COGNITIVE SCIENCE SOCIETY July 25-26, 2005, Stresa, Italy, pp. 87-95

TOWARD SOCIAL MECHANISMS OF ANDROID SCIENCE
A COGSCI 2005 WORKSHOP

The Body-in-Motion and Social Scaffolding: Implications for Human and Android Cognitive Development

Jessica Lindblom (jessica.lindblom@his.se)
School of Humanities and Informatics, University of Skövde, Box 408
541 28 Skövde, SWEDEN

Tom Ziemke (tom.ziemke@his.se)
School of Humanities and Informatics, University of Skövde, Box 408
541 28 Skövde, SWEDEN

Abstract

Embodiment has become an important concept in many areas of cognitive science during the past two decades, but as yet there is no common understanding of what actually constitutes embodied cognition. Much focus, not least in embodied AI and robotics, has been on what kind of 'bodily realization' is necessary for embodied cognition, but crucial factors such as the role of social interaction and the "body-in-motion" have still not received much attention. We argue that the intertwining of social scaffolding and self-produced locomotion behavior is fundamental to the development of joint attention activities and a 'self' in the human child. We also discuss the implications of the social dynamics of bodily experience for android science. We further argue that keeping scientific and engineering perspectives apart, but also understanding their relation, is important for clarifying the objectives of android science, and not least for the public perception of this type of research. In particular, we address these issues from the perspectives of cognitive modeling and human-robot interaction.

Introduction

Theories of *embodied cognition* offer a radical shift in explanations of cognition, and can be viewed as a Copernican revolution against standard computationalist cognitive science. Roughly speaking, the embodied approach stresses that our cognitive processes depend on experiences that come from having a body with particular sensorimotor capabilities that interact with the surrounding physical and social world. Accordingly, much research in AI has shifted focus to robots, in particular to 'human-like' robot bodies in attempts to construct 'human-like' AI. However, despite nearly two decades of research under the banners of embodiment and embodied cognition, there is still no common understanding of what actually constitutes embodied cognition, or what kind of body it might require (e.g., Anderson, 2003; Chrisley & Ziemke, 2003; Clark, 1999; Núñez, 1999; Wilson, 2002, Ziemke, 2003).

Consequently, it is equally unclear whether or not humanoid/android robotic embodiment is actually 'human-like' in any sense relevant to embodied cognition. The lack of agreement has resulted in some severe oversimplifications of the role of the body in cognition. Firstly, the discussion has mostly focused on the "static" body in itself, i.e. what kind of physical or software 'realization' of the body is necessary for human intelligence. The crucial role of the *body-in-motion,* however, has received very little attention, although research in anthropology has shown the relevance of locomotion experience for human cognition (e.g., Farnell, 1999; Sheets-Johnstone, 1999). Secondly, the focus has mostly been on the relation between the individual body and individual cognitive processes, but the view of the mind as first and foremost *social* has largely been neglected (cf. Vygotsky, 1978), even in social and epigenetic robotics research. Despite claims to the contrary, current theories of embodied cognition and AI only peripherally address the role of embodiment in *social* interaction (Lindblom & Ziemke, 2003), although nearly two thirds of the information in a social situation is considered to be derived from so-called "non-verbal signs" or "body-language" (Burgoon, Buller, & Woodall, 1996). The distributed cognition approach proposed by Hutchins (1995), for instance, treats social interactions as directly observable cognitive events, as well as the materials involved in these interactions. However, whereas the body is regarded as one of the medium of the information flow, the main focus is not in particular on the role of the body in social interaction, but the transformation of information through different media at a more general level. Theories of cultural and social cognition, on the other hand, still mainly overlook the *bodily* aspects of social interaction (Rogoff, 2003; Tomasello, 1999).

This paper aims to extend the present discussions of embodiment, emphasizing that the "*body-in-motion*" is crucially relevant to the emergence of the capacity for joint attention and understanding others as intentional agents, which are central building blocks of human social interaction.

The rest of the paper is structured as follows. The next section emphasizes the role of socio-cultural factors for the development of human cognition, primarily following Tomasello's (1999) lines of argument. Next, we discuss different theoretical standpoints that stress the crucial relevance of "body-in-motion" and social scaffolding for cognition**.** We also present some empirical evidence that stresses the importance of self-produced locomotor behavior for the onset of joint-attention abilities in the human infant. Then we address implications of social dynamics of bodily experience for androids from two perspectives, namely cognitive modeling and human-robot interaction.

