Minding the Body
Interacting socially through embodied action

by

Jessica Lindblom
Abstract

This dissertation clarifies the role and relevance of the body in social interaction and cognition from an embodied cognitive science perspective. Theories of embodied cognition have during the past two decades offered a radical shift in explanations of the human mind, from traditional computationalism which considers cognition in terms of internal symbolic representations and computational processes, to emphasizing the way cognition is shaped by the body and its sensorimotor interaction with the surrounding social and material world. This thesis develops a framework for the embodied nature of social interaction and cognition, which is based on an interdisciplinary approach that ranges historically in time and across different disciplines. It includes work in cognitive science, artificial intelligence, phenomenology, ethology, developmental psychology, neuroscience, social psychology, linguistics, communication, and gesture studies. The theoretical framework presents a thorough and integrated understanding that supports and explains the embodied nature of social interaction and cognition. It is argued that embodiment is the part and parcel of social interaction and cognition in the most general and specific ways, in which dynamically embodied actions themselves have meaning and agency. The framework is illustrated by empirical work that provides some detailed observational fieldwork on embodied actions captured in three different episodes of spontaneous social interaction in situ. Besides illustrating the theoretical issues discussed in the thesis, the empirical work also reveals some novel characteristics of embodied action in social interaction and cognition. Furthermore, the ontogeny of social interaction and cognition is considered from an embodied perspective, in which social scaffolding and embodied experience play crucial roles during child development. In addition, the issue what it would take for an artificial system to be (socially) embodied is discussed from the perspectives of cognitive modeling and technology. Finally, the theoretical contributions and implications of the study of embodied actions in social interaction and cognition for cognitive science and related disciplines are summed up. The practical relevance for applications to artificial intelligence and human-computer interaction is also outlined as well as some aspects for future work.
Acknowledgements

Looking back at the past years, it is quite clear that the effort behind writing a thesis is a collective, and indeed a social experience.

First of all, I wish to express my sincere thanks to my excellent supervisor Professor Tom Ziemke for believing in my capability from the very beginning and for all the support and encouragement over the years.

I also wish to thank my colleagues and friends at the University of Skövde, especially Tarja Susi, Henrik Svensson, Jana Rambusch and Maria Nilsson, for sharing my time as a doctoral student. We have discussed all kinds of research matters, and you have given me insightful comments on research ideas, as well as been supportive and encouraging. Special thanks also to Beatrice Alenljung, for being such a good teacher, mentor, colleague and friend throughout the years.

I also want to express my gratitude to my faculty opponent, Professor Shaun Gallagher, and wish to thank the examination committee: Professor Lars-Erik Berg, Professor Nils Dahlbäck, and Docent Agneta Gulz. A special thanks to my co-supervisor Professor Hanna Risku for helpful comments on earlier drafts and for your encouragement. I also wish to thank Vera Lindroos for reading and commenting on this thesis as well as improving my English.

Thanks to all people that allowed me to use the video-recording from our visit in Montana as data for the empirical study, especially Bob BlackBull. Thanks also to my cousin Christine Olsson for designing the cover and her friendly support.

I also want to thank my friends and family outside the academic world. Here I particularly include those who spent some time with me when I needed to get away from the writing, including four-legged friends. The time spent with the horses has been a solace for the soul.

Last but not least, I owe most to my family, for being there, and I dedicate this thesis to my grandparents.
List of Publications

**Journal papers:**


**Book chapters:**


**Conference and workshop papers:**


**Other conference and workshop presentations:**


# Table of Contents

1. **Introduction** .......................................................................................................................... 1  
   1.1 Motivations and Aim........................................................................................................... 5  
   1.2 The Road Taken ................................................................................................................. 7  
   1.3 Body, Embodiment and Embodied Cognition: Common Misunderstandings ..................... 9  
   1.4 Thesis Road Map .............................................................................................................. 14  

2. **Body and Mind – A Historical Perspective** ...................................................................... 17  
   2.1 The Roots of Dualism - from Plato to Descartes ............................................................... 18  
   2.2 Darwin’s Work and its Legacy ........................................................................................... 19  
   2.3 Behaviorism ....................................................................................................................... 27  
   2.4 Beyond the Bounds of Behaviorism .................................................................................. 30  
      2.4.1 The field of ethology and the work of von Uexküll .................................................... 30  
      2.4.2 From the Gestalt school to Gibson and Merleau-Ponty ............................................ 35  
      2.4.3 The genetic epistemology of Piaget .......................................................................... 41  
      2.4.4 The cultural and historical approach of Vygotsky ...................................................... 46  
      2.4.5 Dewey and Bartlett .................................................................................................... 52  
      2.4.6 Mead’s theory of the socially interactive mind ......................................................... 55  
   2.5 The Fall of Behaviorism and the Rise of a Science of Mind ............................................. 61  
   2.6 Computationalism: The Disembodied Paradigm ............................................................... 66  
   2.7 Criticisms of Computationalism and the Inception of Connectionism ............................ 69  
   2.8 In Summary ....................................................................................................................... 74  

3. **Embodied Cognitive Science** ............................................................................................ 75  
   3.1 Putting Brain, Body and World Together Again .............................................................. 76  
   3.2 The Nature of Embodied Cognition ............................................................................... 78  
      3.2.1 Six views of embodied cognition ............................................................................. 83  
      3.2.2 Four complementary aspects of embodied cognition .............................................. 91  
      3.2.3 Some additional notions of embodiment ................................................................. 93  
   3.3 Body in Motion ................................................................................................................. 100  
   3.4 Toward the Social Dimension of Embodiment ............................................................... 104  

