Published in D. Zahavi, T. Grünbaum, J. Parnas (eds.): *The structure and development of self-consciousness: Interdisciplinary perspectives.* John Benjamins, Amsterdam, 2004, 35-63. Please quote only published version.

Dan Zahavi
Danish National Research Foundation: Center for Subjectivity Research
University of Copenhagen

The embodied self-awareness of the infant: A challenge to the theory-theory of mind?

The aim of the following contribution is to discuss whether recent findings in developmental psychology, findings concerning infantile self- and other-experience, might challenge a view held by advocates of the theory-theory of mind, namely the view that both self-awareness and intersubjectivity presuppose a theory of mind.

1. Theory of mind

The term "theory of mind" was originally introduced by Premack and Woodruff in a seminal paper on intentionality in primates:

> In saying that an individual has a theory of mind, we mean that the individual imputes mental states to himself and to others (either to conspecifics or to other species as well). A system of inferences of this kind is properly viewed as a theory, first, because such states are not directly observable, and second, because the system can be used to make predictions, specifically about the behavior of other organisms (Premack & Woodruff, 1978: 515).

The phrase "theory of mind" was consequently used as shorthand for our ability to attribute mental states – such as intentions, beliefs, and desires – to self and others and to interpret, predict, and explain behavior in terms of mental states.[1] However, although Premack and Woodruff took it for granted that it was the possession and use of a theory that gave the individual the capacity to attribute mental states, the contemporary debate is split on this issue. On one side, we have the theory-theory of mind, and on the other the simulation theory of mind. The theory-theorists claim that the ability to explain and predict behavior is underpinned by a folk-psychological theory dealing with the structure and functioning of the mind. We attribute beliefs to others by deploying theoretical knowledge. There is, however, disagreement among the theory-theorists about whether the theory in question is innate and modularized (Carruthers, Baron-Cohen), or whether it is acquired in the same way as scientific theories are acquired (Gopnik, Meltzoff). Most claim that there is some innate basis, but as Gopnik points out, it is necessary to distinguish between modularity nativism and starting-state nativism. The *theory-formation* theory, which takes the child to be a little scientist, who is constructing and revising theories in the light of incoming data, can accept a certain nativism, but such initial structures are taken to be defeasible. They can be changed and will be changed by new evidence (Gopnik, 1996: 171). Thus, for the theory-formation theory, there is a striking similarity between the acquisition of scientific knowledge and the child's increasing ability to adopt the intentional stance and mind-read, i.e., his or her ability to interpret behavior in terms of an agent's mental state. The same cognitive

processes are responsible for scientific progress and for the development of a child's understanding of the mind (Gopnik, 1996: 169). In contrast, the *modularists* claim that the core of the folk-psychological theory is hardwired. As they point out, if the theory were merely the product of scientific investigation, why is it culturally universal and why do all children reach the same theory at the same age (Carruthers, 1996a: 23). According to the modularists, the theory is forged by evolution and innately given, and although it might need experience as a trigger, the theory of mind module will not be modified by experience.

Whereas the theory-theorists claim that we employ a theory about the psychological when we predict and explain the behavior of others, the simulationists claim that we possess no such theory, or at least none complete enough to underpin all our competence with psychological notions (Heal, 1996: 75). Whereas the theory-theorists make use of what Gordon has called a cold methodology and argue that our understanding of others chiefly engages intellectual processes, moving by inference from one belief to the other, the simulationists employ a hot methodology and argue that our understanding of others exploits our own motivational and emotional resources (Gordon, 1996: 11). Thus, according to the simulationists, what lies at the root of our mind-reading abilities is not any sort of theory, but rather an ability to project ourselves imaginatively into another person's perspective, simulating his or her mental activity with our own.

In the following, my focus will be on the theory-theory.[2] What precisely do the theory-theorists mean by theory? We have already seen that the view differs. Some take the theory of mind to be a theory in a very literal sense and compare it to a scientific theory.[3] Others regard it in a more extended sense and compare it to a set of rules of symbol manipulation instantiated in an innate module. Some take the theory in question to be explicit, to be something the agent is conscious of, others consider it to be more or less implicit and tacit, and to be something that operates on a subpersonal level.

This issue is of crucial importance not only in order to understand what is actually at stake, but also in order to properly evaluate the debate between the theory-theory and the simulation-theory. Unfortunately, however, not everybody has been careful to spell out what precisely they mean when they say that our understanding of the mental is underpinned by a theory. Generally speaking, however, many theory-theorists have tended to construe theory in a rather loose sense, in order to increase the plausibility of their own position, but the danger they thereby run is to become vulnerable to what is known as the "promiscuity objection" (cf. Blackburn, 1995). In the end, the notion of theory becomes vacuous, and everything turns out to be theoretical, including cooking, gardening, and fishing. In order to avoid this, some theory-theorists have simply bitten the bullet, and have accepted a strong definition of theory that entails much more than simply some kind of semantic holism.

The theory-theory claims that mastery of mental concepts is constituted by knowledge of a psychological theory. More specifically, our understanding of mental notions depends upon our knowledge of the positions that these notions occupy within the theory. Thus, the notions are thought to receive their sense from the theory in which they are embedded, rather than through some ostensive definition or direct acquaintance. This is probably one of the most characteristic features of the theory-theory: It denies that our reference to mental states such as beliefs and desires is based on any direct experience of such mental states, and instead argues that the concepts in question are theoretical postulates that have been developed through a process of abstract theorizing. To put it differently, since the theory-theorists consider the attribution of mental states to be a question of an inference to best explanation and prediction of behavioral data, they have often

taken mental states to be unobservable and theoretically postulated entities. Leslie vividly articulates such a view in the following passage:

> One of the most important powers of the human mind is to conceive of and think about itself and other minds. Because the mental states of others (and indeed of ourselves) are completely hidden from the senses, they can only ever be inferred (Leslie, 1987: 139; cf. Baron-Cohen, 1995: xvii).

2. Theory-theory of self-awareness

According to many theory-theorists (Gopnik, Carruthers, Frith and Happé) – at least if one takes some of their explicit statements at face value – we come to know our own beliefs and occurrent mental states just like we come to know the beliefs and experiences of others. In both cases, the same cognitive mechanism is in use, in both cases we are dealing with a process of mind-reading, in both cases we are dealing with the application of a theory of mind. Thus, according to what might be labeled the theory-theory account of self-awareness, my access to my own mind depends on the same mechanisms that I use in attributing mental states to others. In both cases, the access, the understanding, and the knowledge are theory-mediated, and the mediating theory is the same for self and for other (Carruthers & Smith, 1996: 3; Gopnik, 1993: 3; Frith & Happé, 1999: 7). Even though we seem to perceive our own mental states directly, this direct perception is an illusion.

The theory-theory predicts that there should be no difference in the development of our ability to attribute mental states to self and other, since the same cognitive mechanism is used in both cases. In other words, the individual's ability to mind-read should be equally good (or bad) regardless of whether the tasks concern his own mental states or the mental states of others.[4] The existence of an extensive parallelism would consequently provide empirical support for the theory-theory. Does such a parallelism in fact exist? In order to investigate the matter, a whole battery of tests has been used.[5] Let me focus on the most well known tests: the *false-belief* tasks.

The two most frequently used false-belief tasks are the location change and the content change tasks. The *Sally-Anne task*, a location change task, is set up in the following manner. The child is confronted with two dolls, Sally and Anne. Sally has a box and Anne has a basket. Sally puts a marble into her box and then goes for a walk. While she is away, Anne takes the marble from the box, and puts it into her own basket. Sally then returns. She wants to play with her marble, but where will she look for it? When four-year old children (and older) are confronted with the question, they typically say that she will look inside her box, since that is where she *falsely believes* it to be hidden. Younger children, however, often point to the basket, indicating that they think that Sally will look for the marble where it really is. They apparently fail to understand that other persons' beliefs may be false (Frith & Happé, 1999: 3-4).