Cultural Cognition and the Nine-Month Revolution

The ability to engage in social interaction is a crucial building block of social life and cognition, and thus one of the foundations for human culture.

Humans "identify" with their conspecifics more deeply than other primates and the human child has a biologically inherited capacity for living culturally (cf. Rogoff, 2003; Tomasello, 1999). Early on human infants display a large number of activity patterns that appear to be species-unique. For instance, the typical rhythm of "burst-pause-burst" during breast-feeding does not occur in other primates. Moreover, human infants show a wide range of facial expressions, rhythmical stereotypes, and complex face-to-face interaction patterns between infant and caregiver that are absent in chimpanzees and gorillas (Hendriks-Jansen, 1996). That means, human infants are "ultra" social already from birth, in a way that other primates are not, and the role of these social interaction patterns is supposed to "hijack" the caregiver's attention in order to create a 'social glue' between infant and caregiver during the infant's development (Hendriks-Jansen, 1996). However, these early uniquely social bonding behaviors alone cannot explain why humans are able to "identify" so strongly with others. There has to be something more. Tomasello (1999) suggests that only humans are also able to understand other persons as *intentional* agents like themselves, i.e. "animate beings who have goals and who make active choices among behavioral means for attaining those goals, including active choices about what to pay attention to in pursuing those goals" (ibid., p. 68). This understanding emerges when human infants begin to participate in various joint attention activities (Tomasello, 1999).

It has been noted that Euro-American children begin to participate in social discourse from about the age of nine months at which point they make their first attempts to share attention with other people, as well as imitatively learn from and through social interactions with them. These newly developed 'joint attention activities' represent the emergence of the unique human social ability to deeply identify with others (Tomasello, 1999).

The range of new social behaviors that emerge at this point in infant development indicate a drastic change in the way the child begins to understand the surrounding (physical and) social world – the so-called "*nine-month revolution*". Before that time, the interaction behaviors of human children are mostly *dyadic*, i.e. two-way interactions between the child and the caretaker. Then by the age of nine months, a set of *triadic* behaviors emerges, involving a coordinated interaction between child, objects, and other people. As a result, a referential triangle of 'shared attention' develops in the child. The referential triangle includes another person and the object or event on which they focus their attention. Tomasello (1999) emphasizes that these new triadic behaviors are the result of the unique human social-cognitive adaptation to identify and understand others as intentional agents. He claims that it is this particular ability, and not any specialized biological adaptations explicitly, that is responsible for many, if not all, of the most unique and essential cognitive functions and processes of human being. However, the question is – *Why* does this revolution of joint attention behaviors occur at the age of nine months?

Tomasello (1999) suggests that the relation between self-understanding and the similar understanding of others as intentional agents as oneself explains the nine-month social-cognitive revolution, since "the hypothesis is that as this new experience of self-agency emerges, a new understanding of others emerges as a direct result" (ibid., p. 70). However, while that describes *what* happens around nine months of age, it does not explain *how* and *why* this shift in understanding actually occurs. Tomasello admits that the personal experiences necessary for this understanding remain unclear, which raises another related question – *How* does this link between self and others emerge? Tomasello emphasizes that in coming to understand others as intentional agents around the age of nine months, another crucial factor enters the scene – the ability to more or less simulate the other person's intentional actions by analogy to one's own actions, and as a result, the self becomes intentional. Tomasello stresses that there is no need for the child to be able to conceptualize before simulating, since it is enough to perceive the other person's intentional actions via an analogy to the self.

Simulation of another individual's point of view is achieved by matching the other person's mental states with a resonance state of one's own, putting oneself in another person's 'shoes' by simulating the behavior of another individual 'off-line', in order to predict or determine the behavior of the other agent. Gallese *et al.* (2002, p. 459) suggested "that the capacity to empathize with others – may rely on a series of matching mechanisms that we just have started to uncover". Such a mechanism may rely on, or be a part of special kinds of visio-motor neurons in the premotor cortex in monkeys, so-called *mirror neurons* (cf., e.g., Gallese & Goldman, 1998; Rizzolatti *et al*., 2002). These neurons are able to respond, for example, both to performed particular hand actions, and when observing the same action while it is performed by other conspecifics (Gallese & Goldman, 1998). For that reason, mirror neurons are supposed to constitute a cortical system, which is able to fit observation and execution of *goal-related* motor actions. Empirical evidence indicates that such a system actually is present in human beings as well, and the functional role of this matching system might be a part of, or a precursor to, a general mind-reading capability. Recent empirical results indicate that mirror neuron activity also correlates with action understanding as well as experiential understanding of others' emotions (Gallese, Keysers & Rizzolatti, 2004; see also Jacob & Jeannerod (2005) for a critique of motor theories of simulation).

