9

The Interactive Practice of Mind

In psychology, philosophy of mind, and recently in the neurosciences, studies of how one person understands and interrelates with another person have been conducted under the heading *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 4. 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 its social environment and gradually learns about people (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 behavior. In the earliest level of development, the 4- to 5-year-old child's theory of mind involves 'first-order 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: 288). Normal children between the ages of 6 and 7 years are able to achieve the second level. The 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. Goldman 1989; Gordon 1986, 1995*a*; and Heal 1986, 1998*a*, *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 behavior. 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 behavior 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 behavior 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 unconscious, 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 approaches 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. In this regard, the experimental and neurological evidence used to support such claims is impressive. I will nonetheless suggest that it is possible to raise questions about certain background assumptions that shape experimental design and that lead to 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.[1] Pragmatic claims may be strong or weak. Some theorists (e.g. Baron-Cohen 1995; Leslie 2000; Tooby and Cosmides 1995) make a very strong pragmatic claim, 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, 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 for theory of mind entails, let me review several of its various formulations. Recently, in an extensive conversation with Paul Ricœur, the neuroscientist Jean-Pierre Changeux proposed that one's relations to others depend on a 'cognitive device' that allows for the representation of the others' 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 Ricœur 2000: 154–7). Two important researchers of this cognitive mechanism, Tooby and Cosmides (1995: p. xvii), 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'. Baron-Cohen (1995: 3) writes: 'it is hard for us to make sense of behavior in any other way than via the mentalistic (or ''intentional'') framework'. And 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). Uta and Christopher Frith (1999) maintain that mental state attribution plays a major role in all social interactions, a conclusion echoed by Alan Leslie (2000). He defines ToMM as a specialized component of social intelligence, but claims that it is necessarily involved 'whenever

[1] 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).

an agent's behavior is attended', for example, 'in conversations and other real-time social interactions' (p. 1236).[2]

In this chapter my intention is not simply to criticize the approaches of theory theory and simulation theory. I will offer an alternative proposal—I'll refer to it as '*interaction theory*'—consistent with the framework developed in previous chapters. The understanding of the other person is *primarily* neither theoretical nor based on an internal simulation. It 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. On this basis, one could justify a weaker pragmatic claim for theory of mind. But neither theoretical nor simulation strategies constitute the primary way in which we relate to, interact with, or understand others. Furthermore, in those cases where we do use theoretical and simulation strategies, these strategies are already shaped by a more primary embodied practice.

What Does Phenomenology Say about Theory of Mind?

I will start with a critical review of the two major theory of mind approaches. 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 simply to brush over these differences. They will motivate a variety of qualifications on the points that I will outline here. These qualifications notwithstanding, and although all the following

[2] Strong claims such as these can be found in too many places to list, but several others are worth noting. Currie and Sterelny (2000: 145) write: 'mind-reading and the capacity to negotiate the social world are not the same thing, but the former seems to be necessary for the latter. . . . our basic grip on the social world depends on our being able to see our fellows as motivated by beliefs and desires we sometimes share and sometimes do not'. Frith and Happé (1999: 2) propose that mind-reading 'appears to be a prerequisite for normal social interaction: in everyday life we make sense of each other's behaviour by appeal to a belief-desire psychology'. Wellman (1993: 31–2) maintains that children at age 4 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'. Karmiloff-Smith (1992: 117) writes: 'young children are spontaneous psychologists. They are interested in how the mind can have thoughts and theories and in how representations mediate between the mind and the world. In order to engage in human interaction . . . to understand their intentions/beliefs/desires, to interpret their statements/gestures/actions . . . each of us relies on . . . a folk theory that enables us to ascribe mental states to ourselves.' And according to Malle (2002, and see 2001): 'Theory of mind arguably underlies all conscious and unconscious cognition of human behavior, thus resembling a system of Kantian categories of social perception—i.e., the concepts by which people grasp social reality.' I thank Matthew Ratcliff for several of these citations.

critical points do not apply to every representative of these richly diverse positions, they 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*.

> *The mentalistic supposition:* The problem of intersubjectivity is precisely the problem of other *minds*. That is, the problem is to explain how we can access the minds of others.

According to this supposition, this is a problem of access because other minds are hidden away, closed in, behind the overt behavior that we can see. This seems to be a Cartesian supposition about the very nature of what we call 'the mind'. The mind is conceived as an inner realm, in contrast to behavior, which is external and observable, and which borrows its intentionality from the mental states that control it. Both theory theory and simulation theory set the problem as one of gaining access to other minds, and their explanations of social cognition are framed in precisely these terms.

Furthermore, theory of mind suggests that we use our knowledge of another person's mind to *explain* or *predict* the other person's behavior. Since, according to theory of mind, 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. For example, Karmiloff-Smith (1992: 138) contends that theory of mind 'involves inferences based on unobservables (mental states, such as belief), a coherent set of explanations and causal links between mental states and behavior which are predictive of future actions'. 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. 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, explicitly to 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 (or 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 practicing philosophical meditations of the sort Descartes practiced.

For some theorists of mind (e.g. Carruthers and Smith 1996; Frith and Happé 1999; Gopnik 1993), even to know our own mind we need to take this conceptual, theoretical attitude toward our own experience, and they discount the idea that we have something like a direct access to our own experience (see Zahavi and Parnas 2003 for a phenomenological critique of this idea). And since we certainly have no direct access to other people, to understand them we must 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. We know that 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'm actually going!' Clearly if I am to take a theoretical stance toward what the other person is experiencing, I need to interpret her behavior 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 some kind of pre-theoretical knowledge that I get through being raised in a social environment. If I were forced to formalize the rule that guided my theoretical stance, it would likely 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 not something that I gain through a theoretical explanation; rather, I know what an exasperated intonation is by means of perceptual experience.

