Shaun Gallagher

The Practice of Mind

Theory, Simulation or Primary Interaction?[1]

Theory of mind explanations of how we know other minds are limited in several ways. First, they construe intersubjective relations too narrowly in terms of the specialized cognitive abilities of explaining and predicting another person's mental states and behaviours. Second, they sometimes draw conclusions about second-person interaction from experiments designed to test third-person observation of another's behaviour. As a result, the larger claims that are sometimes made for theory of mind, namely that theory of mind is our primary and pervasive means for understanding other persons, go beyond both the phenomenological and the scientific evidence. I argue that the interpretation of 'primary intersubjectivity' as merely precursory to theory of mind is inadequate. Rather, primary intersubjectivity, understood as a set of embodied practices and capabilities, is not only primary in a developmental sense, but is the primary way we continue to understand others in second-person interactions.

In psychology, philosophy of mind and, more recently, in the neurosciences, studies of how one person understands and interrelates with another person have been conducted under the heading of 'theory of mind'. Discussions of theory of mind are dominated by two main approaches: theory theory and simulation theory. The major tenets of theory theory are based on well-designed scientific experiments that show that children develop an understanding of other minds around the age of four. One version of theory theory claims that this understanding is based on an innately specified, domain specific mechanism designed for 'reading' other minds (Baron-Cohen, 1995; Leslie, 1991). An alternative version claims that the child attains this ability through a course of development in which the child tests and learns from the social environment (Gopnik and Meltzoff, 1997). Common to both versions of theory theory is the idea that children attain their understanding of other minds by implicitly employing a theoretical stance. This stance involves postulating the existence of mental states in others and using such postulations to explain and predict another person's behaviour. In the earliest level of development, the four- to five-year-old child's theory of mind involves 'first-order

[1] My thanks to an anonymous referee for constructive remarks on a previous version of this paper. I have also benefited from discussions following presentations of this material at York University in Toronto and at the British Society for Phenomenology meeting at Oxford University.

belief attribution' in which she distinguishes her own belief from someone else's belief. The next level involves 'second-order belief attribution', the ability to 'think about another person's thoughts about a third person's thoughts about an objective event' (Baron-Cohen, 1989, p. 288). Normal children between the ages of six and seven years old are able to achieve the second level. The very few autistic children who attain the earliest level, do so late, and they fail to attain the second level.

The second approach, simulation theory, argues that one does not *theorize* about the other person but uses one's own mental experience as an internal model for the other mind (e.g. Gordon, 1986, 1995a; Goldman, 1989; and Heal 1986, 1998a,b). To understand the other person, I simulate the thoughts or feelings that I would experience *if I were in the situation of the other.* I emulate what must be going through the other person's mind; or I create in my own mind pretend beliefs, desires or strategies that I use to understand the other's behaviour. My source for these simulations is not a theory that I have. Rather, I have a real model of the mind at my immediate disposal, that is, I have *my own mind*, and I can use it to generate and run simulations. I simply run through the sequence or pattern of behaviour or decision making that I would engage in if I were faced with the situation in question. I do it 'off line', however. That is, my imaginary rehearsal does not lead to actualizing the behaviour on my part. Finally, I attribute this pattern to the other person who is actually in that situation. According to simulation theory, this process may remain non-conscious, with only an awareness of the resulting understanding or prediction. The process itself, nonetheless, is structured as an internal, representational simulation (Gordon, 1986).

Across both of these approaches in the theory of mind literature one can distinguish two specific kinds of claims. First, *developmental claims* involve the timing and order of development, the importance and balance of innate mechanisms versus experience, and so forth. The experimental and neurological evidence used to support such claims and to justify the theory or simulation interpretations is impressive. It is possible, however, to raise questions about certain background assumptions that shape the design of such experiments and the interpretation of data as supportive of certain aspects of theory of mind. Second, *pragmatic claims* concern the scope of the applicability of theory of mind.[2] Pragmatic claims may be strong or weak. Some theorists (e.g. Baron-Cohen, 1995; Tooby and Cosmides, 1995; Leslie, 2000) make a very strong pragmatic claim for theory of mind, namely that, once formed, theory of mind is our primary and pervasive means for understanding other persons. It is not clear, however, that the experimental evidence used to support the developmental claims counts as evidence to support the strong pragmatic claim. Although I will question the interpretation of the science that informs the developmental claims, in this paper my main target is the strong pragmatic claim — that ordinarily and for the most part theory of mind forms the basis for our understanding of others.

To make clear what the strong pragmatic claim entails, let me review several of its various formulations. Recently, in a long conversation with Paul Ricoeur, the neuroscientist Jean-Pierre Changeux proposed that 'one's relation to others' depends

[2] These specific claims aside, the distinction between developmental aspects and pragmatic aspects of theory of mind is mirrored in recent research suggesting that the development of theory of mind may depend on normal language development (Astington and Jenkins, 1999), but that near perfect performance on theory of mind tasks does not depend on normal language functioning (Varley and Siegal, 2000).

on a 'cognitive device' that allows for the representation of the other's mental states, 'their sufferings, plans of action, [and] intentions'. He specifically cites experiments that support the concept of a theory of mind, and he maintains that it is just this type of mechanism that allows humans to acquire a system of moral values and aesthetic preferences (Changeux and Ricoeur, 2000, pp. 154–7). Two important researchers of this cognitive mechanism, Tooby and Cosmides (1995), suggest that 'humans everywhere interpret the behavior of others in . . . mentalistic terms because we all come equipped with a "theory of mind" module (ToMM) that is compelled to interpret others this way, with mentalistic terms as its natural language' (p. xvii). Baron-Cohen (1995, p. 3) writes: 'it is hard for us to make sense of behavior in any other way than via the mentalistic (or "intentional") framework'. Quoting Dan Sperber he continues: ' "attribution of mental states is to humans as echolocation is to the bat." It is our natural way of understanding the social environment' (p. 4). The conclusion proposed by Uta and Christopher Frith (Frith and Frith, 1999), that mental state attribution plays a major role in all social interactions, is echoed by Alan Leslie (2000), who defines ToMM as a specialized component of social intelligence, but claims that it is necessarily involved 'whenever an agent's behavior is attended', for example, 'in conversations and other real-time social interactions' (p. 1236).[3]

I do not propose simply to criticize the approaches of theory theory and simulation theory without offering something in their place. The alternative that I will propose is that the understanding of the other person is primarily neither theoretical nor based on an internal simulation, but is a form of embodied practice. In explicating this idea I do not want to deny that we do develop capacities for both theoretical interpretation and simulation, and that in certain cases we do understand others by enacting just such theoretical attitudes or simulations. Such instances are rare, however, relative to the majority of our interactions. Theory theory and simulation theory, at best, explain a very narrow and specialized set of cognitive processes that we sometimes use to relate to others (this would constitute a weak pragmatic claim for theory of mind).[4] Neither theoretical nor simulation strategies constitute the primary way in which we relate to, interact with or understand others.

The Embodied Practice of Primary Intersubjectivity

There is good scientific evidence to support the developmental claim that around the age of four children come to recognize that others are capable of having beliefs different from their own. Prior to this, however, the basis for human interaction and for understanding others has already been laid down by certain embodied practices — practices that are emotional, sensory-motor, perceptual and nonconceptual. I want to suggest that these embodied practices constitute our primary access for understanding others, and continue to do so even after we attain theory of mind abilities. Development that is specific to theory of mind happens within a wider framework of interpersonal pragmatics which can be characterized as second-person interactions with other persons perceived as others.

---

[3] Also, Wellman (1993): children at age four begin to 'see people as living their lives within a world of mental content that determines how they behave in the world of real objects and acts', they construe 'people's real-world actions as *inevitably* filtered through representations of the world rather than linked to the world directly' (pp. 31–2).

[4] Concerning a related narrowness of theory theory, see Chandler and Carpendale (1998).

The basic claim that I will defend is that in most intersubjective situations we have a direct, pragmatic understanding of another person's intentions because their intentions are explicitly expressed in their embodied actions. For the most part this understanding does not require us to postulate some belief or desire that is hidden away in the other person's mind, since what we might reflectively or abstractly call their belief or desire is expressed directly in their behaviour. The evidence to support this claim overlaps to some extent with evidence that is sometimes cited for both theory theory and simulation theory. I will review and reinterpret this evidence first, and then go on to discuss evidence that suggests that theory theory and simulation theory are unable to capture the full range of second-person interactions.

Many of those who argue for the theory or simulation approach acknowledge that for either a theoretical stance or a simulation routine to get off the ground some understanding of the context and behaviour of the other person must be had first; otherwise I would have nothing to simulate or to theorize about. This suggests that before I can develop a theory of mind I must already have an understanding of the other and their experience — including the other as the subject of intentional action. Prior to the possibility of knowing the other's mind in either a theoretical or simulation mode, one already requires (a) an understanding of what it means to be an experiencing subject; (b) an understanding of what it means that certain kinds of entities (but not others) in the environment are indeed such subjects; and (c) an understanding that in some way these entities are similar to and in other ways different from oneself. Furthermore, to form a theory about or to simulate what another person believes or desires, we already need to have specific pre-theoretical knowledge about how people behave in particular contexts.

One way to summarize these pre-theoretical conditions is to say, following a formulation suggested by Bruner and Kalmar (1998) concerning our understanding of the self, that the understanding of others in terms of their mental states requires a 'massively hermeneutic' background. This suggests that there is much going on in our understanding of others, in excess of and prior to the acquisition of theoretical and/or simulation capabilities. How do we get this background understanding? Some theorists answer this question by pointing to capabilities in infants and young children that they consider 'precursors' of theory of mind (Baron-Cohen, 1995; Gopnik and Meltzoff, 1997; Meltzoff, 1995; Meltzoff and Prinz, 2001; Nadel and Butterworth, 1999). In contrast, I interpret these capabilities as clues for an alternative approach to the issue of how we understand other people.