In the *Smarties-task*, a content change task, children are shown a candy box. Based on its appearance, children first believe that the box contains sweets, but the box is then opened and is shown to contain pencils. The box is then closed again, and the children are asked what other children, who have not yet seen inside the box, will think it contains. The average 4-year-old answers that other children will think it contains candy, whereas younger children answer pencils. Once again, the result seems to demonstrate that very young children are unable to comprehend that other persons might have false beliefs.

Why is there this interest in children's ability to succeed on false-belief tasks? Because, in order for a child to ascribe false beliefs to others (and to

himself), he must supposedly be able to understand that our beliefs might differ from reality. In order to make sense of Sally's behavior, for instance, the young child has to understand that Sally is acting not on the basis of what is actually the case, but on the basis of a false belief about what is the case. Thus, the young child must be able to understand the difference between reality and our beliefs about reality. She must have beliefs about beliefs; she must be in possession of a theory of mind.

In order to test the existence of a parallelism in the attribution of mental states to self and to other, an ingenious variation of the *Smarties-task* was devised. Children were presented with the deceptive candy box that was full of pencils. They were first asked the above-mentioned questions, and then the following question was added: "When you first saw the box, before we opened it, what did you think was inside it?" Somewhat surprisingly, one-half to two-thirds of the 3-year-olds said that they had originally thought that it contained pencils. They apparently failed to remember their own past false beliefs. Thus, 3-year-old children seem to have as much trouble understanding their own past false beliefs as they have in understanding the false beliefs of others (Gopnik, 1993: 6-8).

According to Gopnik, this finding reveals a striking parallelism between children's understanding of the psychological states of others and their understanding of their own immediate-past psychological states. But this parallelism does not support our commonsense intuition that the process of discovering our own mental states is fundamentally different from the process of discovering someone else's states. According to the commonsense view, there is phenomenological asymmetry between self and other. Whereas we can observe the other's behavior and have to infer his beliefs and desires, we have direct access to our own beliefs and desires and can simply report them. We need not infer their existence; we do not need any theoretical model at all. But for Gopnik, the existence of the parallelism challenges this view. In fact, when children can report and understand the psychological states of others, they can report having had those states themselves, and when they cannot report and understand the psychological states of others, they report that they have not had those states themselves (Gopnik, 1993: 9). In short, there is little evidence that mental states are attributed to self before they are attributed to others, and vice versa.

If our acquisition of beliefs about our own mental states parallels our acquisition of beliefs about the mental states of others, and if the epistemic source is fundamentally the same in both cases, why do we normally tend to believe that there is such a big difference between the two? The explanation offered by both Gopnik and Carruthers is that we have become experts on reading our own minds, and after having reached a certain expertise we tend to see things at once, although what we see is actually the result of a complex theoretical process. We draw on an accumulated theoretical knowledge, but our expertise makes us unaware of the inferential processes and makes us believe that our experience is immediate and non-inferential. In other words, self-knowledge or self-consciousness can be thought of in analogy with the theory-laden perception of theoretical entities in science. Just as a diagnostician can sometimes *see* a cancer in the blur of an x-ray picture, so, too, each of us can sometimes see that we are in a state accorded such-and-such a role by folk-psychological theory (Gopnik, 1993: 11; Carruthers, 1996a: 26; 1996b: 259-260).

Gopnik argues that developmental evidence confirms the existence of a crucial parallelism in the attribution of mental states to others and in the attribution of mental states to self. However, the most significant of the findings presented by Gopnik demonstrates the existence of a parallelism in the attribution of *current* false beliefs to others and in the attribution of *past* false beliefs to self, but it is rather unclear why these findings – puzzling and interesting as they are – should

warrant the kind of sweeping claim made by Gopnik. Apparently, however, Gopnik's idea is that unless you are able to appreciate that it is possible to have mistaken beliefs, you cannot understand what it means to have beliefs or intentional states at all (Gopnik, 1993: 6). It is certainly reasonable to assume that if a child can understand what a false belief is, then she can also understand what a belief is. But is it also reasonable to conclude that unless a child can understand false beliefs she cannot understand beliefs? Certainly, if we are talking about a full-fledged theoretical understanding of beliefs, i.e., of an actual theory of beliefs. Such a theory must involve an understanding and explanation of the possibility of error. If it did not, we would likely say that it was not really a theory of beliefs, or at best, that it was a very inadequate theory of beliefs. However, it is hardly surprising that we have high requirements for what a theory should entail. The question is whether it is appropriate to apply the same strong requirements to a young child, and to claim that the young child doesn't experience himself as having intentional states unless he masters a theory of mind, unless he is capable of attributing false-beliefs to self and others.

In a more recent article, Nichols and Stich (2002) launched a rather damning attack on the theory-theory account of self-awareness. As they pointed out, there are three ways to interpret the theory. First, it could be taken to involve the claim that the only information we have about our own mental states is the kind of evidence that others are also in possession of. In this sense, knowledge of self and knowledge of others would be completely analogous. However, this is a form of pure behaviorism that is hopelessly implausible. Second, the theory could be taken to involve the concession that my access to my own mental states is based on information that is not available in the case of my access to the mental states of others. The problem, however, is that the theory never spells out what precisely this information is. Gopnik refers to first-person psychological experience as "the Cartesian buzz" (Gopnik, 1993: 11), but as Nichols and Stich point out, this is not a very illuminating answer. Finally, the theory-theory account of self-awareness might argue that the additional information that is available in my own case is information about my own mental states. But if this information is available to me from the outset, there is no reason to introduce and involve any theory of mind mechanism (Nichols & Stich, 2002: 12).

I agree with this criticism, but in the following, I will consider a somewhat different challenge to the theory-theory of mind. As we have just seen, the theory-theory claims that self-awareness – in the sense of having access to or being acquainted with one's own mind – is theoretical in nature and that it presupposes a theory of mind. According to the standard view, however, children only gain possession of a theory of mind when they are around 4-years old.[6] It is only at that age that they can pass the classical theory of mind tasks, such as the false-belief task or the appearance-reality task. And as both Baron-Cohen and Frith and Happé have argued, one can test the presence of self-awareness using these classical tests (Frith & Happé, 1999: 5; Baron-Cohen, 1989: 591). As Baron-Cohen for instance puts it, since the ability to understand the appearance-reality distinction involves the ability to attribute mental states to oneself, a failure to pass the task suggests a lack of self-awareness (Baron-Cohen, 1989: 596).[7]

In order fully to understand the theory-theory perspective on self-consciousness, it might be useful to recall that theory-theorists are committed to some version of the higher-order account of consciousness. This commitment is rarely spelled out, but it is crucial to their overall line of argumentation. Carruthers is a theorist who has not only undertaken the trouble of actually spelling out the link between the theory-theory of mind, the higher-order thought theory, and the issues of self-awareness and phenomenal consciousness, but who has also done so with exemplary lucidity and characteristic bluntness. Carruthers takes conscious mental

states, that is, mental states with a distinctive subjective feel to them, mental states that it feels like something to be the subject of, to be mental states of which the subject is aware, and he consequently argues that conscious mental states require self-awareness, or to put it differently, he argues that self-awareness is a conceptually necessary condition for there to be phenomenal consciousness (Carruthers, 1996c: 155). Carruthers considers the self-awareness in question to be a type of higher-order thinking, and he therefore argues that a creature must be able to think about and hence conceptualize its own mental states if these states are to feel like anything to the organism. Thus, to have a phenomenally conscious perception of a surface as green, the creature must entertain the higher-order thought "I am perceiving a green surface." Since mental concepts get their significance from being embedded in a folk-psychological theory of the structure and functioning of the mind, what this ultimately means is that only creatures in possession of a theory of mind are capable of enjoying conscious experiences (Carruthers, 1996c: 158; 2000: 194). As he puts it:

> [I]n order to think about your own thoughts, or your own experiences, you have to possess the *concepts* of thought and experience. And these get their life and significance from being embedded in a folk-psychological theory of the structure and functioning of the mind. So in the case of any creature to whom it is implausible to attribute a theory of mind – and I assume that this includes most animals and young infants – it will be equally implausible to suppose that they engage in conscious thinking. […] If animals (or most animals) lack higher-order thoughts, then by the same token they will lack conscious experiences. For there will be just as little reason to believe that they are capable of thinking about their own experiences, as such. If true, this conclusion may have profound implications for our moral attitudes towards animals and animal suffering (Carruthers, 1996c: 221; Cf. 2000: 194).

