are distinguished by their functional role. A problem for this view is that the functional roles of pretence and belief in a game of make-believe are so close that it is difficult to find any conceptual space between them, as reflection on the case of pretending the tennis ball is an orange, while believing the bowl is a bowl, reminds us. Make-believe would not function unless the belief and the pretence played the same functional role within that context. This is one reason to be sceptical about distinguishing between pretence and belief on the basis of functional role.

However, even if the pretence and belief were distinguished by functional role, it leaves the M-representational hypothesis out of the picture. For example, if, as seems plausible, pretence has a different structure, like that of a scripted behavioural routine, then the ability to play the functional role need not depend on a particular structure. A pretence will be a pretence whether Leslie or Lillard is right about the nature of the realizer states involved. So, while Leslie is right that the children must be able to solve the tracking problem, neither functional role or M-representation seem to explain how they do so.

Currie suggests another way to distinguish between the attitudes which avoids difficulties with M-representation and the problems he sees with functional role characterization of belief and pretence: characterization by normative structure. Attitudes are characterized by asking under what conditions the attitude can be said to have successfully carried out its function. For example, desires are successful when they are satisfied, beliefs are successful when they are true, fear is successful when it causes aversion from dangerous things, imagination when it produces a nonveridical representation, and so on. What of pretence? In the case of make-believe, it is successful when it brings about the imaginative engagement of the participants. It is not a tautology to say that the success condition for the attitude of pretence is to evoke pretence. If this sounds a little strange, recall that pretence is a strange attitude. The purpose of pretence is not to evoke a true or a false belief (see below on the difference between pretence and deception), satisfac-

tion or an affective sensation. A game of make-believe would succeed irrespective of the presence of any of these conditions. Of course pretence may be, and often is associated with the presence of one or more of these types of mental state but they are not essential to its success.

Currie suggests that the reason for the failure he diagnoses in Leslie's attempt to link pretence to belief attribution via M-representation is that pretence is a *sui generis* propositional attitude whose governing norm, pretence, is different from the norm of truth which governs belief. And we could add that this difference would apply irrespective of the truth of M-representation theory. This focus on normative constraints helps to explain another puzzling feature of cognitive development; the relation between pretence, belief and deception. On the M-representation theory of pretence, pretence arrives both too early and too late. Too late because, if Baron-Cohen is right, shared attention has a metarepresentational structure which is available to support the attitude of pretence from about 18 months. Yet pretence does not arrive until about the age of 2. This might support Leslie's allocation of pretence and shared attention to different modules on the basis that shared attention does not require M-representational analysis whereas belief and pretence, because they are indirect attitudes, do. However, if this is the case, why does the child who can play games of pretence have to wait until after she can pass the false-belief test to master deceptive strategies? After all, isn't deception a form of pretence? In this case pretence seems to arrive too early.

The answer, which is obscured by the M-representation theory, seems to be that this is the wrong analysis of the relation between pretence and deception. In fact, to deceive is to create a false belief in the mind of a subject rather than to engage in a pretence. The difference is not merely verbal, it suggests a reason why it is that the child who can pretend at the age of 2 cannot deceive until after she has passed the false-belief test. The ability to deceive requires mastery of the cognitive attitudes because the norm of deception is falsity, just as the norm of belief is truth. So it is not surprising that the child cannot deceive until she has a grip on the notion of truth and falsity, but she can pretend well before that (Sodian and Frith, 1993).

The cognitive attitudes are the last to be developed. The M-representational theory tends to explain this in virtue of the complexity of the representations on which they depend. If, however, we classify attitudes by their normative roles, then we should expect the cognitive attitudes to come on line last. The early survival of the child does not require the making of true judgments but the forming of affective bonds with other humans in order to obtain nourishment and nurture. In other words, the newborn child has a strong need for engagement with the mother satisfied by physical contact which is linked with pleasurable feelings. I tentatively link physical contact with early imitation because imitation plays a similar role in establishing the nurturing relationship. Gopnik suggests another link with affective contact in her explanation of the mechanisms which sustain imitation. She regards it as sustained by the ability to correlate representations of bodily

movement with proprioceptive awareness and affective sensation (Gopnik and Meltzoff, 1993). It is no coincidence that games of imitation, gaze monitoring and protodeclarative pointing all involve affective engagement, maintained by direct physical contact or eye contact. Thus I suggest that the norm involved is affective engagement. Once again this should not be surprising, given the evidence from other primate communities about the role gaze monitoring plays in social interaction.