Pre-theoretical (non-conceptual) capabilities for understanding others already exist in very young children. Children, prior to the age of three, already have a sense of what it means to be an experiencing subject; that certain kinds of entities (but not others) in the environment are indeed such subjects; and that in some way these entities are similar to and in other ways different from themselves. This sense of others is already implicit, at least in a primitive way, in the behaviour of the newborn. We see evidence for it in instances of neonate imitation, which depends not only on a distinction between self and non-self, and a proprioceptive sense of one's own body, but the recognition that the other is in fact of the same sort as oneself (Bermúdez, 1996; Gallagher, 1996; Gallagher and Meltzoff, 1996). Infants are able to distinguish between inanimate objects and people (agents), and can respond in a distinctive way to human faces, that is, in a way that they do not respond to other objects (see

Legerstee, 1991; Johnson, 2000; Johnson *et al*., 1998). Experiments by Meltzoff and Moore (1977, 1994) demonstrate that from birth the action of the infant and the perceived action of the other person are coded in the same 'language', a cross-modal system that is directly attuned to the actions and gestures of other humans. In the case of imitated facial gestures, one does not require an intermediate theory or simulation to translate between one's proprioceptive experience of one's face and the visual perception of the other's face. The translation is already accomplished at the level of an innate body schema that integrates sensory and motor systems (Gallagher and Meltzoff, 1996). There is, in this case, a common bodily intentionality that is shared across the perceiving subject and the perceived other. As Gopnik and Meltzoff indicate, 'we innately map the visually perceived motions of others onto our own kinesthetic sensations' (1997, p. 129).

Should we interpret this intermodal and intersubjective mapping as a primitive form of theorizing or 'an initial theory of action'? Gopnik and Meltzoff (1997) think so. They suggest that infants form a 'plan', an internal representation of what they will do, and then they 'recognize the relationship between their plan to produce the action and the action they perceive in others' (p. 130). On this view, this is the beginning of an inference-like operation that is eventually promoted into a theoretical attitude. But is the motor plan equivalent to a mental state? They suggest it is, although not a very sophisticated mental state. But if, in this case, we ask what a mental state is, it seems to be nothing other than a certain disposition of the body to act intentionally, plus the phenomenal sense of what it is like to do the action. Certainly it does not have the status of an ideational event that intervenes to mediate vision and proprioception. Intermodal experience is characterized as phenomenologically transparent. That is, the sensory-motor process does not require an internal copy that the infant consults in order to know what to do. Although neonates do in fact perfect their imitative actions (improving the match between their gesture and the perceived gesture — therefore indicating that they register the difference between themselves and the other), they need no internal plan to consult since they have a visual model right in front of them, namely the face of the other, as well as a proprioceptive model, namely the gesture that is taking shape on their own face. Even in those cases where the infant has cause to remember the presented gesture in order to imitate it after a delay (see Meltzoff and Moore, 1994), it is difficult to construe a sensory-motor memory as a theory of action.

Accordingly, the body schema does not function as an 'abstract representation' (Gopnik and Meltzoff, 1997, p. 133). If, as Meltzoff himself proposes, the body schema is an innate system designed for motor control, it seems more appropriate to understand it as a set of pragmatic (action oriented) capabilities embodied in the developing nervous system (see Gallagher *et al*., 1998). In the human infant this system accounts for the possibility of recognizing and imitating other humans.

To the capabilities implicit in neonate imitation we need to add a number of other early interactive capabilities that constitute what Trevarthen (1979) has called 'primary intersubjectivity'. Although these aspects of behaviour are sometimes enlisted in the cause of theory theory (see Baron-Cohen, 1995, p. 55; Gopnik and Meltzoff, 1997, p. 131), it is quite possible to understand them as supporting a more immediate, less theoretical (non-mentalistic) mode of interaction. Baron-Cohen (1995), for example, proposes two mechanisms as necessary, but not sufficient, components of a theory of mind mechanism. The first he terms the 'intentionality detector' (ID). He

considers this to be an innate capability that allows the infant to read 'mental states in behaviour' (p. 32). The ID allows the infant to interpret (notably without the intervention of theory or simulation) bodily movement as goal-directed intentional movement. In effect, the infant is capable of perceiving other persons as agents. On the one hand, this mechanism may not be specific enough to limit the attribution of agency to just humans (see Scholl and Tremoulet, 2000). On the other hand, combined with other capabilities, such as imitation of human gestures and eye-tracking (see below), ID is quickly honed to serve intersubjective interpretation. The interpretation fostered by ID, however, does not require advanced cognitive abilities. It is perceptual and, as Scholl and Tremoulet suggest, 'fast, automatic, irresistible and highly stimulus-driven' (p. 299).

Evidence for early, non-mentalistic interpretation of the intentional actions of others can be found in numerous studies. Baldwin and colleagues have shown that infants at ten to eleven months are able to parse some kinds of continuous action according to intentional boundaries (Baldwin and Baird, 2001; Baldwin *et al*., in press). Eighteen-month-old children can comprehend what another person intends to do. They are able to re-enact to completion the goal-directed behaviour that an observed subject does not complete (Meltzoff, 1995; Meltzoff and Brooks, in press). Quite obviously ID provides an understanding of what an intentional state is; in the first place, another's intentional state is simply the other's action or the state of a perceived body. This understanding is non-mentalistic in the same sense that our understanding of our own intentional actions is non-mentalistic. To be precise, we do not interpret our own action on either an abstract, physiological level ('I am activating a certain group of muscles'), or in terms of a mentalistic performance ('I believe P, so I will do X'). Rather, quite naturally, we understand our own actions on the highest pragmatic level possible (see e.g. Jeannerod, 1997; Gallagher and Marcel, 1999). For example, if, as I reach for a cup, someone asks me what I am doing, I do not say, ordinarily, 'I am reaching for a cup'; rather I say, 'I'm taking a drink'. I tend to understand my actions just at that pragmatic, intentional (goal-oriented) level, ignoring possible sub-personal or lower-level descriptions, and also ignoring ideational or mentalistic interpretations, e.g. 'What are you doing?' 'I'm acting on a belief (desire) that I am thirsty'. Likewise, the interpretation of the actions of others occurs at that same pragmatic (intentional) level. We interpret their actions in terms of their goals and intentions set in contextualized situations, rather than abstractly in terms of either their muscular performance or their beliefs.[5]

[5] Do our interpretations of such actions depend on inference? Baldwin and Baird (2001) argue that inference is required to sort out which one of many possible interpretations is correct. They cite Searle's example.

> If I am going for a walk to Hyde Park, there are any number of things that are happening in the course of my walk, but their descriptions do not describe my intentional actions, because in acting what I am doing depends in large part on what I think I am doing. So for example, I am also moving in the general direction of Patagonia, shaking the hair on my head up and down, wearing out my shoes and moving a lot of air molecules. However, none of these other descriptions seems to get at what is essential about this action, as the action it is. (Searle, 1984, p. 58)

According to Baldwin and Baird, to work out the right interpretation of Searle's action we need much more information about him and human behaviour, and on that basis we proceed to make an inferential judgment about his intentions. But clearly, given the situation, Patagonia, bouncing hair

The second mechanism proposed by Baron-Cohen is what he terms the 'eye-direction detector' (EDD). EDD allows the infant to recognize where another person is looking. Obviously, this mechanism is more specific than ID since it is linked to the perception of eyes and faces. It allows the infant to see (1) that the other person is looking in a certain direction and (2) that the other person sees what she is looking at. Does EDD involve an inference in moving from step (1) to step (2)? Baron-Cohen suggests that an inference is necessary to understand that the other person actually sees what she is looking at. Specifically, he points out that the infant experiences its own vision as contingent on opening versus closing its eyes. His suggestion is more in line with simulation theory: 'from very early on, infants presumably distinguish seeing from not-seeing . . . Although this knowledge is initially based on the infant's own experience, it could be generalized to an Agent by analogy with the Self' (Baron-Cohen, 1995, p. 43). But, one could ask, how does seeing differ from looking? Of course *by virtue of experience* we may come to discover that someone can be looking in a certain direction and not seeing something that is located in that direction. But that would seem to be something that we learn rather than a default mode of EDD. *On the face of it*, that is, at a primary (default) level of experience, there does not seem to be an extra step between looking at something and seeing it.[6]

Baron-Cohen makes it clear that ID and EDD separately or together are sufficient to enable the child to recognize dyadic relations between the other and the self, or between the other and the world. The child can understand that the other person *wants* food or *intends* to open the door, that the other can *see* him (the child) or is *looking at* the door. These are basic intentional relations. Of course children do not simply observe others, they interact with others, and in doing so they develop a further capability which Baron-Cohen terms the 'shared attention mechanism' (SAM). Behaviour representative of joint attention begins to develop around nine to fourteen months. The child alternates between monitoring the gaze of the other and what the other is gazing at, checking to verify that they are continuing to look at the same thing. The child also learns to point at around this same time. Phillips, Baron-Cohen and Rutter (1992) show that infants between nine and eighteen months look to the eyes of the other person to help interpret the meaning of an ambiguous event. In such interactions, well before the development of a theory of mind mechanism, the child looks to the body and the expressive movement of the other to discern the intention of the person or to find the meaning of some object. In this kind of second-person interaction two-year-olds are even capable of recognizing pretend behaviour, for example the mother pretending the banana is a telephone (Leslie, 1994).

---

and shoe-wear as such, and molecular movement, simply do not enter into my interpretation, unless I start making abstract, theoretical inferences. Rather, if I see John Searle walking toward Hyde Park, I'm likely to say, 'There's John Searle out for a walk.' Or, 'That guy is heading for the park.' The other interpretations simply do not come up, unless I start making large and abstract inferences. Since I don't see John Searle every day, I may in fact start to wonder what his further intentions are — is he going to philosophize in the park? But if I were to seriously pursue this question I would have to take action — follow him, stop and ask him, ask someone else who might know, etc. Without such action my inferences would be blind.

