Mind in practice : a pragmatic and interdisciplinary account of intersubjectivity

Bruin, L.C. de

4.

Mind Shaping in Early Ontogeny

That many operations of the mind have their natural signs in the countenance, voice and gesture, I suppose every man will admit. The only question is, whether we understand the significations of those signs, by the constitutions of our nature, by a kind of natural perception similar to the perceptions of sense; or whether we gradually learn the signification of such signs from experience, as we learn that smoke is a sign of fire or that freezing is a sign of cold [...] It seems to me incredible, that the notions men have of the expressions of features, voice, and gesture, are entirely the fruit of experience.

- Reid 1983

The mind in action

The previous chapters mainly dealt with intersubjectivity through the theory-colored spectacles of TT and ST. Consequently, we have primarily focused on social encounters in which agents were portrayed as bystanders, merely observing others without actively interacting with them. In such a context, intersubjectivity is primarily about mental state management. The mind is presented as an autonomous spectator, and knowledge of the other mind is considered to be one of its cognitive and conceptual achievements. The body is supposed to facilitate this process, but it is not supposed to play a *constitutive* role.

My own approach, by contrast, is firmly rooted in the pragmatist assumption that the mind is fundamentally shaped by its bodily existence (embodiment) and cannot be understood in isolation from its environment (embedment). It borrows from *enactivism* insofar it subscribes to a conception of the mind as emerging from the intricate web of interactive processes that is characteristic for a *complex system*. Complex systems are self-generating and self-maintaining wholes, which define their boundaries through their interaction with the surrounding world (cf. Varela 1979, Thompson 2007). A system is complex in virtue of the dynamic processes that hold between its sub-systems, and this is why its (emergent) properties cannot be fully explained in terms of these sub-systems alone (cf. Cilliers 2005). In order to understand a complex system, it is necessary to take into account the various interactive processes that describe its organization and define it as a system. In order to understand the complex system that is mind, we must pay attention to the dynamic processes between brain, body and environment that give rise to it. At the same time, however, the mind is more than a coupled system of brain, body and environment in isolation. The mind is stimulated, constrained and co-constituted by *other* coupled systems, and emerges as the result of continuing interactions with *other minds*.

This chapter shows how, at a very basic level and without cognitive and/or conceptual requirements, such interactions can be explained in terms of *second-person practices* (see fig. 4.1).[53]

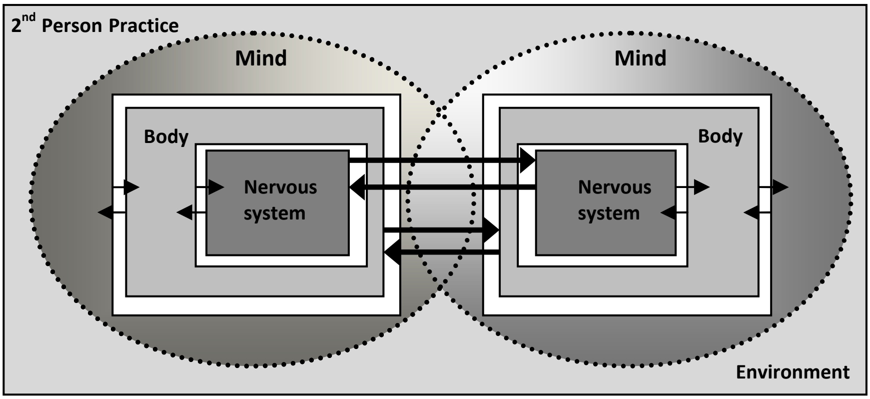


Fig. 4.1 Interacting minds in a second-person practice. Minds dynamically 'co-emerge' as the result of a constant interaction between nervous system, body and environment

[53] I share this starting point with many other enactive approaches to intersubjectivity (e.g., Fuchs and De Jaegher 2009, Gallagher and Zahavi 2008, Hutto 2007, Iacoboni 2003, Ratcliffe 2007, Thompson 2007).

These embodied and embedded ways of dealing with others constitute the base-line for social understanding, and they provide the background knowledge required for our more sophisticated modes of intersubjectivity. There are two ways in which these practices are primary to more advanced forms of social understanding. In the first place, they involve social abilities that come *earlier* in development and may even be partially *innate*. Secondly, they are also primary in the sense that they *continue* to characterize most of our social interactions throughout ontogeny, and remain the *default* mode of how we understand others.

The first part of this chapter shows that many embodied practices are already up and running from the moment we are born. I start by discussing a broad range of empirical findings demonstrating that very young infants are already able to interact with others in a rather sophisticated way.[54] Empirical research on early imitation reveals that neonates manifest a very primitive form of *co-consciousness*, in the sense that they have a proprioceptive awareness of both self and other. During the first year, various embodied practices trigger the infant to develop this awareness into a more advanced action-based understanding of intentional and emotional behavior (section 1). These practices are not self-sufficient. They depend on and are shaped by our bodily existence and various (partly) innate sensory-motor capacities (section 2). At around one year, infants acquire abilities that allow for a more advanced understanding of others in terms of their involvement in pragmatic contexts (section 3). The defining feature of these embedded practices is, as Hobson (2002) puts it, 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' (p.62).[55] Altogether, these practices provide infants by the end of the second year with a large body of pre-theoretical knowledge - the 'know how' required for the more advanced (narrative) modes of intersubjectivity that will be discussed in chapter 5.

---

[54] Some of the empirical evidence that is reviewed in this chapter is also put forward to support TT and/or ST approaches to intersubjectivity. However, I aim to show that it fits more comfortably with a pragmatic story about intersubjectivity, since such a story takes their functioning at face value and looks at what infants are *actually doing in practice*, as supposed to hypothesizing what should be going on in theoretical or simulation terms.

[55] I call these practices 'embedded practices' because they allow for a more advanced, 'situated' form of social understanding.

4.1 Embodied practices

Early sympathizers

By the time we are born our capacities for intersubjectivity are already shaped by our body and its movement. Bodily movement, as Gallagher (2005) aptly puts it, has already been organized in proprioceptive and cross-modal registrations in order to provide the capacity for differentiation between self and non-self. 'Movement and the registration of that movement in a developing proprioceptive system contributes to the self-organizing development of neuronal structures responsible not only for motor action, but for the way we come to be conscious of ourselves, to communicate with others, and to live in the surrounding world' (p.1).

Developmental studies point out that neonates indeed manifest a clear sense of self as a differentiated and situated entity in the world. Rochat and Hespos (1997), for example, have shown that they are already capable of discriminating between external and self-stimulation. In the external stimulation condition of their study, the index finger of the experimenter touched one of the infant's cheeks. In the self-stimulation condition, the infants spontaneously brought one hand to their face, touching one of their cheeks. The study revealed that neonates displayed significantly more rooting responses (i.e., head turn towards the stimulation with mouth open and tonguing) following external stimulation compared to self-stimulation. Neonates are not only able to discriminate between themselves and their environment, but they also respond selectively to other human agents. Despite not yet having acquired the appropriate concept of 'agent' or 'face', they differentiate effectively between agents and non-agents, and faces and non-faces.

It has been shown that very young infants are particularly sensitive to the *emotions* of other people, expressing what Trevarthen (1979) called 'intersubjective sympathy'. For example, Field et al. (1982) have shown that, as soon as 36 hours after their birth, neonates are already capable of discriminating the facial expressions happy, sad, and surprised. They also produce much more reactive crying when they hear the sound of another neonate crying instead of white noise or a synthetic cry (cf. Sagi and Hoffman 1976, Martin and Clark 1987).[56]

Section 4.1 has been written in collaboration with Sanneke de Haan, and I want to acknowledge her for several insights presented here.

A good illustration of the infants' responsiveness to the emotions of others is *affective synchrony*, which begins to occur in mother-infant interactions when infants are around 2-3 months of age (Stern 1985, Trevarthen 1979). Both mother and infant contribute to these affect-sharing episodes, using an increasing repertoire of interactive behaviors. A closer look at these specific social interactions (so-called 'microanalyses') reveals that mothers are highly likely to imitate infant expressions of enjoyment and interest, as well as expressions of surprise, sadness, and anger (Malatesta and Haviland 1982). However, they rarely display negative emotions to their infants. Infant-mother interactions exhibit considerable positive synchrony, partly as a consequence of the mother's contingent matching of positive infant emotional expressions.[57]

Stern (1985) claims that the early interactions between infants and their caregivers are first and foremost directed at the *attunement* of affect. He coins the term 'vitality affect' to clarify how different modalities can have the same 'kinematics' and thus express the same affect. For example, a mother can sooth her baby by saying 'there, there' in a comforting tone of voice, or by re-assuringly stroking the baby's back. The rhythm of speaking and the rhythm of stroking are the same, and in both allow the mother to express the vitality affect of soothing.

Stern emphasizes that we need more than imitation alone to explain what happens in such interactive exchanges.[58] He also notes that the first interactions between infants and caregivers typically entail matching the same vitality affect in the same modality, whereas from roughly 9 months on, caregivers are more inclined to react with the same vitality affect in a *different* modality. However, there is evidence that 5-month-old infants are

---

[56] What is interesting about this example is that neonates do not seem to respond to the sound of their *own* cries (on audiotapes). This supports the claim that there already is some kind of self-other distinction functioning right from birth.

[57] But this also works in the opposite direction. For example, Field et al. (1985) documented how depressed mothers influence their infants through these interactions in a negative way.

[58] Stern (1985) writes: 'For there to be an intersubjective exchange about affect, then, strict imitation alone won't do. In fact, several processes must take place. First, the parent must be able to read the infant's feeling state from the infant's overt behavior. Second, the parent must perform some behavior that is not a strict imitation but nonetheless corresponds in some way to the infant's overt behavior. Third, the infant must be able to read this corresponding parental response as having to do with the infant's own original feeling experience and not just imitating the infant's behavior' (p.139). The mere reproduction of the other's over behavior does not yet give us a clue that the other person really has a similar experience. It is exactly the slight modulation, for instance a change in the modality of expression that reveals the idiosyncrasy of the other and the individuality of their expression.

already able to detect a correspondence between different modalities that specify the expression of an emotion, such as visual and auditory information (Walker 1982; Hobson 1993, 2002). In any case, what is important here is that there appears to be a growing differentiation and complexity in the affect attunement of young infants. As Gopnik and Meltzoff (1997) put it, they increasingly interact with others in ‘a way that seems “tuned” to the vocalizations and gestures of the other person’ (p.131).

Early responders

From very early on children already show responsiveness to goal-directed or intentional behavior.[59] A series of experiments by Leslie (1982, 1988), for example, indicates that by 5 months, infants perceive intentionality and have different expectations about the effects on another object of the actions of a human hand versus an inanimate object. Woodward (1998) agrees. By habituating 5-month-old infants to a hand reaching for one of two objects, she found that they looked longer when the hand reached for the object not previously obtained, regardless of its position. She concluded that the infants were not ‘encoding’ the structural elements of the display (e.g., movement to the left or to the right), but the *goal* of the actor’s reach. This was further supported by a condition where the infants did not look longer when the hand was replaced by a metal rod (which helped to rule out an explanation in terms of a conditioned response, or at least one formed during the habituation phase). By 9 months, infants are able to follow the other person’s eyes and start to perceive various movements of the head, the mouth, the hands, and more general body movements as meaningful, intentional movements (Senju et al. 2006). And at around 10 months, infants have learned to parse specific kinds of continuous action according to intentional boundaries (Baird and Baldwin 2001, Baldwin et al. 2001).