Protoimperative pointing I am inclined to assign a different norm, the satisfaction of desire. Protoimperative pointing does not seem to be essentially intersubjective in the same way as gaze monitoring and protodeclarative pointing. The satisfaction involved comes not from the engagement with the other but from causing the other to bring the relevant object. Thus it is tempting to group it with crying, another activity designed to bring about the satisfaction of desire. So I tentatively exclude protoimperative pointing from the list of behavioural types which are specifically social. I am fortified in this by the fact that autistics, whose deficit is social, seem to have intact protoimperative pointing and the fact that a fairly typical autistic behaviour well past infancy is to push or shove another towards the object of desire, rather than to look at it or point and then check the gaze of the other to see if attention has successfully been directed.

Pretence, I have already mentioned, has a different normative constraint: pretence. Thus it does not fit readily in a precursive relationship with any of the other abilities. In some respects, pretence is an anomaly, an island of conceptual precocity in the young child's social repertoire. However, the incongruity is lessened if we see it as a *sui generis* attitude because in such a case the pressure to link it with belief, with pointing or with eye monitoring is removed. One of the difficulties with pretence seems to be in trying to fit in with other behaviours whose norms are different.

Of course pretence has an important role to play in social development insofar as it is manifest both early as simple object substitution and later in games of imaginative play involving others. And the lack of imaginative play of autistics also suggests a link between pretend play and normal social development. However, if pretence does play an important role, I suggest that it is not in virtue of the M-representational status of the realizer states on which it depends. Rather, I suggest, it draws the pretender into the community by allowing her to rehearse behavioural routines typical of the culture of that community. Work in cognitive anthropology (Fivush 1987, Boyer 1993) suggests that young children learn the behaviour appropriate to specific cultural settings by building mental models called scripts. There are scripts for almost any routine situation and variation from the script (e.g. eating the wrong food for breakfast) provokes confusion. If this is correct, pretence may play a role in setting and revising the allowable parameters for social behaviour, building on games of object substitution. This is only a tentative hypothesis, but it is worth noting that cultural anthropologists like Pascal Boyer have also emphasized the role of stereotypical narratives in explaining tradition, from things like ritual circumcision to the royal wed-

ding. The variations which are tolerated within a tradition turn out to be those which can be accommodated to the existing script, rather than to a set of abstract principles tacitly understood by the participants. So perhaps early pretence has its role in social cognition as the rehearsal of an ability to acquire culture (broadly construed as behaviour typical of a social group) through the construction of scripts. Auden might have been correct when he wrote, in *The Age of Anxiety*, 'human beings are, necessarily, actors who cannot become something until they have first pretended to be it'.

My suggestion is that the cognitive attitudes are the last to be developed, not because they depend on representations of greater structural complexity or require more computing power than early imitation or pretence, but because they are not really needed by the very young child to interpret and predict its social environment. The early needs of the child are for nourishment and nurture; after that it is advantageous for the child to involve herself with her fellows, but all this can be accomplished without the need to make judgments about the truth or falsity of another's representations of the world. So even if the metarepresentation theory is correct, and attributing beliefs is a matter of entertaining M-representations, it does not follow that any of the earlier abilities of the child can be linked in an obvious way to belief attribution, precisely because those earlier abilities obey different norms from the cognitive attitudes.

In suggesting that they obey different norms one can remain neutral over whether they are part of the same modularized ability. I suspect that they are not, but the issue is complicated. Some considerations in favour of modularity are the apparent innateness, automaticity, rapid development and domain specificity of social cognition. These considerations are added to by the fact that autism is present from birth, manifest as delayed development of a variety of social skills which are unconscious and automatic in normal children. However, the key to modularity is *informational encapsulation*.

Once we start to describe an ability in terms of informational encapsulation, normativity enters the picture because to describe the information relevant to the function of a module requires the attribution of a success condition for performance of that module. Consider Marr's putative success condition for human vision, the construction of 3D colour images. His theory is then designed to describe the computational tasks involved in transforming monochrome 2D input to 3D coloured output. The point is that the success of his programme cannot be judged without some antecedent characterization of what is involved in successfully seeing. The problem, difficult enough for vision, becomes more difficult still when we turn to social abilities like grammatical usage. Here we are at the mercy of intuition. A linguistic theory is successful just so far as it comports with the intuitions and usage of native speakers, so linguists in effect try and extract a norm from the behaviour of their informants.