[6] See Leslie and Frith (1988). Their discussion of seeing and not seeing in terms of a geometrical-causal line of sight suggests that the default does not involve a distinction between seeing and looking. Baron-Cohen (1995), who carefully provides evidence for the other aspects of EDD, does not provide evidence for there being an inference between looking and seeing.

There are many more intention-signalling behaviours that infants and young children are capable of perceiving. In addition to the eyes, it is likely that various movements of the head, the mouth, the hands, and more general body movements are perceived as meaningful or goal-directed. Such perceptions are important for a non-mentalistic (pre-theoretical) understanding of the intentions and dispositions of other persons as well as for social reinforcement (see review by Allison, Puce and McCarthy, 2000), and they are operative by the end of the first year (Baldwin, 1993; Johnson, 2000; Johnson *et al.*, 1998). In effect, this kind of perception-based understanding is a form of body-reading rather than mind-reading. In seeing the actions and expressive movements of the other person one already sees their meaning; no inference to a hidden set of mental states (beliefs, desires, etc.) is necessary.

There is also evidence for affective and temporal coordination between the gestures and expressions of the infant and those of the other persons with whom they interact. Infants 'vocalize and gesture in a way that seems "tuned" [affectively and temporally] to the vocalizations and gestures of the other person' (Gopnik and Meltzoff, 1997, p. 131). At five to seven months infants are able to detect correspondences between visual and auditory information that specify the expression of emotions (Walker, 1982). Importantly, the perception of emotion in the movement of others is a perception of an embodied comportment, rather than a theory or simulation of an emotional state. Moore, Hobson and Lee (1997) have demonstrated the emotional nature of human movement using actors with point-lights attached to various body joints.[7] Non-autistic subjects view the abstractly outlined but clearly embodied movement of the actors in a darkened room and are able to identify the emotion that is being represented. The emotional states of others are not, in primary experience, mental attributes that we have to infer. One perceives the emotion in the movement and expression of the other's body.[8]

Given the capabilities that are available under the title of primary intersubjectivity, I propose what in relation to theory theory or simulation theory is a revised, and in some sense enhanced or extended *developmental claim*. Before we are in a position to form a theory about or to simulate what the other person believes or desires, we already have specific pre-theoretical knowledge about how people behave in particular contexts. We are able to get this kind of knowledge precisely through the various capabilities that characterize primary intersubjectivity, including imitation, intentionality detection, eye-tracking, the perception of intentional or goal-related movements, and the perception of meaning and emotion in movement and posture. This kind of knowledge, which is the 'massively hermeneutic' background required for the more conceptual accomplishments of mentalistic interpretation, derives from embodied practices in second-person interactions with others. As a result, before we are in a position to theorize, simulate, explain or predict mental states in others, we

---

[7] As early as five months of age infants show preferential attentiveness to human shape and movement in such displays (Bertenthal, Proffitt and Cutting, 1984). The subjects in Moore, Hobson and Lee (1997) were older children classified as normal, autistic and non-autistic mentally retarded. The results demonstrated that the autistic children had relatively more difficulty in recognizing (or simply failed to recognize) emotional attitudes.

[8] Hobson (1993) provides a strong argument along this line. He cites Merleau-Ponty (1994) who notes the 'simple fact that I live in the facial expressions of the other, as I feel him living in mine' (p. 146). Also see Cole (1998, 1999) on the importance of the face in such contexts.

are already in a position to interact with and to understand others in terms of their gestures, intentions and emotions, and in terms of what they see, what they do or pretend to do with objects, and how they act toward ourselves and others.

I also want to argue for the following *pragmatic claim*. Primary, embodied intersubjectivity is not primary simply in developmental terms. Rather it remains primary across all face-to-face intersubjective experiences, and it subtends the occasional and secondary intersubjective practices of explaining or predicting what other people believe, desire or intend in the practice of their own minds.

What Can Phenomenology Show?

There are significant differences between theory theorists and simulation theorists, as well as between nativist and non-nativist accounts of theory theory. There are also disagreements among both simulationists and theory theorists on the question of implicit versus explicit processes. I do not mean to simply brush over these differences. They will motivate a variety of qualifications on the points that I will outline here. The main qualification is that all of the following critical points do not apply to every representative of these richly diverse positions. Notwithstanding this qualification, the following points do apply to a large part of the literature on theory of mind.

A common and basic assumption implicit to theory of mind accounts is that to know another person is to know that person's mind, and this means to know their beliefs, desires or intentional states. I will refer to this as the *mentalistic supposition*. Furthermore, theory of mind suggests that we use our knowledge of another person's mind to explain or predict the other person's behaviour. Since we have no direct access to another person's intentional states, we either postulate what their beliefs or desires are on the basis of a set of causal-explanatory laws (theory theory) or we project the results of certain simulation routines. There is no requirement that such theorizing or simulating be conscious or explicit. We may learn to engage in such interpretation to the point that it becomes habitual and transparent.

The mentalistic supposition implies that an explicit recognition of another person's beliefs, desires or intentional states is clearly conceptual; and that an implicit recognition is informed by such conceptual knowledge. One requires a concept of belief or desire before one can attribute such things to another person. This conceptual recognition involves an element of abstractness. To discover a belief as an intentional state even in myself requires that I take up a second-order reflective stance and recognize that my cognitive action can be classified as a belief. Indeed, to explicitly recognize that I myself 'have a mind' is already something of a theoretical postulate. This is not to deny that I might have something like a direct access to my own experience, or that this experience can be characterized as self-conscious. I can easily say, for example, 'I feel very good about planning my trip.' But to say that this experience of feeling good is in fact a *feeling*, and that this feeling depends on a *belief* that I will actually take the trip, requires something like a reflective detachment from my phenomenal experience, and the positing of a feeling (or belief) as a feeling (as a belief). It would involve a further postulation that such feelings and beliefs are in some fashion part of what it means to have a mind. This kind of metacognitive theorizing is always possible for the adult human, but for the most part I would suggest that, in practice, this is

not the way we think of ourselves — unless we are practising philosophical meditations of the sort Descartes practised.

Perhaps the fact that we have something like a direct access to our own experience does not require that to know our own mind we take a conceptual, abstract, theoretical attitude toward our own experience. Rather, for theory theory, the idea (the pragmatic claim) is that to understand the other person, to whom we have no direct access, we take just such a theoretical attitude. In order to understand that the other person feels very good about planning her trip, I can only hypothesize that she has a certain set of feelings and beliefs that normally go along with a situation like that. One's theory depends upon and is complicated, however, by what one knows of such situations. Some people do not have good feelings about planning trips; they actually get stressed out. Sometimes they may even say 'I don't believe that I am actually going!' Clearly if I am to take a theoretical stance toward what the other person is experiencing, I need to interpret her behaviour on the basis of what I see and hear, and on the basis of what I know of such things. What I know of such things, however, is not easily summarized. Part of what I know includes the kind of pre-theoretical knowledge that I get through capacities that belong to primary intersubjectivity, as described above. If I were to formalize a rule that guided my theoretical stance, it would probably include aspects of pre-theoretical knowledge. Consider the following formulation. 'When someone is planning a trip and she says something like "I don't believe that I am actually going," with intonations that signal exasperation, she really means that she does believe that she is going and she is not enjoying the planning process.' An exasperated intonation, however, is something that I learn about at the level of primary intersubjectivity.

Do we react to the exasperation in a person's voice by appealing (implicitly or explicitly) to a theory? It seems possible to describe it in this way in cases where the situation is not typical, or when, perhaps, the behaviour of the other person is out of character or out of context, or when we don't know the person, or in cases where we are talking with someone else about a third person. When we do not know the person we may need to run through certain possibilities and perhaps engage in a process of interpretation from a distance, much as a historian might attempt to understand a historical figure — forming a hypothesis on the basis of evidence.[9] Even in cases where we know (or think we know) a person very well, we may express puzzlement about their behaviour. In discussing a friend's behaviour with someone who doesn't know her as well, we may come to devise a theory about why she is acting in a certain way. It seems very possible to describe such cases in terms of a theory of mind. Is this a good description of our ordinary interactions with others?

Simulation theory claims that it is not. It is not clear that we represent, explicitly or implicitly, the sorts of rules (causal–explanatory laws) that would summarize what we know of human situations and that would operate as the bases for a theoretical understanding of the other person. Indeed, we find it difficult even to formulate such rules, and this seems odd if we actually use them all the time (Goldman, 1989). Furthermore, at least on the developmental version of theory theory, there is no way to account for the fact that children as young as three or four years putatively develop

[9] Davies and Stone (1998) consider certain limitations of historical analysis based on simulation, citing Collingwood's claim that historical understanding can be achieved by the re-enactment of the historical character's thought.

the very same theory (a common folk psychology), when theory formation in general usually leads to a diversification of theory (Carruthers, 1996; Goldman, 1989).

Do we, then, simulate the other person's belief? Again, this process itself may remain implicit, with only an awareness of the resulting prediction. The process itself, nonetheless, is structured as an internal, representational simulation (Gordon, 1986). The simulation model is closer than theory theory to what I described above as an embodied practice of primary intersubjectivity. It involves something more like a practiced skill than a theoretical stance. Indeed, there is some suggestion that the result of simulation is not so much a mental model of the other's mind, but a motor adjustment in my own system that allows me insight into the other person's behaviour (Gordon, unpublished MS, cited in Stich and Nichols, 1992; Grezes and Decety, 2001). On the other hand, various descriptions of simulation invoke the idea of predicting behaviour on the basis of hypothetical beliefs and desires that are fed into a cognitive decision-making system (see Carruthers (1996) for a description of an approach that combines theory and simulation along this line). The result of this process is to project or attribute relevant intentional states to the mind of the other person. Like theory theory, simulation theory understands the other person as a collection of such mental states, and often understands the simulation itself as a mental state.