4. **Embodiment and Social Interaction** ............................................................................... 107  
   4.1 Embodied Social Psychology .......................................................................................... 108  
      4.1.1 Four general social embodiment effects .................................................................. 108  
      4.1.2 Social embodiment effects in attitudes, social perception and emotions ............... 111  
   4.2 Phenomenological Experience and Neurological Underpinnings of Embodiment .......... 117  
      4.2.1 Embodied simulations ............................................................................................ 119  
      4.2.2 Mirror neurons and the action-perception linkage ................................................. 125  
   4.3 Embodied Linguistics ....................................................................................................... 129  
      4.3.1 Language as grounded in embodiment ....... Language as grounded in embodiment 130  
      4.3.2 Gesture in language ................................................................................................. 134  
   4.4 Four Fundamental Functions of the Body in Social Interactions .................................. 143  
      4.4.1 The body as a social resonance mechanism ............................................................... 143  
      4.4.2 The body as means and end in social interaction ..................................................... 144  
      4.4.3 Embodied actions as a helping hand in shaping and sharing thoughts .................... 146  
      4.4.4 The body as a representational device ...................................................................... 147
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Characterizing and Studying Social Interaction and Cognition</td>
<td>150</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Social interaction in social insects</td>
<td>151</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Social interaction beyond social insects</td>
<td>153</td>
</tr>
<tr>
<td>5.2</td>
<td>Various Levels and Kinds of Human Social Interaction</td>
<td>156</td>
</tr>
<tr>
<td>5.2.1</td>
<td>The primary level</td>
<td>158</td>
</tr>
<tr>
<td>5.2.2</td>
<td>The secondary level</td>
<td>159</td>
</tr>
<tr>
<td>5.2.3</td>
<td>The tertiary level</td>
<td>160</td>
</tr>
<tr>
<td>5.3</td>
<td>Different Metaphors of Social Interaction</td>
<td>163</td>
</tr>
<tr>
<td>5.3.1</td>
<td>The information transmission metaphor</td>
<td>163</td>
</tr>
<tr>
<td>5.3.2</td>
<td>The dance metaphor</td>
<td>166</td>
</tr>
<tr>
<td>5.4</td>
<td>Revisited: Characterizing and Studying Social Interaction and Cognition</td>
<td>171</td>
</tr>
<tr>
<td>5.5</td>
<td>The Role of the Social Body in Motion in Cognitive Development</td>
<td>179</td>
</tr>
<tr>
<td>5.5.1</td>
<td>Initial cross modal interaction and the primitive ‘self’</td>
<td>181</td>
</tr>
<tr>
<td>5.5.2</td>
<td>The shift from dyadic to triadic interaction and the emergence of ‘self’</td>
<td>182</td>
</tr>
<tr>
<td>5.5.3</td>
<td>Symbolic and reflective interaction enter the scene</td>
<td>185</td>
</tr>
<tr>
<td>5.6</td>
<td>In Summary</td>
<td>189</td>
</tr>
<tr>
<td>6.1</td>
<td>Putting it all Together</td>
<td>192</td>
</tr>
<tr>
<td>6.2</td>
<td>A Framework for the Embodied Nature of Social Interaction and Cognition</td>
<td>194</td>
</tr>
<tr>
<td>6.3</td>
<td>General Remarks</td>
<td>201</td>
</tr>
<tr>
<td>7.1</td>
<td>Case Study: Empirical Design</td>
<td>203</td>
</tr>
<tr>
<td>7.2</td>
<td>Case Study: Analysis and Results</td>
<td>207</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Different herds of Spanish Mustang horses</td>
<td>207</td>
</tr>
<tr>
<td>7.2.2</td>
<td>They own me</td>
<td>219</td>
</tr>
<tr>
<td>7.2.3</td>
<td>Lending a hand in joint action</td>
<td>226</td>
</tr>
<tr>
<td>7.3</td>
<td>Case Study: Discussion and Conclusions</td>
<td>243</td>
</tr>
<tr>
<td>7.3.1</td>
<td>The relation between the theoretical framework and the empirical work</td>
<td>244</td>
</tr>
<tr>
<td>7.3.2</td>
<td>Additional empirical findings and conclusions</td>
<td>245</td>
</tr>
<tr>
<td>7.3.3</td>
<td>Issues concerning the analysis of the empirical data</td>
<td>247</td>
</tr>
<tr>
<td>7.3.4</td>
<td>Synthesis of the theoretical and empirical work</td>
<td>249</td>
</tr>
<tr>
<td>8.1</td>
<td>Reflections</td>
<td>253</td>
</tr>
<tr>
<td>8.2</td>
<td>Implications for Socially Interactive Technology</td>
<td>266</td>
</tr>
<tr>
<td>8.2.1</td>
<td>Scientific modeling perspective</td>
<td>269</td>
</tr>
<tr>
<td>8.2.2</td>
<td>Engineering and human-robot interaction</td>
<td>273</td>
</tr>
<tr>
<td>8.3</td>
<td>The Road behind Us - Contributions and Implications</td>
<td>276</td>
</tr>
<tr>
<td>8.4</td>
<td>The Road Ahead - Future Work</td>
<td>282</td>
</tr>
<tr>
<td>8.5</td>
<td>Closing Remarks</td>
<td>284</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>287</td>
</tr>
<tr>
<td>Appendix</td>
<td></td>
<td>307</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1. The organization of higher behaviour via an mediated act........................................... 48
Figure 2. The "me" is the social self, and the "I" is a response to the "me" in the construction of self................................................................. 60
Figure 3. Schematic illustration of a) structural coupling with the environment respectively b) another organism (TOC) .............................................................. 150
Figure 4. Indirect social interaction mediated through chemical substances in the environment. ........................................................................................................ 152
Figure 5. Image 5a illustrates that much work on socially embodied action has yet focused on isolated parts, and image 5b illustrates the integrated and holistic approach in this thesis that focuses on cross-modal interaction. ...................................................... 195
Figure 6. The including of other persons is not enough for us as observers to grasp what is going on................................................................................................. 196
Figure 7. Embodied actions are situated in cultural and material contexts................................. 197
Figure 8. The dynamical and temporal aspects of socially embodied actions are parts of the ongoing activity................................................................. 198
Figure 9. The main focus of analysis is embodied actions in social interactions and cognition is not blurred, as a way to represent that they are placed in the foreground. The aspects that represent the background are blurred....................................................... 199
Figure 10. The visiting group is sitting at Bob’s veranda while Bob is standing beside the vehicle. ................................................................................................. 207
Figure 11. Just before Bob changes the topic, and the beginning of the utterance “…and now we have” .................................................................................. 208
Figure 12. When Bob utters “…I have’, which continues with “…let’s see…” ....................... 209
Figure 13. When Bob starts saying “we have two herds up in Canada,” and his tapping with the index finger .................................................................................. 209
Figure 14. The touching of the middle finger and the moving of the middle finger during the pronunciation of “two herds” ....................................................................................... 210
Figure 15. The shift of gaze and the pointing gesture, and the continuance of the pointing gesture upward and backward while pronouncing “Salish-Kootenai” ............... 211
Figure 16. The beginning of the utterance concerning the Northern Cheyenne .................... 212
Figure 17. The shift in gaze and the wider movement of the middle finger in the Crow example. ........................................................................................................ 213
Figure 18. The semiotic gesture in the Fort Belknap utterance.................................................... 215
Figure 19. The shift in gaze during the utterance of “for the Assiniboine and the Gros Ventres” ........................................................................................................ 216
Figure 20. The first two images illustrate Bob’s bodily actions when talking about the future of the horses, and the last one shows his altered bodily actions when he has changed the topic........................................................................................................ 217
Figure 21. The change of topic is manifested in both bodily posture and focus of attention. 220
Figure 22. The topic to be discussed, the relationship between Bob and his horses, is expressed through his bodily posture, gesture and focus of attention........................... 220
Figure 23. The gesture performed during the verbalization of horses in the utterance “…how many horses do you own?” ........................................................................... 221
Figure 24. The gesture in form of bodily actions toward and from the thorax performed during the verbalization of “-they own me” ........................................................................ 222
Figure 25. The co-regulated bodily actions performed by Bob and Katrin while Katrin utterers “…How many horses own you?” ................................................................. 223
Figure 26. The bodily actions performed after the spoken dialogue........................................ 224
Figure 27. The fillies kept in the paddock for surveillance, and Bob patting the filly which is unable to rise by herself. ................................................................. 227
Figure 28. Bob’s and the filly’s lifting/rising attempt................................................................. 228
Figure 29. The first part of the initial lifting attempt by Bob, BoChief and Dag ..................... 229
Figure 30. The first part of the unsuccessful lifting attempt by Bob, BoChief and Dag. .......... 230
Figure 31. After the initial, but unsuccessful, lift by Bob, BoChief and Dag. ...................... 230
Figure 32. The first two frames (a-b) show Bob’s shift in pointing from Katrin to the filly, and the following ones (c-f) display Bob’s lifting gesture and then Dag’s accompanying gesture. ................................................................. 231
Figure 33. Dag’s repeated lifting gesture, which now both represents his coming grip as well as their first-hand as well as third-hand perspective. ............................................. 233
Figure 34. The successful lift and Bob’s huge sigh of relief in the final image. ..................... 235
Figure 35. The coordinated and co-regulated actions of Bob and Katrin’s turning around movements. .......................................................... 237
Figure 36. The integration of the different general dimensions of embodied actions in social interaction and cognition in situ...................................................... 250
Figure 37. Map over USA and Canada ................................................................................. 307
Figure 38. Map over Montana and Blackfeet country. ....................................................... 308
List of Tables
Table 1. Relations between embodiment and different kinds of movements. ....................... 138
Table 2. Schemes for classifying different kinds of social interactions ................................. 157
Table 3. Two metaphors of social interaction and cognition .................................................. 169
Good research, as well as increasing knowledge, poses more questions than it solves.