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 behavior 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.[3] Even in cases where we know (or think we know) a person very well, we may express puzzlement about their behavior. In discussing a friend's behavior 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?

[3] 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.

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 3 or 4 years putatively develop the very same theory (a common folk psychology). 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 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 as a motor adjustment in my own system that allows me insight into the other person's behavior (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 behavior 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: 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 action. If, in contrast, we think of communicative interaction as being accomplished in the very action of communication, in the expressive movement of speech, gesture, and the interaction itself,[4] then the idea that the understanding of another person involves an attempt to theorize about an unseen belief, or to 'mind-read', is problematic.

This phenomenologically based criticism is subject to an objection that is often 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 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?

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. Phenomenology tells us that our primary and usual way of being in the world is pragmatic interaction (characterized by action, involvement, and interaction based on environmental and contextual factors), rather than mentalistic or conceptual contemplation (characterized as explanation or prediction based on mental contents).[5]

Phenomenology cannot tell us whether our response to the exasperation in a person's voice involves an implicit (subconscious) theory or simulation routine. But a careful and methodical phenomenology[6] should be able to tell us whether, when

[4] In contrast to someone like Merleau-Ponty (1962), who conceives of thought as being accomplished in speech, Baron-Cohen (1995: 29), defending theory theory, endorses a traditional Augustinian view of language: 'language functions principally as a "printout" of the contents of the mind'. 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).

[5] Heidegger (1968) emphasizes the primacy of our pragmatic interactions with the world. Our primary encounter with things in the world is not as objects to contemplate, but as things that we are already using. Only when something goes wrong do we start to treat them as things that need explanation. For Heidegger and several other phenomenologists like Gurwitsch (1931/1978), our primary encounters with others are in these pragmatic contexts. This is directly related to secondary intersubjectivity, discussed below.

[6] 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 folk theories. A methodological phenomenology

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 non-conscious 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 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. In our everyday and ordinary encounters we rarely look for causal-mentalistic explanations for people's actions. Rather than being occasions for explanation, 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 3-year-olds, prior to attaining a theory of mind, and 'representational theory' or theory of mind. According to Perner, 3-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 certain pre-theoretical embodied capabilities that 3-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 4, but rather is enhanced by further experience.[7]

would include a bracketing of just such folk theories, folk psychology, theories about theory of mind, etc. This is referred to as the phenomenological reduction of the sort practiced by Husserl, and its aim is to attend to experience as it happens.

[7] 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 the more

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

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 is offered: '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. And what causes the thought? Perhaps some cognitive schema that associates this scenario with the influence of a third party (the other man). 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 (and most likely, an emotional overtone) 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 sides or perhaps moving as far away as we could to give the couple some privacy? And 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 behavior, including verbal behavior, but also in phenomenal experience.[8] Thus, given a particular context, one may have

primary processes of social understanding are not replaced by the more mentalistic ones, but that the more primary ones continue to function. Gordon (1995*b*), however, in a gloss on Perner, suggests that what passes as situation theory in adult behavior 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. Yet 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'. I will argue that a good part of the mental stays 'out there' and does not end up hidden away. It becomes embedded in our embodied and communicative practices.

[8] The dispositional theory of belief as found in Ryle and Wittgenstein, is appealing in this context. When dispositions are activated, however, there is something it is like to believe or to act on belief. This view, which Eric Schwitzgebel (2002) calls a 'phenomenal, dispositional account of belief', clearly does not involve a reductionist type of behaviorism, 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 such as Merleau-Ponty. For its implications in the developmental context see Schwitzgebel (1999*a, b*).

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 behaviors (actions, gestures, speech-acts, etc.). We are rarely required to postulate an idealized and abstract mental belief standing behind these behaviors in order to grasp the disposition that is overtly constituted and expressed in the contextualized behavior. In certain contextualized interactions, I need go no further than the person's gestures or emotional expressions to gain my understanding of how it is with that person.

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 subpersonal 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.

To push this idea further, and to make it clear, theorists of mind might claim not only that the process is unconscious but also that the *product* is often unconscious, that is, that we make unconscious predictions and perhaps even unconsciously explain things (if by 'explanation' we just mean discovering causes).[9] If so, the idea of using phenomenology to discover how we interact with others loses its bite. Even if we don't find explanation and prediction in (most of) our phenomenology, we still might be explaining and predicting, theorizing or simulating at an unconscious level.

To this I offer two responses. The second is perhaps the more important, but the first is a matter of vocabulary, terminology, or perhaps description. If in fact, on the non-conscious or subpersonal level, there is some ongoing process that might be described in terms of discovering causes, why call this 'explanation' (likewise, if it is something like a natural inference, why use the term 'theory')? Explanation (or theory) seems to mean (even in our everyday psychology) a process that involves reflective consciousness. The term 'prediction' also seems to me to describe a reflective conscious act (something much more than a pre-reflective anticipation that is part of the structure of experience). It sounds odd, for example, to say that the motor system has developed a theory (or a prediction or an explanation), that allows it to anticipate where a particular movement will end up. 'Explanation' and 'prediction' are personal-level terms.