In the situation of talking with someone else about a third person, it seems possible to describe our attitude toward the person under discussion as theoretical or as involving a simulation of the other person's mental states. But does the same description capture the dynamics of our interaction with our interlocutor? That is, in a second-person conversational situation, although we may indeed tacitly follow certain rules of conversation, our process of interpretation does not seem to involve a detached or abstract, third-person quest for causal explanation. Nor does it seem to be a theory-driven interpretation that takes the other person's words as evidence for a mental state standing behind what he has just said. Even if we are trying to read between the lines and we reach the conclusion that the person we are conversing with believes the wrong thing concerning the other person, our understanding of this is poorly described as resulting from formulating a theoretical hypothesis or running a simulation routine about what he believes. We do not posit a theoretical entity called a belief and attribute it to him. We do not interact with him by conceiving of his mind as a set of *cogitationes* closed up in immanence (Merleau-Ponty, 1962, p. 353).

Both theory theory and simulation theory conceive of communicative interaction between two people as a process that takes place between two Cartesian minds. It assumes that one's understanding involves a retreat into a realm of *theoria* or *simulacra*, into a set of internal mental operations that come to be expressed (externalized) in speech, gesture or interaction. If, in contrast, we think of communicative interaction as being accomplished in the very action of communication, in the speech, gesture and interaction itself,[10] then the idea that the understanding of another person involves an attempt to theorize about an unseen belief, or to mind-read, is problematic.

---

[10] Here I follow Merleau-Ponty (1962), who conceives of thought as being accomplished in speech. In contrast, a leading theorist of mind, Baron-Cohen (1995), endorses a traditional Augustinian view of language: 'language functions principally as a "printout" of the contents of the mind' (p. 29). It follows that 'in decoding speech we go way beyond the words we hear or read, to hypothesize about the speaker's mental states' (p. 27).

This phenomenologically based criticism is subject to an objection that is typically raised at this point. Is an appeal to phenomenology in this context justified? Theory theorists and simulation theorists often claim that the employment of a theory of mind or simulation routine is unconscious and that what we experience or seemingly experience is not a good guide for what is really going on in such cases (e.g. Goldman and Gallese, 2000). On this account we should think of the theory or simulation routine as somehow programmed into the very structure of our experience of others. If that is the case and our engagement in a theory or simulation procedure is not always explicit or conscious does this mean that our phenomenology is simply wrong? Or does it mean rather that what we capture in phenomenological reflection is something else?

In principle, phenomenology would not be able to say whether a subpersonal cognitive routine is operative; but it would be able to say whether my normal experience of the other person is best characterized as *explanation* and *prediction*, the kind of interpretations that both theory theory and simulation theory posit. I suggest that what phenomenology tells us is that explanation and prediction are specialized and relatively rare modes of understanding others, and that something like evaluative understanding about what someone means or about how I should respond in any particular situation best characterize most of our interactions. The kind of phenomenology I have in mind here is close to a Heideggerian existential phenomenology. It tells us that our primary and usual way of being in the world is pragmatic (characterized by action, involvement and interaction based on environmental and contextual factors), rather than mentalistic or conceptual (characterized as explanation or prediction based on mental contents).[11]

Both theory theory and simulation theory construe our encounter with other people in terms of explaining or predicting the other's beliefs, desires and behaviours. Phenomenology cannot tell us whether our response to the exasperation in a person's voice involves an implicit (sub-conscious) theory or pretend belief. But a careful and methodical phenomenology[12] should be able to tell us whether, when we hear the exasperated voice, our usual response involves formulating an explanation or predicting what the person will do next? Our encounters with others are in fact not normally occasions for theorizing or simulating if such nonconscious procedures are cashed out phenomenologically as explaining or predicting on the basis of postulated mental states. Rather, pragmatic interaction and evaluative understanding take up most of our effort. Only when second-person pragmatic interactions or our evaluative

[11] Heidegger's famous description of the carpenter's hammer is a relevant example. For the most part the carpenter's experience of the hammer is a pragmatic one. She hammers without thinking of the hammer as an object, using it rather as an extension of her body. Her relationship to it is not theoretical or conceptual, but fully caught up in a complex set of pragmatic activities. Only when something goes wrong with the hammering, or when the hammer breaks, does she regard the hammer as an object and as something to be explained. A theory of the hammer is experientially secondary to its use (Heidegger, 1968). The suggestion is not that one's relation to another person is equivalent to one's relation to equipment but that, as in the case of the hammer, one's relation to others is not primarily theoretical or conceptual, but is first of all an interactive one.

[12] In contrast to non-methodical introspection. This qualification is meant to head off the standard reply that introspective reports are notoriously suspect guides to what subjects are doing even at the conscious level, since they are infected (as it were) by what one of my referee's called 'local politics (currently popular psychological views, tried and true folk-notions, and so forth and so on)'. In the method of a phenomenological reduction of the sort practised by Husserl, care is taken to systematically eliminate such prejudices.

attempts to understand break down do we resort to the more specialized practices of third-person explanation and prediction.

The distinction between explanation and evaluation is an important one to make in this context.[13] In our everyday and ordinary encounters we rarely look for causal-mentalistic explanations for people's actions. Rather than being folk psychological, in the sense of involving a folk theory, our encounters are primarily occasions for interactions and evaluations. My action, or the action of another, may be motivated in part by the fact that the situation is just such that this is the action that is called for. In such cases, an action is not caused by a well-formed mental state, but is motivated by some aspect of the situation, as I experience and evaluate it.

One way to understand what I mean by evaluation is to reframe a distinction made by Perner (1991) in his explication of theory theory. He distinguishes between 'situation theory', employed by three-year-olds prior to attaining a theory of mind, and 'representational theory' or theory of mind. According to Perner, three-year-olds employ some aspect of the environment plus some understanding of desire, but are unable to comprehend the concept of the other's belief. One should note, however, that the environment, or the situation, is not something that the child, or the adult, objectively confronts as an outside observer. The notion of situation should be understood to include the experiencing subject (that is, oneself) and the action of that subject. Our involvement in a situation is not as a third-person observer developing a situation theory, as if we were not part of the situation ourselves. Our interaction with another human being is not equivalent to a detached observation (or explanation) of what that person is doing. The notion of evaluation signifies an embedded cognitive practice that relies on those pre-theoretical embodied capabilities that three-year-olds have already developed to understand intersubjective situations. Even to the extent that evaluation becomes reflective, it is more like an embedded reflection on possible actions (Gallagher and Marcel, 1999) than a detached consideration of mental states. Rather than drawing up a theory about a particular situation, or taking an objective, observational stance toward the other person, we have the capacity for measuring it up in pragmatic terms. This capacity does not disappear when the child reaches the age of four, but is rather enhanced by further experience.[14]

Consider the following example that Baron-Cohen (1995, p. 28) cites from Pinker (1994):

---

[13] See (Jacobson, 2000). What Jacobson says of eliminative materialism and more generally of functionalism, can equally be said of theory theory and simulation theory: 'each take the *defining role* of folk-psychological terms to be in causal discourse while it is instead in significant part in evaluative discourse'. On the relation between theory and explanation, see Schwitzgebel (1999a).

[14] Perner (1991) goes on to suggest that theory of mind doesn't actually replace situation theory. It simply amends it to cover problem cases. Even as adults 'we stay situation theorists at heart. We resort to a representational theory only when we need to.' Barresi and Moore (1996) also argue that more primary processes of social understanding are not replaced by the more mentalistic ones, but that the more primary ones continue to function. I disagree with Gordon (1995b) who, in a gloss on Perner, suggests that what passes as situation theory in adult behaviour is really a sophistication in simulating and attributing beliefs and intentions which becomes manifest only when there is a problem. The sophistication of our simulation abilities, he contends, simply makes it seem as if we are not simulating. Gordon does suggest that prior to the development of simulation abilities, the mental, in some sense, 'is already "out there" in the environment, though not yet conceptualized as mental'. My point is that a good part of the mental does not end up hidden away. It becomes embedded in our embodied and communicative practices.

Woman: I'm leaving you.

Man: Who is he?

Overhearing this bit of discussion, the task, according to Baron-Cohen, is to explain why the man utters this phrase. The explanation: 'the man must have thought [formed a belief] that the woman was leaving him for another man'. A certain thought or belief causes the man to say what he says. What causes the thought? Perhaps some cognitive schema that associates this scenario with the influence of a third party. If indeed an explanation is needed this may be a good folk-psychological one, but the question to start with is whether, upon overhearing this bit of conversation, we would be motivated *to explain it* rather than to comprehend it in an evaluative way. From our perspective, as interlopers who are listening in, the thought expressed in the man's words does not have the status of a belief in his head; the thought is already given to us in the words and we have no need to posit a belief over and above them. Would we not already have a pre-theoretical understanding of what was meant, and instead of formulating an explanation would we not be taking some stance or action — choosing up sides or perhaps moving as far away as we could to give the couple privacy? In reality, the man himself may have no such discrete belief. He may have blurted out the question as a question that had never before dawned on him, because he saw something like shame or defiance in the woman's eyes.

Theory of mind conceptualizes beliefs and other intentional states as discretely representational. There are good reasons, however, to view beliefs as dispositions that are sometimes ambiguous even from the perspective of the believer. To have a belief is not to have an all-or-nothing mental representation, but to have some more or less complete set of dispositions to act and to experience in certain ways. Dispositions are actualized, not only in overt behaviour, including verbal behaviour, but also in phenomenal experience.[15] Thus, given a particular context, one may have a disposition to feel upset or to perceive things as grating, depending on a variety of circumstances. For our understanding of other people, I am suggesting that we rarely need to go beyond contextualized overt behaviours (actions, gestures, speech-acts, etc.). We are rarely required to postulate an idealized and abstract mental belief standing behind these behaviours in order to grasp the disposition that is overtly constituted and expressed in the contextualized behaviour. In certain contextualized interaction I need go no further than the person's gestures or emotional expressions to gain my understanding of how it is with that person.