James D. Steel, 2004
Chapter 1

First of all, there is the belief that, in talking about human cognitive activities, it is necessary to speak about mental representations and to posit a level of analysis wholly separate from the biological or neurological, on the one hand, and the sociological or cultural, on the other.

Gardner, 1987

By using the term embodied we mean to highlight two points: first, that cognition depends upon the kinds of experiences that come from having a body with various sensorimotor capacities, and second, that these individual sensorimotor capacities themselves are embedded in a more surrounding biological, psychological and cultural context.

Varela, Thompson and Rosch, 1991

1. Introduction

What is the role and relevance of the body in social interaction and cognition? There is no single, simple answer to this question. As the introductory quotes reveal, in cognitive science there are completely different views of how to consider the issue. Is it the case that the body has no role at all, other than a mere implementation of a computational process as stated above by Gardner or does it play a crucial role in the shaping of the mind as the quote by Varela et al. suggests? The traditional view of social interaction in cognitive science has been that agents relate to each other in much the same way as they relate to other parts of the external world, that is by having more or less explicit internal representations of each other, which then are manipulated internally (cf. e.g. Augoustinos & Walker, 1995; Fiske & Taylor, 1984; Kunda, 1999). The most common, as well as still dominant, view of the role of the body in social interaction and cognition, in cognitive science, is as a trivial ‘appendage’ to the real intellectual language and mind. Therefore, bodily aspects are frequently addressed in terms of non-verbal communication, nonverbal behavior, or body language. However, it has been estimated that almost two thirds of the meaning in a social situation are conceived from these ‘non-verbal signs’ (cf. e.g. Burgoon, Buller & Woodall, 1996), whereas speech has been estimated to account for merely some few percent (cf. e.g. Mehrabian, 1972).

Nevertheless, how to describe and define the significance of the body in social interaction and cognition depends on one’s theoretical orientation.
According to my own point of view, which resembles that of Varela et al., this issue can be considered from an embodied cognitive science perspective. During the past two decades, theories of embodied, situated and distributed cognition have offered a radical shift in explaining the human mind. One might say a Copernican revolution within the cognitive sciences - from the traditional cognitivist perspective, (cf. Gardner’s quote above) which considers cognition in terms of internal symbolic representations and computational processes - to emphasizing the way cognition is shaped by the body and its sensorimotor interaction with the surrounding world (cf. e.g. Clark 1997, 1999a, 1999b; Dreyfus, 1972/1992; Gibbs, 2006; Hutchins, 1995; Lakoff & Johnson, 1999; Suchman, 1987; Varela, Thompson & Rosch, 1991). This is a reaction against the cognitivists’ computer metaphor of mind, which is a centralized view of cognition taking place inside the skull with the body only serving as some kind of input and output device, i.e. a physical interface between an internal program (cognitive processes) and an external world. Thus, embodiment has become a much discussed concept (Anderson, 2003, in press; Wilson, 2002; Ziemke, 2003) which many regard, together with situatedness, to be the defining feature of a new approach to the study of cognition. Usually referred to as ‘embodied cognitive science’ it portrays a much more complex picture of the mind.