Carruthers consequently holds the view that animals (and infants under the age of three) lack phenomenal consciousness, lack a dimension of subjectivity. In his view, they are blind to the existence of their own mental states; there is in fact nothing it is like for them to feel pain or pleasure (Carruthers, 1998: 216; 2000: 203). Carruthers concedes that most of us believe that it must be like something to be a young infant, a cat, or a camel, and that the experiences of these creatures have subjective feels to them, but he considers this common-sense belief to be quite groundless (Carruthers, 1996: 223).[8]

If a theory of mind is required for self-awareness, any creature that lacks such a theory will also lack self-awareness. Is it true, however, that infants lack self-awareness during the first 3-4 years of life? To suggest that an infant only becomes self-aware when he is in possession of a theory of mind, or to mention some other traditional candidates, when he masters the use of the first-person pronoun, or when he is able to recognize himself in the mirror, in my view, is to operate with an unacceptably narrow definition of self-awareness.[9] It is also a suggestion that a number of prominent developmental psychologists have criticized. However, if it could be shown that infants are in possession of self-awareness before they acquire a theory of mind, the theory-theory would be in trouble.

3. Developmental counter-evidence

Let us take a closer look at some of the empirical findings that are discussed in the work of Stern, Neisser, Butterworth, and Rochat. These developmental psychologists argue that the infant is in possession of self-experience from birth, and they all reject the view, originally defended by Piaget, according to which the infant initially lives in a kind of a dualistic fusion where there is as yet no distinction between self, world, and other (Piaget & Inhelder, 1969: 22). Thus, according to this once widely held view, the infant was initially supposed to exist in a "state of undifferentiation, of fusion with mother, in which the 'I' is not yet differentiated from the 'not-I' and in which inside and outside are only gradually coming to be sensed as different" (Mahler, Pine & Bergman, 1975: 44).

If we start with Stern, he argues that theory and language *transform* and *articulate* the infant's experience of self and other; they do not constitute it. Already from birth onward, the infant gains possession of different pre-reflective and pre-linguistic "senses of self". Stern concedes that the sense of self initially available to the infant is basic, but he lists four types of experiences that are present at around 3 months of age. There is *self-agency*, that is, the sense of authorship of one's own actions; there is *self-coherence*, that is, the sense of being an integrated, non-fragmented whole; there is *self-affectivity*, that is, the experience of subjective feelings; and finally there is *self-history*, that is, the having of a sense of endurance, of being in continuity with one's own past (Stern, 1985: 71). These four experiences are all basic types of self-experience and, according to Stern, they are not merely cognitive constructs, but rather lived, existential counterparts to the objectifiable, verbalizable self.

It would lead us too far astray to discuss Stern's analyses of all four types in detail, but let me focus on his account of self-agency or authorship of actions. How does a child distinguish between her own movements/actions and the movements/actions of others, and what enables her to experience her*self* as an agent? Stern distinguishes between two *experiential* invariants: 1) The sense of volition that precedes a motor act, and 2) the proprioceptive feedback that does or does not occur during the act (Stern, 1985: 76). The child typically encounters three different types of action: self-willed action of self, other-willed action of other, and other-willed action of self. Further, the child is able to distinguish between the three precisely because of the presence or absence of invariants 1 and 2. If the experience of the action contains both volition and proprioceptive feedback, we are dealing with a self-willed action of self. If neither is present, we have an other-willed action of other. And if the proprioceptive feedback is present, but the experience of volition is absent (as in the case where the mother is moving the hand of the infant), we have an other-willed action of self.

Just like Stern, Neisser, Butterworth, and Rochat also reject the view that self-awareness has a late developmental onset. In a well-known article from 1988, Neisser distinguished five different selves: the ecological self, the interpersonal self, the extended self, the private self, and the conceptual self (Neisser, 1988: 35). The most basic and primitive of these is the ecological self, that is, the individual understood as an active agent in the immediate environment. When and how are we aware of the ecological self? According to Neisser, this occurs whenever we perceive. Following Gibson, Neisser takes perception to involve information about the relation between the perceiver and the environment. In this sense, all perception involves a kind of self-sensitivity; all perception involves a co-perception of self and of environment (cf. Gibson, 1986: 126). As perceivers, we are embedded and embodied agents. We see with mobile eyes that are set in a head that can turn and that is attached to a body that can move from place to place; in this sense a stationary point of view is only the limiting case of a mobile point of view (Gibson, 1986: 53, 205). But every movement of the perceiver produces a systematic flow pattern in the visual field, which provides us with awareness of our

own movements and postures. Thus, proprioception (or kinaesthesis) is richly intermodal; it is neither attached to a unique sense-organ, nor is it to be identified with a specific body sense. Rather, it is a mechanism of self-sensitivity, common to all perceptual systems. It can be obtained through vision or audition, as well as through the muscles and joints.

Employing the Gibsonian notion of affordance, Neisser writes that any given situation affords some actions and not others. We see at a glance whether objects are within reach, whether doors are wide enough to walk through, or chairs are the right height to sit on. Moreover, this perception is "body-scaled", that is, the distance that matters is not measured in centimeters, but in relation to our own bodily dimensions and capabilities (Neisser, 1993: 8). For instance, a young infant (a few weeks old) can discriminate between objects that are within his reach and objects that are outside his reach. The infant is far less inclined to reach out for an object that is outside his reach. But, of course, for the infant to be able to make this distinction, he must be aware of the position of the object in relation to *himself*. That is, the infant has to be in possession of *self-specifying information.* Even very young infants pick up the information that specifies the ecological self. They respond to the optical flow, discriminate between themselves and other objects, and easily distinguish their own actions and their immediate consequences from events of other kinds. They perceive themselves (among other things), they perceive where they are, how they are moving, what they are doing, and whether a given action is their own or not. These achievements appear already in the first weeks and months of life, and, according to both Butterworth and Neisser, they testify to the existence of a primitive and irreducible form of self-awareness (Neisser, 1993: 4; Butterworth, 2000: 24).

According to Rochat, newborn infants (24 hours old) can discriminate between double touch stimulation combined with proprioception and single touch of exogenous origin. All healthy infants have an innate rooting response. When the corner of an infant's mouth is touched, the infant turns her head and opens her mouth toward the stimulation. By recording the frequency of rooting in response to either external tactile stimulation or tactile self-stimulation, it was discovered that newborns showed rooting responses almost three times more frequently in response to the external stimulus. Rochat thus concludes that even newborns can pick up the intermodal invariants that specify self- versus nonself-stimulation, and that they thereby have the ability to develop an early sense of self (Rochat, 2001: 40-41). Infants are in possession of proprioceptive information from birth and as Rochat argues, proprioception is "the modality of the self par excellence" (Rochat, 2001: 35). Thus, long before they are able to pass any mirror self-recognition tasks, not to speak of any false-belief tasks, infants have a sense of their own bodies as organized and environmentally embedded entities. They have an early sense of their own bodies, and hence an early perceptually-based sense of themselves (Rochat, 2001: 41). Following in the footsteps of Neisser and Gibson, Rochat calls this early sense of self the infant's ecological self (Rochat, 2001: 30-31).

For Rochat, the ecological self is clearly a bodily self, and he argues that the infant's self-experience is initially a matter of the infant's experience of his own embodied self. It is through their early body exploration that infants specify themselves as differentiated agents in the environment, eventually developing an explicit awareness of themselves. More precisely, infants have an inborn inclination to investigate their own bodies. This inclination forms the cradle of self-perception and constitutes the developmental origin of self-knowledge (Rochat, 2001: 29, 39, 74).