Even if explaining and predicting another person's intentional states and behaviours are structured as theories or simulations, a more basic question is whether our ordinary attempts to understand other people are best characterized as explanations and predictions. Those who defend theory of mind might reply that even if our relations with others phenomenologically *seem* to be pragmatically interactive, they are, in fact, implicitly matters of theorizing or simulating. Even if we are aware of only direct evaluative responses, such responses may be the result of busy sub-personal

[15] This view, a 'phenomenal, dispositional account of belief' (Schwitzgebel, in press), clearly does not involve a reductionist type of behaviourism, as one finds in the usual interpretation of Ryle (1949). Schwitzgebel's excellent account, framed in a purely analytic exposition, is quite consistent with phenomenological accounts found in theorists like Merleau-Ponty. For its implications in the developmental context see Schwitzgebel (1999b)

mechanisms that have the structure of theory or simulation. In this case, controlled experimentation (rather than phenomenology) is the only way to investigate such cognitive mechanisms. Thus, we clearly need to examine the scientific evidence in support of this claim.

The Science of Other Minds

Both theory theory and simulation theory claim the support of good science. Theory theory appeals to classic false-belief tests in developmental psychology for its justification. Simulation theory has recently received support from neuroscience. If one is going to challenge either of these approaches, it is important to consider the scientific evidence and to indicate whether the challenge puts the scientific evidence into question, or whether a new theory would be consistent with the established evidence. I can not review all of the scientific evidence for either of these approaches here, but I will look at a representative sampling and try to indicate certain limitations in the empirical data consistent with my remarks in the previous sections.

1. False-belief experiments

In the 'standard' false-belief task a subject is asked about the thoughts and actions of another person or character who lacks certain information that the subject has. For example, the subject knows that a candy box actually contains pencils. Someone else (this could be a puppet or a real person) enters the room. The question that is posed to the subject is 'What will the other person say is in the candy box?'. Four-year-olds generally answer correctly that the other person will think that there are candies in the box. Three-year-olds are unable to see that the other person may falsely believe that there are candies in the box. So three-year-olds answer that the other person will say there are pencils in the box (see e.g. Perner, Leekam and Wimmer, 1987). False belief tests can be made more or less complicated.

In a series of experiments often sited in support of theory theory, Heinz Wimmer and Josef Perner (1983) investigated a subject's competence in representing another person's belief when that belief differs from what the subject knows to be true. In four experiments children between the ages of three and nine were divided into three groups: three- to four-year olds, four- to six-year olds, and six- to nine-year olds. Each child was told stories that involved, first, a cooperative situation and then a competitive situation. For example, a kid named Maxi puts a piece of chocolate in a blue cupboard and then goes out to play. While he is gone, and without his knowledge, the chocolate is moved into a green cupboard. In the cooperative version of the story Maxi, upon returning, cooperates with another character in obtaining the chocolate. In the competitive version Maxi is in competition with an antagonist. All stories are told up to the point where the main characters look for the hidden object. At this time, each subject is asked to indicate (a) where the chocolate actually was located (the reality question), (b) where Maxi would look for the chocolate (the belief question) and (c) where Maxi would tell the other character to look.

All age groups were able to answer the reality question correctly. Answers to the other questions generally varied in relation to the age of the subjects. When asked where Maxi would look for the object (the belief question) most of the four- to five-year-olds chose the green cupboard incorrectly. However, most of the six- to

nine-year-olds chose the blue cupboard, correctly, despite the fact that the object was really in the green cupboard. When asked, in the competitive version, where Maxi would say the object was hidden, most of the subjects who answered correctly on the belief question were able to create a deceitful utterance required for the competitive versions of the stories. These subjects understood that Maxi would deceive his competitor purposely. Most of the same subjects were also able to create a truthful utterance for the cooperative versions of the stories.

Why were the youngest subjects unable to correctly ascribe a wrong belief to Maxi? A second experiment was designed to answer this question. The same stories were used as in the previous experiment, but with several modifications. A memory question (Do you remember where Maxi put the chocolate?) was asked when the subject answered incorrectly to the belief question. Also, subjects were reminded of what Maxi did before he went outside before being asked the belief question. The results showed an improvement of the five- to six-year-olds in their responses to the belief question. The three- to four-year-olds were unable to correctly ascribe a wrong belief even with the modifications.

Wimmer and Perner concluded from these and several other experiments that children age six and above are able to cope with representational complexities. Four- to six-year-old children have the ability to represent wrong beliefs, but are sensitive to modifications in the task. Few in the three- to four-year-old group are able to represent wrong beliefs or another person's absence of knowledge. Most children who are able to represent wrong beliefs are also able to construct deceitful utterances. Children between the ages of four and six are able to demonstrate inferential skills.

These experiments, and many others based on the same experimental paradigm (see e.g. Baron-Cohen, Leslie and Frith, 1985) are often cited as evidence for the development of a theory of mind at around four years of age. As Stich and Nichols (1992) point out, however, theory theory, as well as simulation theory, are compatible with but do not necessarily entail the Maxi experiments (see Gordon, 1995b). So these experiments cannot be used to support one approach over another. Indeed, a number of authors argue that subjects who fail false-belief tests do not necessarily fail them because they lack a theory of mind. It may be that the intellectual processing involved in the testing is simply too complicated.[16] Furthermore, the false-belief paradigm does not capture all there is to say about children's abilities to understand others. Bloom and German (2000), who generally support a theory approach, cite various aspects of primary intersubjectivity as already providing such capabilities prior to age four. They conclude, rightly, that the false-belief test is 'an ingenious, but very difficult task that taps one aspect of people's understanding of the minds of others' (p. B30).

---

[16] Leslie and Thaiss (1992) show that when photographs are used to represent mental states four-year-olds do worse than their performance on the standard false-belief tests. If it were a matter of picturing mental states as representations, the four-year-old should do equally well on the photograph test (see Leslie, 2000). Three-year-olds fail both the photograph tests (in which false beliefs are not at stake) and the false-belief tests, suggesting not that children have problems with beliefs *per se*, but with the complexity of the problems (Bloom and German, 2000). Furthermore, Siegal and Beattie (1991) and Surian and Leslie (1999) have shown that three-year-olds are capable of passing false-belief tests if the wording of the questions is modified. This suggests that 'normally developing children's performance on false-belief problems is limited by processing resources rather than by inability to represent belief states in others' (Leslie, 2000, p. 1242). Bloom and German (2000) and Barresi and Moore (1996) present similar arguments.

The fact that these experiments are designed to test one aspect of how people understand the minds of others is both their strength and their weakness. The experiments clearly show that something new happens at age four, and that what happens is somewhat consistent with certain assumptions that are shared by both theory theory and simulation theory. The experiments are designed to test whether children at certain ages have acquired an ability to explain or predict the behaviour of others. But, as indicated above, explaining and predicting are very specialized cognitive abilities, and do not capture the fuller picture of how we understand other people.[17]

Two other important limitations of false-belief tests in relation to theory of mind should be pointed out. First, subjects are asked to predict the behaviour of others with whom they are *not* interacting. Based on a third-person observation, the child is asked to predict what the other person will do. Can the results of these experiments be used to characterize second-person ('I–you') interaction?[18] If second-person interaction is the primary and ordinary way of encountering the other person, can we be certain that results based on third-person observation can truly characterize our understanding of others? It is interesting to note that in the three-year-old subject's second-person interaction with the experimenter, the subject does not seem to have difficulty understanding the experimenter in the way that she seems to misunderstand the third person about whom she is asked. It is not at all clear that how we interact with another person directly in a second-person relationship can be captured by activities in the category of third-person observation.

Second, false-belief experiments, like the one conducted by Wimmer and Perner, are designed to test a conscious, metarepresentational process. That is, in such experiments, the subjects are not only provided with the task of explaining or predicting, but they are asked to perform these tasks consciously, and in a reflective manner. In contrast, many theorists claim that theory of mind mechanisms are sub-personal, operating below the level of consciousness. In effect, the experimental design simply does not address the issue of how theory of mind mechanisms function non-consciously.

There are thus at least three factors that limit the conclusions that can be drawn from such experiments for theory of mind, and especially for the pragmatic claim that theory of mind characterizes all of our interpersonal interactions.

(1) The experiments explicitly call for the specialized cognitive activities of explaining and predicting.

(2) The experiments involve third-person perspectives rather than second-person interactions.

(3) The experiments involve conscious processes and do not address theory of mind mechanisms that operate non-consciously.

---

[17] Stich and Nichols (1992) suggest, concerning these experiments, 'the explanation of the data offered by the experimenters is one that presupposes the correctness of the theory-theory'. One could further suggest that the kinds of questions that are asked, and the kinds of answers that are sought in these experiments, are framed by theory of mind's contention that explanation and prediction are primary ways of interpreting other's minds.

[18] For more on the concept of second-person interaction, and its irreducibility to first-person and/or third-person perspectives, see Gomez (1996) and Reddy (1996), as well as the previous section.

It might seem that the following experiment could address the second limitation. In Wimmer, Hogrefe and Sodian (1988), two children face each other and each answers questions about what they know or about what the other child knows concerning the contents of a box into which one of them has looked. Children of three and four years of age answer correctly about their own knowledge, but incorrectly about the other child's knowledge, even when they see the other child has looked into the box. Although this seems closer to second-person interaction, the children are not really interacting on the cognitive level that is being tested. That is, questions are posed by the experimenter (with whom the children are interacting), but they call for third-person explanation or prediction of the other person with whom they are not interacting.