By taking a situated, distributed and embodied perspective, it has been suggested that the external environment can be used as a kind of extension of our mind, since these external structures function to complement our individual ‘skin and skull’. For example, Clark (1997, p. 180), states “[w]e are masters at structuring our physical and social worlds so as to press complex coherent behaviors from these unruly resources. We use intelligence to structure our environment so that we can succeed with less intelligence”. This is in stark contrast to mainstream cognitive science, which has viewed context, history and culture as “murky concepts” (Gardner, 1987, p. 41) that would only cause problems in the effort to find the ‘essence’ of individual cognition. Instead, it was argued, these aspects could be addressed and integrated when cognitive science had achieved an understanding of the central inner mechanisms of individual cognition (Gardner, 1987). Hutchins (1995), however, pointed out that there are unnoticed costs involved when we disregard culture, context and history, which he considers important factors in the development of individual intelligence. In addition, Tomasello (2000) for instance, hypothesized that if a human child grew up from birth without any cultural contacts, and no exposure to human artifacts the child would not develop the cognitive skills that are considered the hallmarks of human intelligence.

In order to exemplify the close interrelatedness between the so-called ‘biological’ and ‘cultural’ aspects, one can use Ingold’s (2000) example of learning to walk as an illustration. It is commonly argued that walking is an innate human capability, but Ingold does not categorize human walking as either biological or cultural. A child learns to walk according to the standard manner of its social and cultural environment, which is reinforced by biological aspects. Some cultures encourage children to start walking at a
very early age, as in Western societies, and therefore different physical scaffolds are used to encourage their motoric development. Other cultures actively delay their children’s initial walking attempts, and actually hinder their motoric development. Furthermore, different ways of walking are culture-dependent (e.g. Rogoff, 2003). Therefore, there is no one ‘natural’ or ‘pure’ biological way of walking, as one might assume. This means, the human skill of walking can be viewed as not only ‘biological’ in the sense of being a part of the functions of the individual human’s biology, but also a result of the child’s involvement in a social and physical world during normal development. An illustrative example is found in Maturana and Varela (1987). They describe the case of two Hindu girls in India\(^1\), who were taken from a wolf pack with which they have lived in without human contact. (The girls were five respectively eight years old at that time). When the girls were ‘rescued’ from the family of wolves, they moved on all fours, not knowing how to walk on two legs. The girl that survived the breaking-up from the wolf pack subsequently learned to walk on two legs as ‘ordinary’ humans, through human support. Consequently, instead of continuing the struggle between biology and culture, one should consider their interrelatedness from a socially embodied perspective, since “the former can only ground the latter and thus can never explain it” (Varela, 1994, p. 171). As pointed out by Rogoff (2003, p. 65), “the either/or questions are as pointless as asking whether people rely more on their right leg or their left leg for walking”. While this means that embodiment constrains (e.g. how we can move our hands), cultural norms affect (e.g. how to gesture in a certain cultural setting), but do not determine, the structure of socially embodied interactions.

The use of strategies such as taking advantage of external structures to coordinate action and cognitive behavior might be considered another and complementary way of explaining intelligent behavior, instead of merely a focus on mental representations of explicit knowledge. These external structures function as a kind of supportive framework or scaffolding, i.e. external resources to support and simplify cognitive activity for an individual agent (cf. e.g. Clark, 1997; Hendriks-Jansen, 1996; Wood et al., 1976). In a broad sense, the human body plus these external factors result in the ‘mind’, the boundary of which extends further into the world than cognitive science initially assumed. Accordingly, it has been argued that cognition is not an activity of the mind alone, since the mind is ‘leaking’ out to the environment, to use Clark’s (1997) vocabulary. Instead, cognition is distributed across the agent, the actual situation and its resources. This has led to the claim that the environment is a part of the cognitive system (cf. e.g. Clark, 1997, 1998; Clark & Chalmers, 1998; Hutchins, 1995; Susi, Lindblom & Ziemke, 2003; Thelen & Smith, 1994; Wertsch, 1998; Wilson, 2002). It is therefore very difficult to determine what the ‘border’ is between our senses and the world, since it is impossible to draw a sharp line between what goes on ‘inside’ the mind and what takes place in the world. Culture and language are considered our most significant scaffolds (Clark, 1997).

If we leave aside for a moment the main characteristics of situated, embodied and distributed approaches of cognition and focus on embodiment, we notice that historically, there are several reasons for the widespread neglect of the body in mainstream cognitive and social sciences. On the one hand, it is a consequence of the Platonic-Cartesian heritage, which has resulted in the view of the mind as the internal locus of rationality, thought, language and knowledge (for criticisms of that view, cf. e.g. Damasio, 1995, 1999, 2003; Farnell, 1995, 1999; Johnson & Rohrer, 2007; Sheets-Johnstone, 2003), which is supported by the Christian disregard of the flesh as the locus of sinful desire and irrationality (but see also Barbour2, 1999). Moreover, the opposite dimensions have been mapped on each other, resulting in the dualisms of, for instance, mind/body, mental/behavior, reason/emotion, and subjective/objective. On the other hand, researchers commonly overlook the role of the body because they are afraid of slipping into biological reductionism, and therefore they generally tend, or prefer to view mind as superior to and independent of the body (see e.g. Rose & Rose, 2000; Segerstråle, 2003; Segerstråle & Molnar, 1997). In short, the dichotomy between mind and body has in turn produced a disjunction between verbal and so-called nonverbal aspects of interaction. While dictionary definitions of the concept ‘nonverbal’ usually refer to the absence of words, this has unfortunately been interpreted synonymously with the absence of mind (Farnell, 1999), which according to Varela (1994) is a ‘theoretical failure of nerve’. Therefore, as Agar (2005, personal communication) phrases it, “[w]e need to find a cure for the Cartesian sickness”. The traditional dichotomy of mind vs. body is challenged by embodied cognitive science. Furthermore, Ingold (2000) emphasizes that body and mind are not two separate things, but rather two ways of describing the same process, i.e. the activity of the human agent in its physical and social environment. Similarly, Gallagher (2005) points out that an embodied approach attempts to redraw the map, “to develop a vocabulary, which is to say, a discursive or explanatory framework, that helps us to understand how the body shapes the mind” (ibid., pp. 243-244). It should be pointed out, however, that there are different opinions within embodied cognitive science concerning to what extent cognition is considered to be embodied. Clark (1999a), for instance, distinguishes between simple embodiment and radical embodiment. In simple embodiment, the traditional foundation of cognitive science (i.e., information-processing and computationalism) is preserved, and the nature of embodiment is merely considered a constraint of the ‘inner’ organization and processing. Radical embodiment, on the other hand, goes much further and treats the facts of embodiment as a fundamental shift in the explanation of cognition that is “profoundly altering the subject matter and theoretical framework of cognitive science” (ibid. p. 348). In should be emphasized, that I in this thesis, the ‘radical’ view using Clark’s vocabulary, is the chosen approach.