Terminology aside, however, my second point is about evidence. What evidence could be offered to support the idea that an explanation or prediction is taking place on the subpersonal, non-conscious level? I suppose that to show that some subpersonal process is a process of explanation, one would have to show, at the very least, that this process is a process of inferring causes. Obviously, neurons are firing at a subpersonal level, but how we get from neurons to inference-making is not clear. Although it seems possible to go the other way, that is, in some cases to go

[9] My thanks to Eric Schwitzgebel for raising this point in private correspondence.

from a description of experience to the claim that we are making non-conscious inferences, one would have to provide a description of what that experience is. But if the theorist of mind did offer a description of that experience, their own rejection of the validity of what phenomenology can show would undermine their claim. As a result, everything would depend on the results found in controlled experimentation. 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. I cannot review all 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 section.

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 who doesn't know that the box contains pencils (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. A 3-year-old is unable to see that the other person may falsely believe that there are candies in the box. So 3-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 cited 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 3 and 9 were divided into three groups: 3–4-year-olds, 4–6-year-olds, and 6–9-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 is 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 4–5-year-olds chose the green cupboard incorrectly. However, most of the 6–9-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 5–6-year-olds in their responses to the belief question. The 3–4-year-olds were unable correctly to ascribe a wrong belief even with the modifications.

Wimmer and Perner concluded from these and several other experiments that children age 6 and above are able to cope with representational complexities. children of 4–6 have the ability to represent wrong beliefs, but are sensitive to modifications in the task. Few in the 3–4-year-old group are able to represent false beliefs or another person's absence of knowledge. Most children who are able to represent false beliefs are also able to construct deceitful utterances. Children between the ages of 4 and 6 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 4 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 1995*b*). So these experiments cannot be used to support one approach over another. Others 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.[10] Furthermore, the false-belief paradigm does not

[10] Leslie and Thaiss (1992) show that when photographs are used to represent mental states, 4-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 4-year-old should do equally well on the photograph test

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 presupposed capabilities already developed prior to age 4. 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).

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 4, 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 behavior of others. But, as I suggested above, explaining and predicting are very specialized cognitive abilities, and do not capture the fuller picture of how we understand other people.[11]

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 behavior 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. If *second-person* interaction is the primary and ordinary way of encountering the other person, and I suggest that it is, then we cannot be certain that results based on third-person observation truly characterize our understanding of others. In this respect, it is simply not valid to use the results of these experiments to characterize second-person ('I–you') interaction, or to appeal to this evidence as justification for what I have called the strong pragmatic claim for theory of mind.[12] It is interesting to note that in the 3-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 direct interaction 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

(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 3-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: 1242). Bloom and German (2000) and Barresi and Moore (1996) present similar arguments.

[11] 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.

[12] 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).

experiments, the subjects are not only provided the task of explaining or predicting, but they are asked to perform these tasks consciously, and in a reflective manner. The science of false-belief tests does not provide any evidence for the claim that theory of mind processes are implicit or subpersonal. The experimental design does not address the issue of how theory of mind mechanisms function non-consciously. In this regard the science simply cannot count against the phenomenology.

Thus, there are 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 our interpersonal interactions.

1. The experiments explicitly test 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.

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 one 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 3 and 4 years of age answer correctly about their own knowledge, but incorrectly about the other child's knowledge, even when they see that 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, the 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 is not enough to have looked inside the box; they 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 1995*b*). Her knowledge, one might say, is already in her action. As Gareth Evans (1982) pointed out, 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. When a 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 4–5 years of age 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. 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 6-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.

Mirror Neurons: Who, Where, and When

A different sort of scientific evidence has recently been cited in support of simulation theory, namely, the specific operations of mirror neurons or shared neural representations (see Ch. 8). The proposal is that these operations constitute an internal simulation of another person's behavior. Since mirror neurons or shared representations 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, they 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 behavior into a mental plan of that behavior 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 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: 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.[13] 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 behavior. Nonetheless, the process thought to prefigure a more mature simulation routine is still 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 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 my own first-person model mediates my understanding of another person's behavior. Goldman and Gallese (2000: 256) 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'. Consider two possible interpretations of this idea. One interpretation is that the internal representation is literally a re-presentation, or copy, of the originally perceived action. In other words, two things are happening: I *perceive* an action, and I *simulate* it in terms of my own possibilities. In this case, simulation is secondary to and distinguishable from perception. On the second interpretation, since mirror neurons involve extremely good examples of intermodal perception (translating vision into proprioceptive and body schematic registrations), the most parsimonious simulationist account is that when we perceive another person's actions, that perception registers in the mirror system as already a first-person model of what such actions would be if they were the perceiver's own actions.[14]

[13] Third-person perspectives are often employed in experimental situations. 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). See Gallagher (2003*a*) for discussion.

[14] Frederique de Vignemont (2004) has recently proposed a third possibility. Mirror neurons may encode intentional action by default in strictly neutral terms—neither in first- nor third-person perspectives. Whether the action registers as such, it then requires specification as to whether it is my action or someone else's action. This is accomplished by non-overlapping brain areas connected with the 'Who' system—e.g. the right inferior parietal cortex for third-person perspective, and the anterior insula for

On either interpretation, however, the subject seemingly reads off the meaning of the other, not directly from the other's actions, but from the internal simulation of *the subject's own* 'as if' actions. This view suggests that 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, at the level of explaining or predicting the other person's behavior, it is also not clear that the neurological picture supports it. We may be able to see this by considering some of the brain-scan experiments that were done in connection with the 'Who' system and shared representations discussed in Ch. 8 (Georgieff and Jeannerod 1998, and various other studies). These studies use the terminology of simulation theory. Subjects are asked to simulate their own first-person movements or the movements of others in a third-person perspective. In this case, simulation means mental simulation—an explicit imaginative enaction of the movement—consciously imagining oneself or the other doing the action. The neuronal representations responsible for explicit action simulation are in large part the same neuronal representations that are activated in the case of observing action and in performing action (Blakemore and Decety 2001; Grezes and Decety 2001). In the experimental situation when I am asked to observe or to simulate an action performed by someone else, imaging results show significant overlap for observation and simulation in the supplementary motor area (SMA), the dorsal premotor cortex, the supramarginal gyrus, and the superior parietal lobe.