However, the idea of simulating the other person's view for understanding that other people also have intentions results in the question – *How* does the child create and distinguish between its own first-hand experience of own actions and third-hand experiences of actions performed by others? This is an underestimated problem in theories of mind-reading, and has not received enough attention, despite the fact that the ability to shift between first-hand and third-hand perspectives is an essential aspect of social cognition (Jacob & Jeannerod, 2005).

The 'explanation' offered by Tomasello (1999) is that the time when the child starts to understand that other persons have intentions and goals like themselves, is a result of our species' "ultra" social ability. On the contrary, we suggest that neither our "ultra" social ability nor simulation theories alone are able to explain how this intentional understanding emerges in the child. Instead, we argue that self-initiated and self-experienced locomotion behavior is another missing piece in the puzzle for the emergence of the social understanding of the self. In the following section we elaborate this hypothesis in more detail.

"Body-in-Motion" and Social Scaffolding

Trevarthen (1977) pointed out that one reason for the neglect of the moving body in psychological research was that the actual movement patterns of humans were as difficult to observe before the invention of cinephotography as were the planets before the development of the telescope. Psychology therefore became more of a static science of perception, cognition and action than a science of dynamic interactions. On the other hand, when researchers actually pay attention to embodied movement, it often appears that, as Farnell (1995) puts it, the moving body has lost its mind. However, a shift in the study of human body movements has occurred more recently, from a distal observer's description of behavior to the stance of viewing body movements as dynamically embodied actions (Farnell, 1995, 1999).

The French philosopher Merleau-Ponty (1908-1961) strongly emphasized that the mind was essentially embodied and constantly interacting with the world, arguing that bodies are deeply 'cognitive' in themselves (cf. Dreyfus, 1992; Loren & Dietrich, 1997; Priest, 1998). On the other hand, Sheets-Johnstone (2003) emphasizes that although Merleau-Ponty is viewed as the "knight of the Body", he overlooked the deeply engrained role of self-experienced movement in embodied beings. She claims that the core of being is the relation between the body and movement, emphasizing that "consciousness does not arise in *matter*, it arises in organic forms, forms that are *animate*" (ibid. p. 43). The human infant is not born inanimate, but already moving, and has to catch herself in the tactile-kinesthetic apprenticeship of her own body. That means, there is a need to discover how we actually "put ourselves together" (Sheets-Johnstone, 2000). On the other hand, what both Merleau-Ponty and Sheets-Johnstone overlook, in our opinion, is the first and foremost *social* nature of the human mind, i.e. the apprenticeship of body-in-motion is not an individual enterprise (Varela, 1994). Embodiment is more than the organism or the "packaging", more than the experience of doing - there is the movement itself, which is more than just manipulating limbs, since "the body is both a means and the end of communicational intentions" (Varela, 1994, p. 168), and this primacy of the body-in-motion entails both language and gesture.

Varela (1994) suggests that a reliable theory of embodiment has to acknowledge the dynamic nature of human action, including the person that enacts the body, all kinds of physical and social actions, as well as meaning accomplished through actions. The main idea is that neither bodies nor minds themselves have intentions; it is only a person, a "self" or an *intentional* agent, in Tomasello's vocabulary, that has intentions. The point Varela wants to make is that the "enactment" of the body is a social act, and in order to direct oneself, you have to consider how others will act and react in response to your own actions.

How then does a movement become transformed into an intention or an embodied action? There are so-called "objective" descriptions of observed bodily movements, but they are un-convincing since they do not consider the non-observable social situation at hand, which actually is what gives the meaning to the visible embodied actions. By using the term 'action' instead of 'movement', Farnell (1995) highlights that socially embodied actions are a set of movements that have agency, meaning or intentions for the actual person or agent in view of the fact that "*bodies do not move and minds do not think – people just do*" (ibid., p. 14).