Around the age of fifteen to eighteen months, the child becomes able to perform symbolic actions, and it acquires some linguistic competence. That the child becomes able to assume a detached perspective on itself can be seen for

instance from its behavior before a mirror. Prior to this age, the child presumably does not realize that it sees itself in the mirror. If one marks the face of a child with rouge without her knowledge and she subsequently looks in a mirror, a younger child will point to the mirror and not to herself. But after the age of eighteen months, the child will touch the rouge on her own face. Since the confrontation with the mirror motivates a *self-directed* behavior, it is assumed that the child now recognizes what she sees in the mirror as her own reflection (Lewis & Brooks-Gunn, 1979: 33-46; Stern, 1985: 165).

However, although this recognition testifies to the existence of self-awareness, its absence certainly does not imply a lack of self-awareness. Not only is the recognition of one's own reflection by no means a primitive and basic type of self-awareness, on the contrary, we are dealing with a rather sophisticated type of representationally mediated self-identification, where the self-awareness in question takes place across distance and separation. We identify "that other" as ourselves. Moreover, the child would not be able to perform this identification, which presumably takes place through the perfect match between his own bodily movements and the movements of the mirror image, if he were not already aware of his *own* bodily movements. In short, in order to recognize oneself in the mirror, one must already be in possession of bodily self-awareness.

All of these authors are pointing to a dimension of bodily self-experience that is in place long before the infant is capable of solving any theory of mind tasks. Insofar as the theory-theory wants to uphold the view that all self-awareness is theoretically mediated, it is confronted with a serious problem. Let us not forget, however, that the theory-theory of mind defends a double thesis. It is not only claiming that self-awareness is theoretically mediated, it is also claiming that intersubjectivity is theoretically mediated. After all, the whole idea is that any reference to minded beings (be it to oneself or to others) involves a process of mind-reading, involves an application of a theory of mind. Given this situation, it is natural to ask whether the theory-theory treatment of intersubjectivity might also be beset with related empirical and conceptual difficulties.

4. Embodiment and intersubjectivity

Infants are in possession of a form of bodily self-awareness long before they are in possession of a theory of mind, long before the infant is able to pass any theory of mind tasks. Moreover, they are certainly also capable of social interaction at this early stage. Whereas we, in adult life, occasionally make inferential attributions of mental states to other people, such attributions cannot be considered the basis of the smooth and immediate interpersonal interaction – often called primary intersubjectivity – found in young infants (Trevarthen, 1979). In some respects, the period between two and six months might be classified as the most social period in one's life. The social smile is already in place, and the child has a clear preference for perceiving other subjects rather than inanimate objects (Stern, 1985: 63, 72; Spitz, 1983: 98-124). Although an infant initially has very little command over her own locomotion, she has an almost fully developed control over her eye-movements, and can function as a social partner through her gaze. By controlling her own direction of gaze, she can regulate the level and amount of social stimulation. And through gaze behaviors, such as averting her gaze, shutting her eyes, staring past, becoming glassy-eyed, etc., to a large extent she can initiate, maintain, terminate, and avoid social contact (Stern, 1985: 21).

2-3 month-old infants will engage in "protoconversations" with other people by smiling and vocalizing, and will demonstrate a capacity to vary the timing and intensity of communication with their partners. The purpose of this early interaction

seems to be the interaction itself, with the participants affectively resonating to one another (Fivaz et al., 2004). When a mother mirrors the infant's affect, the infant will reciprocate and show sensitivity to the affective mirroring of the mother. In fact, infants clearly expect people to communicate reciprocally with them in face-to-face interactions, and to work actively with them in order to sustain and regulate the interaction. If the mother is asked to remain immobile and unresponsive, the infant will react by ceasing to smile, and will exhibit distress and attempt to regain her participation.

Infants are clearly reacting differently to mere objects and other subjects from the very start. Whereas objects are simply toys to be looked at and manipulated, the faces, voices, and bodily movements of other people are treated as special social parameters (Legerstee, 1999: 217, 220-221). Infants are also able to interpret the bodily movements of others as goal-directed and intentional, in short, they have the capacity to perceive others as agents. And there is nothing inferential about this early social interaction; rather, it is a form of intersubjectivity based on the infant's intuitive grasp of the expressive gestures of other individuals.

Around the age of nine months, a change occurs, insofar as the infant starts to realize that it can share experiences of the world with others. This change in the infant's experience of self and other is evinced from the infant's attempt to share joint attention, intentions, and affective states (Stern, 1985: 128). As Rochat writes:

> Research shows that by nine months infants begin to treat and understand others as "intentional agents", somehow explicitly recognizing that like themselves, people plan and are deliberate in their actions. So, for example, infants will start sharing their attention toward objects with others, looking up toward them to check if they are equally engaged. They will start to refer to other people socially, and in particular to take into consideration the emotional expression of others while planning actions or trying to understand a novel situation in the environment (Rochat, 2001: 185).

Infants of nine months can follow the eye-gaze or pointing finger of another person, and when they do so, they often look back at the person and appear to use the feedback from his or her face to confirm that they have in fact reached the right target. In other words, they seek to validate whether joint *attention* has been achieved. Similarly, they might show objects to others, often looking to the other person's eyes, to check whether he or she is attending. As for the sharing of *intentions*, it is most obvious in protolinguistic requests for help. Such requests suggest that the infant apprehends the other as someone who can comprehend and satisfy her own intentions. Similarly, they might respond to simple verbal requests by others, or shake their heads to express refusal. Thus, intentions have become shareable experiences (Stern, 1985: 129-131). Finally, the sharing of *affection*, or interaffectivity, which is presumably the first and most basic form of subjective sharing, can also be witnessed. If an infant is placed in a situation that is bound to generate uncertainty, for instance, by being approached by a new, unusual, and highly stimulating object, such as a bleeping and flashing toy, he will look toward his mother for her emotional reaction, essentially to see what he should feel in order to help resolve his own uncertainty. If the mother shows pleasure by smiling, the infant will continue his exploration; if she shows fear, the infant will turn back from the object and perhaps become upset (Stern, 1985: 132). A vivid example of this is the famous "visual cliff" experiment. Infants aged 12 months are placed on one side of a "visual cliff", i.e., an apparent sudden drop beneath a transparent surface. On the other side of the cliff, the infant's mother and an

attractive toy are placed. When the infant notices the drop-off, she will typically look spontaneously at her mother's face. If the mother poses a happy face, most infants will cross to the deep side; if the mother poses a fearful expression, the infants will freeze or even actively retreat. It is noteworthy that the mother's mere presence is not enough, rather her emotional reaction, as perceived through her expressions and behavior, has a decisive influence (Hobson, 1991: 47). In other words, the infant appears to recognize that another person's expression has meaning with reference to an environment common to both of them. The gestures and utterances of the caretaker are perceived as being both emotionally expressive and as being directed to something in the infant's world (Hobson, 1993: 38, 140-141). Thus, Hobson concludes that infants

> ...have direct perception of and natural engagement with person-related meanings that are apprehended in the expressions and behaviour of other persons. It is only gradually, and with considerable input from adults, that they eventually come to conceive of 'bodies' on the one hand, and 'minds' on the other. (Hobson, 1993: 117)

Are embodied self-experience and the experience of others linked? Some philosophers have argued that unless self-experience is embodied, intersubjectivity is neither possible nor comprehensible. To put it differently, if we adopt what McCulloch has recently called a behavior-rejecting mentalism (McCulloch, 2003: 94), i.e., if we deny that embodiment and bodily behavior have any essential role to play in experience and cognition, if we deny that embodiment and environmental embedding are essential to having a mind, we will have a hard time escaping solipsism.