Baron-Cohen (1995) has proposed to explain this early responsiveness to intentional action in terms of what he calls an ‘intentionality detector’ (ID): a perceptual device that

[59] My use of the term ‘intentional’ here is in line with Hutto’s (2007) description of ‘intentional attitudes’. According to Hutto, preverbal infants display intentional attitudes insofar as they selectively respond to certain aspects of their environment. However, intentional attitudes should not be confused with *propositional* attitudes. The latter are exclusively employed by those beings that have mastered certain linguistic constructions and practices, including the ability to represent and reason about complex states of affairs in truth-evaluable ways.

enables neonates to distinguish animate from inanimate objects. He argues that the ID is activated 'whenever there is any perceptual input that might identify something as an agent [...] This could be anything with self-propelled motion. Thus, a person, a butterfly, a billiard ball, a cat a cloud, a hand, or a unicorn would do' (p.33). The ID is supposed to be a kind of device that allows the infant to read 'mental states in behavior' by interpreting 'motion stimuli in terms of the primitive volitional mental states of goal and desire' (p.32). Baron-Cohen thinks that goals and desires are *primitive* mental states because they are minimally required to make sense of the universal movement of all animals: approach and avoidance. This is how he puts it: 'If you see an animal moving, be it an amoeba, a mouse, or a British prime minister, all you need to refer to in order to begin to interpret its movement are these two basic mental states' (ibid.).

However, as I already pointed out in previous chapters (cf. chapter 1.3 and 2.1), there are serious problems with the idea of locating mental states at the sub-personal level. Moreover, the question is whether it is *necessary* to do so. Do we really need to postulate primitive mental states such as desires and goals in order to make sense of the infants' responsiveness to intentional action? Gallagher (2001) thinks not. He suggests that the ID allows the infant to perceive intentional movement in a non-mentalistic way, and approvingly cites Scholl and Tremoulet (2000), who claim that the ID is 'fast, automatic, irresistible and highly stimulus-driven' (p.299).

A similar, but somewhat more advanced version of the ID is what Baron-Cohen (1995) calls the 'eye-direction detector' (EDD). The EDD is more specific than the ID since it is linked directly to the perception of faces, in particular the eyes. According to Baron-Cohen, the first function of the EDD consists of the detection of eye-like stimuli. Whenever the EDD detects eye-like stimuli, it 'fixates on these for relatively long bursts and starts to monitor what the eyes do' (p.39). The EDD builds on the idea that young infants already have a natural preference for looking at the eyes of other persons over looking at other parts of their face. For example, it has been shown that, at the age of 2 months, infants look almost as long at the eyes as at the whole face, but significantly less at other parts of the face (cf. Hainline 1978; Maurer and Barrera 1981, 1985).

Baron-Cohen suggests that the EDD has a second function as well: it enables the infant to determine whether the eyes it is looking at are directed at itself or at something else. There is some evidence that infants are already able to do this at a very young age. For example, it has been shown that 6-month-old children look approximately two and a half minutes longer at a face looking at them than at a face looking away (Butterworth 1991, Vicera and Johnson 1995).

The third function of the EDD, according to Baron-Cohen, is to 'infer from its own case that if another organism's eyes are directed at something, then that organism sees that thing' (1995, p.39). Such an inference is necessary in order to understand that the other person actually sees what he or she is looking at. However, Gallagher (2001) has argued that this assumption is mistaken, because it is only by virtue of e*xperience* that the infant comes to discover that someone could be looking in a certain direction without actually seeing something. This is something we learn rather than a default mode of the EDD: '*on the face of it*, that is, at a primary (default) level of experience, there does not seem to be an extra step between looking at something and seeing it' (p.89, italics in original).

In a certain sense, however, this seems to be precisely what Baron-Cohen is proposing. He suggests that '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' (p.43, italics added). What is problematic here is precisely the assumption that the infant comes to distinguish between seeing and not-seeing on the basis of its *own* experience, and consequently has to generalize this on the basis of an analogy. This shows that Baron-Cohen not only assumes that young infants already possess mental concepts, but also that they are able to make inferences over them on the basis of an analogy. However, as Hutto (2007a) points out, basic one-to-one interactions such as the above are not rightly characterized as involving an analogical comparison with others, or the neutral observation of outward behavior followed by cold inferences that the other is in such and such mental state. This is not only because these abilities come with severe developmental constraints, but also because there is a much more pragmatic explanation available, as we will see in a few sections.

There is also a terminological problem with Baron-Cohen's approach. An important drawback of notions such as 'detector', 'device' and 'mechanism' is that they invite a *mechanical* description of what goes on during these interactions. The notion of responsiveness is much more appropriate because it emphasizes the *interactive* nature of our involvements with others. It is often taken for granted that children need to posses certain individual abilities *before* they are able to participate in embodied practices. But this assumption is problematic insofar it obscures the fact that these abilities often develop in and through the kind of interactions they are supposed to precede and explain. Therefore, the quest for the 'underlying mechanisms of change' (Striano and Reid 2006) that motivates much infant-research seems to be misguided to the extent that it is aimed at pin-pointing the individual 'pre-cursors' of our 'full-fledged' interactive abilities. Such a linear and individually centered account of the acquisition of our social know-how does no justice to the intersubjective dynamics of development, in which the mechanisms themselves are subject to dramatic change as well.

Early imitators

So far I have not paid attention to *imitation* - an ability that is crucial to infants' development, since it provides them with numerous new opportunities to explore the field of intersubjectivity. The body of research on imitation is impressive. Meltzoff and Moore (1983), for example, have shown that one hour after they are born, neonates already imitate a variety of facial gestures such as mouth-opening and tongue-protrusion. Slightly older infants, with greater neuromuscular control, can imitate more specific behaviors such as tongue protrusion to one side (Meltzoff and Moore 1995). Although their first imitative attempts lack a high degree of accuracy, infants learn to correct and improve their gestural performance over time. This allows them to increasingly fine-tune and sophisticate their interactions with others.

I should point out that the second-person interactions in which imitative behavior is embedded are better characterized in terms of embodied resonance than in terms of pure mirroring – again because of the mechanical and reflex-like connotation of these latter terms. Tomasello (1999), for instance, has suggested that young children are 'imitation machines' (p.195). However, such a mechanical view cannot explain why infants are more likely to imitate after they have been attended to by the experimenter, as Csibra and Gergely (2009) have shown in recent experiments. The notion of embodied resonance, by contrast, allows us to account for the individual modulations infants bring to bear in their interactions. They do not completely *merge* into each other, but instead mutually *tune in* to each other. Their individual modulations attest to their autonomy: for perfect contingency you only need a mirror, but for genuine social interaction you need another person.[60]

Research shows that infants from 3 months on prefer these slight modulations (e.g., time-delay) in their embodied responses, except for autistic children who continue to prefer perfect contingency (Gergely 2001). Whereas *perfect* contingency only reflects one's own agency, *imperfect* contingency suggests the influence of another person and thus interpersonal contact. Given that normal infants are still exploring their sense of agency during this period, it seems natural to assume that they are mainly interested in finding out what they *themselves* effectuate. However, as soon as their sense of agency has reached a certain level of sophistication, a pure reflection on their own deeds probably becomes a bit boring - especially compared to the novelty that is introduced by interactions with other persons. Autistic children, however, continue to prefer perfectly contingent feedback to modulated feedback. Gergely (2001, p.418) explains this in terms of the 'faulty switch' of a postulated 'contingency detection module', which leads to symptomatic difficulties in social interactions. Although there is still an ongoing debate on the underlying mechanism(s) of autism, I am skeptical whether this talk about modules will bring us any further. But given their difficulties in social interaction and problems in dealing with novelty, it is not surprising that both the suggestion of another person and the possibility of interpersonal contact are less attractive to autistic children.

Meltzoff and Moore (1994) have investigated nine characteristics of early imitation in infants under 2 months:

1. Infants imitate a range of acts
2. Imitation is specific (tongue protrusion leads to tongue not lip protrusion)
3. Literal newborns imitate
4. Infants quickly activate the appropriate body part
5. Infants correct their imitative efforts
6. Novel acts can be imitated
7. Absent targets can be imitated

[60] As De Jaegher and Di Paolo (2007) remark, participatory sense-making is only participatory as long as the participants remain autonomous. Otherwise it would be merely one person forcing a sense upon another, a one-way interaction (see also Fuchs and De Jaegher 2009).

8. Static gestures can be imitated
9. Infants recognize being imitated [61]

They point out that there is an interesting developmental change in the infants' expression of imitative behavior. Although their abilities to imitate are in place right from the off, infants still need a lot of practice to pull of the more advanced modes of imitation that come later in development. For example, neonates imitate novel acts, but research on older infants reveals a generative imitation of novelty that is beyond the scope of younger infants (Bauer and Mandler 1992, Barr et al. 1996). More in general, there seems to be a progression in imitation from pure body actions, to actions on objects, to using one object as a tool for manipulating other objects. The question is: how can we explain this progression in imitative skills?

This is where Meltzoff and Moore (1994) offer us the `active intermodal mapping' (or AIM) hypothesis (see fig. 4.2). The basic idea behind the AIM hypothesis is that imitation is essentially a 'matching-to-target' process. The active nature of this matching process is captured by a 'proprioceptive feedback loop'. The loop allows the infant's motor performance to be evaluated against the perceived target and serves as a basis for correction. This process is facilitated by a 'supramodal perceptual system' that translates visual input into motor output, and lets perception and action communicate with each other within the same 'language'. It enables the infant to recognize a structural equivalence between its own acts and the ones it sees. A successful matching between perception and action is what grounds its apprehension that the other is, in some primitive sense, 'like me'. Gopnik and Meltzoff (1997) propose to explain this intermodal and intersubjective mapping as a primitive form of *theorizing*.

[61] Notice that the imitation described in these experiments cannot be a matter of *reflex behavior* or *release mechanisms*. Reflex and release mechanisms are highly specific, and no such mechanism could exist for imitation in general. Yet the range of behaviors displayed by the infants in these studies would require the unlikely assumption of distinct release mechanisms for each kind of behavior: tongue protrusion, mouth openings, lip protrusion, head movement, finger movement, as well as smile, frown, and so forth. Importantly, the studies that show imitative behavior after a delay clearly indicate the involvement of memory. It should also be remarked that the infants improve or correct their imitative response over time. They get better at the gesture after a few practices. Neither delayed reaction nor improved performance is compatible with a simple reflex or release mechanism.