The role of social interactions for the transformation of movements into intended actions was illustrated already by Vygotsky in the mid-1930s, when he explained the essential role of social interactions for the development of pointing in the child (Vygotsky, 1978). Initially, it is only a simple and incomplete grasping movement directed towards a desired object, and is only constituted by the child's bodily movements, and nothing more. When the caretaker assists the child, the meaning of the situation itself changes. The child's 'failed' reaching attempt provokes a reaction, not from the desired object, but from another person. The individual gesture 'in itself' becomes a gesture 'for-others'. The caretaker interprets the child's reaching movement as a kind of pointing gesture, resulting in a socially meaningful communicative act, whereas the child at the moment is not aware of its communication ability. After a while, however, the child becomes aware of the communicative function of the performed movements, and then begins using referential gestures towards other people, rather than to the object of interest that initially was the child's primary focus. For that reason, "*the grasping movement changes to the act of pointing"* (Vygotsky, 1978, p 56). Kozulin (1986) pointed out that it is essential to note that the child herself is the last person who 'consciously' grasps the 'new' meaning of this pointing gesture.

That means, the social surrounding functions as a social scaffold for the development of pointing, where the initial quite simple bodily movement becomes an *intentional action.* Thus, our embodiment constrains while cultural customs affect, but do not determine, the organization of social interactions (Farnell, 1999).

Self-Produced Locomotion Behavior

The experience of self-produced locomotion behavior is a rather neglected factor, despite the fact that research has shown its significance in the child's social as well as emotional development (Campos *et al,* 2000). It should be stressed, however, that locomotion is not necessarily a causal factor in itself. Instead, the child's cognitive and emotional development emerges from the *experiences* that result from the child's own locomotion behavior. When the human child starts to locomote voluntarily, i.e. crawling and creeping, these behaviors produce a wide range of changing experiences in the infant's social and emotional development (Campos *et al*., 2000). The role and relevance of this new social interaction situation should not be disregarded. It becomes necessary for the child to adapt to the new situation, paying close attention, both to their environment as well as to its self-produced movement with respect to the environment. As a result, some pervasive consequences occur, which in turn, affect the physical and social world around the child, in particular the *interaction* between the child and its surroundings (Campos *et al*., 2000).

In discussing differences between children with and without self-produced locomotion experience, Campos et al. use an analogy based on a French saying, which states that "when the finger points at the moon – the idiot looks at the finger". On the whole, their empirical data suggests that children without self-produced locomotion experience perform like the 'idiot' in the French saying, whereas children with locomotion experience are able to follow, to various degrees, referential gestures towards a distal target.

Hence, their proposal is that crawling is the cradle of the "social referencing phenomenon", since it is mainly after the child starts to crawl that she receives social signals that have an obvious distal referent. When the child begins to locomote there is a sudden increase of the behavioral pattern of checking back and forth to the caregiver. This behavior is a crucial feature of the "information-seeking" aspect of social referencing, which makes it possible for the child to understand how the regulation of social interaction is affected by distal communication. Hence, it is via these regulations of interaction that the child develops a shared meaning with its caregiver, and at around 9 months the child is able to respond to gestural communication when the target is absent from its own visual field (Campos *et al*., 2000). That means, at that point in time the child is able to differentiate between its own visual field and the gesturer's visual field. In other words, the child displays a beginning for *perspective taking.*

This social ability then develops further and encompasses communicational signs from others, which make it possible for the child to grasp that other people also have intentions. For example, creeping and crawling infants appear both to be more attentive and actively searching for communicative signals from the experimenter while performing Piaget's well-known "A-not-B-task" (Campos *et al*., 2000). In addition, empirical research shows that infants with locomotion experience perform better on tasks assessing the tendency to follow referential gestural communication than pre-locomotor children (i.e. gaze-following, head-turn, and pointing). Similar results were shown in studies on Chinese children (they begin to locomote later than Euro American children due to cultural factors) and infants with motor disabilities (Campos *et al.*, 2000). In sum, empirical evidence show that there is a significant developmental change in referential gestural communication around the age of 9 months, and that self-produced locomotion experience is involved in that particular shift.

The crucial issue here is the role and relevance of self-produced locomotion behavior for the emergence of the "nine-month revolution", the point in time when the child begins to understand that others are intentional beings like itself. Hence, that point in development when children begin to understand themselves and others as *intentional* agents, around nine months of age (in European-American children), 'coincides' with the onset of self-produced locomotor behavior. We suggest that this is in fact no coincidence at all. Instead, it is primarily through the experience of self-produced locomotion and the subsequent experience of literally *perceiving* the (physical and social) world and *acting* upon it from different perspectives, depending on one's own embodied action, that infants develop the capacity of understanding others as having different perspectives and own intentions. That means, when children begin to locomote by themselves, they acquire an individual experience of the surrounding world through their own actions and perceptions.