What does the argument look like? If my own self-experience, in the first instance, is of a purely mental nature, if my embodiment does not figure in my self-acquaintance from the very start, we need to understand how I will ever be inclined to attribute selfhood to others. Why should I even so much as think that there are other selves? Had subjectivity been an exclusive first-person phenomenon, were it only present in the form of an immediate and unique inwardness, I would only know one case of it – my own – and would have had no reason to ascribe it to others, and to recognize other bodies as embodied subjects. To quote Merleau-Ponty and Davidson:

> If the sole experience of the subject is the one which I gain by coinciding with it, if the mind, by definition, eludes "the outside spectator" and can be recognized only from within, my *cogito* is necessarily unique, and cannot be "shared in" by another. Perhaps we can say that it is "transferable" to others. But then how could such a transfer ever be brought about? What spectacle can ever validly induce me to posit outside myself that mode of existence the whole significance of which demands that it be grasped from within? Unless I learn within myself to recognize the junction of the *for itself* and the *in itself*, none of those mechanisms called other bodies will ever be able to come to life; unless I have an exterior others have no interior. (Merleau-Ponty, 1945: 427-428)

> If the mental states of others are known only through their behavioral and other outward manifestation, while this is not true of our own mental states, why should we think our own mental states are anything like those of others? (Davidson, 2001: 207)

The basic problem is as follows: If my body does not figure essentially in my self-ascription of (some) mental terms and if my ascription of mental terms to others is essentially based on their bodily behavior and expression, what should then guarantee that we are in fact applying the same concepts to ourselves and to others? The different uses of the concepts threaten the unity of their meaning (cf. Avramides, 2001: 135, 224). The proper way to respond to this skeptical challenge is by abandoning the radical divide between the subject's mind and body, and one way to do so is by appealing to the notion of *action*. Action – at least according to one venerable philosophical tradition – joins mind and body, or more precisely, action is prior to the artificial division between mind and body.

It could be argued, of course, that any account of the mind has to take subjectivity and the first-person perspective seriously, and that a focus on behavior and action will consequently lose what is essential to the mind. However, as Avramides points out, this worry is simply misguided. There is nothing reductive in the reference to action, since subjectivity figures centrally in the concept. Action is the action of subjects; it is the action of minded individuals (Avramides, 2001: 286). We must respect the difference between the first-person and the second- and third-person perspectives and we should recognize the difference between self- and other-ascription. But too much focus on this difference or asymmetry can lead to the mistaken view that only my own experiences are given to me, and that the behavior of the other shields his experiences from me and makes their very existence hypothetical (Avramides, 2001: 187).

Merleau-Ponty has been very explicit in linking the issues of embodiment and intersubjectivity. In his view, subjectivity is essentially embodied. To exist embodied is, however, neither to exist as pure subject nor as pure object, but to exist in a way that transcends both alternatives. It does not entail a loss of self-awareness; on the contrary, self-awareness is intrinsically embodied self-awareness, but it does entail a loss or perhaps rather a release from transparency and purity, thereby permitting intersubjectivity. As Merleau-Ponty writes: "The other can be evident to me because I am not transparent for myself, and because my subjectivity draws its body in its wake" (Merleau-Ponty, 1945: 405). To put it differently, since intersubjectivity is a fact, there must exist a bridge between my self-acquaintance and my acquaintance with others; my experience of my own subjectivity must contain an anticipation of the other, must contain the seeds of alterity (Merleau-Ponty, 1945: 400-401, 405, 511). If I am to recognize other bodies as embodied foreign subjects, I have to be in possession of something that will allow me to do so. But as Merleau-Ponty points out, when I experience an other and when I experience myself, there is in fact a common denominator. In both cases, I am dealing with *embodiment*, and one of the features of my embodied subjectivity is that it per definition comprises an *exteriority*. When I go for a walk, or write a letter, or play ball – to use Strawson's examples (Strawson, 1959: 111) – I am experiencing myself, but in a way that anticipates the manner in which I would experience an other, and an other would experience me. This is not to say that a focus on embodiment and action eradicates the difference between self-ascription and other-ascription, between a first-person perspective and a second-person perspective, but it conceives of the difference in such a manner that their relationship becomes more intelligible. Thus, Merleau-Ponty can describe embodied self-awareness as a presentiment of the other and the experience of the other as an echo of one's own bodily constitution. In short, it is because I am not a pure interiority, but an embodied being that lives outside itself, that transcends itself, that I am capable of encountering and understanding others who exist in the same way (Merleau-Ponty, 1960: 213, 215, 221; 1964: 74).

The idea is not to reduce consciousness as such to intentional behavior. Rather, the idea is simply that bodily behavior, expression, and action are essential

to (and not merely contingent vehicles of) some basic forms of consciousness. Mental states do not simply serve to explain behavior; rather some mental states are directly apprehended in the bodily expressions of people whose mental states they are. Or as Hobson also puts it: "We perceive bodies and bodily expressions, but we do so in such a way that we perceive and react to the mental life that those physical forms express" (Hobson, 2002: 248; Cf. 1993: 184). More generally, there seems to be something very problematic about claiming that intersubjective understanding is a two-stage process of which the first stage is the perception of meaningless behavior, and the second an intellectually-based attribution of psychological meaning. On the contrary, in the face-to-face encounter, we are neither confronted with a mere body, nor with a hidden psyche, but with a unified whole. When I see another's face, I *see* it as friendly or angry, etc., that is, the very face expresses these emotions. To quote Wittgenstein:

> We do not see facial contortions and *make the inference* that he is feeling joy, grief, boredom. We describe a face immediately as sad, radiant, bored, even when we are unable to give any other description of the features (Wittgenstein, 1980: § 570).
>
> In general I do not surmise fear in him – I *see* it. I do not feel that I am deducing the probable existence of something inside from something outside; rather it is as if the human face were in a way translucent and that I were seeing it not in reflected light but rather in its own (Wittgenstein, 1980: § 170).

A similar view has been advocated by both Merleau-Ponty and Scheler, who argue that the affective and emotional experiences of others are given for us *in* expressive phenomena. Anger, shame, hate, and love are not only qualities of subjective experience, but also types of behavior or styles of conduct, which are visible from the outside (Merleau-Ponty, 1964: 52-53; Scheler, 1973: 254).

This does not rule out that some mental states are covert, of course, but not all mental states can lack an essential link to behavior, if intersubjectivity is at all to get off the ground.[10]

Our experience and understanding of others are not infallible, but there is a decisive difference between our everyday uncertainty about what precisely others might be thinking about, and the nightmare vision of the solipsist. Although we might be uncertain about the specific beliefs and intentions of others, this uncertainty does not make us question their very existence. In fact, as Merleau-Ponty points out, our relation to others is deeper than any specific uncertainty we might have regarding them (Merleau-Ponty, 1945: 415).

5. Phenomenological misgivings

There are good reasons (philosophical as well as empirical) for maintaining that body-awareness constitutes genuine self-experience. Unfortunately, however, and contrary to expectations, the accounts offered by Rochat, Butterworth, Neisser, and Stern are not always sufficiently clear on this.

In the introduction to his book *The Infant's World*, Rochat suggests that there are three fundamentally different and contrasted classes of experiences: the experience of self, of objects, and of other people (Rochat, 2001: 27). I wholeheartedly agree with this division, which very much fits the received view in phenomenology. Unfortunately, however, Rochat does not really respect his own division. He very soon starts to talk of the body as an object of exploration, and

speaks of self-perception as a question of differentiating one's own body from *other* objects in the environment (Rochat, 2001: 34, 37).

As for Butterworth, he has argued that proprioceptive self-awareness should be distinguished developmentally from higher-order consciousness or reflective self-awareness (where the self is the object of one's own cognition). However, Butterworth still speaks of the ecological self as being the object of one's own perception, and of primary consciousness as the state of being aware of the self as a thing or an object situated in the physical and social environment (Butterworth, 2000: 19-20).

We find the very same take in Neisser, who repeatedly talks of the self as an object (Neisser, 1988: 35, 39, 40). Although Neisser concedes that the ecological self is *per se* not an object of thought, he nevertheless considers it an object of perception (Neisser, 1988: 41, 56).

If we return to Stern's multi-faceted analysis of the infant's self-experience, we come across a similar objectivistic strain. Stern occasionally makes it sound as if an infant's self-experience is a result of her ability to discriminate herself from others, and that this is merely an instance of her general ability to discriminate between different entities. He claims that the infant, far from being a *tabula rasa*, is predesigned to perceive the world in a highly structured fashion. Just as she very early is able to perceive and organize different stimuli into different natural categories, the infant has inborn capabilities that enable her to discriminate different gestalt constellations of stimuli in such a way that she can keep self and other separate. When the infant feels the caress of her mother, hears the voice of her father, and sees her own hand, she is not overwhelmed by a surge of unstructured sensations, but is able to distinguish between herself, her father, and her mother as three distinct entities. She recognizes that the behavior of different persons is differently structured; she distinguishes one agent from another (Stern, 1983: 56-62), and is thereby ultimately able to discriminate the invariant structure that characterizes her own self-generated actions and experiences from the patterns belonging to the movement and actions of particular others (Stern, 1985: 7, 65, 67).