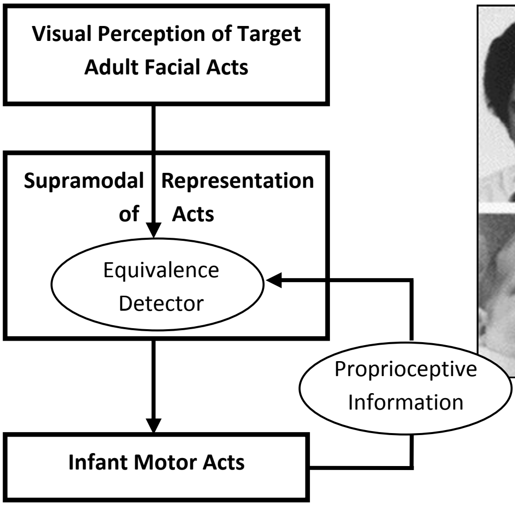


Fig. 4.2 The AIM hypothesis

Fig. 4.3 Neonate Imitation (Meltzoff and Moore 1977)

This lies at the beginning of an inference-like operation that is eventually promoted into a theoretical attitude. Meltzoff (2002) gives us a more comprehensive description of what this implies in terms of development:

(i) Innate equivalence between self and other. Infants can imitate and recognize equivalences between observed and executed acts. This is the 'starting state', as documented by motor imitations in newborns (fig. 4.3).
(ii) Self learning. As infants perform particular actions they have certain mental experiences. Behaviors are regularly related to mental states. For example, when infants produce certain emotional expressions and bodily activities, such as smiling and struggling to obtain a toy, they also experience their own mental states. Infants register this systematic relation between their own behaviors and underlying mental states.
(iii) Others in analogy to the self. When infants see others acting similarly to them, they project that people are having the same mental experience as they themselves when performing those acts. They use the behavior-mental states mappings registered through

their own experience to make inferences about the internal states of others.[62]

Meltzoff (2002) proposes that infants gradually learn to understand others by using knowledge of how they feel when they produce an expression to infer how another feels. He argues that infants 'imbue' the acts of others with 'felt meaning', because they are able to recognize the similarities between their own acts and those of others. 'Their experience of what it feels like to perform acts provides a privileged access to people not afforded by things. It prompts infants to make special attributions to people not made to inanimate things that do not look or act like them' (p.35).

The problem is that Meltzoff's account (just like that of Baron-Cohen) presupposes all the traditional ingredients of a *mindreading* account of intersubjectivity: mental concept mastery, inferential abilities, and the analogical argument. It is highly improbable, however, that these requirements are already within the reach of young infants (cf. Bermudez 1998, Gordon 2004). Moreover, it is not clear *why* we need them to explain the basic form of social understanding that these children are capable of. As we will see in the next sections, it is very well possible to give an explanation of the matching-to-target process that underlies imitation in *sub-personal terms*, without having to refer to mindreading or mental state management.

Body image and body schema

So far I have discussed a number of embodied practices that provide young infants with a basic but effective social understanding of others. I have emphasized that these interactions should not be interpreted in terms of *mindreading*. Rather, as Hutto (2007a) claims, 'we react directly to the attitudes of others as expressed bodily and we do so because of our natural predisposition, some of which gets reformed by experience and enculturation. It cannot be stressed enough that on this model the intervening cognition that makes this possible is not fueled by representations of the behavior or mental states of others' (p.115). But of course, the important question then becomes how we can further articulate such a 'direct reaction' to the attitudes of others without appealing to

[62] See Tomasello (1999) for a similar view. Tomasello claims that 'children make the categorical judgment that others are 'like me' and so they should work like me as well' (pp.75-6).

mindreading procedures, mental concept mastery, or analogical inferences.

With respect to early imitation, the question is how to explain the fact that children are able to successfully match their *perception* of the other person with their own imitative *action*. This is even more puzzling in cases of facial imitation in which infants are not able to perceive their *own* action. Bermudez (1998) formulates the problem as follows: 'Facial imitation involves matching a seen gesture with an unseen gesture, since in normal circumstances one is aware of one's own face only haptically and proprioceptively. If successful facial imitation is to take place, a visual awareness of someone else's face must be apprehended so it can be reproduced on one's own face' (p.125).

What is needed here is something that allows for a dynamic *co-constitution* of perception and action and explains their common coding, without requiring some kind of inferential/conceptual process to mediate between them. The problem with the proposals discussed above is the appeal to 'internal representations' or 'behavior-mental states mappings' in their explanation of such an action-perception loop. Gallagher (2005) argues that a supramodel system that integrates action and perception should not be explained in terms of 'abstract representations', but rather as a set of pragmatic (action-oriented) capabilities embodied in the developing nervous system. These capabilities constitute what he calls the *body schema*: a 'system of sensory-motor capacities' that functions without reflective awareness or the perceptual monitoring in an immediate and close to automatic fashion. This body schema makes it possible for children to develop a *body image*. A fully developed body image consists of a set of intentional states and dispositions such as perceptions, attitudes, and beliefs about one's own body.[63] It involves a form of reflexive and self-referential intentionality that allows me to experience my body as 'mine'. In case of neonate facial imitation, however, the infant does not yet possess a body of beliefs, attitudes or conceptions about its body, nor a visual perception of its own face. The only aspect of the body image available to the infant at this stage in development,

[63] Studies involving the notion of body image frequently distinguish three elements: (a) the subject's perceptual experience of his/her own body; (b) the subject's conceptual understanding of the body in general; and (c) the subject's emotional attitude toward his/her own body (cf. Cash and Brown 1987, Gardner and Moncrieff 1988, Powers et al. 1987). Although body schema and body image usually function synchronously, a few cases have been described in which one of them is dysfunctional. For instance, patients suffering from deafferentation have no proprioception from the neck down and can be said to have a defective body schema. In order to be able to move, they depend on their body image, and simple actions such as walking and holding a cup therefore require a great amount of concentration. Cases of hemi-neglect, in which patients consistently ignore one side of their body, can be interpreted as a sign of a defective body image.

according to Gallagher, is the *proprioceptive awareness* (PA) of its own body. PA is a primitive form of consciousness or pre-reflective awareness that informs the infant about the location of its limbs and its overall posture (without the aid of visual perception). Gallagher argues that this PA enables the newborn to 'know' that its own face is in some way equivalent to the visually presented face it is imitating.

More is needed to explain how PA is related to visual perception and the body schema, however. Therefore, Gallagher also puts forward the notion of *proprioceptive information* (PI), which consists of non-conscious and sub-personal, physiological information that updates the motor system about the position of body parts and movement of the body in general. Importantly, he argues that PA and PI are two sides of the *same* coin that is proprioception.[64] With this 'dual nature' of proprioception on the table, Gallagher is now able to explain how cross-modal communication between vision and proprioception is at the same time a communication between sensory and motor aspects of behavior. Since PI and PA depend on the same physiological mechanisms (the body schema), there is 'an immediate connection, a close interactive coordination, between proprioceptive information, which updates motor action at the level of the body schema, and proprioceptive awareness, as a pre-reflective, performative accompaniment to that action' (2005, p.76). And because PA and vision are intermodally linked, there is also a link between vision and PI, or more generally between sensory/perceptual and motor activities.

Early facial imitation, according to Gallagher, depends on both PA and PI. What the infant sees 'gets translated into a proprioceptive awareness of her own relevant body parts; and PI allows her to move those parts so that her proprioceptive awareness matches up to what she sees' (ibid.). But this translation is not really a translation or a transfer, because it is 'already accomplished' and 'already intersubjective'.

One of the drawbacks of Gallagher's proposal is that it promotes embodiment, but at the same time lacks in neurophysiological detail. As Edelman (1992) already made clear, 'it is not enough to say that the mind is embodied; one must say how' (p.15). It must be admitted, however, that Gallagher himself acknowledges this. He points out that 'recent studies in neuroscience suggest that there are specific neurophysiological mechanisms that can account for the intermodal connections between visual perception and motor behavior. These are mechanisms that operate prenoetically, as general conditions of

[64] I actually prefer the term *kinaesthesia* over *proprioception*, since this place a greater emphasis on *motion* instead of perception. However, to avoid confusion I will follow Gallagher in his use of the term proprioception.

possibility for motor stability and control, but are also directed related to the possibility of imitation' (2005, p.77). I will take a look at these findings in the next section. First, however, I wish to comment briefly on Gallagher's notion of body image as a form of primitive, pre-reflective awareness.

The proprioceptive awareness we witness in neonates can be considered to be the first manifestation of what we call the *mind*. However, Gallagher shows that it is nothing like the isolated, bodiless and static spectator that is usually presupposed by TT or ST. On the contrary, the mind as proprioceptive awareness, as a primitive body image, is structured and shaped by the body and its movement. It emerges as the result of *perception in action* - not in isolation, but through a continuous process of interaction with other minds. From the very moment of its conception, the mind can be seen as the expression of a self-consciousness that is at the same time already a *co-consciousness* (see fig. 4.4). Therefore, 'experientially, and not just objectively, we are born into a world of others' (Gallagher and Meltzoff 1996, p.226). ST and TT often argue that we need inferential and conceptual abilities to read the other mind, assuming that this is a prerequisite for intersubjectivity. But Gallagher shows that right from the moment of birth, children are *already* interacting with other minds. These interactions shape their minds in various ways, and provide them with a solid basis for future participation in more advanced social practices.

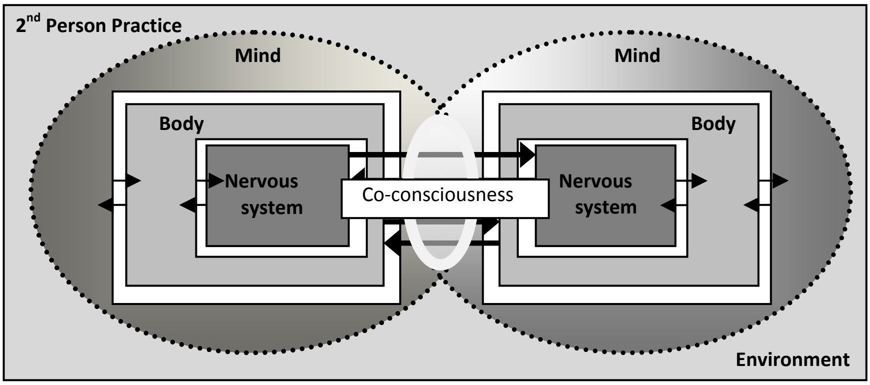


Fig. 4.4 Minds are already co-conscious from the moment of birth. Co-consciousness operates in between the semi-permeable bounds of embodied minds

4.2 Motor models and direct resonance systems

Motor models for basic adaptive feedback control

The challenge is to give a more detailed explanation of the relation between proprioception, body schema and body image, and demonstrate how they are embodied. In this section, I show how *functional motor models* can point us in the right direction.

Let us start by considering the very minimal and primitive body image that was introduced in the previous section. Gallagher (2000) calls this a 'minimal self', which he defines as a basic 'consciousness of oneself as an immediate subject of experience' (p.15).[65] He argues that the minimal self encapsulates two modalities of experience: (i) a sense of *ownership* (SO), the sense that I am the one who is undergoing an experience, and (ii) a sense of *agency* (SA), the sense that I am the one who is the initiator or source of the action.[66] How can we explain the relation between such a primitive body image and the body schema?

First, we need to know something about the motor theory of intentional action. This theory attempts to capture the dynamics of intentional action in terms of 'inverse', 'sensory-feedback' and 'forward' models (Blakemore and Decety 2001, Blakemore et al. 2001, Wolpert et al. 2001). The *inverse model* is important for motor control (see fig. 4.5). It consists of a simple sequence of steps, according to which a so-called 'planner' selects the appropriate motor commands given a desired goal (in terms of sensory states).