As a result, the child distinguishes between itself and the surrounding world, a distinction from which a primitive "self" emerges. Consequently, when the child can put itself in another person's physical position, the child becomes able to relate both to the other person's perspective and its own situation. This perspective-taking is grounded in the experiences of self-produced locomotion behavior, which might be a fundamental aspect for distinguishing between first-hand and third-hand experiences. This emerging understanding is bootstrapped through socially scaffolded bodily experience, which gives the child access to the actual meaning of the social-communicative situation. Subsequently, that understanding of perspective-taking might be used during *embodied simulations,* making it possible for the child to simulate "off-line" what it would be like to be in the other person's situation, based on its own self-produced locomotor experience. In that sense, the sensorimotor and social dynamics of bodily experience function as a crucial driving force in cognitive development, as discussed in more detail elsewhere (Lindblom & Ziemke, 2005).

If we direct our attention back to androids, the following question can be raised – *What* implications, if any, do the (social) dynamics of bodily experience have for androids?

Implications for Android Science

As several authors have pointed out, robotics research, or artificial intelligence research in general, can be viewed from at least two different, though intertwined, perspectives: that of engineering, mostly concerned with the design of interactive systems, and that of science, mostly concerned with the understanding of natural systems. This is, obviously, also the case for android research, which on the one hand can be viewed as an approach to human-robot interaction (HRI) or, more generally, building better socially interactive technology. On the other hand, it can be considered as an approach to developing powerful tools for cognitive-scientific modeling (e.g. MacDorman & Ishiguro, 2004). The question addressed in previous sections, i.e. the role of bodily experience/dynamics in the development of the self and the capacity to recognize others as intentional agents is, in our opinion, relevant to both of these perspectives, as will be elaborated in the following. We believe that keeping the different perspectives apart, but also understanding their relation, is important for clarifying the objectives of android science, and not least for the public perception of this type of research. First, let us have a look at the scientific modeling perspective.

Scientific Modeling Perspective

Cognitive modeling and AI research have a long-standing tradition of considering, very roughly speaking, two different, but in some cases overlapping approaches to building artificial minds (or models of mind). One that puts together relatively complex systems with certain cognitive capacities more or less manually (constructing robots, writing computer programs, etc.), and, another one that builds somewhat simpler, but adaptive systems that at least to some degree self-organize their own cognitive capacities (using various computational learning techniques, etc.). This principal distinction of approaches can be traced back at least to Turing's seminal 1950 paper on *Computing Machinery and Intelligence*, in which he realized the difficulties of attempting to program an adult-like artificial mind and envisioned as a possible alternative so-called "child machines", equipped with "the best sense organs that money can buy", whose education "could follow the normal teaching of a child". Again, android research can be said to combine elements of both approaches, since obviously to some degree physical robot bodies need to be pre-built, before computational learning/development can start.

We can then further ask whether android science is intended to be, in Searle's (1980) terms, a "strong AI" that builds actual human-like minds with actual mental properties, i.e. actual (original) intentionality etc., or is it intended to be a "weak AI" that builds better models of human minds than what might be possible with other types of tools/models, such as non-android humanoids or non-robotic computational models.

When it comes to "strong AI", we have argued elsewhere in detail (e.g., Lindblom & Ziemke, 2003; Ziemke, 2001a, 2001b) that humanoid/epigenetic robotics, as compared to other robotic or computational modeling, is not making any major steps forward towards actual artificial minds, because it still makes an essential distinction between the physical body that is pre-built and the computational mind supposed to be developed in social interaction with, e.g., human caretakers. The example discussed in the previous section, of the role of crawling/creeping in the development of self and understanding the intentionality of others, illustrates the basic problem (a 'catch 22', one might say) of all robotic approaches to "strong AI", including android robotics with a "strong AI" ambition. If cognitive development (of self, intentionality, etc.) in some sense is dependent upon or emerges from bodily development, and the development of the necessary skills for controlling the body, then the division into physical body and computational mind, or into a physical construction phase and a (socially scaffolded) computational development phase is not a viable approach to developing actual artificial minds with original intentionality, and so on.

When it comes to android science as a "weak AI", i.e. the modeling of human minds rather than building artificial ones, the crucial question is perhaps whether or not androids are more powerful tools than other robots, e.g. non-android humanoids. This is, of course, an empirical question, and time will tell which approach provides the best models, but let us speculate a little anyway. A second question then is exactly what it is that android science provides a better model–of? MacDorman & Ishiguro, (2004), for example, define an android as "an artificial system that has humanlike behavior and appearance and is capable of sustaining natural relationships with people".