---

2 Barbour points out that the dichotomous concept of man in Christianity is a result of the Greek dualism of body and soul and actually not supported by the biblical view.
1.1 Motivations and Aim
While the body’s role in cognitive processes has received much attention in recent discussions under the banners of embodiment, embodied cognition and embodied cognitive science, and a large variety of notions and levels of embodiment and embodied cognition have been developed, there is no common understanding of what actually constitutes embodied cognition, and subsequently what kind of ‘body’ it might require (cf. e.g. Anderson, 2003, in press; Brooks et al., 1998; Chrisley & Ziemke, 2003; Clark, 1997, 1999a, 1999b; Dautenhahn, Ogden & Quick, 2002; Gallagher, 2005; Gibbs, 2006; Lakoff & Johnson, 1999; Núñez, 1999; Riegler, 2002; Rohrer, 2007; Sharkey & Ziemke, 2001; Varela, Thompson & Rosch, 1991; Wilson, 2002, Ziemke, 2001a, 2001b, 2002, 2003). Much research, so far, has considered the interaction between the individual agent and its environment, while from the artificial intelligence (AI) perspective it has been discussed what kind of body an artificial system might need to count as an embodied cognizer. Furthermore, although embodied cognitive science pays attention to both the socio-cultural embedding of cognitive processes and their bodily basis, current theories of embodiment need to move beyond the present emphasis on the interactions between the individual and the physical environment, to interactions between agents and their social environment (e.g. Anderson, in press; Lindblom & Ziemke, 2003, 2005a, 2005b, 2007; Riegler 2002; Semin & Smith, 2002; Sinha & Jensen, 2000; Ziemke, 2002, 2003). Somewhat surprisingly, however, the distributed cognition approach proposed by Hutchins (1995), for instance, treats social interactions as directly observable cognitive events, along with the materials involved in these interactions, while the body is merely regarded as one of the media of the information flow. This means, the main focus is not particularly on the role of the body in social interaction, but on the transformation of information through different media generally and subsequently there is still little work on the role of embodiment in social interaction (cf. Alač, 2005a, 2005b; Alač & Hutchins, 2004; Hutchins, 2006).

Thus, the research aim is to clarify the role and relevance of embodiment in social interaction and cognition. In doing so, I do not intend to bridge the ‘gap’ between, e.g., verbal vs. nonverbal interaction. Instead, I recharacterize the issue and develop a thorough and integrated understanding that supports and explains the relationships that actually exist, which originate from the embodied nature of social interaction and cognition. I argue that embodiment is the part and parcel of social interaction and cognition in the most general and specific ways, and in which dynamically embodied actions themselves have meaning and agency. Accordingly, I highlight the active and dynamical aspects of social interaction and cognition, in comparison to the more static and mental view usually stated in mainstream cognitive science. Therefore, since concepts such as body, mind, self, social interaction, cognition and culture are each crucial in bringing forth human social cognition they should not be considered as separate entities. They develop, constrain and function in relationship to each other, resulting, metaphorically speaking, in a full circle where there is no explicit beginning or end.
Furthermore, it should be noted that investigating the role of embodiment in social interaction and cognition is not just another variation of the old philosophical mind-body problem, but it also has several implications for AI and socially interactive technology (cf. Lindblom & Ziemke, 2002, 2003, 2005b, 2006; Ziemke & Lindblom, 2006). While the concepts of embodiment and social interaction are commonly used in the research of socially interactive robotics, their roles are still poorly understood. For instance, some social interaction capacities have been implemented in humanoid and android robots (cf. e.g. Brooks et al., 1988; Ishiguro, 2006, Kozima, 2000; Kozima & Yano, 2001; MacDorman & Ishiguro, 2004), but there are, at present, no coherent and unified theories about how to implement social interaction skills in robotic systems. This has resulted in a mismatch of different conceptual approaches and implementations (cf. e.g. Breazeal, 2002; Fong & Nourbakhsh, 2003; Fong, Nourbakhsh, & Dautenhahn 2003). Hence, a deeper understanding of how embodiment matters in human social interactive cognition is required from a cognitive science perspective, which then subsequently, can be applied and/or modified to the domain of socially interactive robotics (cf. Lindblom & Ziemke, 2006, Ziemke & Lindblom, 2006).

Moreover, the integration of deeper theoretical knowledge concerning embodiment in social interaction is also requested from the area of human-computer interaction (HCI), particularly in the subfield of computer-supported cooperative work (CSCW). Within CSCW, it has been recognized that technology is constantly used in a social context, but there is presently no thorough understanding of how to build computer systems that actually support the stark human sensitivity to social context and interactions. This lack of understanding is captured in the notion of the so-called social-technical gap, which refers to what current technology ideally should support socially and what it presently actually does support (cf. e.g. Ackerman, 2002; Erickson & Kellogg, 2002). Therefore, our socially embodied interactions in the real world are in sharp contrast to what is offered in contemporary technology. Apparently, one of the main reasons for this contrast is that many aspects of embodied social interaction, which has not been considered essential to cooperation and social interaction, become ‘lost’ in current technology, as current designs of information and communication technology are still dominated by outdated information processing models of human communication. In short, in order to build better technology and reduce the social-technical gap, we need a more thorough theoretical understanding of embodiment and social interaction before discussing its implications in the fields of cognitive modeling, and the design of socially interactive technology. This means that the request addressed earlier in this subsection, i.e. the need for a better theoretical understanding of embodiment in social interaction also has several implications with regard to the fields of socially interactive robotics in particular, and socially interactive technology in general.
Thus, the research presented in this dissertation is motivated by both a scientific interest in the role and relevance of embodiment in human social interaction and cognition, as well as its potential implications to socially interactive technology.