The more complex and wide ranging the behaviour the less likely the relevant information is to be encapsulated. It is one of the triumphs of linguistics that an ability like language can (allegedly) be unified in terms

of the representation of a very few principles which can subsume all the relevant information. It is of course one charge against linguistics that the unity and simplicity are achieved by retreat to too high a level of theoretical abstraction, by postulating principles which are so general that they are almost vacuous.

I think that the linguist can meet this charge, precisely because we do have a strong intuitive grip on the domain of explanation and what seems to count as the information to be regimented by a good explanation. However, in the case of social cognition the charge is more difficult to answer. Exactly what is the success condition for social cognition? Interpretation and prediction of intentional behaviour is the answer widely accepted. But this norm seems to operate at too high a level to guarantee informational encapsulation. Surely all the information relevant to navigating the social world cannot be processed by the same device. When my psychiatrist tells me that this paper is late because I have a subconscious desire to castrate my father, is she using the same mechanism which allows someone to respond instantly and automatically to someone's smile, to follow a gesture, get a joke, to feel embarrassed or sexually aroused? It seems unlikely, yet these are all capacities on which interpretation and prediction of others' behaviour depend.

To characterize all these abilities as having the same success condition, social recognition and response, is a bit like saying that the welding, machining and assembling which take place in a factory all have as their normative goal the production of a car. In one sense this is true, the output of all these activities is ultimately linked to that goal. In another sense it would be ridiculous to specify the norm governing the installation of a tail light as the production of a car.

So a factory might be a modular device for the production of a certain make of car, but it fulfils the normative requirements on car production because it is a place where a variety of other modularized capacities combine. I suspect that the same is true of social recognition and response. It depends on a lot of separate capacities, each governed by different norms, functioning in concert to enable the organism to navigate its social environment. Note that Leslie himself appears to agree when he divides the ToMM into a variety of submodules in order to explain the autistic and normal performance on the false belief test. Note also that he complicates matters still further by suggesting that modular capacities whose norm is not specifically social, such as the NID, can be used, by the small minority of autistics who eventually succeed, to pass the false belief test. Tager-Flusberg (1993) makes a version of a similar move when she suggests that perhaps autistics use the representative device of language to navigate their social world. So it seems that devices which are not specifically social can be pressed into service to meet the goals of social cognition. Furthermore, it seems that Leslie is prepared to regard social cognition as subtended by multiple modular capacities, individuated by different norms. So it turns out that the real difference between Leslie and myself is that I regard social

cognition as more fragmented than he, in that I am prepared to decompose an apparently unified ability into a variety of relatively autonomous ones.

4. Explaining Development

If this approach is correct, three questions remain. To answer them would be to develop a complete theory of development. Here I shall merely indicate the lines such a theory could take consistently with the empirical and philosophical work in the area.

Is there a single factor involved in the domain specific deficit of autism? Autism is less likely to be the result of a single modular incapacity than a failure to harness to the task of interpreting behaviour a variety of abilities whose norms are independent. I have suggested elsewhere that this is because the autistic child, even from birth, never experiences the normal affective responses to others. She doesn't seek affection, imitate or monitor gaze. For her, the actions and reactions of her fellow humans do not provoke any automatic affective responses. By contrast, almost from birth, the normal human spends much of her time looking at others' faces, checking them for response and reinforcing, through the affective responses this engagement generates, a disposition to social involvement. Thus I suggest that the difference between the autistic and the normal person (of the same IQ) is not so much a lack of some specifically social capacity as the failure to develop her capacities within an intensely social context. The autistic child tends to play, to use language and even to pretend in ways which are vastly deviant from the norm. Take pretence for example. It used to be thought that autistics did not pretend or play imaginatively. However, their deficits in these areas are a tendency to routine and stereotype with no open-ended extension or social involvement, rather than a lack of pretend play *per se*. Since they are divorced from the social arena, it is not a surprise that they tend to be fascinated by the concrete physical world and tend to endless repetitive behaviour. Similarly their language deficit is not a semantic or syntactic one but pragmatic (Tager-Flusberg, 1993). They do not infer the thoughts and feelings of others to interpret the nonliteral aspects of speech, and their conversation tends to be repetitive, stereotypical and not to involve others. I suggest that the language case is an instance similar to autistic pretence: not the complete absence of an ability but its deformation due to its detachment from the social context in which it is normally developed and employed.