It is quite obvious that a human-like appearance makes a crucial difference from the engineering (or human-robot interaction) perspective (discussed in more detail below), i.e. it makes a crucial difference for the android's potential human users or collaborators. But does it make a difference to the scientist modeling human cognition and behavior? We have argued above that, when it comes to the development of self, intentionality, etc., androids have shortcomings just as other, non-android robots. That means, they might very well be more likely to be attributed with human-like intentionality by humans (including human cognitive modelers), but they are unlikely to develop (and thus serve as a model of) human-like intentionality themselves.

Hence, the major contribution of android robotics as a "weak AI" is, in our opinion, that they can be used in more realistic experiments of social interaction with actual humans, just because they elicit more natural responses, i.e. responses more similar to those in actual human-human social interaction. In that sense, the contribution of androids to cognitive-scientific modeling might not so much lie in themselves actually being better models (of human self, intentionality, etc.), but in their contribution to developing better models of human-human social interaction.

Human-Robot Interaction Perspective

Let us then also consider the "engineering" perspective, i.e. android robotics as an interactive design approach to more useful human-robot social interaction, or more generally, improved socially interactive technology (cf., e.g., Benyon, Turner & Turner, 2005; Picard, 1998; Preece, Rogers & Sharp, 2002). It is here that we believe the main contribution of android science lies, since there are some major benefits in designing technology that supports the remarkable human sensitivity of social interaction. Today, many of the outwardly visible and recognizable patterns of joint attention are mimicked and displayed in robotic systems (cf. e.g., Imai, Ono & Ishiguro, 2003; Kozima & Yano, 2001), which humans are able to grasp rather easily. Hence, the more natural the interaction with androids or any kind of technology , the more useful it will be for humans in their daily life, and androids then might be a more suitable design solution than mechanical-looking robots. In our opinion, the main benefits and drawbacks are the following.

Generally, we consider that similarities in bodily-shape, appearance and expressive behavior of androids offer a number of advantages for social interaction between humans and technology. For that reason, the user's acceptance of human-robotic systems might be accomplished much more easily with androids. Moreover, the human-like morphology is also well suited to function in the human physical environment, as it is situated in our homes, manipulates our tools, and performs real physical tasks. For those reasons, it might feel more enjoyable and fun to interact with these robots. Moreover, human-android interaction might provide an increased subjective feeling of the quality of the interaction itself, and this will have tentative advantages concerning user satisfaction (cf. Benyon, Turner & Turner, 2005). Given that humans are "ultra" social animals, and consequently experts in social interaction, the need for costly training programs in order to educate humans to use interactive systems would hopefully decrease. However, future evaluations of HRI will shed more light on these issues.

It should be noted, however, that human daily life is a highly complex web of tasks and social skills. In order to fully function as 'truly' human-like social interaction partners, androids have to "understand" our intentions if they are to assist and help us properly. This in turn requires that androids are able to interpret our intentions by recognizing bodily movements and accomplishing joint attention.

The optimal solution would be to engage robots in various sorts of social learning processes, such as imitation and co-operative learning, and to teach the robot in the same way as when instructing another person or a child. Hence, there seems to be a demand for a kind of "socialization" or "enculturation" process of androids, similar to the epigenesis of human children (Lindblom & Ziemke, 2003; Zlatev, 2001). However, in our opinion, it is difficult to achieve the human kind of social learning in androids, because the quest for socially guided learning in androids suffers from the same shortcomings as discussed above for cognitive modeling in the sense of a "strong AI", i.e. the lack of 'intrinsic' intentionality. Although androids appear to express emotions and perform basic joint attention behaviors they do not actually experience these abilities themselves. The actual experience lies in the eyes of the beholder – namely the human user. That means, what actually "looks" like an intentional understanding in the robot, from an observer's point of view, has no correspondence in the robotic system itself. We propose that all of the implemented social learning techniques in current robotic systems are merely more or less reactive models with quite basic 'built in' responses, rather than human-like developmental models. It is, so far at least, just the human observer who interprets the behavior as 'intelligent' and 'meaningful'.