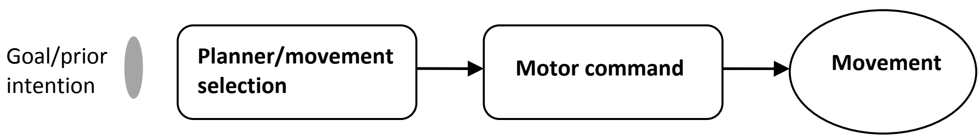


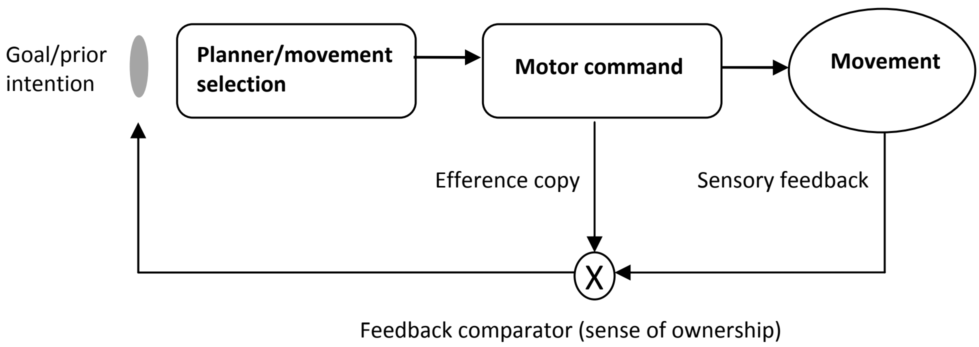
Fig. 4.5 Inverse model

[65] Gallagher seems to have borrowed the term minimal self from Strawson (1999).
[66] In normal voluntary or willed action, SO and SA are intimately intertwined and often indistinguishable. However, Gallagher (2000) argues that there are a number of situations in which it becomes possible to distinguish between them, namely in cases of involuntary movements, unbidden thoughts, schizophrenic experiences such as thought insertion. In these cases, according to Gallagher, the sense of agency is lacking but the sense of ownership is retained in some form.

This motor command is then sent to the muscles, and this leads to movement.

The sensory-feedback model (see fig. 4.6) is an extension of the inverse model, because it contains an extra flow of proprioceptive information. When a motor command is sent to the muscles, an *efference copy* of this signal is sent to a self-monitoring system (or *comparator*), which compares it to *re-afferent sensory feedback* about the movement actually made. Feedback might include visual and proprioceptive inputs resulting from movements of one's own hands, or movement through space, or manipulation of objects. When there is indeed a *match* between efference copy and sensory feedback, the feedback comparator model delivers a sense of ownership (SO) for the action. Gallagher (2005) explains this as follows: 'Exteroceptive sense modalities (such as touch or vision) provide information about both the environment and the moving subject (tactile and visual proprioception). Such information comes into a complex intermodal relationship with somatic proprioception to form coordinated and intermodal sensory feedback. That sensory feedback coordinates with efferent copies of motor commands in the nervous system, verifying that it is the subject who is moving rather than the environment' (p.106). In this way, the sensory-feedback model is able to generate a 'non-observational and pre-reflective differentiation between self- and non-self' (p.175).

Fig. 4.6 Sensory-feedback model

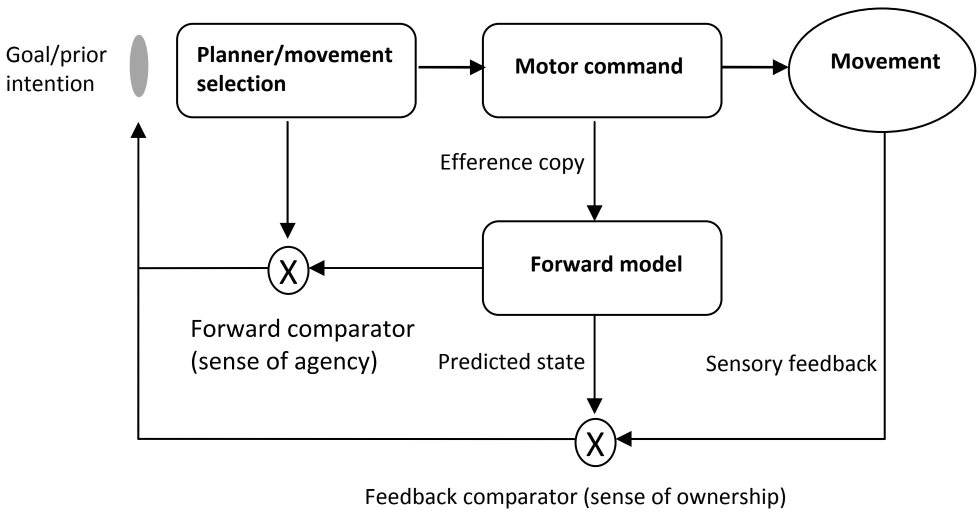


The sensory-feedback model is adaptive because it allows us to adjust ourselves to changing environmental conditions and compensates for exogenous disturbances: in the presence of different exogenous events, different outputs are needed to achieve the target.

This makes it possible to explain, for example, how we are able to correct our movement on the basis of sensory feedback about our actual movement. The model also sheds new light on the neonate ability to monitor and correct their imitations, in the sense that it shows that there need not be an *explicitly* recognized (cognitive) match between the infant's visual perception of the other's face and the proprioceptive awareness of its own face.

The sensory-feedback model is important for motor control, and explains how we are able to adjust our movements on the basis of sensory feedback. However, such an adjustment can only take place after the delays associated with sensory transmission. The so-called *forward model* bypasses these delays (and thus allows for better movement control) by positing a motor program that runs a slightly different sequence (see fig. 4.7). This time, the efference copy of the motor command is also sent to a *forward comparator*, which compares it to motor intentions and, when necessary, makes automatic corrections to movement *prior* to sensory feedback. Over time an association is established between efference copy and subsequent input, so that in effect a copy of the motor output signals comes to evoke the associated input signal.

Fig. 4.7 Forward model of goal-directed action



It can then operate as a *simulation* of feedback, to predict the consequences of output on input.[67] For example, the forward model might enable me to predict the sensory consequences of my act of reaching for and grasping a glass of water.

The forward model is responsible for generating a conscious sense of agency for action (Georgieff and Jeannerod 1998, Jeannerod 1994). Moreover, it can also account for the attenuated experience of the sensory consequences of one's own actions, compared to the sensory experience of exogenous changes in one's environment: the sensory consequences of one's own actions are predictable (from the efference copy of one's motor instructions), and therefore worth less perceptual attention than sensory changes exogenously produced. This could explain why in normal situations, proprioceptive awareness (PA) is attentively recessive and does not take center-stage in consciousness. This is because the forward model continues to function on the basis of proprioceptive information (PI), allowing one's body to work in a quite automatic way that does not require explicit monitoring. This may be different, however, in situations that require attentiveness to bodily movement. In infancy, for example, proprioception may be more centrally attended to when children learn how to walk. Early imitation also requires more focused propriospecific awareness. Gallagher and Meltzoff (1996) suggest that newborns use the proprioceptive experience of their own invisible movements to copy the movements of others. It not only helps them to monitor, correct and improve their imitations on the fly, but also allows them to memorize these imitations.[68] This shows that the forward model is not only important for motor control but also for motor *learning*.

The above models provide us with a functional architecture of the body schema, and they make it possible to explain how proprioception enables neonates to develop a primitive body image. But how is this architecture implemented at the neurobiological level? What kind of processes could facilitate the social interactions mentioned in the previous sections? And how do they provide young infants with co-consciousness, i.e. with an awareness of both self and other?

---

[67] Remark that this makes it possible to defend a (very weak) notion of simulation at the functional level. But such a notion depends on the intelligibility of forward models of goal-directed action, and it certainly does not satisfy a definition of simulation as the manipulation of pretend mental states.

[68] According to Gallagher and Meltzoff (1996), infants 'have the capacity to act out what they see in the face of the adult - they recognize what they see as one of their own capabilities' (p.223).

A neural architecture for imitation?

This is where the discovery of mirror neurons could be relevant. Mirror neurons are a class of visuomotor neurons that show activity in relation to both specific actions performed by self and matching actions performed by others. They 'mirror' the behavior of the other, as though the observer *himself* were acting (e.g., Rizzolatti et al. 1996, 2000). Mirror neurons appear to be involved in a larger cortical system (a 'mirror neuron system') that automatically 'duplicates' the observed action in the observer's motor system. This allows for an immediate, automatic and almost reflex-like understanding of others, without further inferential or conceptual requirements. Gallese (2001) gives the following explanation: 'when we observe goal-related behaviors […] specific sectors of our premotor cortex become active. These cortical sectors are those same sectors that are active when we actually perform the same actions. In other words, when we observe actions performed by other individuals our motor system 'resonates' along with that of the observed agent [...] action understanding heavily relies on a neural mechanism that matches, in the same neuronal substrate, the observed behavior with the one [the observer could execute]' (pp.38-9). What is attractive about the mirror neuron system is that it might explain how perception and action are dynamically co-constituted, and how action understanding emerges from the space that perception and action *share*.

Evidence for the existence of a mirror neuron system (MNS) was first discovered in the brain of the macaque monkey, comprising three cortical regions that exhibited the required functional properties and connectivity patterns: the superior temporal sulcus (STS) in the superior temporal cortex, area F5 in the inferior frontal cortex, and area PF in the posterior parietal cortex (Keysers and Perrett 2004).[69]

[69] In the early nineties, Perrett et al. (1989) demonstrated that neurons in the superior temporal sulcus (STS), which normally respond to moving biological stimuli (such as hands, faces and bodies) respond to these stimuli only when they are engaged in *goal-oriented actions* (see also Perrett et al. 1990, Perrett and Emery 1994). For example, some of them fired when the macaque saw a hand reaching toward an object and grasping it, but did not do so when the hand merely reached toward the object, without trying to grasp it. The investigators concluded from these observations that STS neurons probably code the perception of a meaningful interaction between an object and an intentional agent. The properties of these STS neurons seemed to be limited to the visual domain, since there was no association between the neuronal responses in STS and motor behavior. Another line of research, however, initiated by Rizzolatti et al. (2001), found parietal neurons with visual responses similar to the ones observed in STS *but this time with motor properties* (di Pellegrino et al. 1992, Gallese et al. 1996). These neurons were located in the ventral

Early experiments designed to detect the existence of a human MNS were motivated by the idea that if such a system existed, then the motor area it encompassed had to be active during both the execution and observation of a goal-directed grasping task. However, experimenters soon realized that instead of monkeyish grasping, *human imitation* offered a much more promising paradigm (Grafton et al. 1996, Rizzolatti et al. 1996). Imitation involves both the observation and execution of an action, and thus fitted perfectly with the properties of the system they were looking for. The investigators hypothesized that instances of imitation would yield an amount of mirror neuron activity approximately equal to the sum of activity during observation and execution. If they could identify brain areas that showed such a double amount of activity, then this would support the existence of a MNS in humans.