Of course to distinguish the simulation from the observation, the subject's mental simulation in these experiments cannot happen simultaneously with an occurrent observation of another's action. It seems reasonable to assume that some of the same brain areas that are activated in the conscious imaginative enaction of mental simulation would be the same areas activated in the more covert, subpersonal simulation. In the observation mode the specified brain areas are activated. However, there is no evidence in the observation mode for a secondary activation of the overlapping areas that would count as a subpersonal 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 my perception of that action. In opposition to the first interpretation of simulation theory outlined above, which claims that simulation is an extra step over and above the perception of an action, 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 a simulation are part and parcel of the activation

first-person perspective, as discussed in the previous chapter. For the simulationist this offers the possibility that the simulation is carried out in neutral terms and then applied to the right person.

that corresponds to the original perception. The point is that there is no evidence that perception and simulation are two separate processes. Rather, one could say, in effect, perception of action is already an understanding of the action; there is no extra step involved that could count as a separate simulation routine.

One might object that there is certainly an articulated process involved in the observation of the other's action. There is the visual perception that involves processing in the visual cortex, and there is an additional activation of the shared representation areas in the frontal and parietal cortexes. This may be a definitional problem about where to draw the line between perception and other cognitive processes. Where in the brain does perception end and some other process begin? The 'Where' question, in this sense, is not easily answerable—nor is the 'When' question. Perception is not a momentary event. Even if we conceive it as a very short visual glance, if it attains semantic significance it will take some small amount of time. One might register this in temporal frameworks of working memory or the retentionally-protentionally constituted specious present. It is not clear, however, that we can even ask the question about how long a perception lasts, or how short is the shortest possible perception. Brain processes may be measured in msecs, but perceptual experience involves experienced time.

If simulation is not separate from perception, this may or may not support the second interpretation—namely, that the perception of another person's action registers as already a first-person model of what that action would be if it were the perceiver's own action. But if perception already has a simulation structure, it is not clear why one needs an 'internal' model at all. The required model *is* the action of the other, and it is already being perceived. Why would one need to 'read off' the meaning of an action on an internal 'as if' model, indirectly, when one is observing that very action performed by the other? On this non-simulationist view, mirror neurons and shared representations are not primarily the mediators of simulation but are the enactment of direct intersubjective perception.

There is growing consensus that mirror neurons and the related brain areas that are activated for self-movement and perception of another person's movement play an important role in neonate imitation and the infant's ability to perceive intentions (Blakemore and Decety 2001; Chaminade, Meltzoff, and Decety 2002; and Decety *et al*. 2002; Gallagher 2001*b*). 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 prenoetic and perceptual levels that are sufficient for intersubjective interaction.

Interaction and Intersubjectivity

There is good scientific evidence to support the developmental claim that around the age of 4 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 non-conceptual. I want to suggest that these embodied practices constitute our primary access for understanding others, and continue to do so in large measure 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* embodied interactions with other persons perceived as others.

The basic claim that I will defend is that in most intersubjective situations we have a direct understanding of another person's intentions because their intentions are explicitly expressed in their embodied actions, and mirrored in our own capabilities for action. For the most part this understanding does not require the postulation of 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 behavior. The evidence to support this claim overlaps to some extent with evidence that has been 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 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 behavior 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 person and his/her experience—including an understanding of the other as the subject of intentional action. Prior to the possibility of knowing another person's mind in either a theoretical or simulation mode, one already requires

1. an understanding of what it means to be an experiencing subject;
2. an understanding that certain kinds of entities (but not others) in the environment are indeed such subjects;
3. an understanding that in some ways these entities are similar to and in other ways different from oneself; and
4. a 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, 2002; Nadel and Butterworth 1999).

In contrast, I take these capabilities as clues for an alternative approach to the issue of how we understand other people.

Primary Intersubjectivity

Pre-theoretical (non-conceptual) capabilities for understanding others already exist in very young children. Children, prior to the age of 3, 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 behavior of the newborn. We see evidence for it in instances of neonate imitation. As we have seen (Ch. 3), neonate imitation depends not only on a distinction between self and non-self, and a proprioceptive sense of one's own body, but also on 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 and human bodies, that is, in a way that they do not respond to other objects (see S. C. Johnson 2000; S. C. Johnson *et al.* 1998; Legerstee 1991; Meltzoff 1995). Following Meltzoff and Moore (1977, 1994), we have seen that from birth, actions of the infant and the perceived actions of others are coded in the same 'language', in 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. There is, in this case, a common bodily intentionality that is shared across the perceiving subject and the perceived other. As Gopnik and Meltzoff (1997: 129) indicate, 'we innately map the visually perceived motions of others onto our own kinesthetic sensations'.[15]

Should we interpret this intermodal and intersubjective mapping as a primitive form of theorizing or an 'initial theory' of action? Gopnik and Meltzoff (p. 130) answer in the affirmative. 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'. 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

[15] This idea is reminiscent of Husserl's (1973) analysis of the link between kinesthetic experience and perception. On the neurological level it is also supported by the recent research on mirror neurons and shared activation patterns for self-movement and perception of movement. Petit (1999) points out the relation between Husserl's analysis and the results of research on mirror neurons. Here we seem to have good agreement between phenomenology, neuroscience, and developmental psychology.