These remarks may seem to claim that the autistic deficit is entirely the result of inability to experience the normal range of affective responses to others. I do think that the deficits I describe are plausibly explained in this way but I have not argued here that all the autistic deficits result from affective disengagement (though elsewhere I have defended a version of this thesis). However, it is not part of the thesis I present here that there is an alternative *unifying* explanation to the M-representation theory. I think that social recognition and response depend on a variety of abilities which

develop independently and at different times. Some of these may be precursive of each other (perhaps early imitation, gaze monitoring and protodeclarative pointing). But I am committed to the view that pretence is a *sui generis* ability relatively unconnected with earlier and later aspects of social cognition.

I have retreated to something very like the 'stage setting' hypothesis proposed by an early proponent of a nonmodular version of the 'theory of mind', Alison Gopnik. I do not agree with her that the abilities involved in passing the false-belief theory are theoretical in any important sense of the term, or that belief attribution is developed via a series of theoretical precursors. Nonetheless, in her explanation of early imitation, she points out the importance of early affective involvement in embedding the child in the social context in which most of her cognitive abilities will subsequently develop.

A second response to the demand for a unifying explanation is to note that there is reason to suggest that deviant behaviours other than autism can potentially be explained at the neural level, bypassing the need to formulate an account of the structure of the concepts involved and logical links between them. For example, some mental illnesses such as schizophrenia are manifest as a set of bizarre behaviours. One possible way to explain the behaviour of the schizophrenic might be to construct a theory of the schizophrenic 'world view' as manifest in their actions and to compare it to the set of concepts relied on by the rest of us in the situations in which schizophrenics cannot cope. However, schizophrenia might be the result of something as simple as an excess of dopamine or the level of a neurotransmitter which causes the brain to malfunction in a specific way.

Or imagine that the development of attention (the ability to represent relevant items of information together) depended on the release of a simple protein at a specific stage of development. If that protein was never released, then all the cognitive skills which depend on working memory would be affected. The behavioural outcome would be a disturbance to a wide range of capacities with some isolated skills left unaffected. The developmental profile of such a person would also look bizarre compared to the normal case, but one cannot help feeling that an investigation which tried to establish a set of precursor relationships between various cognitive skills normally exhibited would end up being misplaced.

I am not claiming that autism is a similar case, just suggesting, firstly, that the analysis of folk concepts may in fact mislead in the explanation of mental malfunction and, secondly, that the abandonment of such analyses does not leave us without explanatory resources. Perhaps, in the autistic case, nothing will be lost if we sacrifice the unity and simplicity of the modular hypothesis and recognize that social cognition depends on a variety of abilities functioning in concert.

How do the variety of abilities involved in social cognition function in concert if they do not depend on the maturation of a single module?

A unified account of the development of social cognition suffers from the

problem that the candidate precursors are governed by norms which are essentially noncognitive. In attempting to explain development we are left with a variety of alternatives if we discount the idea, which I hope now seems unattractive, that there is a single unified M-representation module whose telos is to enable the child to attribute beliefs.

The first is that there might be some other single mental capacity which is involved. Simulation theory is the main alternative here and much of what I have said would be conducive to simulation theorists. The essential feature of simulation theory is that it is a version of the mental model approach to inference. Instead of representing psychological concepts or generalizations and subsuming behaviour under them, the child uses a model which embodies, rather than explicitly represents, psychological truths. The difference is like that between predicting a boat's performance by using hydrodynamic theory or by using a model boat which embodies those truths. In the latter case one does not need to know anything about Froude's laws to make the same predictions. One can predict what will happen without being able to explain why it happens. In the case of human behaviour the child uses herself as a model in order to predict what others will do. One advantage of the simulation theory is that in characterizing it this way it is normatively neutral: a child can simulate affective response, as in Gopnik's account of early imitation, attention, or the more cognitive attitudes, without knowing anything about the nature of the mechanism (i.e. her own psychology) which sustains them.

This is not the place to rehearse the debate over simulation theory, but merely to note one of its advantages. However, in any such debate, one crucial issue is left: 3- and 4-year-old performance on the false-belief test. It marks a clear developmental break. It seems to me that a simulationist would need to explain the 4-year-old's ability as the result of simulating her own response in a different situation, compared to the 3-year-old's inability to abstract from her own situation. The issue is whether this can be done without letting a version of metarepresentation back in. However, even if we allow a version of hybrid theory which allows metarepresentation to explain the difference between 3- and 4-year-olds, it does not presuppose precursive relationships between belief attribution and the earlier abilities which depend on tuning a simulator to a variety of different tasks which do not require that level of abstraction.