This implies that androids do not interact socially through their lived bodies in the sense that humans do, but rather via their bodily appearance. That means, there is a sharp contrast between our socially embodied interactions in the lived world and what is offered in contemporary androids, and, hence, the truly human-like appearance of androids is 'naturalistic' only from the human observer's point of view. At best, this bodily appearance enhances the interactive experience for the human, making it more enjoyable and pleasurable.

What consequences will that 'intrinsic' lack have on the view of HRI? On the one hand, we are able to "interpret" the android's behavior, but the other way around would be harder, since the robot hardly is able to understand and interpret *our* intentions. On the other hand, a very close human-like bodily appearance may offer a too promising impression about its intelligent capabilities, and the user may therefore be disappointed and not experience the system as useful and enjoyable to interact with.

However, if we weigh the pros and cons of using androids in HRI, we believe that regardless of whether android robots will be truly intelligent in the 'strong' sense, it is simply a fact that this type of technology allows humans to become more socially situated in the world of technical artifacts. That means, the real strength of HRI, or development of artifacts in general, in our opinion, is not its role as a "strong" robotic AI, but rather its potential to facilitate a more "natural" human-technology interaction, allowing humans to interact with artifacts in the same way they interact with each other.

Conclusions

This paper extends current theories of embodied cognition by including the role and relevance of *body-in-motion* for broadening the social mind. Crucial to the embodiment of cognition, according to this account, is perhaps not so much the physical realization of the static body, or its interactions with the environment as such. Instead, we have emphasized the elementary and intertwined relation between the experiences of one's own moving body and its interplay with the physical and social environment.

We suggest that the problem when constructing android systems, and humanoid robots in general, is that the major goal typically is to construct the end result, e.g. walking androids with certain cognitive and communicative capacities. For that reason, the importance of bodily and cognitive development occurring in parallel is often largely overlooked (cf. Lindblom & Ziemke, 2003; Vygotsky, 1978). Instead, it would be more interesting and cognitively plausible from an epigenetic point-of view, to actually build "infant androids", which then would be able to develop physically and cognitively like a human child.

But, for obvious technical reasons that is not possible with current technology. The point we want to make is that, in order to produce a human-like android mind, it is not enough to construct a human-like robot body and then let it develop cognitive capacities, because having a human-like bodily shape is not the same as human-like embodiment. Embodied cognition is not the sum of (physical) "bodily-shape" and (computational) "cognitive abilities" but emerges from *embodied cognitive development* shaped through the dynamics of (socially scaffolded) bodily experience of the physical and cultural world. To conclude, we believe that keeping the different perspectives apart, but also understanding their relation, is important for clarifying the objectives of android science, and not least for the public perception of this type of research.

Acknowledgments

The authors would like to thank Tarja Susi and Henrik Svensson for discussion and helpful comments.

References

Anderson, M. L. (2003). Embodied cognition: a field guide. *Artificial Intelligence,* 149, 91-130.

Benyon, D., Turner, P. & Turner, S. (2005). *Designing interactive systems – people, activities, contexts, technologies.* London: Addison-Wesley.

Burgoon, J. K., Buller, D. B. & Woodall, W. G. (1996) *Nonverbal communication: the unspoken dialogue.* New York: McGraw-Hill.

Campos, J. J., Anderson, D. I., Barbu-Roth, M. A., Hubbard, E. M., Hertenstein, M. J., & Witherington, D. (2000). Travel broadens the mind. *Infancy, 1(2),* 149-219.

Chrisley, R. & Ziemke, T. (2003). Embodiment. In *Encyclopedia of cognitive science.* London: Macmillian.

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Science, 3(9),* 345-351.

Dreyfus, H. L. (1992). *What computers still can't do – a critique of artificial reason*. Cambridge, MA: MIT Press.

Farnell, B. (1995). *Do you see what I mean? Plains Indian sign talk and the embodiment of action.* Austin: University of Texas.

Farnell, B. (1999). Moving bodies, acting selves. *Annual Review of Anthropology, 28,* 341-373.

Gallese, V. & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences,* 2(12), 493-501.

Gallese, V., Ferrari, P.F., Kohler, E. & Fogassi, L. (2002) The eyes, the hand, and the mind: behavioral and neurophysical aspects of social cognition. In M. Bekoff, C. Allen and M. Burghardt (eds.) *The cognitive animal – empirical and theoretical perspectives on animal cognition* (pp. 451-461). Cambridge, MA: MIT Press.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition*. Trends in Cognitive Sciences, 8 (9),* 398-403.