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 (a mental simulation) that the infant consults in order to know what to do. Neonates, as we have noted, perfect their imitative actions. They improve the match between their gesture and the perceived gesture. They therefore register the difference between themselves and the other. But to do this 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', as sometimes claimed (Gopnik and Meltzoff 1997: 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. 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 behavior are sometimes enlisted in the cause of theory theory (see Baron-Cohen 1995: 55; Gopnik and Meltzoff 1997: 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 behavior' (p. 32). The ID allows the infant to interpret bodily movement as goal-directed intentional movement. Notably this is possible without the intervention of theory or simulation, which, all evidence indicates, are more sophisticated abilities that develop later. 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 understanding of others fostered by ID, however, does not require advanced cognitive abilities. It is perceptual, and as Scholl and Tremoulet (2000: 299) suggest, 'fast, automatic, irresistible and highly stimulus-driven'.

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 (1995: 43) 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'. But, if we take this situation beyond the simple eyes open/eyes closed contrast, 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. I sometimes look, but don't see. But that is something we learn by experience rather than a default mode of EDD. *On the face of it*, that is, at a primary (default) level of experience, from the perspective of the child who is observing the event, there does not seem to be an extra step between looking at something and seeing it.[16]

There are many more intention-signaling behaviors 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; S. C. Johnson 2000; S. C. 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: 131). At 5–7 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

[16] 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 at this age. Also see Meltzoff and Brooks (2001).

joints.[17] 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. As early as 5 months of age infants show preferential attentiveness to human shape and movement in such displays (Bertenthal, Proffitt, and Cutting 1984). 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 and especially in the face.[18]

Secondary Intersubjectivity

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). Behavior representative of joint attention begins to develop around 9–14 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. This marks the beginnings of what Trevarthen terms 'secondary intersubjectivity'.

Trevarthen shows that around the age of 1 year, infants go beyond the person-to-person immediacy of primary intersubjectivity, and enter into contexts of shared attention—shared situations—in which they learn what things mean and what they are for (see Trevarthen and Hubley 1978). Peter Hobson nicely summarizes this notion of secondary intersubjectivity. 'The defining feature of secondary intersubjectivity is that an object or event can become a focus *between* people. Objects and events can be communicated about. . . . the infant's interactions with another person begin to have reference to the things that surround them' (Hobson 2002: 62). Children do not simply observe others; they are not passive observers. Rather they interact with others, and in doing so they develop further capabilities in the contexts of those interactions.

[17] The subjects in Moore, Hobson, and Lee (1997) were older children classified in three groups 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 (see below for more on autism).

[18] Hobson (1993) provides a strong argument along this line. He cites Merleau-Ponty (1964), who notes the 'simple fact that I live in the facial expressions of the other, as I feel him living in mine' (p. 146). One could also cite Scheler (1954): 'For we certainly believe ourselves to be directly acquainted with another person's joy in his laughter, with his sorrow and pain in his tears, with his shame in his blushing, with his entreaty in his outstretched hands. . . . And with the tenor of this thoughts in the sound of his words. If anyone tells me that this is not ''perception'', for it cannot be so, in view of the fact that a perception is simply a ''complex of physical sensations'' . . . I would beg him to turn aside from such questionable theories and address himself to the phenomenological facts.' Also see Cole (1998, 1999) on the importance of the face in such contexts.

Further 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 10–11 months are able to parse some kinds of continuous action according to intentional boundaries (Baldwin and Baird 2001; Baldwin *et al.* 2001). Children of 15–18 months can comprehend what another person intends to do. They are able to re-enact to completion the goal-directed behavior that an observed subject does not complete (Meltzoff 1995; Meltzoff and Brooks 2001). Quite obviously ID provides an understanding of what an intentional state is; in the first place, however, another's intentional state is not a private mental state, but simply the other's action or the state of a perceived body.

The child also learns to point around this same time. Phillips, Baron-Cohen, and Rutter (1992) show that infants between 9 and 18 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 pragmatic meaning of some object. In this kind of second-person interaction 2-year-olds are even capable of recognizing pretend behavior, for example the mother pretending the banana is a telephone (Leslie 1994).

This understanding is non-mentalistic in the same sense that our understanding of our own intentional actions is non-mentalistic. As I've suggested in previous chapters, when questioned about what we are doing, we do not interpret our own actions (e.g. getting a drink or greeting a friend) on either an abstract, physiological level ('I am activating a certain group of muscles'), or in terms of a mentalistic performance ('I am acting on a belief that I'm thirsty, or that this is a familiar human being capable of communication'). Rather, quite naturally, we understand our own actions on the highest pragmatic level possible. I tend to understand my actions at just that pragmatic, intentional (goal-oriented) level, ignoring possible subpersonal or lower-level descriptions, and also ignoring ideational or mentalistic interpretations. 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.

Do such interpretations, even in the adult, 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 an example proposed by John Searle.

> 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: 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 behavior, and on that

basis we proceed to make an inferential judgment about his intentions. But clearly, given the situation, Patagonia, the simple physical facts of bouncing hair and footwear, and molecular movement simply do not enter into my interpretation, even as possibilities, unless I start to make 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 seriously to 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.