1.2 The Road Taken

In order to investigate and analyze the role and relevance of embodiment in social interaction, the chosen approach consists of three interrelated parts; i) theoretical work that results in a conceptual framework, ii) empirical work which illustrates parts of the theoretical framework, and iii) their implications to cognitive science and socially interactive technology. The main part of the research is theoretical work based on an extensive literature analysis, which is used to situate the resulting framework in its historical context and to serve as its foundation. However, there is no single methodology or discipline that alone can provide the full picture of the task to be accomplished in this thesis, and therefore an interdisciplinary approach, which combines work and insights from a number of different disciplines, is used. The research literature discussed in this thesis is wide-ranging, flowing not only vertically through time but also horizontally across disciplines. It involves and addresses different research methodologies and disciplines, such as artificial intelligence, phenomenology, ethology, cognitive science, developmental psychology, neuroscience, social psychology, communication, gesture studies, and linguistics. However, it should be noted that I am aware of the potential risks of such a strategy, since I cannot claim to be a specialist in all of these disciplines, their specific terminologies and theories. Although these different disciplines at first glance may not seem to have much in common, they offer highly complementary rather than alternative views, which help us gain deeper as well as broader views of how crucial the body and its physiological processes are in socially interactive cognition. Therefore, the hand that holds the analytic lens throughout the thesis takes an embodied outlook. The combination of theories and knowledge from different disciplines results in a theoretical framework. Crick and Koch’s (2003) definition of framework is used in this thesis. They claim a

...framework is not a detailed hypothesis or set of hypotheses; rather, it is a suggested point of view for an attack on a scientific problem, often suggesting testable hypotheses... a good framework is one that sounds reasonably plausible relative to available scientific data and that turns out to be largely correct (ibid., p.119, my emphases).

In other words, the framework acts as a guide, for explaining or clarifying the issue, rather than being correct in detail (Crick & Koch, 2003). My framework for the embodied nature of social interaction and cognition in particular, builds on the prior work of cultural-historical ideas (cf. e.g. Mead, 1934; Rogoff, 2003; Shanker & King, 2002; Tomasello, 1999; Vygotsky, 1934/1978), and while my own research is directly descended from this line of work, it is cross-fertilized with current theories of embodied cognitive science. What distinguishes the empirical work as well as philosophical ideas of earlier and contemporary scholars (cf. e.g. Farnell,
Generally speaking, the framework acts as a guide for the empirical part of this thesis, the purpose of which is to illustrate the framework in a naturalistic setting. Given the explanatory and descriptive nature of this thesis, a qualitative research approach is used for the empirical part. While qualitative research focuses on the qualities of entities and the processes of meaning in a (social) activity, quantitative research usually stresses the measurement and analysis of causal relationships between variables, and not the process of how meaning is constructed (Denzin & Lincoln, 2000). In order to illustrate (parts of) the framework’s potential, a case study consisting of several episodes of different kinds of embodied social interaction was conducted. Due to the topic’s situated nature, the study was carried out in different everyday social activities. The case study also offers additional empirical evidence in support of the framework. Moreover, the episodes of the case study also validate and expand the proposed framework, since they provide additional findings and unforeseen phenomena that consequently enhance our understanding of the role of embodiment in social interaction and cognition. The final result bears on the synthesis of the theoretical work and empirical findings, and offers a more thorough understanding of embodiment for socially interactive cognition. Hence we discuss its implications to cognitive science and socially interactive technology. It should be noted, however, that bridging the gap concerning the lack of understanding is not the same as filling it in.

Furthermore, due to the interdisciplinary nature of this thesis, it is worth mentioning that there are some problems with vocabulary in this thesis. While the aim is to move beyond the traditional dichotomies of mind/body, nature/nurture, individual/social, and so on, sometimes these concepts are nevertheless applied because they are accepted and commonly used terms. It should be noted, however, that they usually function in this thesis as abstractions, and should therefore not be interpreted literally. Hence, we should be aware of the potential possibility of falling into the trap set by the philosophical underpinnings of our own cultural background when using and reflecting over these terms.
1.3 Body, Embodiment and Embodied Cognition: Common Misunderstandings

Generally speaking, within cognitive science, the concepts of body, embodiment, and embodied cognition are sometimes used interchangeably, which has resulted in some confusion concerning their individual meaning and interrelatedness. In order to clarify the issue, nine of these interrelated misunderstandings are briefly disentangled below. The section ends with some clarifications of these concepts and a characterization of the role and relevance of embodiment in social interaction and cognition.

To consider the embodiment of cognition and social interaction is not the same as ‘burying the mind in the body’. Instead, establishing the differentiation between cognitive and physiological processes, as usually done in mainstream cognitive science, is not a suitable approach. Instead, taking an embodied stance actually means that we suppose the functions that control and guide our bodily actions and processes are also used in cognition and social interaction. Although embodiment has become a crucial concept in many areas of cognitive science in recent years, the term obviously still means different things to different people, which in turn has generated some general misunderstandings concerning the ways in which embodiment matters to cognition. In order to introduce the concept of embodiment, and to briefly clarify what embodiment means for the reader unfamiliar with the embodied approach in cognitive science, I have decided to briefly disentangle some issues concerning embodiment that I have encountered in different situations. These issues have been discussed, for example, in introductory courses of embodied and situated cognition that I have given lessons in, at scientific conferences, e.g., the Annual Meetings of the Cognitive Science Society, doctoral courses that I have participated in, discussions with colleagues, and in some reviews of my submitted papers.