A theory theory interpretation of this experiment is that these children use different mental processes to assess what they themselves know as opposed to what the other child knows. To answer about their own knowledge the children use an 'answer check procedure'. 'They simply check to see whether they have an answer to the embedded question in their knowledge base, and if they do they respond affirmatively' (Stich and Nichols, 1992). According to this account they do not know that they know the contents of the box until they find a belief or piece of knowledge in their own cognitive system. To say that they know what is in the box, it would not be enough to have looked inside the box; they would also have to look inside their own minds. They have to 'check' with themselves in something like a metarepresentative introspection (Leslie, 1988).

It seems more likely, and much more parsimonious, however, that their answer about what they know is based simply on looking inside the box rather than looking inside their own mind. The child looks inside the box and is then asked whether she knows what is in the box. Her positive answer is based on the fact that she just saw what was inside the box, rather than on an introspective discovery of a belief about the contents of the box (see Gordon, 1995b). Her knowledge, one might say, is already in her action. If a subject is asked 'Do you believe that p?' the subject does not start searching in her mind for the *belief that p*. Rather, she straightforwardly considers whether p is or is not the case (see Evans, 1982). In cases when the child does not know what is in the box, her failure to acknowledge that another child who has looked inside the box does know would be surprising only to someone who would expect her to think theoretically, in terms of intentional states abstracted from her own actions. What is not surprising, however, is that the subject has no problem understanding the question put to her by the experimenter with whom she is interacting. Nor is there any indication that she is surprised by the possibility that someone else may or may not have knowledge.

Children aged four to five years have progressed to the point of having the ability to tell correctly what another child who has seen the transfer of a piece of candy from one box to another knows about the contents of the second box. In this part of the experiment, however, both children (the subject and the other) have seen the transfer together. One could still say that their knowledge is in their action. But the same age group fails to understand that in certain circumstances the other child, without visual knowledge, might know the same fact by inference. Again, this would be surprising only if the subject understood the other child in terms of having abstract mental states. The same experiments show that a six-year-old child is capable of precisely this realization and has thus attained some advanced part of a theory of mind. Yet to show that

a child attains a theory of mind at some specific point in development, such that they can consciously explain or predict what someone with whom they are not interacting knows, is not to demonstrate that the child's primary understanding of others is based on theory of mind capabilities. These same children, we would assume, were able to play together and communicate prior to learning that knowledge and beliefs can be caused by inference as well as by direct perceptual access.

2. Mirror neurons

A different sort of scientific evidence has recently been cited in support of simulation theory, namely the proposal that the specific operations of mirror neurons can contribute to a simulation model of how we understand others. Mirror neurons, located in the premotor cortex (area F5) of the macaque monkey and, as evidence suggests, in the premotor cortex and Broca's area in the human (see Fadiga *et al*., 1995; Rizzolatti *et al*., 1996; Grafton *et al*., 1996), respond *both* when a particular motor action is performed by the subject *and* when the subject observes the same goal-directed action performed by another individual. Mirror neurons thus constitute an intermodal link between the visual perception of action or dynamic expression and the first-person, *intra*subjective, proprioceptive sense of one's own capabilities.

Simulation theorists suggest that mirror neurons help us to translate our visual perception of the other person's behaviour into a mental plan of that behaviour in ourselves, thus enabling an explanation or prediction of the other person's thoughts or actions. Mirror neurons facilitate the creation of pretend ('off-line') actions (motor images) that correspond to the visually perceived actions of others (Gallese and Goldman, 1998). Mirror neurons, of course, are part of the motor system, so that the 'plan' that is generated is a motoric one. This, it is argued, at least prefigures (or is a primitive kind of) mental simulation, and as such it supports simulation theory rather than theory theory. 'The point is that [mirror neuron] activity is not mere theoretical inference. It creates in the observer a state that matches that of the target [person]' (Gallese and Goldman, 1998, p. 498).

This approach addresses some of the limitations found in the false-belief experiments. First, the activation of mirror neurons can be thought to be most appropriately the result of specific second-person interactions, although they also operate in third-person perspectives on how others interact.[19] Second, studies of mirror neurons are clearly studies of non-conscious, automatic processes that may or may not be experienced at a conscious level, although they surely shape conscious behaviour. Nonetheless, the process described as prefiguring a more mature simulation routine is described in a fashion similar to the theory theory approach, as resulting in the specialized cognitive activities of explaining, predicting and 'retrodicting'. Indeed, only by describing the activity as involving a representational 'plan' (Goldman and Gallese (2000) reject the idea of a non-representational intentionality) can simulation

[19] In experimental situations, of course, third-person perspectives are often employed. That is, the observation of the other person is conducted in a detached rather than interactive setting. This difference is usually ignored. For example, Ruby and Decety (2001) use the term 'third-person simulation' to signify the motor simulation of another person's action (in contrast to 'first-person simulation' of one's own action), without considering whether interactive observation might be different from detached observation, or for that matter whether the simulation of another's action could itself take the form of egocentric simulation (that is, I simulate the other's action as if it were my own) or allocentric simulation (I simulate the other's action as if it were her action performed where she is).

theorists claim that mirror neuron activity prefigures the more developed representational processes involved in explaining and predicting.

The implication of this representationalist view is that the understanding of the other's behaviour is mediated by a model of ourselves. Goldman and Gallese (2000) suggest that mirror neurons rely on an 'internal representation of goals, emotions, body states and the like to map the same states in other individuals' (p. 256). On the simulation account it would not be enough to see another person's actions and for them to register in the mirror system; the activation of the neurons must generate an extra copy of the actions as they would be if they were the perceiver's own actions. We then read off the meaning of the other, not from *her* actions, but from the internal simulation of *our own* 'as if' actions. This view suggests that in this regard, the subject who understands the other person is not interacting with the other person so much as interacting with an internally simulated model of himself pretending to be the other person. In effect, in contrast to the eclipse of second-person interaction by third-person observation in false-belief tests, here second-person interaction is reduced to a first-person internal activity.

Not only is this interpretation not phenomenologically parsimonious, it is also not clear that the neurological picture supports it. Neuronal patterns (representations) responsible for either implicit or explicit action simulation are in large part the same neuronal patterns that are activated in the case of observing action and in performing action (Grezes and Decety, 2001). In an experimental situation I may be asked to execute an action, simulate an action or observe an action performed by someone else. There is significant overlap between action execution, simulation and observation in the supplementary motor area (SMA), the dorsal premotor cortex, the supramarginal gyrus and the superior parietal lobe. Mental simulation is, in addition, associated with activation in the ventral premotor cortex, which may indicate a linguistic contribution. Observation of action is associated with additional activation in the temporal pathway, consistent with visual processing. Grezes and Decety suggest that other non-overlapping areas may be responsible for distinguishing our own agency from the agency of others (see Ruby and Decety, 2001). There is, however, following the observation of another person's action, no evidence for a secondary activation of the overlapping areas that would count as an internal copy (simulation) over and above the original activation generated by the observation. In other words, if I observe another person perform action X, then there is activation in the relevant brain areas that corresponds to the observation. There is no evidence that there is something like a second activation of those same areas that would correspond to an internal copy or simulation of action X. The neurological underpinnings of what could count as the simulation are part and parcel of the activation that corresponds to the original observation. In effect, perception of action is already an understanding of the action; there is no extra step involved that could count as a simulation routine.

On this view, mirror neurons are not primarily the mediators of simulation (although they may play an important role in simulation, which is always a possibility for the subject), but of direct intersubjective perception and direct action. In principle there is no reason to think that mirror neurons do not function at birth.[20] If they do,

[20] The fact that at birth mirror neurons may be unmyelinated would not prevent them from functioning. The lack of myelination would simply slow their activation.

they may play a role in neonate imitation. To imitate a facial gesture that it sees, however, the infant has no need to simulate the gesture internally. It is already simulating it on its own face. Its own body is already in communication with the other's body at a perceptual level.

Conclusion

Some of the empirical evidence generally cited in support of theory of mind reflects an underlying theoretical bias shared by both theory theorists and simulation theorists. Namely that a normal understanding of others amounts to the explanation and prediction of their behaviour by ascribing to them specific mental states. Given a different theoretical conception of how we understand others, that is by employing capabilities of primary intersubjectivity, new experiments may be designed and old ones may be reinterpreted in ways that would offer important qualifications to theory of mind.

In regard to the developmental claim, I have argued that the picture is more complicated than that presented in theory of mind approaches, and that an embodied practice of mind begins much earlier than the onset of theory of mind capabilities. That this is an embodied practice, and that in the capabilities that characterize primary intersubjectivity the intentions and emotions of other persons are perceptually interpreted in movements, gestures, postures, facial expressions and contextualized behaviours — such facts go directly against the mentalistic supposition that guides theory theory and simulation theory. Developing a sophisticated understanding of others depends, first and foremost, on building the capacity for the embodied practices that come to be manifested in everyday encounters. Capacities for the simulated and theoretical understanding of others (a more specialized set of cognitive abilities) depend on the development of these more basic practices.

In regard to the pragmatic claim sometimes made for theory of mind, I have argued that understanding others in everyday life does not usually involve either taking a theoretical stance or deploying a simulation routine. It depends instead on a capacity for embodied practice that begins early (and is likely to be partially innate) and continues throughout normal (non-pathological) experience. Thus, in contrast to the strong pragmatic claim for theory of mind, namely that it is our primary and pervasive means for understanding other persons, I would substitute a strong pragmatic claim for primary intersubjectivity. It is not just primary in developmental terms: it continues to characterize most of our interpersonal interactions, and it forms the basis for the more specialized mentalistic interpretations of how others perform in the practice of their own minds.