Given the capabilities that are available under the title of primary and secondary intersubjectivity, I propose what is, in relation to theory theory or simulation theory, a revised, and in some sense enhanced or extended *developmental claim*. Before we are in a position to form a theory about someone, 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 and secondary intersubjectivity, including imitation, intentionality detection, eye-tracking, the perception of meaning and emotion in movement and posture, and the understanding of intentional or goal-related movements in pragmatic contexts. This kind of knowledge is the 'massively hermeneutic' background required for the more conceptual accomplishments of mentalistic interpretation. It derives from embodied practices in second-person interactions with others long before we reach the age of theoretical reason. As a result, before we are in a position to explain or predict the behavior of others, to mentalize or mind-read, to theorize or simulate, we 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 in the pragmatically contextualized activities of everyday life.

This expanded developmental claim undermines the strong pragmatic claim made for theory of mind, and supports a strong pragmatic claim for interaction theory. 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 relatively rare intersubjective practices of explaining or predicting what other people believe, desire, or intend in the practice of their own minds.

Autism, Central Coherence, and Interaction Theory

Theory of mind has provided an important framework for understanding autism. It is one of the most widely discussed approaches to explaining the cognitive and

behavioral aspects of this developmental disorder (e.g. Baron-Cohen 1995; Baron-Cohen, Leslie, and Frith 1985; Frith and Happé 1999). Autistic children clearly demonstrate impairment of certain social abilities, and especially inadequate development in the mentalistic understanding of others. Proponents of theory of mind 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 syndrome children to the performance of autistic children. The results are quite dramatic. Baron-Cohen (1989) shows that autistic children, more advanced in mental age than normal, and Down 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 pretense, this view is also consistent with impairments in pretend play in autistic subjects.

This is a bare bones and oversimplified version of the theory of mind account, but I think it captures its essential features. Even ignoring the problems that we have considered in the previous sections, it is generally acknowledged that there are some significant 'internal' problems or limitations in this account of autism. They are internal in the sense that they are problems that appear when one accepts the general terms of the theory of mind account. For example, if theory of mind is to be an account that captures the definitive nature of autism, it is problematic that a significant percentage of autistic individuals are capable of passing false belief and other 'theory of mind' tests. Happé (1995), for example, points out that the range of autistic children who pass such tests varies across different studies from 15 to 60 per cent (Reed and Paterson 1990, and Prior, Dahlstrom, and Squires 1990, respectively). This suggests that some autistic subjects seemingly do possess a theory of mind. Another problem involves the fact that although the theory of mind approach is capable of addressing some of the major cognitive symptoms of autism, especially those involving social cognition and communication, it is unable to explain other symptoms, most of them non-social symptoms, characteristically found in many autistic individuals, for example: restricted range of interest, obsessive concern for sameness, preoccupation with objects or parts of objects, high cognitive ability for rote memory, echolalia, non-semantic form perception, and a variety of sensory and motor behaviors such as oversensitivity to stimuli and repetitious and odd movements (see Happé 1995: 113 ff.).

To the extent that these non-social symptoms of autism show the limits of theory of mind accounts, they also show the limits of interaction theory, or any theory that focuses on just the social aspects, to explain all there is to explain in autism. We need to face up to this fact by developing an account of the social symptoms that is not inconsistent with a broader account that would explain the non-social symptoms. In this concluding section I want to map out a general account in which the problem of specialized cognitive functions related to theory of mind appears at the end of

a long line of effects that are more basic and start with disruptions in some basic sensory-motor processes.

Unfortunately, there is still no consensus about what happens in the brain of the autistic subject. Recent research on apoptosis (the natural pruning of the excess of neuronal cells with which we are born) suggests that the normal timing of this process is disrupted in the autistic brain (see e.g. Courchesne, Carper, and Akshoomoff 2003; Fatemi and Halt 2001; Fatemi *et al.* 2001; Margolis, Chuang, and Post 1994). If that is the case, it is likely that many and diverse neurological problems affecting many different parts of the brain, and different kinds of dynamic processing in the brain, could result. It is not surprising, then, to find abnormalities in the neuronal processes that underlie face recognition (the fusiform gyrus (Pierce *et al.* 2001)), emotion perception (amygdala and limbic system (Bachevalier 2000; Bauman and Kemper 1994)), and many other sensory, motor, and cognitive problems that can result from a variety of brain abnormalities. Since the neurological picture remains unclear, however, we need to look to behavioral indicators to find the first clues about autism.

A variety of basic sensory-motor problems exist in autistic children between ages 3 and 10 years (see Damasio and Maurer 1978; Vilensky *et al.* 1981) and even before that in infants who are later diagnosed as autistic. Teitelbaum *et al.* (1998) studied videos of infants who were diagnosed as autistic around age 3 years. Movement disturbances were observed in all the infants as early as age 4–6 months, and in some from birth. These include problems in lying, righting, sitting, crawling, and walking, as well as abnormal mouth shapes. They involve delayed development, as well as abnormal motor patterns, for example, asymmetries or unusual sequencing in crawling and walking.