Iacoboni et al. (1999) found two areas that satisfied this condition, and also seemed to correspond anatomically to the macaque mirror areas. The first was located in the pars opercularis of the inferior frontal gyrus (in the inferior frontal cortex), the second in the posterior parietal cortex.[70] Together with the superior temporal sulcus (STS), coding for the perception of an observed intentional action, these areas could form a *blueprint* for the mirror neuron system. The STS, however, showed a somewhat unexpected pattern of activity. Although it yielded greater activity for action observation compared to control visual tasks and for imitation compared to control motor tasks (as was to be expected), there was also greater activity for imitation compared to action observation.

---

premotor cortex of the monkey (called area ‘F5’). Rizzolatti et al. (2001) also found that the posterior parietal cortex (PPC) of the macaque (area ‘PF’) contained mirror neurons almost identical to the ones described in F5. The areas PF and F5 appeared to be anatomically connected (Rizzolatti et al. 1998). Furthermore, evidence was found for a link between the STS neurons and the posterior parietal cortex (Seltzer and Pandya 1994). Together, the three cortical regions of the macaque brain (STS in the superior temporal cortex, area F5 in the inferior frontal cortex, and area PF in the posterior parietal cortex) seemed to have the functional properties and connectivity patterns required to instantiate a whole circuit for action recognition - a mirror neuron system.

[70] Iacoboni proposed a division of labor between the frontal and the posterior parietal mirror areas, inspired by single-cell studies (Kalaska et al. 1983, Lacquaniti et al. 1995) and neuroimaging data (Decety et al. 1997, Grèzes et al. 1998): the frontal mirror areas code the goal of the imitated action and the posterior parietal mirror areas code the associated movements. He claimed that certain experiments provide evidence for this idea. Koski et al. (2002), for example, demonstrated a modulation of activity in inferior frontal mirror areas during imitation of goal-oriented action, with greater activity during goal-oriented imitation compared to non goal-oriented imitation. See also the next section for a discussion of other experiments that might support such a proposal.

Mapping the above neural circuit onto the functional inverse and forward motor models described in the previous section allows us to make sense of the functional processes that underlie imitation (see fig. 4.8), and also helps us to understand the mentioned unexpected STS activity. Let us start with an observer who perceives the action of another agent. First, so-called *canonical* neurons in the superior temporal sulcus code an early visual 'description' of the perceived action (Perrett et al. 1990) and send this information to posterior parietal mirror neurons. This privileged flow of information is supported by robust anatomical connections between superior temporal and posterior parietal cortex (Seltzer and Pandya 1994). Second, the posterior parietal cortex codes the precise kinesthetic aspect of the movement of the agent (Kalaska et al. 1983, Lacquaniti et al. 1995) and sends this information to inferior frontal mirror neurons.[71]

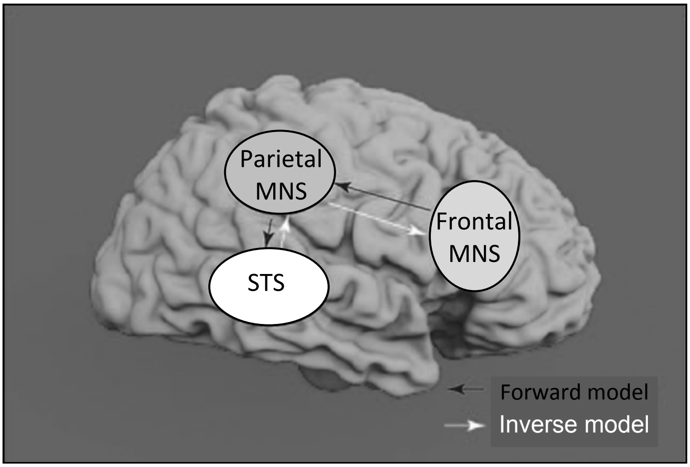


Fig. 4.8 A functional model of imitation (Iacoboni 2005):

1) The STS provides a higher-order visual 'description' of the observed action (inverse model)
2) This description is fed into the fronto-parietal mirror neuron system, where the goal of the action and the motor specifications to achieve it is coded (inverse model)
3) Copies of the motor imitative plan are sent from the fronto-parietal mirror neuron system to the STS, where there is a match between the predicted sensory consequences of the planned imitative action and the visual description of the observed action (forward model)

[71] Anatomical connections between these two regions are well documented in macaque monkeys (Sakata et al. 1973).

Third, the inferior frontal cortex of the observer codes the *goal* of the action. There is some neurophysiological (Umilta et. 2001, Kohler et al. 2002, Keysers et al. 2003) and imaging data (Koski et al. 2002) in support of this role for inferior frontal mirror neurons. This three-step process can be captured by means of a *forward* model, which uses the STS visual description of the action as input and the goal of the action as output.

Fourth, efferent copies of motor plans are sent from the parietal and frontal mirror areas of the observer back to the superior temporal cortex (Iacoboni et al. 2001), such that a matching mechanism between the visual description of the observed action and the predicted sensory consequences of the planned imitative action can occur. And fifth, if there is a positive match between the visual description of the observed action and the predicted sensory consequences of the planned imitative action, this forward/inverse model is reinforced by a 'responsibility signal' (Haruno et al. 2001) that assigns high responsibility for imitating the desired action. The observer is now ready to imitate the action of the other agent.

The mirror neuron system and action understanding

The above blueprint of the MNS might help us to give a plausible explanation of infant imitation. But we need to be careful here. To start with, there are different ways to make sense of imitation. The most restrictive definition of imitation requires the execution of a *novel* action (that is learned by observing another do it) and, in addition to novelty, also involves some understanding of the *means/ends structure* of that action: you have to be able to copy the other’s means of achieving her goal, not just her goal, or just her movements. Research on human imitation has shown that infants of 13-14 months are able to do this. But although the human MNS resembles the system found in the macaque brain, macaque monkeys are not able to imitate in the strict sense. They only have the capacity for action *emulation*. In action emulation, you observe another person achieving a goal in a certain way, find that goal attractive and attempt to achieve it yourself by whatever means.[72]

[72] Note that the reproduction of an observed action may be the same whether it is performed by imitation or emulation (cf. Czibra 2007). If the observer has effectors and biological constraints similar to that of the model, it is likely that she will emulate the outcome of the model’s action by

One important question is whether the human MNS by itself is able to facilitate imitative behavior in the strict sense of the word. As Hurley (2008) points out, Iacoboni's model is *in theory* able to explain how we understand the means/end structure of an action, because it distinguishes between the neural coding of *goals* and *movements.* When an observer perceives an agent moving in a goal-directed way, his inferior frontal mirror neurons encode the *goal* of the observed action, and this provides him with an understanding of the *intention* of the action. In addition, his posterior parietal mirror neurons encode the *movements* associated with the observed action, and this provides him with an understanding of *how* to achieve the goal by means of the observed movements. Linking these two processes in the right way could pave the way for imitative learning (in the strict sense).

However, in practice it turns out that, besides the MNS, imitative learning involves other brain areas as well. Molnar-Szakacs et al. (2005), for example, found that the imitation of novel actions yields additional activation of the dorsolateral prefrontal cortex (BA46) and cortical areas that are involved in motor preparation: the dorsal premotor cortex, the mesial frontal cortex and the superior parietal lobule. They argued that the activity in BA46 seemed to reflect the selection of motor acts that are 'appropriate' for the task that is executed (cf. Rowe et al. 2000).

In order to find out to which extent the MNS supports the understanding of action *intentions*, we have to explicate the notion of intention first. Gallese and Goldman (1998) originally proposed that the MNS might explain intentional action understanding in terms of *propositional attitudes.* They hypothesized that in case of a plan 'externally generated' in the brain of the observer, the latter's mirror neuron system would *retrodict* the 'target's mental state' (i.e., the agent's intention) by 'moving backwards from the observed action' (pp.495-6). However, the ability to understand actions in terms of propositional attitudes is a rather advanced mode of social interaction. Hutto (2009) remarks that it is a 'sophisticated high level capacity; it involves being able to answer a particular sort of 'why'-

---

means of the same behavior, i.e. she will faithfully reconstruct the observed action. This is why, in studies of imitation, unusual or inefficient goal-directed actions are demonstrated to participants in order to test whether they tend to *emulate* the outcome by their own (more efficient) means, or really *imitate* the observed action (Meltzoff 1988, Gergely et al. 2002, Horner and Whiten 2005). Meltzoff (1988), for example, tested whether infants are capable of imitation by demonstrating an unusual action to them, in which the model switched on a box-light by pushing it with his forehead. If the infants emulated the outcome, then they would have used a simpler action to achieve the same goal, such as pushing the box with their hands (cf. section 4 of this chapter).

question by skillfully deploying the idiom of mental predicates (beliefs, desires, hopes, fears, etc.)' (p.10).

Nevertheless, Iacoboni et al. (2005) have tried to demonstrate that the MNS contributes to the understanding of the 'why' of an action as well. In their experiment, they presented subjects with a series of short movies, which were labeled the 'context' condition, the 'action' condition and the 'intention' condition. In the context condition, subjects would see objects (a tea-pot, a mug, cookies, etc.) arranged either as if before tea (the 'drinking' context) or as if after tea (the 'cleaning' context). In the action condition, subjects would see a human hand grasp a mug either with precision grip or using a whole-hand prehension with no other contextual elements present. In the intention condition, the grasping actions were embedded in the two scenes used in the context condition, the drinking context and the cleaning context. Here, the context cued the intention behind the action.[73]

Iacoboni et al. (2005) found significant differences between the intention condition and the action and context conditions in the human brain areas known to have mirror properties. They showed that, compared to the action condition, the intention condition yielded significant signal increases in visual areas (STS) and the dorsal part of the pars opercularis of the inferior frontal gyrus. Importantly, they also found increased activity in the pars opercularis.

The experimenters argued that this means that Iacoboni's model is basically correct in assuming that the MNS does not simply enable movement recognition ('that's a grasp'), but also is critical for understanding of the *goal* of an action. According to them, the experiment shows that the MNS not only enables the observer to understand *what* the agent is doing (by generating motor activation associated with the same *movement* in the observer), but also *why* the agent is doing this (by generating motor activation associated with a similar *goal* for the observer).[74] They conclude that 'the role of the mirror neuron system in coding actions is more complex than previously shown and extends from action recognition to the coding of intentions' (p. 532).

---

[73] The 'drinking' context suggested that the hand was grasping the cup to drink. The 'cleaning' context suggested that the hand was grasping the cup to clean up. Thus, the intention condition contained information that allowed the understanding of intention, whereas the action and context conditions did not (since the action condition was ambiguous, and the context condition did not contain any action).

[74] Cf. Iacoboni et al. (2005), see also Rizzolatti and Craighero (2004).