Another possibility is that the false-belief task marks the difference between the ability to model human psychology and to make explicit some generalizations about human psychology. An analogy might be someone who, after using a model boat in a variety of different situations over a long period, is finally able to work out some of Froude's laws in a rudimentary form. I think this is an attractive option but note that, once again, it seems to leave belief attribution as a relatively isolated ability. In fact it is conducive to an account which treats belief attribution as an instance of a more generalized representational capacity for the formation and evaluation of hypotheses. For instance, it may be as a result of the development of gen-

eralized cognition, or the arrival of language, that the child is able to make explicit what was previously inexplicit or to imaginatively deploy her simulator in a wider range of situations. So false-belief attribution need not be the result of the culmination of development of a single module, but may be one more instance of an ability being applied to the problem of navigating the social world.

One argument against this sort of account is that it cannot explain the autistic failures on the false-belief test. The reason is that the false-belief test is supposed to test for a specifically social, as opposed to generalized cognitive deficit (recall the differential autistic performance on the false-photo task which led to the modification of the original metarepresentation hypothesis). However, if autism is partly the result of lack of affective, and consequently social, engagement from very early on, it is unsurprising that the autistic child does not develop the same abilities as the normal child. Not because social cognition depends on a single unifying mechanism, but because the autistic child never acquires and deploys any of its developing abilities in specifically social contexts. Good examples appear to be pretence and the typical autistic language deficit. These are not abilities whose primary norms are social engagement but they are normally developed within a social context. The result in the case of autistics, who are not anchored in the social world, is that their pretence and language use is deviant, though presumably the processing on which their ability to pretend and to use language depends is unimpaired. Similarly, if what is required to pass the false-belief test is a simulator tuned to appropriate circumstances, or a generalized cognitive ability, it would not be surprising if the autistic child failed the test, not so much because it lacked these abilities but because of their deviant development in the absence of specifically social input. This hypothesis is consistent with the cases in which the autistic of normal intelligence eventually passes the false-belief test. Despite their success at this task, they still attract the classification of autism due to their difficulties with social recognition and response. Typically this is explained as a consequence of the malfunction, at an early stage, of the social cognition module, resulting in their having to rely on other means to cope with their social environment. The alternative explanation is that the failure to pass the false-belief test may be the result of an ability which is intact but which is never employed by autistics in the normal way, precisely because it has not been developed, as in the normal case, through social interaction. A parallel might be someone who had to learn language from a book or a computer rather than via normal social engagement. It would not be surprising if they had a delayed and deviant pattern of language acquisition.

These are only suggestions. The purpose of the paper is to continue to try to break the stranglehold the M-representation theory has on research in the area and to suggest that focus on the normative aspects of the abilities involved in social interaction might be productive. The projects are related because it turns out that the inability of the M-representation hypothesis to forge a precursive link between pretence and belief attribution is a direct

result of the fact that pretence does not obey a cognitive norm and thus does not require the conceptual structure imputed to the cognitive attitudes. Pursuing this line of enquiry, it turns out that a similar problem exists for the rest of the attitudes claimed as precursors of belief attribution by versions of the M-representation theory. Once we fragment the unity of social cognition this way, attention turns to explaining the integration of these different submodules in the project of social interaction. This is a research task for the future, but I suggest that attention to the role of affect yields some initially plausible hypotheses.

Department of General Philosophy
Sydney University

References

Baron-Cohen, S. 1993: From Attention-Goal Psychology to Belief-Desire Psychology: The Development of a Theory of Mind, and its Dysfunction. In Baron-Cohen et al. (eds.), 1993.

Baron-Cohen, S., Leslie, A. and Frith, U. 1985: Does the Autistic Child Have a 'Theory of Mind'? *Cognition*, 21, 37–46.

Baron-Cohen, S., Tager-Flusberger, H. and Cohen, D. (eds.) 1993: *Understanding Other Minds: Perspectives from Autism*. Oxford University Press.

Baron-Cohen, S. and Swettenham, J. 1996: The Relationship Between SAM and ToMM: Two Hypotheses. In P. Carruthers P. K. Smith (eds.), *Theories of Theories of Mind*. Cambridge University Press.

Boyer, P. (ed.) 1993: *Cognitive Aspects of Religious Symbolism*. Cambridge University Press.

Currie, G. 1986: Fictional Truth. *Philosophical Studies*, 50, 195–212.