These ways of describing and accounting for self-experience, however, are beset with a major problem. Even if an infant is able to distinguish between different entities in such a way that no confusion takes place, this does not answer the key question: How does the infant sense that one of these experiential configurations is *itself*? The answer given is not satisfactory. Although both Stern and Rochat acknowledge that the infant's (direct and immediate) experience of proprioception and volition is of crucial importance (Rochat, 2001: 89; Stern, 1983: 65), they still make it sound as if self-awareness is a question of discriminating correctly between two types of objects. But this is to commit the mistake of equating self-experience with object-identification, as if the infant were first confronted with certain experiences that he then subsequently succeeded in identifying as his own.

Why is it problematic to conceive of the embodied self as an object, and of embodied self-awareness as a kind of object-awareness? To put it very simply, for something to be given as an object is for it to be given as something that transcends the merely subjective. For something to be given as an object of experience is for it to differ from the subjective experience itself. However, if this is so, if object-awareness always involves a kind of epistemic divide, if object-awareness always entails a distinction between the subject and the object of experience, object-awareness cannot help us understand self-awareness. After all, self-awareness is precisely supposed to acquaint us with our own subjectivity; it is not supposed merely to acquaint us with yet another object of experience. Perhaps it could be objected that there surely are cases where I am confronted with a

certain object, and then recognize that the object in question is in fact myself. This is true of course, but this kind of objectified self-recognition can never constitute the most fundamental form of self-awareness. Why not? Because in order for me to recognize a certain object as myself, I need to hold something true of it that I already know to be true of myself. The only way to avoid an infinite regress is by accepting the existence of a non-objectifying self-acquaintance. To quote Sidney Shoemaker:

> The reason one is not presented to oneself "as an object" in self-awareness is that self-awareness is not perceptual awareness, i.e., is not a sort of awareness in which objects are presented. It is awareness of facts unmediated by awareness of objects. But it is worth noting that if one were aware of oneself as an object in such cases (as one is in fact aware of oneself as an object when one sees oneself in a mirror), this would not help to explain one's self-knowledge. For awareness that the presented object was φ, would not tell one that one was oneself φ, unless one had identified the object as oneself; and one could not do this unless one already had some self-knowledge, namely the knowledge that one is the unique possessor of whatever set of properties of the presented object one took to show it to be oneself. Perceptual self-knowledge presupposes non-perceptual self-knowledge, so not all self-knowledge can be perceptual. (Shoemaker, 1984: 105)

This reasoning holds true even for self-knowledge obtained through introspection. That is, it will not do to claim that introspection is distinguished by the fact that its object has a property that immediately identifies it as being me, since no other self could possibly have it, namely the property of being the private and exclusive object of precisely my introspection. This explanation will not do, since I will be unable to identify an introspected self as myself by the fact that it is introspectively observed by me, unless I know it is the object of *my* introspection, i.e., unless I know that it is in fact *me* that undertakes this introspection. This knowledge cannot itself be based on identification if one is to avoid an infinite regress (Shoemaker, 1968: 561-563).

To recapitulate, the problem with the account offered by Stern, Rochat, and Butterworth is that they conceive of the embodied self as an object, and of embodied self-awareness as a kind of object-awareness. What is the alternative? Perhaps, phenomenological writings on self-awareness and embodiment can get us further.

According to the phenomenologists, self-awareness should be construed very broadly. In contrast to what is claimed by the theory-theory, self-awareness is not something that only comes about the moment I construct a theory about the cause of my own behaviour, a theory that postulates the existence of mental states. Nor is it something that only comes about the moment one scrutinizes one's experiences attentively, not to speak of it being something that only comes about the moment one recognizes one's own mirror image, or refers to oneself using the first-person pronoun, or is in possession of identifying knowledge of one's own life story. Rather, literally all the major figures in phenomenology defend the view that the experiential dimension is as such characterized by a tacit self-awareness. They consequently take it to be legitimate to speak of self-awareness as soon as I am not simply conscious of a foreign object, but acquainted with the experience of the object as well, for in such a case my consciousness reveals itself to me. Thus, self-awareness is taken to be a question of having first-personal access to one's own consciousness; it is a question of the first-personal givenness or manifestation of

experiential life. Most people are prepared to concede that there is necessarily something "it is like" for a subject to undergo a conscious experience (to taste ice cream, to feel joy, to remember a walk in the Alps). But insofar as there is something it is like for the subject to have the experience, the subject must in some way have access to and be acquainted with the experience. Moreover, although conscious experiences differ from one another – what it is like to smell crushed mint leaves is different from what it is like to see a sunset or to hear Lalo's *Symphonie Espagnole* – they also share certain features. One commonality is the quality of *mineness*, the fact that the experiences are characterized by first-personal givenness. That is, the experience is given (at least tacitly) as *my* experience, as an experience *I* am undergoing or living through. First-personal experience presents me with an immediate and non-observational access to myself. All of this suggests that we are dealing with a (minimal) form of self-awareness, that is, (phenomenal) consciousness is taken to entail a (weak) sense of self-awareness. To put it differently, self-awareness is taken to be a necessary condition for phenomenal consciousness. Unless a mental process is self-conscious, there will be nothing it is like to undergo the process, and it therefore cannot be a phenomenally conscious process.[11]

The claim that there is a close link between consciousness and self-awareness is less exceptional than might be expected. In fact, it might be argued that such a claim is part of current orthodoxy, since higher-order theories typically take the difference between conscious and non-conscious mental states to rest upon the presence or absence of a relevant meta-mental state. To put it differently, (intransitive) consciousness has frequently been taken to be a question of the mind directing its intentional aim at its own states and operations. Thus, higher-order theories have typically taken self-directedness to be constitutive of (intransitive) consciousness.

But one might share the view that there is a close link between consciousness and self-awareness and still disagree about the nature of the link. And although the phenomenological take might superficially resemble the view of the higher-order theories, we are ultimately confronted with two radically divergent accounts. In contrast to the higher-order theories, the phenomenologists explicitly deny that the self-awareness that is present the moment I consciously experience something is to be understood in terms of some kind of reflection, or introspection, or higher-order monitoring. It does not involve an additional mental state, but is rather to be understood as an intrinsic feature of the primary experience.[12]

Of course, this is not to deny that there are also far more complex forms of self-awareness that are both theory- and language dependent and intersubjectively constituted, but the primitive self-awareness that is part and parcel of phenomenal consciousness is independent of such conceptual sophistication.

The phenomenological analysis of self-awareness complements the argumentation provided by the developmental psychologists, since it explicitly tackles an issue they largely remain silent about, namely the nature of experience and phenomenal consciousness. Most of the developmental evidence presented by Stern, Rochat, Neisser, and Butterworth is obviously behavioral in nature. However, in order for a creature to be in possession of self-awareness, it is not sufficient that the creature in question behaves in a certain way. It also has to be in possession of experiences, and it must behave as it does because it has the experiences it has. To put it differently, any reasonable ascription of self-awareness cannot bypass a discussion of the relationship between the experiential dimension and self-awareness, but this is precisely what the phenomenological tradition can provide.