These findings seem to indicate that the MNS plays an important role in the understanding of intentional action. However, we have to remark that the above experiment still deals with *intentional* instead of *propositional* attitudes. In this respect, Iacoboni et al. seem to be interested in a notion of intentionality that is far more basic than the one that Gallese and Goldman (1998) were originally after. There are other problems as well. According to Jacob (2005), for example, it is possible that the enhanced activity found in the pars opercularis of the observer is not so much the *output* but the *input* of the mirroring process. Csibra (2007) has pointed out that this problem reveals a more general tension between two conflicting claims made by those who defend an all-inclusive approach to intentional action understanding in terms of mirror neurons: namely, that action mirroring somehow is thought to represent *both* low-level resonance mechanisms *and* high-level action understanding. This tension arises from the fact that 'the more it seems that mirroring is nothing else but faithful duplication of observed actions, the less evidence it provides for action understanding; and the more mirroring represents high-level interpretation of the observed actions, the less evidence it provides that this interpretation is generated by low-level motor duplication' (p.447).[75]

I am not sure whether Csibra's criticism hits home. Actually, I think that his ideas about what it means to understand the intention behind an action are quite different from those of Iacoboni et al. - perhaps they come closer to Hutto's notion of propositional attitudes. But Csibra is probably right that we need more than just the MNS in order to

[75] Csibra (2007) illustrates this point by discussing another experiment on intention understanding in monkeys by Fogassi et al. (2005). In this study, single cell recordings were used in an attempt to prove that the MNS enables the observer (a monkey) to discriminate between two tokens of an intentional act of grasping, one for the purpose of eating and another for the purpose of placing. Csibra (2007) points out that the slight kinematic variation that was found (though not reported) could explain the activation difference across the mirror neurons. If we assume that the observed actions included the same kinematic differences as in the monkeys' actions, and the monkeys' parietal mirror neurons were sensitive to these parameters, then their activation represents a low-level mirroring phenomenon. According to Csibra, however, in this case nothing suggests that the monkeys understood the *intention* behind the observed action. But if we accept that the selectivity of the mirror neurons was independent of the kinematic parameters and reflected a true form of intentional understanding based on contextual cues, as Fogassi et al. (2005) would have it, then there is no evidence that this intentional understanding is based on low-level mirroring. Therefore, Csibra (2007) concludes that 'one cannot have one's cake and eat it too: the discharge of a set of MNs cannot represent the activation of the observer's motor system at low and high levels at the same time' (p.447).

account for findings on intentional action understanding such as those of Iacoboni et al. (2005).

Originally, mirror neurons were simply supposed to fire during both the execution and the observation of one and the same motor act, and the resulting match was thought to be responsible for action understanding. But many MNS advocates now recognize that if Iacoboni's model of action understanding is correct, then we need to explain the relationship between the neural coding of intention and movement *given a certain context*. According to Iacoboni et al. (2005), this requires an 'additional mechanism', one that involves neurons that are 'perceptually triggered by a given motor act', but whose discharge commands the execution of a different motor act 'functionally related to the former and […] part of the same action chain' (p.533). They suggest that the coding of action intentions is probably based on the activation of a neuronal chain formed by mirror neurons coding the observed motor act and by 'logically related' mirror neurons coding the motor acts that are most likely to follow the observed one, given a certain context. However, Jacob (2005) correctly points out that the idea of a motor chain that consists of 'logically related' mirror neurons sounds pretty much like a classical *inference* system that translates between perception and action. Such a mechanism only seems to be required because the MNS by itself cannot account for the complexity of mapping an observed movement onto an underlying intention.

In other words: it is not yet clear to which extent the MNS is involved in the context-sensitive understanding of goal-directed action, and this is subject to further research. However, a somewhat simplified version of Iacoboni's blueprint could still be used in order to explain how we are able to *anticipate* the actions of others. If we use a two-step inverse-forward model, it is possible to explain action anticipation in the following way: first, the STS feeds a visual description of the perceived action into the fronto-parietal mirror neuron system, where the kinaesthetic aspect of the perceived movement is coded in terms of a motor plan (inverse model). Second, efferent copies of this motor plan are sent back to the STS in order to predict the sensory consequences of this action (forward model). This process allows the observer to predict and anticipate the agent's next move, i.e., his next motor sequence.

Interestingly, the idea that the MNS might enable action *anticipation* in accordance with this simplified inverse-forward model perfectly fits with Gallese and Goldman's (1998) original suggestion that the ability to anticipate and predict the next move of conspecifics is crucial, since this move might be 'cooperative, non-cooperative, or even threatening' (pp.495-6). However, my proposal requires a much weaker (non-conceptual) notion of intentionality, and therefore offers a much more pragmatic explanation of the kind of intentional action understanding that young children display.[76]

Some additional considerations

The ability to anticipate the other's movement can be regarded as enabling a very early, action-oriented stage in the development of our understanding of other minds. But the MNS could also help us to explain our occasional feelings of *empathy* - the 'subjective experience of similarity between the self and others' (Decety and Jackson 2004, p.71).When we perceive the emotions and sensations of other agents, our MNS might trigger our motor system to resonate along with that of the observed agent in a direct fashion, and make it possible to 'put ourselves in the other's shoes' in order to experience what they are feeling in a non-inferential and non-conceptual way.

There is evidence that this is indeed what happens. Wicker et al. (2003), for example, showed that mirror processes play an important role in our experiencing of the emotion of disgust. They scanned participants both during their own experiences of disgust and during observation of other people's faces expressing disgust. The participants were scanned while viewing movies of individuals smelling the contents of a glass (disgusting, pleasant, or neutral) and forming spontaneous facial expressions. The same participants were also scanned while inhaling disgusting or pleasant odorants through a mask. The experimenters found that the same areas, the left anterior insula and the right anterior cingulate cortex were preferentially activated both during the experience evoked by disgusting odorants and during observation of other people's disgust-expressive faces. Gallese (2001, 2004) found similar evidence for the relevance of mirroring processes with respect to the sensation of pain.

I wish to close this section with a final observation about the importance of imitation and the involvement of the MNS. Full-fledged imitation allows young children to copy both

[76] Remark that the fact that action anticipation can be explained in terms of an inverse/forward model does not provide any support for ST, since it is not possible to draw a strict line between the observation of an action and something that counts as a simulation. Nor does this kind of action anticipation involve any mental state management. See also chapter 2.4.

the means and the goal of the actions they observe, and this is arguably a great way to generate new possibilities for intersubjective understanding without having to appeal to an innate theory of mind. In his discussion of Fodor's nativism, Meltzoff (2005) remarks that: 'Fodor is correct that solipsism and blank-slate empiricism are too impoverished to characterize the human starting state. However, this does not mean that adult commonsense psychology is implanted in the mind at birth or matures independent of experience. Here is an alternative to Fodor's creation myth. Nature designed a baby with an imitative brain; culture immerses the child in social play with psychological agents perceived to be "like me". Adult commonsense psychology is the product' (p.77).

At the same time, however, we have seen that we should not expect an easy explanation as to precisely what it involves. Although the MNS might be a step in the right direction, there are certainly more aspects that need attention. One of them concerns the role of *inhibition* in the imitation of behavior. In Iacoboni's model, for example, there needs to be a positive match between the visual description of the observed action and the predicted sensory consequences of the planned imitative action, otherwise the perceived action will not be imitated. But this already presupposes a form of inhibition that is quite advanced. How are these inhibitory processes related to the MNS? This requires further research. Another important aspect has to do with the experimental set-ups that are used to investigate the role of the MNS in our intersubjective engagements. Currently, most experiments on the MNS involve the passive third-person observation of another agent. But this is clearly not representative for most of our everyday social encounters, which are better characterized in terms of active second-person engagements. Future research on the MNS definitely has to take this into account.

4.3 Embedded practices

Shared attention

The previous sections demonstrated that neonates and very young infants are well able to individuate other persons and interact with them in a *dyadic* way. But arguably these 'primary intersubjective' capacities (Trevarthen 1979) do not yet involve a very strong notion of intentional understanding on behalf of the young infant. It is generally accepted that such a notion only starts to emerge when infants start perceiving and interacting with other agents in a world-involving way, entering the realm of 'secondary intersubjectivity' (ibid.). The embedded practices that are characteristic for secondary intersubjectivity are *triadic*, in the sense that they involve a referential triangle of child, adult, and the environment - an outside object or event to which they share attention (see fig. 4.9).

Shared attention not only involves infants attending to the same objects or event at the same time, however. It also requires that they mutually recognize that their attending has a *common focus*. This makes it quite different from the forms of reciprocal imitation described in the previous sections. In situations of shared attention, according to Hutto (2007a), 'I see what the other is attending to, I see that they are attending to it, and I see that they are attending to both the object and to my attending. Only in this way is the object recognized as a common focal point' (p.126).

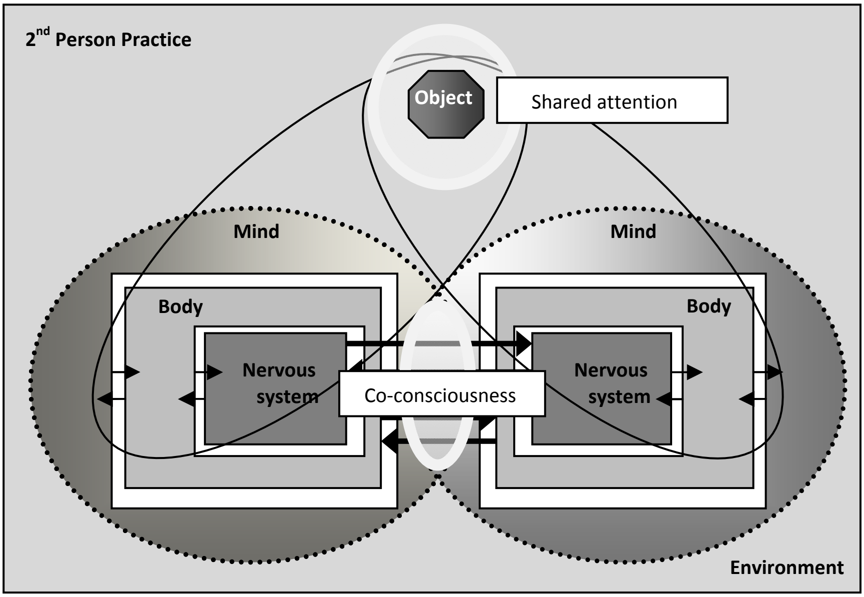


Fig. 4.9 Shared attention in secondary intersubjectivity

From 6 months of age onwards, infants are already capable of perceiving other people as being directed towards objects, first in their grasps of objects, later also when they gaze and point at (distant) objects (Woodward 2005). Yet, as Tomasello et al. (2005) point out, infants' object-directed understanding of others merely has the effect that they 'expect the adult to be consistent in his interactions with the same object over a short span of time [...] they do not have any understanding of the internal structure of intentional actions' (pp.678-9). But by 12 months of age, experimental findings suggest that infants 'can (1) interpret others' actions as goal-directed, (2) evaluate which one of the alternative actions available within the constraints of the situation is the most efficient means to the goal, and (3) expect the agent to perform the most efficient means available' (Gergely and Csibra 2003, p.288). However, Reddy (2003) has argued that shared attention can be said to arrive much *earlier* as long as it is not defined with respect to an outside object but rather to the child itself (since the child is already aware that itself can be an object of attention). Interestingly, she suggests that we have to pay more attention to the *second-person interaction*s between the infant and the experimenter to discover this.

Baron-Cohen (1995) has proposed that our capacity for shared attention might be facilitated by a 'shared attention mechanism (SAM)'. He argues that 'SAM's key function is to build […] triadic representations. Essentially triadic representations specify the relations among an Agent, the Self, and a (third) Object. [...] Included in a triadic representation is an embedded element which specifies that Agent and Self are both attending to the same object' (pp.44-5).