Firstly, it should be stressed that embodiment does not imply that the body alone, without the brain, is responsible for cognitive processes. To cite an anonymous reviewer of Lindblom (2006) - “if all cognition is ‘just’ a coupling of the body with the world, we wouldn’t need a brain”. Instead, embodiment stresses the interplay between the environment, brain and the sensorimotor processes of the rest of the body, which are pivotal for cognitive activity to take place. In other words, embodiment does not exclude the brain, but instead includes it with the rest of the body in cognitive processes, hence widening the ‘bodily’ cognitive processes to not only depend on the ‘gray matters between the ears’. As Lakoff and Johnson (1999, p. 265) characterize it,

...the mind is embodied, not in any trivial sense (e.g., the” wetware” of the brain runs the “software” of the mind), but in the deep sense that our conceptual systems and our capacity for thought are shaped by the nature of our brains, bodies, and bodily interactions. There is neither no mind separated from and independent of the body, nor are there thoughts that have an existence independent of our bodies and brains.
The second common misunderstanding regards the nature of the interplay between the brain and the rest of the body, in which the brain is generally viewed as the organ being in charge which is ‘running’ the rest of the body. Metaphorically speaking, Chiel and Beer (1997) emphasize that the relationship between the brain and the body should be viewed as a group of players in a jazz improvisation, rather than seeing the brain as the conductor of an orchestra (i.e. the rest of the body). This means that the role of the whole ‘bodily’ nervous system “is not so much to direct or program behavior as they shape it and evoke the appropriate patterns of dynamics from the entire coupled system” (Chiel & Beer, 1997, p. 555). Similarly, Morris (2005, p. 67) emphasizes that “[t]he sort of brain we have could not do the sort of work that it does if it were centrally organized; it depends on being uncoordinated, on neurons firing on their own, yet doing work, because they are roped together through interactions that stretch across space and lag in time”. In other words, the relationship between the body and brain functions as a distributed dynamical process which is not centrally organized. Hence, although theories of embodied cognition reject the dualistic divide between body and mind, we should not replace it with a cut between the brain and the rest of the body. In order to avoid that brain-body split, there is a need to deepen the understanding of the physiological processes of the body and its interactions with the environment, which leads us to another misunderstanding of embodiment.

The commonly used image of the brain’s and the body’s interaction with each other rests on an ‘electronical’ metaphor. This common and still dominant metaphor suggests that the neurons in the brain and the central neural system form a communication network system that resembles a telephone system with intricately crisscrossed wiring. The impact of this metaphor was due to the technology and tools available in the earlier days, which allowed researchers to view and study the brain from an electronical perspective (Pert, 1997). However, new technological inventions led to new scientific discoveries that resulted in the less known chemical metaphor (Pert, 1997) of bodily interaction. This metaphor stresses that the interaction is based on biochemical links rather than on electronical signals (cf. e.g. Damasio, 1999, 2003; Pert, 1997). The biochemical interaction system represents a second interaction system that operates on a much longer time scale and over much greater distances than the electronical one (Pert, 1997). From an evolutionary perspective, the biochemical interaction system is much older and basic than the electronical system, given that significant chemical substances such as endorphins, for instance, were produced in cells long before neurons and brains were even developed. Moreover, the electronical metaphor holds that a connection is either ‘on’ or ‘off’, which refers to whether the cell discharges electricity or not. Biochemical substances like peptides, on the contrary, flow and move around continuously through extracellular space, not following any specified channels, and therefore they operate more fluently and dynamically over longer distances (Pert, 1997). Consequently, the scientific discoveries of the mid-1980s offer a chemical metaphor of the body’s internal communication system which is used throughout the organism’s body to interact across the
endocrine, neurological, gastrointestinal, and emotional systems of the whole body (i.e., brain plus body). Thus, to make an analogy, the shift from an electronical metaphor to a chemical metaphor of bodily interaction can be compared to the shift from atom physics to quantum theory.

It has also been questioned that if our cognition is grounded in embodiment, why do not certain physical handicaps necessarily lead to equivalent mental handicaps. Anderson (in press), for example, argues that many of our basic spatial concepts such as ‘Up’, ‘Down’, ‘Forward’, and ‘Backward’ are genuinely tied to the experiences of movements within a gravity system. Although paralyzed people do not move around in a practical sense they are indeed experiencing the same forces of gravity as other people, and use these concepts in a socially shared surrounding. Moreover, Gibbs (2006) points out that these individuals, to various degrees, experience a wide range of tactile-kinesthetic information from hand, eye, mouth, and head movements as well as sensations that arise from other bodily functions. Furthermore, the above described biochemical substances operate in paralyzed people as well, shaping and regulating bodily sensations and feelings. Indeed, paralyzed or physically handicapped people can still benefit from diverse action patterns and bodily sensations, which then might be sufficient for cognition to be grounded in embodiment.

When discussing cultural differences of cognition in general, it is also commonly argued that culture differs, but the human body, its sensorimotor functions and its interactions with the surroundings are rather ‘fixed’ across the world, that is, culture varies but bodies are stable. However, this is an oversimplification, since ‘culture’ and ‘body’, and their interactions are not uni-directed but have intertwined relationships. Instead, it is preferable to say that the body has dynamical potential to be reshaped by so-called cultural influences as well as vice versa. Greenfield (2002), for instance, claims that newborns of the Zinacantec people of Maya in Mexico are born with longer visual attention spans and are able to attentively observe their surroundings for much longer periods than Euro-American infants in the United States. This might constitute the basis for subsequent observational learning in ontogeny, which is much more frequent in Zinacantec culture than Euro-American culture. Moreover, Zinacantec newborns are less physically active than Euro-American infants, showing lower rates of motor activity. Most generally, the Zinacantec’s worldview is holistic which suggests how ‘culture’ impacts ‘biology’, given that their visual attention covers whole scenes rather than just some selected spots moment-by-moment. Moreover, the bodily movements of people in Zinacantec culture are smoother and more controlled than in Western culture. Thus, ‘culture’ reinforces ‘biology’ as much as ‘biology’ reinforces ‘culture’, which means that the divide between ‘culture’ and ‘biology’ is an artificial abstraction, because humans are “biologically cultural” (Rogoff, 2003).

A sixth misconception concerns the common and still dominant textbook view of the ways our senses and motor system function. In a broad sense, they are described as rather passive organs that only have the function of
input channels to the brain which does the cognitive work, and thereafter sends output signals to the motor system. Instead, a process such as visual perception, for instance, is an active and dynamical procedure involving muscle coordination in the course of making action-perception couplings with the environment. Gislén et al. (2006), for instance, investigated why children of some tribes of sea gypsies in Southeast Asia, who dive after seashells for a living, are able to see sharply under water, whereas European children do not. Apparently, this ability did not rely on some 'biological' adaptation of the eye lens alone, since Swedish children could learn to see sharply under water after some practice. It eventuated that the ability to see sharply under the water was the result of a sensorimotor adaptation of the muscles that control the lens of the eye. That is, the cultural pressure and specialization of making their living have resulted in an adaptation of the muscles involved in visual perception. Thus, culture and biology mutually define and shape each other in many instances of cognition, in the sense that cultural customs affect, but do not determine the organization of interactions under the constraints of embodiment (cf. Farnell 1999; Rogoff, 2003).