Hendriks-Jansen, H. (1996). *Catching ourselves in the act – situated activity, interactive emergence, evolution, and human thought.* Cambridge, MA: MIT Press.

Hutchins, E.(1995). *Cognition in the wild.* Cambridge, MA: MIT Press.

Imai, M., Ono, T. & Ishiguro, H. (2003). Physical relation and expression: joint-attention for human-robot interaction. *IEEE Transactions on Industrial Electronics, 50 (4),* 636-643.

Ingold, T. (2000). Evolving skills. In H. Rose and S. Rose (Eds.), *Alas, poor Darwin: arguments against evolutionary psychology*. New York: Harmony Books.

Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends in Cognitive Sciences, 9 (1),* 21-25.

Kozima, H., & Yano, H. (2001). A robot that learns to communicate with human caregivers. In *Proceedings of the First international Workshop on Epigenetic Robotic.* Lund University Cognitive Studies, vol. 85, Lund: Sweden.

Kozulin, A. (1986). Vygotsky in context. In L.S. Vygotsky's *Thought and language.* Cambridge, MA: MIT Press.

Lindblom, J. & Ziemke, T. (2003). Social situatedness of natural and artificial intelligence: Vygotsky and beyond. *Adaptive Behavior, 11(2), 79-96.*

Lindblom, J. & Ziemke, T. (2005). Body-in-Motion: broadening the social mind. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society.* July, 21-23, Stresa, Italy.

Loren, L. A. & Dietrich, E. (1997). Merleau-Ponty, embodied cognition, and the problem of intentionality. *Cybernetics and Systems,* 28(5), 345-358.

MacDorman, K. F. & Ishiguro, H. (2004). The study of interaction through the development of androids. IPSJ SIG Technical Reports 2004-CVIM-146 (pp. 69-75), 2004(113). November 11-12, 2004. Tokyo,Japan.

Núñez, R. (1999). Could the future taste purple? *Journal of Consciousness Studies, 6(11-12),* 41-60.

Picard, R. W. (1998). *Affective computing*. Cambridge, MA: MIT Press.

Preece, J., Rogers, Y. & Sharp, H. (2002). *Interaction design – beyond human-computer interaction.* New York: Wiley.

Priest, S. (1998). *Merleau-Ponty*. London: Routledge.

Rizzolatti, G., Fadiga, L., Fogassi, L. & Gallese, V. (2002). From mirror neurons to imitation: facts and speculations. In A. N. Meltzoff and W. Prinz (Eds.), *The imitative mind – development, evolution, and brain bases.* Cambridge: Cambridge University Press.

Rogoff, B. (2003). *The cultural nature of human development.* New York: Oxford University Press.

Searle, J. (1980) Minds, brains and programs. *Behavioral and Brain Sciences*, 3, 417-457. Sheets-Johnstone, M. (1999). *The primacy of movement.* Amsterdam: John Benjamins.

Sheets-Johnstone, M. (1999). *The primacy of movement.* Amsterdam: John Benjamins.

Sheets-Johnstone, M. (2003). Answering the challenges of animation: response to Crease's review essay. *Phenomenology and the Cognitive Sciences, 2,* 84-93.

Tomasello, M. (1999). *The cultural origins of human cognition.* Cambridge, MA: Harvard University Press.

Trevarthen, C. (1977). Descriptive analysis of infant communicative behaviour. In H.R. Schaffer (ed.), *Studies in mother-infant interaction Proceedings of Loch Lomond Symposium* (pp. 227-270). New York: Academic Press.

Turing, A. (1950) Computing machinery and intelligence. *Mind,49,* 433-460.

Varela, C. R. (1994). Harré and Merleau-Ponty: beyond the absent moving body in embodied social theory. *Journal for the Theory of Social Behavior, 24(2),* 167-185.

Vygotsky, L. S. (1978). *Mind in Society: the development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & review, 9(4),* 625-636.

Ziemke, T. (2001a) Are robots embodied? In *Proceedings of the First international Workshop on Epigenetic Robotic.* Lund University Cognitive Studies, vol. 85, Lund: Sweden.

Ziemke, T. (2001b) The construction of 'reality' in the robot. *Foundations of Science*, 6(1), 163-233.

Ziemke, T. (2003). What's that thing called embodiment? In R. Alterman and D. Kirsch (eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (pp. 1305-1310). Mahwah, NJ: Lawrence Erlbaum.

Zlatev, J. (2001). The epigenesis of meaning in human beings, and possibly in robots. *Minds and machines, 11,* 155-195.