Again, the question is why we would need *mental representations* to explain what happens in these cases. As Hutto (2006) argues: 'There is no reason to suppose that [...] shared attention involves making full-fledged propositional attitude ascriptions. Seeing another's seeing does not involve representing the other's *cognitive* take, it only requires recreatively imaging the other's *perceptual* one' (p.192, italics in original). What underpins the mutual connectedness that is characteristic for shared attention is probably much better explained in terms of the MNS.

Pre-linguistic behavior representative of shared attention includes the systematic use of communicative gestures for instrumental purposes such as pointing and gaze alteration (Butterworth and Grover 1990, Butterworth and Jarrett 1991). Infants not only flexibly and reliably look where adults are looking (gaze following), but also try to obtain emotion cues from others to assist in their own assessment of an uncertain or ambiguous situation – this is called ‘social referencing’ (e.g., Rosen et al. 1992). Some have argued that the latter ability requires a ‘rudimentary ability to impute mental states of self and other’, and on top of that, a basic understanding that ‘one mind can be interfaced with another’ (Bretherton 1991, p.57). However, it is very well possible to give an explanation of social referencing without assuming that the infant has to attribute mental states to others. For example, observing the mother's emotion expression may induce the corresponding emotion in the infant who can then proceed to appraise the ambiguous situation on the basis of its *own* felt emotion. What is problematic about the appeal to infant's instrumental use of others to achieve its goals as evidence for attributing mental states to persons is this: while this kind of behavior indeed indicates that the infant is an intentional agent, it clearly does not imply that it must perceive the other person as an agent whose actions are caused by intentional mental states that the infant manipulates through its communicative gestures.

Basically, what happens in shared attention is that children’s ability to perceive affordances, i.e. to see objects in the environment as inviting or as enabling certain kinds of actions, is re-centered to the perspective of the other. Instead of seeing things as affording something for *themselves*, they now see them as affording something for *others* as well. This lays a foundation for the more advanced modes of perspective taking that are characteristic for narrative practice.

Further developments

Perceiving the other as an intentional agent who is responsive to a growing array of affordances in their environment allows the infant to anticipate, cooperate and coordinate in increasingly complex practices. In situations of shared attention, several behaviors often come together, enabling infants to ‘tune in’ to the attention and behavior of adults toward outside objects and events (cf. Tomasello 1999). Amongst others, there is a further development of imitative behavior - infants begin to act on objects in the way adults are acting on them. Already at 6 months of age infants can reproduce others’ actions on objects (Barr et al. 1996). However, it is not until the age of 13-14 months that there is evidence of *imitative learning*. Imitative learning consists of reproducing the intentional actions of others, including both the goal at which they are aiming and the behavior or strategy by means of which they are attempting to accomplish that goal. For example, in a study by Meltzoff (1988), infants of 14 months old observed an adult bend at the waist and touch his head to a panel, thus turning on a light. The infants followed suit even though they might also have turned on the light by simpler means (e.g., with their hands) - implying that they were indeed reproducing the adult's action. Moreover, they did not turn the light on in this odd way unless they had seen the model do it first (see also Meltzoff 2004, Gergely et al. 2002). Similarly, 16-months-old infants imitatively learn from a complex behavioral sequence only those behaviors that appear intentional, ignoring those that appear accidental (Carpenter et al. 1998). They do not just mimic the limb movements of other persons; they attempt to reproduce other persons' intentional action. By 18 months, infants are able to re-enact to completion the goal-directed behavior that an observed subject does not complete (e.g. pulling apart miniature dumbbells), but they will not re-enact the target act when it is performed by a mechanical device (Meltzoff 1995, Meltzoff and Brooks 2001).[77]

Also emerging in the second year is the infant's capacity for *pretend play.* The main characteristic of pretend play is that children pretend an object to be something else (Leslie 1987, Garvey 1990, Lillard 2002). For example, a child who is pretending a pile of sand is fantastic chocolate cake might call it cake, mimic eating it, and perhaps even say: 'Yum-yum, what delicious cake!' Notably, the child will not actually go so far as to eat the sand, since it is clearly aware of the cake's real identity. Around this age, children also become capable of recognizing pretend behavior of *others*. For example, when the mother pretends the banana is a telephone, the child is able to pick it up, hold it up to his ear and mouth and say: 'Hi. How are you? [Brief pause] I'm fine. OK. Bye'. Besides the substitution of objects, pretend play can also involve imagined objects, or roles and situations.

It has been argued that pretend play presupposes a capacity for 'secondary representation'. Perner (1991), for example, claims that young infants are initially only capable of entertaining 'primary representations' that represent 'the world as it is'. Such a representational system is not sufficient to facilitate pretend play, however. The child's primary representation of a banana, for example, cannot also incorporate a representation of this banana as a telephone. This requires the capability to entertain 'secondary

[77] Interestingly, infants are more prone to imitate an unfulfilled goal if the action is marked linguistically as purposeful, e.g. 'Let's put this on here. There we go!', but not if it is marked as accidental, e.g. 'Let's put this on here. Whoops!' (Carpenter et al. 1998). Infants not only try to reproduce the intentional *actions* of others, but also pieces of *language* (Tomasello et al. 1996).

representations', which add the ability to model hypothetical situations, and makes it possible for the child to simultaneously entertain multiple mental models.

However, Lillard (1993) found that children first understand pretending only as an action, and only much later come to see it as involving mental representations. In fact, most studies of pretense involve pretense with actions (cf. Flavell et al. 1987, Wellman and Woolley 1990). In these experiments, children perform correctly by directly referring to the *actions themselves*, rather than by *mentally representing* them. In other words, they initially seem to interpret pretense in terms of action alone. This emphasizes the embodied, situated and enactive character of early pretend play.

What is problematic about the notion of mental representation is precisely that it is usually associated with a number of *opposite* features, such as context-independency (representational contents are self-sufficient and exportable to different situations), objectivity (representations depict in a isomorphic way how the world and the actions are structured) and abstractness (representations provide neutral depictions valid under any possible perspective, not from situated points of view).

Shared attention and language acquisition

Shared attention is an important precursor to the development of *linguistic practices*. Infants pick up language socially by using it in pragmatic context, and by noticing what others do with it, through sharing interests, pretend play and imitative learning (Bates et al. 1975, Ninio and Snow 1996). There is much evidence that shared attention is strongly associated with the picking up of words in the infant's second and third year (Locke 1993, Rollins and Snow 1998, Tomasello 1988). Its onset not only consistently precedes the emergence of referential language in the second year of life, but the ability to engage in shared attention during infant-mother interactions also predicts the infant's word comprehension and word production (Carpenter et al. 1998).[78]

New words also prepare children for more sophisticated social interaction. Eilan (2005) observes that the 'first words emerge during the thirteenth month, on average, and

[78] For example, gaze following, an important prerequisite for shared attention, predicts vocabulary between the first and the second year (Morales et al. 1998), and shared attention bids have been shown to make a unique contribution to language development at 30 months (Morales et al. 2000).

from then on until the end of the second year, attentional behaviors become progressively more sophisticated - for example, we find progressively sensitive checks of where the adult is looking, before, during and after pointing initiated by the infant, or showing of objects to adults, the bouts of attending together to an object become longer and able to sustain the beginning of extended play with, and conversations about the object(s) attended to' (p.5). The growing ability to use linguistic signs provides children with new modes of expression and enables more advanced forms of understanding others than those of the purely embodied and embedded form.

Although it is often assumed that young children acquire language through ostensive definition (adults stop what they are doing, hold up objects, and name these objects for them), this is empirically not the case. In general, for the vast majority of words in their language, children must find a way to learn them in the ongoing flow of social interaction, sometimes from speech not even addressed to them. Tomasello et al. (1993) call this kind of imitative learning *cultural learning* because the child is not just learning things *from* other persons; it is also learning things *through* them in the sense that it must know something of the adult's perspective on a situation to learn the active use of this same intentional act. The idea is that children only come to understand a symbolic convention by learning to understand their communicative partner as an *intentional agent*, one with whom one might share attention, since 'a linguistic symbol is nothing other than a marker for an intersubjectively shared understanding of a situation' (Tomasello 1999, p.516).

The development of linguistic practices does not only depend on the embodied and embedded practices described in the sections above, but it also takes them to the next level. Language starts to provide 'an immensely delicate and useful way of pointing' (Heal 2005), exponentially extending the ways in which infant and other can explore the world together, adding rationales of increasingly complex structure to the world of the infant, possible reasons that they and others may act upon. How this happens is the topic of the next chapter.

4.4 Social understanding without cognitive or conceptual requirements

Summary

In this chapter I have discussed a number of embodied and embedded practices through which infants learn to deal with others in a direct, i.e. non-conceptual and non-inferential fashion. These interactions contextualize our engagements with other minds and provide us with the 'know how' that is required for more advanced (conceptual) modes of intersubjectivity. Importantly, they are best and most parsimoniously explained without reference to theory, simulation or mindreading. I very much agree with Hutto (2007a), who stresses that our 'nonverbal acts of intersubjective responding are not prosecuted by the deployment of theory, inferential reasoning, or projective simulation. We can be sure of this because no ascriptions are made to others on the basis of their observed behavior - there is no need to bridge an imagined gap between self and other; indeed the very idea of such a gap existing at this level is problematic' (p.115). Embodied and embedded practices do not presuppose higher order cognitive abilities or advanced mental state manipulation skills. Rather, they structure and scaffold these later developments.[79]

I have shown that in their 'ordinary' second-person interactions with others, children do not put themselves in the observer position – they are not passively standing at the side thinking about how to access other minds or trying to find explanations for others' behavior. Rather, they actively respond to them in various embodied and embedded ways (see fig. 4.10). Gallagher (2007) hits the nail on the head when he claims that 'what we call social *cognition* is often nothing more than social *interaction*. What I perceive in these cases does not constitute something short of understanding. Rather my understanding of the other person is constituted within the perception-action loops that define the various things that I am doing with or in response to others' (p.540, italics added). These perception-action loops are structured and shaped by our bodily existence and various (partly) innate sensory-motor capacities. Mirror neuron processes show how perceived

[79] Nor do these practices merely function as developmental precursors to a Theory of Mind. There is only one study that reports an association between pretend play in 33-months-old children and their success in passing a series of false-belief tasks 7 months later (Youngblade and Dunn 1995). Two other studies, however, fail to reveal a similar longitudinal association between pretense and Theory of Mind development (Charman et al. 2000, Jenkins and Astington 2000). With regard to imitation, there is no evidence for an association between early imitation and the later development of a Theory of Mind (Charman et al. 2000).

behavior and responsive actions can become intelligible *together*, that is, in the same process. They allow for a dynamic co-constitution of perception and action, and do not require inferential or conceptual process to mediate between them. At the same time, however, our intersubjective abilities cannot be explained in (or reduced to) purely neurobiological terms. Although brain processes are without a doubt important for explaining how infants are able to understand others, they would not occur unless these infants were acting within a broader social context. This context has to be taken into account in order to do justice to the interactive nature of intersubjectivity.