Just these kinds of sensory-motor processes have been shown to be important in explaining some basic aspects of social cognition. Here the evidence that a subject's understanding of another person's actions and intentions depends to some extent on a mirrored reverberation in the subject's own motor system is relevant. The neurology of 'shared representations' for intersubjective perception (Georgieff and Jeannerod 1998) suggests that problems with our own motor or body-schematic system could significantly interfere with our capacities for understanding others. Accordingly, it is possible that developmental problems involving sensory-motor processes may have an effect on the capabilities that make up primary intersubjectivity, and therefore the autistic child's ability to understand the actions and intentions of others.[19]

Importantly, however, the disrupted development of these sensory-motor processes may contribute not only to deficiencies in primary intersubjectivity, but are likely to offer some explanation of the other sensory-motor symptoms of autism:

[19] Much more work is needed in this regard. An easy objection to this idea is that there are many individuals with severe sensory-motor problems who do not show autistic symptoms in regard to social development. We could imagine, however, that some specific, early-developed sensory-motor problems may interfere with capacities to interact with others, while other sensory-motor problems may not. So it would be important to find out more about the nature of these problems—more, at least, than the study of videotapes may reveal.

oversensitivity to stimuli, repetitious and odd movements, and, possibly, echolalia. In addition, studies that focus on the perception of emotion show that autistic children do not understand the embodied behavior of other persons in the same way that normal children would. Autistic children have difficulties in perceiving the bodily expression of emotion in others, and in imitating certain stylistic aspects of actions performed by others, especially those stylistic aspects indicative of emotional state (Moore, Hobson, and Lee 1997). They also have problems in understanding the other person as a self-oriented agent (Hobson and Lee 1999). Some autistic children, for example, attempt to perform an imitative action within the spatial framework of the experimenter's body rather than their own, and thus demonstrate a sensory-motor confusion between egocentric and allocentric spatial frameworks.[20]

There are, however, other cognitive problems in autism. Uta Frith (1989) and Francesca Happé (1995) have developed a proposal meant to supplement the theory of mind approach, since the latter leaves many symptoms unexplained. Frith (1989) suggested that autism involves an imbalance in the integration of information, and specifically in integrating parts and wholes. She refers to this as a problem with 'central coherence'. Perception and understanding are normally shaped by gestalt principles. In autism these gestalt principles seem to break down. Happé emphasizes the idea that autistic cognition focuses on parts rather than on the broader contexts that provide meaning for the parts. Autistic subjects thus have difficulty seeing things in their context; they treat them as non-contextualized, in an impoverished or abstract way. Happé cites a clinical example: 'A clinician testing a bright autistic boy presented him with a toy bed, and asked the child to name the parts. The child correctly labeled the bed, mattress and quilt. The clinician then pointed to the pillow and asked, "And what is this?" The boy replied, "It's a piece of ravioli" ' (Happé 1995: 117). Indeed, the pillow did resemble a piece of ravioli, out of context, but ordinarily one would see it as a pillow in the context of the bed.

This problem of central coherence permeates autistic cognition and can generate a variety of symptoms and test results, including what might be regarded as positive effects (unusual talents for remembering word-strings or unrelated items, echoing nonsense, sorting faces by accessories, recognizing faces upside down) and negative effects (unusual weakness for remembering sentences and related items, sorting faces by emotion, recognizing upright faces). Problems with central coherence also affect perceptual experience. In contrast to normal test subjects, for example, autistic children are better able to find embedded figures in complex backgrounds—for them, the background context does not interfere with their search abilities, as it does for non-autistic subjects.

[20] 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*b*). 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.

If we characterize these gestalt problems of central coherence to be problems that involve understanding context, then it is clear that such problems may interfere with the capabilities that make up secondary intersubjectivity—intersubjective capabilities that depend on understanding others and interacting with them *in contexts*—contexts that are pragmatic, but also social. Seeing another person move in a certain way could mean many different things if it is done outside any particular context. Imagine, for example, that you see my right arm, with open hand, drop through the air, but nothing else that would provide the context for what it means, then it could mean many different things. It might be part of a gesture that means 'hello' or 'goodbye'; it might mean 'get out of here'; it might be that I intend to make an important point by bringing my hand down hard on the desk in front of me. Without the context, my intention is simply not clear to anyone who would be watching me, or trying to interact with me.

Problems with central coherence can contribute to an explanation of other non-social problems as well. Specifically, we would expect someone with a central coherence problem to manifest certain non-social symptoms found in autistic subjects: restricted range of interest, obsessive concern for sameness, preoccupation with objects or parts of objects, high cognitive ability for rote memory, and non-semantic form perception.

Theorists of mind might suggest that there may be a connection between central coherence and metarepresentation, so that a deficit that affects central coherence may affect the capacity for metarepresentation, which is seemingly important for attaining a theory of mind. Happé (p. 124) notes, however, that one can find weakness in central coherence even in autistic subjects who pass theory of mind tasks. This loosens the tie between central coherence and theory of mind.

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. With respect to autism, however, 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 the 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. An alternative explanation might be that it is a difference between knowing that the other person has limited knowledge (a cognitive state) and knowing how the other person will act. Understanding action may require just

those kinds of shared sensory-motor representations of the other's body that interaction theory predicts may be problematic in the autistic individual.

To see the difference between a theory of mind approach, supplemented with considerations of central coherence, and a fuller account that includes interaction theory, we can compare two diagrams (Figs. 9.1 and 9.2). Figure 9.1 shows the idea that problems of central coherence may interfere with the functioning of meta-representation, and sketches what Happé terms the 'exciting suggestion' that there may be two possible cognitive deficits that underlie autism. But this simply ignores the evidence for other more basic and non-cognitive problems.