Postscript on Autism

A specific developmental delay in the theory of mind mechanism has been an important element in recent explanations of autism. Autistic children demonstrate impairment of certain social abilities. Specifically, autistic children show inadequate development in the mentalistic understanding of others. Proponents of the theory of mind approach link these social impairments to delayed development of the cognitive abilities associated with the theory of mind mechanism. Experiments in support of this view are based on the standard false-belief tasks, comparing the performance of normal and Down's Syndrome children to the performance of autistic children. In such tests, children are asked to judge or predict what other people (or puppets) in a

story believe or how they will act when one of the characters has a false belief. The results are quite dramatic. Baron-Cohen (1989) shows that autistic children, more advanced in mental age than normal and Down's Syndrome children who pass the test, are unable to recognize the significance of false belief. Leslie and Frith (1988) suggest that autistic children are specifically impaired in their capacity for metarepresentation, and this in turn impedes their formulation of a theory of mind. To the extent that metarepresentation is also necessary for pretence, this view is also consistent with impairments in pretend play in autistic subjects.

Metarepresentation involves taking a view on oneself as if upon another person, and on some accounts it develops only as an internalization of an already established social interaction. On this view, however, with respect to autism, the etiological order is not clear. Rather than understanding a deficit in metarepresentation as the cause of problems in social interaction, it seems just as feasible to understand a deficit in metarepresentation as the result of more primary problems in social interaction. Furthermore, there is good evidence to suggest that in autism the deficiency in social interaction is not confined to cognitive dimensions. In some limited respects the autistic's cognitive understanding of others can be at age level. For example, the autistic child may be able to say correctly that the other person does not know that a sought-for object is in a particular location. In spite of that understanding, the same child will predict that the person in question will look for it there — an incorrect response to the false-belief task. Leslie and Frith (1988) explain this as based on an independence between understanding that the other has limited knowledge and the understanding of false belief — in effect, a difference between knowing two different cognitive states. Might it not also be explained as a difference between knowing that the other person has limited knowledge (a cognitive state) and knowing how the other person will act? The action will require a certain kind of movement of the other's body and it may be just that which confuses the prediction.

There is evidence to suggest that across emotional and perceptual dimensions the autistic child does not understand the embodied behaviour of the other person in the same way that a normal child would. Autistic children, for example, have difficulties in perceiving the bodily expression of emotion in others (Moore, Hobson and Lee, 1997) and in imitating certain stylistic aspects of actions performed by others, especially those stylistic aspects indicative of emotional state. They also have problems in understanding the other person as a self-oriented agent (Hobson and Lee, 1999). Some autistic children attempt to perform the imitative action on the experimenter's body rather than on their own, and thus demonstrate a sensory-motor confusion between egocentric and allocentric spatial frameworks.[21]

---

[21] This is a tentative conclusion based on reviewing videotape of the Hobson and Lee experiments. The autistic child does not represent his own body in the action of the other. This would also interfere with any attempt at simulation. In such cases it is as if the autistic child's mirror neurons are not working properly (see Gallagher, 2001). Also, Ohta (1987) notes a pattern of 'partial imitation' of manual gestures in a significant proportion of autistic subjects. For example, subjects positioned face-to-face with the model produced gestures that reversed the orientation of the hands. Barresi and Moore (1996) suggest that such problems can be caused by a failure of intermodal integration of first-person (proprioceptive) information and third-person (visual) information. In the failed imitation, third-person, visual information, predominates. As a result the autistic person fails to attain the capacity for shared intentional experience normally evident in infants at the end of their first year.

Rutter and Bailey (1993) object to theory of mind explanations based on the fact that autism appears at the end of the first year of life, that is, prior to the normal developmental timeframe for theory of mind. Baron-Cohen's (2000) response to this objection, whilst admitting that the pre-theory of mind aspects of primary intersubjectivity are already amiss in autistic children, interprets such primary intersubjective practices as 'infancy precursors to theory of mind' (p. 1251). If we view embodied practices of primary intersubjectivity as necessary conditions rather than as precursors to theory of mind, then the objections of Rutter and Bailey remain cogent. Autistic problems involving various aspects of social interaction, including emotional and motor-sensory aspects, as well as the developmentally later cognitive aspects, are likely to be the result of earlier disruptions in primary intersubjectivity.

References

Allison, T., Puce, Q. and McCarthy, G. (2000), 'Social perception from visual cues: Role of the STS region', *Trends in Cognitive Science*, **4** (7), pp. 267–78.

Astington, J.W. and Jenkins, J.M. (1999), 'A longitudinal study of the relation between language and theory-of-mind development', *Developmental Psychology*, **35**, pp. 1311–20.

Baldwin, D.A. (1993), 'Infants' ability to consult the speaker for clues to word reference', *Journal of Child Language*, **20**, pp. 395–418.

Baldwin, D.A. and Baird, J.A. (2001), 'Discerning intentions in dynamic human action', *Trends in Cognitive Science*, **5** (4), pp. 171–8.

Baldwin, D.A. *et al*. (in press), 'Infants parse dynamic action', *Child Development*.

Baron-Cohen, S. (1989), 'The autistic child's theory of mind: A case of specific developmental delay', *Journal of Child Psychology and Psychiatry*, **30**, pp. 285–98.

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

Baron-Cohen, S. (2000), 'The cognitive neuroscience of autism: Evolutionary approaches', in *The New Cognitive Neurosciences* (2nd edn.), ed. M.S. Gazzaniga (Cambridge, MA: MIT Press), pp. 1249–57.

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

Barresi, J. and Moore, C. (1996), 'Intentional relations and social understanding', *Behavioral and Brain Sciences*, **19** (1), pp. 107–54.

Bermúdez, J. (1996), 'The moral significance of birth', *Ethics*, **106**, pp. 378–403.

Bertenthal, B.I., Proffitt, D.R. and Cutting, J.E. (1984), 'Infant sensitivity to figural coherence in biomechanical motions', *Journal of Experimental Child Psychology*, **37**, pp. 213–30.

Bloom, P. and German, T.P. (2000), 'Two reasons to abandon the false belief task as a test of theory of mind', *Cognition*, **77**, pp. B25–B31.

Bruner, J. and Kalmar, D.A. (1998), 'Narrative and metanarrative in the construction of self', in *Self-Awareness: Its Nature and Development*, ed. M. Ferrari and R.J. Sternberg (New York: Guilford Press), pp. 308–31.

Carruthers, P. (1996), 'Simulation and self-knowledge: A defence of theory-theory', in *Theories of Theories of Mind*, ed. P. Carruthers and P.K. Smith (Cambridge: Cambridge University Press), pp. 22–38.

Chandler, M.J. and Carpendale, J.I.M. (1998), 'Inching toward a mature theory of mind', in *Self-Awareness: Its Nature and Development*, ed. M. Ferrari and R.J. Sternberg (New York: The Guilford Press), pp. 148–90.

Changeux, P. and Ricoeur, P. (2000), *What Makes Us Think?*, tr. M.B. DeBevoise (Princeton, NJ: Princeton University Press).

Cole, J. (1998), *About Face* (Cambridge, MA: MIT Press).

Cole, J. (1999), 'On "being faceless": Selfhood and facial embodiment', in *Models of the Self*, ed. S. Gallagher and J. Shear (Exeter: Imprint Academic), pp. 301–18.

Davies, M. and Stone, T. (ed. 1995a), *Folk Psychology: The Theory of Mind Debate* (Oxford: Blackwell Publishers).

Davies, M. and Stone, T. (ed. 1995b), *Mental Simulation: Evaluations and Applications* (Oxford: Blackwell Publishers).

Davies, M. and Stone, T. (1998), 'Folk psychology and mental simulation', in *Contemporary Issues in the Philosophy of Mind*, ed. A. O'Hear, Royal Institute of Philosophy, Supplement 42 (Cambridge: Cambridge University Press).