Currie, G. 1995: Imagination and Simulation: Aesthetics Meets Cognitive Science. In M. Davies and T. Stone (eds), *Mental Simulation*. Oxford: Basil Blackwell.

Currie, G. 1996: Simulation-Theory, Theory-Theory and the Evidence from Autism. In P. Carruthers and P. Smith (eds), *Theories of Theories of Mind*. Cambridge University Press.

Currie, G. 1997: The Paradox of Caring. In M. Hjort and S. Lavers (eds), *Art and the Emotions*. Oxford University Press.

De Waal, F. 1982: *Chimpanzee Politics*. London: Jonathan Cape.

Evans, G. 1982: *The Varieties of Reference*. Oxford University Press.

Fivush R., Gray, J. and Fromhoff, F. 1987: Two-year-olds Talk About the Past. *Cognitive Development*, 2, 393–409.

Frith, U. 1989: *Autism: Explaining the Enigma*. Oxford: Basil Blackwell.

Frith, U. and Happe, F. 1994: Autism: Beyond Theory of Mind. *Cognition*, 50, 115–32.

Goldman, A. 1993: The Psychology of Folk Psychology. *Behavioral and Brain Sciences*, 16, 16–28.

Golomb, C. and Kuester, R. 1996: On the Transition from Pretence Play to Reality: What Are the Rules of the Game? *British Journal of Developmental Psychology*, 14, 203–17.

Gopnik, A. and Meltzoff, A. 1993: The Role of Imitation in Understanding Persons and Developing a Theory of Mind. In Baron-Cohen et al. (eds.), 1993.

Gopnik, A. and Wellman, H. 1995: Why the Child's Theory of Mind Really Is a Theory. In M. Davies and T. Stone (eds.), *Folk Psychology*. Oxford: Basil Blackwell.

Harris, P. 1989: *Children and Emotion*. Oxford: Basil Blackwell.

Humphreys, N. 1983: *Consciousness Regained*. Oxford University Press.

Kaplan, D. 1975: Quantifying In. In D. Davidson and J. Hirtikka (eds.), *Words and Objections*. Dordrecht: Reidel.

Leslie, A. 1987: Pretense and Representation: The Origins of 'Theory of Mind'. *Psychological Review*, 94, 412–26.

Leslie, A. 1988: The Necessity of Illusion: Perception and Thought in Infancy. In L. Weiskrantz (ed.), *Thought without Language*. Oxford: Clarendon Press.

Leslie, A. 1991: The Theory of Mind Impairment in Autism: Evidence for a Modular Mechanism of Development. In A. Whiten (ed.), *Natural Theories of Mind*. Oxford: Blackwell.

Leslie, A. 1994: Pretending and Believing: Issues in the Theory of ToMM. *Cognition*, 50, 211–38.

Leslie, A. and Frith, U. 1988: Autistic Children's Understanding of Seeing, Knowing and Believing. *British Journal of Development Psychology*, 6, 315–24.

Leslie, A. and Roth, D. 1993: What Autism Teaches us about Metarepresentation. In S. Baron-Cohen et al (eds.), 1993.

Leslie, A. and Thaiss, L. 1992: Domain Specificity in Conceptual Development: Neuropsychological Evidence from Autism. *Cognition*, 43, 225–51.

Lewis, D. 1978: Truth in Fiction. Reprinted in *Philosophical Papers*, vol. 1, 1985. Oxford: Basil Blackwell.

Lillard, A. 1993: Pretend Play Skills and the Child's Theory of Mind. *Child Development*, 64, 348–71.

Pinker, S. 1994: *The Language Instinct*. London: Allen Lane.

Salmon, N. and Soames, S. (eds.) 1988: *Propositions and Attitudes*. Oxford University Press.

Sodian, B. and Frith, U. 1993: The Theory of Mind Deficit in Autism: Evidence from Deception. In Baron-Cohen et al. (eds.), 1993.

Tager-Flusberg, H. 1993: What Language Reveals About the Understanding of Other Minds in Children with Autism. In Baron-Cohen et al. (eds.), 1993.

Walton, K. 1982: Pictures and Make-Believe. *Philosophical Review*, 82, 283–319.

Wellman, H. 1993: Early Understanding of Mind: The Normal Case. In Baron-Cohen et al. (eds.), 1993.

Zaitchek, D. 1990: When Representations Conflict with Reality: The Preschooler's Problem with False Beliefs and False Photographs. *Cognition*, 35, 41–68.