What about embodiment; how would the phenomenologists account for embodied self-awareness? Well, as Michel Henry once pointed out, a

phenomenological clarification of the body must take its departure in the original givenness of the body (Henry, 1965: 79). But how precisely is the body originally given? When I am watching a football match, I am normally not paying attention to the turn of my head when I follow the motions of the players, nor to the narrowing of my eyes when I attempt to discern the features of the goalkeeper. When I give up and reach for my binoculars, the movements of my hand remain outside the focus of my consciousness. When I am occupied with objects and directed at goals, my perceptual acts and their bodily roots are generally passed over in favor of the perceived, i.e., my body tends to efface itself on its way to its intentional goal. This is fortunate, because if we were aware of our bodily movements in the same way in which we are aware of objects, our bodies would make such high demands on our attention that it would interfere with our daily lives. However, when I execute movements without thinking about them, this is not necessarily because the movements are non-conscious, mechanical, or involuntary; rather, they might simply be part of my functioning intentionality, they might simply be immediately and pre-reflectively felt, as both Henry and Merleau-Ponty have argued (Henry, 1965: 128; Merleau-Ponty, 1945: 168). Thus, even if my movements might be absent as thematic intentional objects, this does not have to entail that they are experientially absent in any absolute sense.

Under normal circumstances, I do not need to perceive my arm visually in order to know where it is. If I wish to grasp the fork, I do not first have to search for the hand, since it is always with me. Whereas I can approach or move away from any object in the world, the body itself is always present as my very perspective on the world. That is, rather than being simply yet another perspectivally given object, the body itself is, as Sartre points out, precisely that which allows me to perceive objects perspectivally (Sartre, 1976: 378; Merleau-Ponty, 1945: 107). The body is present, not as a permanent perceptual object, but as myself. Originally, I do not have any consciousness *of* my body as an intentional object. I do not perceive it; *I am it*. Sartre even writes that the lived body is invisibly present, precisely because it is existentially lived rather than known (Sartre, 1976: 372). This is also why Husserl repeatedly has emphasized how important it is to distinguish between *Leib* and *Körper*, that is, between the pre-reflectively lived body, i.e., the body as an embodied first-person perspective, and the subsequent thematic experience *of* the body as an object (Husserl, 1973: 57).

In short, phenomenologists take pre-reflective body-awareness to be a question of how (embodied) consciousness is given to itself not as an *object*, but as a *subject*. Whereas Bermúdez has recently claimed that "somatic proprioception is a form of perception" that takes "the embodied self as its object" (Bermúdez, 1998: 132), the phenomenologists would argue that *primary* body-awareness is not a type of object-consciousness, is *not* a perception of the body as an object at all (cf. Gallagher, 2003), but on the contrary a genuine form of self-experience.[13]

6. Conclusion

Let me by way of conclusion briefly summarize the results. According to the theory-theory, it is not only my understanding of other people's mental states that involves a theory of mind. My access to my own mind also depends upon such a theory. In both cases, the same cognitive mechanisms are in use, in both cases we are dealing with a process of mind-reading, in both cases we are dealing with the application of a theory of mind. According to the standard view, however, children acquire a theory of mind at around the age of 4. It is only at this age they can pass the classical theory of mind tasks. Consequently, the theory-theory of mind argues that children lack proper self- and other-experience, during the first 3-4 years of life.

This view is faced with a number of both empirical and conceptual difficulties. Empirical findings strongly suggest that infants are in possession of a form of bodily self-awareness long before they are in possession of a theory of mind. Infants have an early sense of their own bodies; they are in possession of a dimension of bodily self-experience long before they are capable of solving any theory of mind tasks. Moreover, they are also capable of sophisticated social interaction that early.

At this stage, philosophy enters the picture. Phenomenologically inclined philosophers (including the later Wittgenstein) have not only argued that intersubjectivity presupposes embodiment and bodily self-experience, they have also – contrary to what has occasionally been claimed by developmental psychologists – denied that primary self-awareness is a type of *object*-consciousness. So to conclude, a recommended strategy is to combine the empirical findings of Stern, Neisser, Butterworth, and Rochat with the theoretical considerations of Husserl, Sartre, Merleau-Ponty, and Wittgenstein. Jointly they constitute a serious challenge to the theory-theory of mind.[14]

NOTES

[1] The very choice of term is consequently quite revealing. It clearly indicates that psychological competence is taken to consist in the possession and use of a theory.

[2] This (critical) focus on the theory-theory of mind should not be taken as an implicit endorsement of simulationism. There are problems with the simulation theory as well. Not the least its reliance on some kind of argument from analogy seems problematic (for an extensive criticism of the argument from analogy, cf. Avramides, 2001). Ultimately, one needs to realize that there are other options available than the choice between theory-theory and simulation theory.

[3] Gopnik and Wellman have compared the transition that occurs between the three-year old and the four-year old child's understanding of mind to the transition between Copernicus' *De Revolutionibus* and Kepler's discovery of elliptical orbits (Gopnik & Wellman, 1995: 242).

[4] Occasionally, some of the theory-theorists have been cautious enough to admit that this parallelism might not hold true for all kinds of mental states, but there is no general agreement about what should count as the relevant exceptions.

[5] For an informative overview of these tests, cf. Baron-Cohen, 2000.

[6] This is also granted by modularity nativism. Although this version of the theory-theory argues that the theory is innate, it still concedes that the theory needs a certain amount of experience as a trigger.

[7] Certain core features in infantile autism have frequently been interpreted as a result of a mind-blindness, i.e., they have been explained by reference to a damaged or destroyed theory of mind mechanism. But if autists lack a theory of mind, and if a theory of mind is required for self-awareness, then autists should be "as blind to their own mental states as they are to the mental states of others" (Carruthers, 1996b: 262; cf. Frith & Happé, 1999: 1, 7). Thus, we find Baron-Cohen arguing that autistic subjects are "unaware of their own mental states" (Baron-Cohen, 1989: 595), and Frith and Happé proposing that persons with autism can only judge their own mental states by their actions (Frith & Happé, 1999: 11), i.e., denying that autists have a direct, immediate, or non-inferential access to their own mind.

[8] Although Carruthers is, in general, unequivocal about denying conscious experiences to young infants (cf. Carruthers, 1996: 221; 2000: 202-203), he occasionally leaves a door open for a different conclusion. As he writes at one

point, it might be that infants are capable of discriminating between their experiences (and hence capable of enjoying conscious experiences), even while still being incapable of conceptualizing them (Carruthers, 1996: 222).

[9] The problem is exacerbated by the fact that although "self-awareness" is an ambiguous term, the theory-theorists rarely *define* what precisely they mean when they speak of self-awareness (cf. Zahavi & Parnas, 2003).

[10] It could be objected that although this might hold true for humans, it does not necessarily hold true for all intelligent life. Would it for instance be nonsensical to imagine intersubjectivity between brains-in-vats or between disembodied angels? Is the very idea of telepathy incoherent? (Thanks to Galen Strawson for this objection.) A substantial reply would lead too far. But let me, on the one hand, simply confess that I am not all that convinced that it is legitimate to draw substantial philosophical conclusions from the fact that certain scenarios are imaginable. Is our imagination always trustworthy, does it always attest to metaphysical possibility, or might imaginability not occasionally reflect nothing but our own ignorance (for a more extensive discussion cf. Wilkes, 1988; Parnas & Zahavi, 2000)? On the other hand, I would insist that if something like intersubjectivity is possible between brains-in-vats or angels, then it is a kind of intersubjectivity that is utterly different from the one we are familiar with.

[11] For more recent defenses of this position, cf. Flanagan, 1992; Zahavi, 1999; 2002; 2003; Kriegel, 2003.

[12] Let me forestall a possible objection, namely that this definition of self-awareness is too broad and that it simply includes too much. That is, since it doesn't match our everyday or folk-psychological notion of self-awareness (that tends to link the notion with our ability to recognize or identify ourselves in a thematic way), the present use of the term is inappropriate. I don't think this objection carries a lot of weight. From a conceptual point of view, there are no intrinsic problems whatsoever in using the term "self-awareness" to designate a situation where consciousness is aware of itself, or given to itself. Secondly, it is a simple fact that many of the classical philosophical theories of self-awareness as well as the more recent contributions by such thinkers as Brentano, Husserl, Sartre, Henry, Henrich, Frank, etc. have precisely been discussions of this broad notion. For a more extensive discussion, cf. Zahavi, 1999.

[13] For a more extensive overview of different phenomenological investigations of the body, cf. Zaner, 1964; Leder, 1990; Waldenfels, 2000.