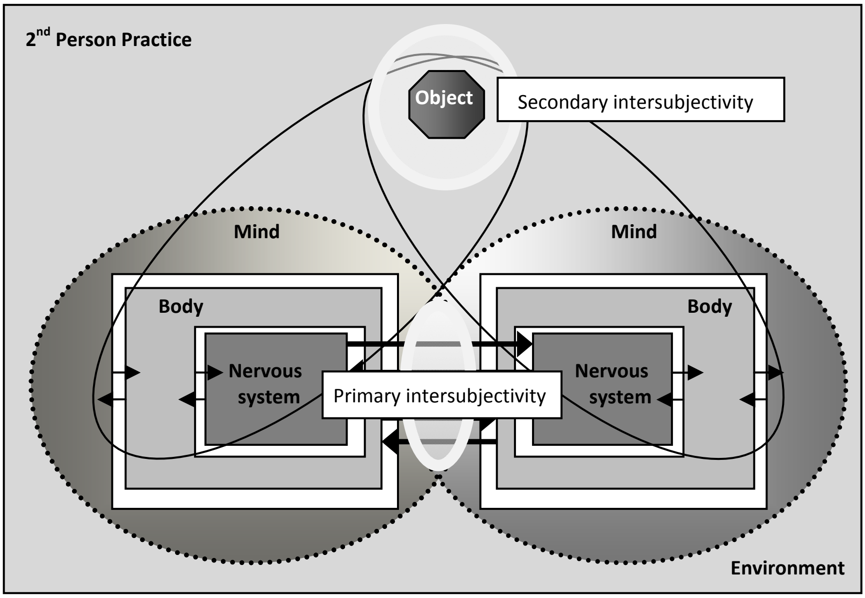


Fig. 4.10 Primary and secondary intersubjectivity

Primary intersubjectivity:

imitation
intentionality detection
action anticipation
eye tracking
movement tracking
emotion understanding

Secondary intersubjectivity:

shared attention
pointing / gaze alteration
social referencing
agency detection
pretend play
advanced imitation

Direct perception

In the remainder of this chapter, I wish to address the question whether we need to further articulate the embodied and embedded practices described in the previous sections, and if so, how. Gallagher (2001) has proposed the term 'body reading' in order to stress the perception-based nature of understanding that is characteristic for these practices. He claims that during our intersubjective engagements, it is very 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 [...] 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' (p.90, italics added).[80]

In his recent writings, Gallagher has further extended these ideas into a theory of 'direct perception' (cf. Gallagher 2007). His starting point is the observation that TT and ST approaches to intersubjectivity somehow seem to assume that perception is by itself not *sufficient* for social interaction. Something more is needed in order to understand our fellow human beings, and this is the reason why these positions appeal to mindreading procedures. According to Gallagher, the problem is that TT and ST start with a notion of perception as 'third-person observation', rather than something that happens in the context of *interaction*. As a result, we are not actively involved with others, but we stand at 'the margins of the situation.' However, Gallagher argues, this idea of perception as mere observation leaves TT and ST with an extremely impoverished idea of what perception actually consists in when it comes to perceiving other people. 'If I were to remain with only this perception I would be totally perplexed or at least puzzled about the other person's behavior. I see what the other person does, but until I call forth some theory, or until I run

[80] Hutto (2007) points out that the reading metaphor is misleading here, since it retains an 'intellectual connotation' that misrepresents what goes on in basic intersubjectivity and ignores that, although infants are not *reading* minds, they are 'immediately responsive to "other minds" nonetheless' (p. 116). I agree, but I think what is more problematic is the emphasis on *individual perception*.

through a simulation routine, I seem not to have any sense of what that person is up to' (p.536).[81]

Gallagher's own approach, by contrast, depends on a very rich notion of perception that builds on the idea that we have a direct perceptual grasp of the other's intentions, feelings, etc. The kind of perception Gallagher has in mind is 'direct' in the sense that nothing is added to it. When we see the actions and expressive movements of other persons, we are able to *directly* perceive their meaning. We do not need to consult a folk psychological theory or run a complicated simulation routine.

The question is why we have to place so much emphasis on the role of direct perception. Although I welcome Gallagher's rejection of mindreading when it comes to explicating low-level embodied practices, I think Hutto (2007a) has a point in claiming that it is 'more correct to say that we are directly *moved* by another's psychological situation rather than that we directly *perceive* it' (p. 116, italics added). The problem with the notion of direct perception is that, despite its success in overcoming the TT/ST mindreading legacy, it seems to encourage an interpretation of second-person embodied practices in terms of *individual perceptual* capacities. And this brings back the old idea that intersubjective understanding is primarily a one man (or woman) spectator sport, a social 'know-how' that is modeled on the first-person perspective of the individual agent.

Perception and the intrasubjective bias

Many of the problems that trouble TT and ST approaches to intersubjectivity can be traced to their commitment to a strong form of internalism, and the (Cartesian) ideal to model our understanding of others on the mind/brain of the individual agent. If we are to avoid these problems, we have to reject the idea that intersubjectivity is primarily a *personal* achievement, and maintain a clear focus on the second-person nature of social interaction. The challenge for a pragmatist approach to intersubjectivity is to look beyond the embodiment of individual subjects in order to properly conceive of the embedded and interactive nature of our understanding of others. This is important because it can put a

[81] Importantly, Gallagher recognizes that this kind of perception is not *completely* impoverished, since it is still smart enough to allow the agent to distinguish between an object in the surrounding environment and another agent. But this is not sufficient for the kind of social understanding that TT and ST are after: some 'extra cognitive tools' are required.

stop to simplistic 'just so stories' about intersubjectivity, and prevents all too easy 'explanations' of the acquisition of our social abilities.

What is problematic about the notion of direct perception is precisely that it seems to discourage us to look beyond individual embodiment. Gallagher's (2000) ideas about the *minimal self* seem to confirm this worry. The word 'minimal' is usually employed to denote the most limited case we can come up with. Therefore, the notion of a minimal self seems to suggest that we can just slice up the self into small pieces in order to find its most minimal part. However, it is very probable that in the end this part will refer to an *individual agent*. The quest for a minimal self easily runs the risk of slicing off the *social* dimension, thereby leaving us with the primacy of an impoverished first-person perspective. This in turn enhances the idea that we need some kind of 'self-sufficient self' before we can engage in social interaction. To avoid these mistakes, it is probably much better to speak of a 'basic self' instead of a 'minimal self' – one that is thoroughly social and relational. Since our experience is always intentional and directed at something, it is necessarily *relational.* And since many of our interactions are with other persons, this relatedness is also a social relatedness. At its most basic stage, the self is always already 'co-conscious'. Moreover, it always already finds itself (or is 'thrown', to use the Heideggerian terminology) in a social *practice.* My wariness of direct perception precisely stems from the conviction that our interactions with others cannot be explained in terms of capacities that are purely *individual or intrasubjective*. Of course I am not denying that there are such things as individual capacities, but the important question is how we acquire them.

Directness versus development

Whereas the term 'perception' seems to be unsuitable for an account of social interaction because it suggests the primacy of the first person perspective, the term 'direct' has the drawback of suggesting that social interaction is never *problematic*, since there is always an immediate and direct understanding of the other. For TT and ST accounts of intersubjectivity, our understanding of other minds is always indirect and deeply problematic. Consequently, we need a lot of theoretical back-up in order to survive our social encounters, and our success on this score is measured by our ability to predict others' behavior. Gallagher, by contrast, seems to suggest that social sense-making is in principle easy and effortless from the very moment we are born. He argues that what is important about direct perception is not 'what directness means, but how smart, how richly informed, it is. The smarter the perception is, the more work it does; the dumber it is, the more it requires extra cognitive processes (theory, simulation) to get the job done' (p.538). Unsurprisingly, the kind of perception Gallagher wishes to promote is very smart and richly informed.[82] He claims that 'practically speaking, direct perception, etc. delivers what I need to interact with others most of the time. In the broad range of normal circumstances there is already so much available in the person's movements, gestures, facial expressions, and so on, as well as in the pragmatic or social context, that I can grasp everything I need for understanding in what is perceptually available' (Gallagher 2007, p.540).

But how and where did direct perception become so incredibly smart? This problem appears to be a direct consequence of Gallagher's intention to model intersubjective knowledge on first-person perception. Admittedly, Gallagher states that one of the sources of intelligent perception is social experience, which fine-tunes our sensory-motor neuronal systems. Also, he acknowledges that direct perception gains in intelligence as infants develop, acquire language, conceptual competency and narrative competency. 'There is no doubt that advances associated with language and concept acquisition will transform perceptual experience, and specifically along lines that are pragmatic and intersubjective, some of which are already traced out in early non-conceptual experience' (p.538).

But it is not clear *how* this works. This is probably why Gallagher stresses that even creatures without much experience, infants being the paradigm example, already display the kinds of skills that are representative for smart perception. According to Gallagher, infants have an *inborn drive* for social interaction.

However, we should be very careful with such an appeal to innateness. The fact that something might be innate does certainly not imply that we do not have to come up with a proper *explanation* of it. Literally, the demarcation line for what is innate and what is not depends on the instance of birth. However, although birth marks a fundamental transition point in the infant's development, the exact timing is relatively arbitrary - the proper time of

[82] Gallagher account of direct perception has many similarities with Gibson's (1979) account of direct perception. But Gallagher also remarks that, despite the fact that he favors a Gibsonian-style account of perception without inference or representation, he does not deny that the organism has something to contribute to the shaping of perceptual information. In other words, Gallagher's strategy is to show how we can follow Gibson's lead and deny the necessity of inference, while at the same time allowing for internal processing to explain how we perceive environmental properties.

birth has a bandwidth of some weeks. Besides, the infant's development does neither start nor stop there. 'Innate' does not stand in opposition to 'development'; it only indicates an earlier development in the womb. And even in this stage, the development of the fetus is not a stubborn mechanical unfolding: it depends on 'favorable circumstances' and the fetus can be severely handicapped if it is deprived of these. Even a fetus cannot be regarded separately from the special 'environment' it interacts with. In other words, to invoke the magical word 'innateness' does not free one from the job of explaining the developmental 'how'. Neither can it be taken to insure some monadic individual capacity, for the development prior to birth is just as much a relational process as it is after birth. Therefore, I think that De Jaegher (2009) is correct in her observation that 'working out a detailed account of social interaction's role in interpersonal understanding is the central element of the story of social cognition. It will allow the issue to move away from the terms of the debate set by TT and ST and followed by direct perception [...] and towards a story that explicitly connects meaning and social interaction' (p.538).

A clear focus on development helps us to take into account the phenomenological fact that social understanding in fact can be *difficult*. The interpersonal abyss as assumed by TT and ST certainly does not do justice to our pervasive experiences of mutual contact and immediate understanding of the persons close to us. On the other hand, social understanding is not always smooth and direct. This may be easier to appreciate if we take into account that, from a developmental perspective, social misunderstandings are not considered to be essentially problematic. Rather, they offer crucial *opportunities for learning*. Social learning takes place in especially those situations in which our perception is not direct, and where we are uncertain of how to proceed. De Jaegher (2009) suggests this when she says that 'Failures in understanding another's behavior are not exceptional. On the contrary, they form part and parcel of the ongoing process of social understanding. More even, misunderstandings are the pivots around which the really interesting stuff of social understanding revolves. In these instances where coordination is lost, we have the potential to gain a lot of understanding' (p.540). Discontinuity in social interaction thus leads to learning and eventually opens up new venues for intersubjective understanding.