If, instead, we consider the effects that both sensory-motor problems and problems of central coherence may have on primary and secondary intersubjectivity, as well as their connections to the non-social symptoms, we could develop a fuller theory as represented in Fig. 9.2.

Here we see that sensory-motor problems may lead to symptoms that are both social (in primary intersubjectivity) and non-social. Problems with central coherence may also lead to symptoms that are both social (including problems in both primary and secondary subjectivity) and non-social. We may also ask whether there is any connection between sensory-motor problems and the problems with central coherence. In general I have been arguing that there are good reasons to think that body-schematic processes are closely related to perceptual and cognitive abilities, but the precise nature of the autistic sensory-motor problems needs to be studied further before any clear answer can be given in regard to their relations to either central coherence or primary intersubjectivity.

The status of theory of mind in this account of autism is left open. In contrast to an autistic deficiency in theory of mind, as argued by theory of mind proponents, high-functioning autistic individuals may actually employ theorizing strategies as a way to compensate for the loss in the capacities of primary and secondary intersubjectivity. If they are not able to perceive the intentions or emotions in the other person's bodily comportment, they may resort to a purely intellectual mentalizing to develop hypotheses about what motivates others to do what they do. Pursuing this suggestion, Zahavi and Parnas (2003) cite accounts of strategies used by high-functioning autistic individuals. A high-functioning autistic person such as Temple

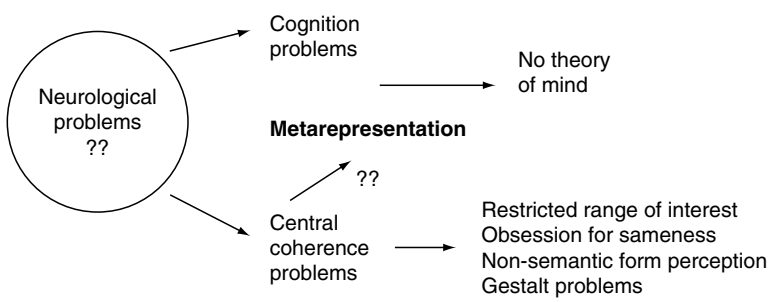

FIG. 9.1. Theory-theory account of autism.

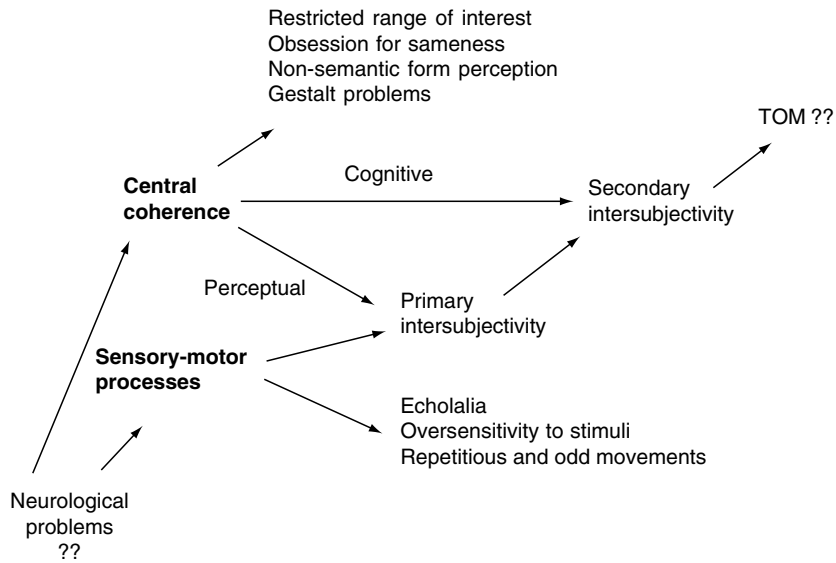

FIG. 9.2. A more comprehensive account of autistic symptoms.

Grandin, for example, uses a variety of strategies to make up for a loss of a natural intersubjectivity. She reads about people, and observes them, in an attempt to arrive at the various principles that would explain and predict their actions in what she describes as 'a strictly logical process'. As Zahavi and Parnas suggest, 'Grandin's compensatory way of understanding others perfectly resembles how *normal* intersubjective understanding is portrayed by the proponents of the theory-theory' (pp. 67–8). She decodes emotional behavior. As Oliver Sacks explains, she lacks an 'implicit knowledge of social conventions and codes'. 'This implicit knowledge, which every normal person accumulates and generates throughout life on the basis of experience and encounters with others, Temple seems to be largely devoid of. Lacking it, she has instead to ''compute'' others' intentions and states of mind, to try to make algorithmic, explicit, what for the rest of us is second nature' (Sacks 1995: 258).[21] Perhaps, however, it is something more basic than a *second* nature, to the extent that primary intersubjectivity may come along for most of us as part of our innate genetic endowment as humans.

[21] Another high-functioning autistic person, Jared Blackburn, puts it this way: 'Those Autistic people who are very intelligent may learn to model other people in a more analytical way, however, as part of adapting to society. For those who are skilled in this, it may become very accurate, and make a few Autistic people seem to have exceptional insight into people. However, even for them there is a social disability, because this accuracy is at a great cost in terms of speed and efficiency, and is maybe virtually useless in practical situations (which involve 'real-time' interaction and fast interpretation and response). Thus, given time I may be able to analyze someone in various ways, and seem to get good results, but may not pick-up on certain aspects of an interaction until I am obsessing over it hours or days later. So in practical situations, I have impaired social cognition, with problematic results, while I may seem to have good insights into people at other times' (Blackburn *et al.* 2000).