[14] This study has been funded by the Danish National Research Foundation.

REFERENCES

Avramides, A. (2001). *Other minds*. London: Routledge.

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

Baron-Cohen, S. (1995). *Mindblindness. An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Baron-Cohen, S. (2000). The cognitive neuroscience of autism: Evolutionary approaches. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (second edition) (1249-1257). Cambridge, MA: MIT Press.

Bermúdez, J. L. (1998). *The paradox of self-consciousness*. Cambridge, MA: MIT Press.

Blackburn, S. (1995). Theory, observation and drama. In M. Davies & T. Stone (Eds.), *Folk psychology: The theory of mind debate* (274-290). Oxford: Blackwell.

Butterworth, G. (2000). An ecological perspective on the self and its development. In D. Zahavi (Ed.), *Exploring the self. Philosophical and psychopathological perspectives on self-experience* (19-38). Amsterdam: John Benjamins.

Carruthers, P. (1996a). Simulation and self-knowledge: A defence of theory-theory. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (22-38). Cambridge: Cambridge University Press.

Carruthers, P. (1996b). Autism as mind-blindness: An elaboration and partial defence. In P.
Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (257-273). Cambridge: Cambridge University Press.

Carruthers, P. (1996c). *Language, thought and consciousness*. Cambridge: Cambridge University Press.

Carruthers, P. (1998). Natural theories of consciousness. *European Journal of Philosophy*, *6*, 203-222.

Carruthers, P. (2000). *Phenomenal consciousness. A naturalistic theory*. Cambridge: Cambridge University Press.

Carruthers, P. & P. K. Smith (1996). Introduction. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (1-8). Cambridge: Cambridge University Press.

Davidson, D. (2001). *Subjective, intersubjective, objective*. Oxford: Oxford University Press.

Fivaz et al. (2004). Threesome intersubjectivity in infancy. In D. Zahavi, T. Grünbaum, & J. Parnas (Eds.), *The structure and development of self-consciousness: Interdisciplinary perspectives*. Amsterdam: John Benjamins.

Flanagan, O. (1992). *Consciousness reconsidered.* Cambridge, MA: MIT Press.

Frith, U. & F. Happé (1999). Theory of mind and self-consciousness: What is it like to be autistic? *Mind & Language*, *14*, 1-22.

Gallagher, S. (2003). Bodily self-awareness and object perception. *Theoria et Historia Scientiarum*, *7*, 53-68.

Gibson, J. J. (1979/1986). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, *16*, 1-14.

Gopnik, A. (1996). Theories and modules: Creation myths, developmental realities, and Neurath's boat. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind*, (169-183). Cambridge: Cambridge University Press.

Gopnik, A., & H. M. Wellman (1995). Why the child's theory of mind really *is* a theory. In M. Davies & T. Stone (Eds.), *Folk psychology: The theory of mind debate* (232-258). Oxford: Blackwell.

Gordon, R. M. (1996). 'Radical' simulationism. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (11-21). Cambridge: Cambridge University Press.

Heal, J. (1996). Simulation, theory, and content, In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (75-89). Cambridge: Cambridge University Press.

Henry, M. (1965). *Philosophie et phénoménologie du corps.* Paris: PUF.

Hobson, R. P. (1991). Against the theory of 'Theory of Mind'. *British Journal of Developmental Psychology*, *9*, 33-51.

Hobson, R. P. (1993). *Autism and the development of mind*. Hove: Psychology Press.

Hobson, R. P. (2002). *The cradle of thought.* London: Macmillan.

Husserl, E. (1973). *Zur Phänomenologie der Intersubjektivität II*. Husserliana XIV. Den Haag:

Martinus Nijhoff.
Kriegel, U. (2003). Consciousness as intransitive self-consciousness: Two views and an argument. *Canadian Journal of Philosophy, 33*, 103-132.
Leder, D. (1990). *The absent body*. Chicago: University of Chicago Press.
Legerstee, M. (1999). Mental and bodily awareness in infancy. In S. Gallagher & J. Shear (Eds.), *Models of the self* (213-230). Exeter: Imprint Academic.
Leslie, A. M. (1987). Children's understanding of the mental world. In R. L. Gregory (Ed.), *The Oxford companion to the mind* (139-142). Oxford: Oxford University Press.
Lewis, M. & J. Brooks-Gunn (1979). *Social cognition and the acquisition of self*. New York: Plenum Press.
McCulloch, G. (2003). *The life of the mind: An essay on phenomenological externalism*. London: Routledge.
Merleau-Ponty, M. (1945). *Phénoménologie de la perception.* Paris: Éditions Gallimard.
Merleau-Ponty, M. (1960). *Signes.* Paris: Éditions Gallimard.
Merleau-Ponty, M. (1964). *Le visible et l'invisible.* Paris: Tel Gallimard.
Neisser, U. (1988). Five kinds of self-knowledge. *Philosophical Psychology*, *1*, 35-59.
Neisser, U. (1993). The self perceived. In U. Neisser (Ed.), *The perceived self: Ecological and interpersonal sources of self-knowledge* (3-21). New York: Cambridge University Press.
Nichols, S. & S. Stich (2002). Reading one's own mind: A cognitive theory of self-awareness. http://ruccs.rutgers.edu/ArchiveFolder/Research%20Group/Publications/Room/ room.html.
Parnas, J. & D. Zahavi (2000). The link: Philosophy-psychopathology-phenomenology. In D. Zahavi (Ed.), *Exploring the self* (1-16). Advances in Consciousness Research. Amsterdam-Philadelphia: John Benjamins Publishing Company.
Premack, D. & G. Woodruff (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *4*, 515-526.
Rochat, P. (2001). *The infant's world*. Cambridge, MA: Harvard University Press.
Sartre, J.-P. (1943/1976). *L'être et le néant.* Paris: Gallimard.
Scheler, M. (1973). *Wesen und Formen der Sympathie.* Bern/München: Francke Verlag.
Shoemaker, S. (1968). Self-reference and self-awareness. *The Journal of Philosophy*, *LXV*, 556-579.
Shoemaker, S. & R. Swinburne (1984). *Personal identity.* Oxford: Blackwell.
Spitz, R. A. (1983). *Dialogues from infancy*. *Selected papers*. R. N. Emde (Ed.). New York: International Universities Press.
Stern, D. N. (1983). The early development of schemas of self, other and 'self with other'. In J. D. Lichtenberg & S. Kaplan (Eds.) *Reflections on self-psychology* (49-84). Hillsdale: Analytical Press.
Stern, D. N. (1985). *The interpersonal world of the infant.* New York: Basic Books.
Strawson, P. F. (1959). *Individuals.* London: Methuen.
Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech. The beginning of interpersonal communication* (321-347). Cambridge: Cambridge University Press.
Waldenfels, B. (2000). *Das leibliche Selbst. Vorlesungen zur Phänomenologie des Leibes*. Frankfurt am Main: Suhrkamp.
Wilkes, K. V. (1988). *Real people. Personal identity without thought experiments*. Oxford: Clarendon Press.
Wittgenstein, L. (1980). *Remarks on the philosophy of psychology II*. Edited by G. H. von Wright & H. Nyman. Oxford: Blackwell.
Zahavi, D. (1999). *Self-awareness and Alterity: A phenomenological investigation*. Evanston: Northwestern University Press.
Zahavi, D. (2002). First-person thoughts and embodied self-awareness. Some reflections on the relation between recent analytical philosophy and phenomenology. *Phenomenology and the Cognitive Sciences, 1*, 7-26.
Zahavi, D. (2003). Phenomenology of self. In T. Kircher & A. David (Eds.), *The self in neuroscience and psychiatry* (56-75). Cambridge: Cambridge University Press.
Zahavi, D. & J. Parnas (2003). Conceptual problems in infantile autism research: Why cognitive science needs phenomenology. *Journal of Consciousness Studies 10*, 155-176.
Zaner, R. M. (1964). *The problem of embodiment. Some contributions to a phenomenology of the body*. The Hague: Martinus Nijhoff.