Evans, G. (1982), *The Varieties of Reference*, ed. J. McDowell (Oxford: Oxford University Press).
Fadiga, L. *et al*. (1995), 'Motor facilitation during action observation: A magnetic stimulation study', *Journal of Neurophysiology*, **73**, pp. 2608–11.
Frith, C.D. and Frith, U. (1999), 'Interacting minds — A biological basis', *Science*, **286**, pp. 1692–5.
Gallagher, S. (1996), 'The moral significance of primitive self-consciousness', *Ethics*, **107**, pp. 129–40.
Gallagher, S. (2001), 'Emotion and intersubjective perception: A speculative account', in *Emotions, Qualia and Consciousness*, ed. A. Kazniak (Cambridge, MA: MIT Press and Naples: Instituto Italiano per gli Studi Filosofici), pp. 95–100.
Gallagher, S., Butterworth, G., Lew, A. and Cole, J. (1998), 'Hand–mouth coordination, congenital absence of limb, and evidence for innate body schemas', *Brain and Cognition*, **38**, pp. 53–65.
Gallagher, S. and Marcel, A.J. (1999), 'The self in contextualized action', *Journal of Consciousness Studies*, **6** (4), pp. 4–30.
Gallagher, S. and Meltzoff, A.N. (1996), 'The earliest sense of self and others: Merleau-Ponty and recent developmental studies', *Philosophical Psychology*, **9**, pp. 213–36.
Gallese, V. and Goldman, A.I. (1998), 'Mirror neurons and the simulation theory of mind reading', *Trends in Cognitive Science*, **2**, pp. 493–501.
Goldman, A.I. (1989), 'Interpretation psychologized', *Mind and Language*, **4**, pp. 161–85; reprinted in *Folk Psychology: The Theory of Mind Debate*, ed. M. Davies and T. Stone (Oxford: Blackwell Publishers, 1995).
Goldman, A.I. and Gallese, V. (2000), 'Reply to Schulkin', *Trends in Cognitive Science*, **4** (7), pp. 255–6.
Gomez, J.C. (1996), 'Second person intentional relations and the evolution of social understanding', *Behavioral and Brain Studies*, **19** (1), pp. 129–30.
Gopnik, A. and Meltzoff, A.N. (1997), *Words, Thoughts, and Theories* (Cambridge, MA: MIT Press).
Gordon, R.M. (1986), 'Folk psychology as simulation', *Mind and Language*, **1**, pp. 158–71; reprinted in *Folk Psychology: The Theory of Mind Debate*, ed. M. Davies and T. Stone (Oxford: Blackwell Publishers, 1995).
Gordon, R.M. (1995a), 'Simulation without introspection or inference from Me to You', in *Mental Simulation: Evaluations and Applications*, ed. M. Davies and T. Stone (Oxford: Blackwell Publishers, 1995).
Gordon, R.M. (1995b), 'Developing commonsense psychology: Experimental data and philosophical data', APA Eastern Division Symposium on Children's Theory of Mind, 27 December 1995 (http://www.umsl.edu/~philo/Mind_Seminar/New%20Pages/papers/Gordon/apakids9.htm).
Grafton, S.T. *et al*. (1996), 'Localization of grasp representations in humans by PET: II: Observations compared with imagination', *Experimental Brain Research*, **112**, pp. 103–11.
Grezes, J. and Decety, J. (2001), 'Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis', *Human Brain Mapping*, **12**, pp. 1–19.
Heal, J. (1986), 'Replication and functionalism', in *Language, Mind, and Logic*, ed. J. Butterfield (Cambridge: Cambridge University Press); reprinted in *Folk Psychology: The Theory of Mind Debate*, ed. M. Davies and T. Stone (Oxford: Blackwell Publishers, 1995).
Heal, J. (1998a), 'Co-cognition and off-line simulation: Two ways of understanding the simulation approach', *Mind and Language*, **13**, pp. 477–98.
Heal, J. (1998b), 'Understanding other minds from the inside', in *Current Issues in Philosophy of Mind*, ed. A. O'Hear (Cambridge: Cambridge University Press).
Heidegger, M. (1968), *Being and Time*, trans. J. Macquarrie and E. Robinson (New York: Harper and Row).
Hobson, P. (1993), 'The emotional origins of social understanding', *Philosophical Psychology*, **6**, pp. 227–49.
Hobson, P. and Lee, A. (1999), 'Imitation and identification in autism', *Journal of Child Psychology and Psychiatry*, **40**, pp. 649–59.
Jacobson, A. (2000), 'The soul unto itself: Self-knowledge and a science of the mind', *Arobase: Journal des lettres et sciences humaines*, **4** (1–2), pp. 100–24.
Jeannerod, M. (1997), *The Cognitive Neuroscience of Action* (Oxford: Blackwell Publishers).
Johnson, S.C. (2000), 'The recognition of mentalistic agents in infancy', *Trends in Cognitive Science*, **4**, pp. 22–8.
Johnson, S. *et al*. (1998), 'Whose gaze will infants follow? The elicitation of gaze-following in 12-month-old infants', *Developmental Science*, **1**, pp. 233–8.
Legerstee, M. (1991), 'The role of person and object in eliciting early imitation', *Journal of Experimental Child Psychology*, **51**, pp. 423–33.
Leslie, A. (1988), 'Some implications of pretense for mechanisms underlying the child's theory of mind', in *Developing Theories of Mind*, ed. J. Astington, P. Harris and D. Olson (Cambridge: CUP).
Leslie, A. (1991), 'The theory of mind impairment in autism: Evidence for a modular mechanism of development?', in *Natural Theories of Mind: Evolution, Development and Simulation of Everyday Mindreading*, ed. A. Whiten (Oxford: Blackwell).

Leslie, A. (1994), 'ToMM, ToBy, and Agency: Core architecture and domain specificity', in *Mapping the Mind: Domain Specificity in Cognition and Culture*, ed. L. Hirschfeld and S. Gelman (Cambridge: Cambridge University Press).
Leslie, A. (2000), ' "Theory of mind" as a mechanism of selective attention', in *The New Cognitive Neurosciences*, ed. M. Gazzaniga (Cambridge, MA: MIT Press), pp. 1235–47.
Leslie, A. and Frith, U. (1988), 'Autistic children's understanding of seeing, knowing and believing', *British Journal of Developmental Psychology*, **6**, pp. 315–24.
Leslie, A. and Thaiss, L. (1992), 'Domain specificity in conceptual development: Neuropsychological evidence from autism', *Cognition*, **43**, pp. 225–51.
Meltzoff, A.N. (1995), 'Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children', *Developmental Psychology*, **31**, pp. 838–50.
Meltzoff, A.N. and Brooks, R. (In press), '"Like me" as a building block for understanding other minds: Bodily acts, attention, and intention', in *Intentions and Intentionality: Foundations of Social Cognition*, ed. B.F. Malle *et al*. (Cambridge, MA: MIT Press).
Meltzoff, A. and Moore, M.K. (1977), 'Imitation of facial and manual gestures by human neonates', *Science*, **198**, pp. 75–8.
Meltzoff, A. and Moore, M.K. (1994), 'Imitation, memory, and the representation of persons', *Infant Behavior and Development*, **17**, pp. 83–99.
Meltzoff, A.N. and Prinz, W. (2001), *The Imitative Mind; Development, Evolution, and Brain Bases* (Cambridge: Cambridge University Press).
Merleau-Ponty, M. (1962), *Phenomenology of Perception*, trans. C. Smith (London: Routledge and Kegan Paul).
Merleau-Ponty, M. (1964), *The Primacy of Perception*, trans. W. Cobb (Evanston, IL: Northwestern University Press).
Moore, D.G., Hobson, R.P. and Lee, A. (1997), 'Components of person perception: An investigation with autistic, non-autistic retarded and typically developing children and adolescents', *British Journal of Developmental Psychology*, **15**, pp. 401–23.
Nadel, J. and Butterworth, G. (1999), *Imitation in Infancy* (Cambridge: Cambridge University Press).
Ohta, M. (1987), 'Cognitive disorders of infantile autism: A study employing the WISC, spatial relationship conceptualization, and gesture imitation', *Journal of Autism and Developmental Disorders*, **17**, pp. 45–62.
Perner, J. (1991), *Understanding the Representational Mind* (Cambridge, MA: MIT Press).
Perner, J., Leekam, S.R. and Wimmer, H. (1987), 'Three-year olds' difficulty with false belief: The case for a conceptual deficit', *British Journal of Developmental Psychology*, **5**, pp. 125–37.
Phillips, W., Baron-Cohen, S. and Rutter, M. (1992), 'The role of eye-contact in the detection of goals: Evidence from normal toddlers, and children with autism or mental handicap', *Development and Psychopathology*, **4**, pp. 375–83.
Pinker, S. (1994), *The Language Instinct* (Baltimore: Penguin).
Reddy, V. (1996), 'Omitting the second person in social understanding', *Behavioral and Brain Sciences*, **19** (1), pp. 140–1.
Rizzolatti, G. *et al*. (1996), 'Localization of grasp representations in humans by PET: I. Observation versus execution', *Experimental Brain Research*, **111**, pp. 246–52.
Ruby, P. and Decety, J. (2001), 'Effect of subjective perspective taking during simulation of action: A PET investigation of agency', *Nature Neuroscience*, **4** (5), pp. 546–50.
Rutter, M. and Bailey, A. (1993), 'Thinking and relationships: Mind and brain', in *Understanding Other Minds: Perspectives from Autism*, ed. S. Baron-Cohen and H. Tager-Flusberg (Oxford: Oxford University Press).
Ryle, G. (1949), *The Concept of Mind* (New York: Barnes and Noble).
Scholl, B.J. and Tremoulet, P.D. (2000), 'Perceptual causality and animacy', *Trends in Cognitive Science*, **4** (8), pp. 299–309.
Schwitzgebel, E. (1999a), 'Children's theories and the drive to explain', *Science and Education*, **8**, pp. 457–88.
Schwitzgebel, E. (1999b), 'Gradual belief change in children', *Human Development*, **42**, pp. 283–96.
Schwitzgebel, E. (in press), 'A phenomenal, dispositional account of belief', *Nous*.
Searle, J.R. (1984), *Minds, Brains, and Science* (Cambridge, MA: Harvard University Press).
Siegal, M. and Beattie, K. (1991), 'Where to look for children's knowledge of false beliefs', *Cognition*, **38**, pp. 1–12.
Stich, S. and Nichols, S. (1992), 'Folk psychology: Simulation or tacit theory?', *Mind and Language*, **7**, pp. 35–71
Surian, L. and Leslie, A. (1999), 'Competence and performance in false belief understanding: A comparison of autistic and three-year-old children', *British Journal of Developmental Psychology*, **17**, pp. 141–55.
Tooby, J. and Cosmides, L. (1995), 'Foreword' to S. Baron-Cohen, *Mindblindness: An Essay on Autism and Theory of Mind* (Cambridge, MA: MIT Press), pp. xi–xviii.

Trevarthen, C. (1979), 'Communication and cooperation in early infancy: A description of primary intersubjectivity', in *Before Speech*, ed. M. Bullowa (Cambridge: Cambridge University Press).
Varley, R. and Siegal, M. (2000), 'Evidence for cognition without grammar from causal reasoning and "theory of mind" in an agrammatic aphasic patient', *Current Biology*, **10**, pp. 723–6.
Walker, A.S. (1982), 'Intermodal perception of expressive behaviors by human infants', *Journal of Experimental Child Psychology*, **33**, pp. 514–35.
Wellman, H.M. (1993), 'Early understanding of mind: The normal case', in *Understanding Other Minds: Perspectives from Autism*, ed. S. Baron-Cohen, H. Tager-Flusberg and D.J. Cohen (Oxford: Oxford University Press).
Wimmer, H. and Perner, J. (1983), 'Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception', *Cognition*, **13**, pp. 103–28.
Wimmer, H., Hogrefe, J. and Sodian, B. (1988), 'A second stage in children's conception of mental life: Understanding informational access as origins of knowledge and belief', in *Developing Theories of Mind*, ed. J. Astington, P. Harris and D. Olson (Cambridge: Cambridge University Press), pp. 173–92.