Another misconception is when an empirical study describing or noticing some kind of embodied action, e.g. gesture, it is viewed as being as study of embodied cognition. This idea might originate from the fact that there is lot of research today that investigates and analyzes gestures and other bodily actions. The question is - are all these previously conducted studies of gestures and other embodied actions the same as embodied cognition? My answer is that these approaches should not primarily and directly be interpreted as embodied cognitive science, although they at first glance seem to consider 'embodied actions'. But as an anonymous reviewer of Lindblom (2006) wrote “...is there any other kind of embodied actions?”. This question is verified, since obviously, independent of research discipline, all our movements of the body are in a fact embodied. That is, it would be impossible to perform a pointing gesture in real life without a physical realization. The crucial point here is that there are different theoretical ways of interpreting and providing meaning to a particular embodied action. Some researchers study gestures and other embodied actions from a semiotic perspective (e.g. Goodwin, 2003), others investigate manual gesture as a complement to speech in utterance production (e.g. Kendon, 2004). This does not imply, however, that they have embodied cognitive science as their theoretical foundation. From an embodied cognitive science perspective, an embodied action is not considered to be a manifestation of an internal cognitive process (as in mainstream cognitive science) but rather as an element of cognitive activity (e.g. Gallagher, 2005, 2007b). Thus, from an embodied perspective, an embodied action is a form of cognition, and not an expression or output of internal cognitive processes (Farnell, 1995, 1999). However, to complicate the issue a little bit more, there are different interpretations to what degree cognition is embodied within the area of situated and embodied cognitive science, whereas some are more radical, like Gallagher (2005) and other less radical, like Hutchins (1995). Thus, studies that analyze and investigate gestures and other embodied actions are
not by default embodied cognitive science research, since they can belong to other research disciplines.

Another misunderstanding is that embodiment only concerns physical interaction and lower level aspects of cognition, but does not have implications for higher-level issues such as abstract concepts, mathematics and the social mind. That is, embodiment is argued to be too limiting to be the foundation for more advanced forms of cognition in both a practical and a literal sense. However, in their work with metaphors, Lakoff and Johnson (1999), for instance, illustrate that many concepts of bodily movements are represented by verbs such as ‘grasp’, ‘pull’, ‘lift’, ‘tap’ and ‘punch’, i.e., about physical actions, which then serve as the basis for more abstract concepts like ‘grasping an idea’. This means that mental concepts are deeply metaphorical, as kinds of second order models of the physical world, based on analogies between concrete domains and conceptual/abstract ones. Similarly, Lakoff and Núñez (2000) discuss arithmetical metaphors as being based on the embodied experience of collecting, constructing and moving entities, however their work is quite provocative and speculative (see Gibbs, 2006). Although there are cultural variations of bodily metaphors, abstract thoughts and language are nevertheless suggested to be deeply rooted in embodied experience, given that many of our metaphoric concepts have their roots in different physical actions.

The final misconception discussed here is that our embodiment is too limited to give rise to all the different concepts, phenomena and cognitive processes that encompass our human ‘being-in-the-world’. To use a quite provocative analogy, assume that much of our human intellect rests on the pillars of the human alphabet, which in Euro-American cultures consists of less than 30 letters that in turn can be combined into various constellations. In general, the combinations of these 30 letters actually serve as the basis for abstract thought and human intelligence. The human body, in contrast, consists of more than 400 muscles, which can adopt a huge variety of postures and movements, which in turn result in various kinds of emotional, sensory-motor and tactile-kinesthetic experiences of the body and its interactions with the physical and social environment. Considered from this perspective, embodiment might not be too limited. Instead, we are able to embody so many phenomena that we actually lack appropriate concepts for describing them all, albeit different cultures may focus more on certain embodied experiences than on others.

Considering the above presented misunderstandings of the concepts of body, and embodiment, let me now briefly describe how I broadly view these issues, which can be characterized in the following way, without aiming to supply any definitions in the strict sense.

The body is the physiological and biological entity that provides the basic foundation for an autonomous agent to be alive, survive and reproduce. It involves different bodily parts, organs (including the brain), and different
bodily functions such as bio-chemical, neurological, and sensori-motoric processes.

**Embodiment** refers to the experiences that arise from the living body in its interactions with a material/physical as well as a social and cultural world. It also refers to how an autonomous agent acts upon these experiences via different means of dynamical action-perception loops that subsequently emerge into different kinds of embodied action patterns which create and maintain the embodied agent’s own understanding and meaningfulness.

In my own work, I emphasize that **embodied cognition** is the interdisciplinary study of the mind and intelligence, in which there is no major cut between cognitive and physiological processes as usually done in computationalism and traditional explanations of the mind. Most generally, embodied cognition means that the same processes which underlie/emerge in embodied action patterns are also used in all kinds of cognition.

More specifically, embodied cognition includes the person that enacts the body, all kinds of physical and social actions, as well as meaning accomplished through actions. The human mind and self are relational by their nature. They develop through the emergence of social interactions, and these processes can be understood only in light of the opportunities and constraints of bodily mechanisms and dynamically embodied action patterns, which are cognitive in themselves - having crucial roles in the shaping and sharing of the human subjective experiences of its acting body. Thus, the enactment of the social mind and self is a social act, and in order to direct yourself, you have to consider how others will act and react in response to your own actions, in which the socio-cultural spheres function as scaffolds for higher level cognition. This means, that the physical, material, social, and cultural environments are parts of an extended cognitive system.

In summary, all these issues belongs under the same embodiment umbrella, and I hope that the above disentanglements and characterizations can serve as the basic foundation for how to consider embodied cognition throughout the remainder of this thesis.

### 1.4 Thesis Road Map

However, before continuing, I would also like to clarify what this thesis is not about. The meaning of the term ‘social embodiment’ in this thesis, is totally different than its meaning within the field of theology, where in a broad sense it refers to the experience of ‘being one with God or Jesus’. Furthermore, the dissertation does not explicitly address comparative studies between humans and other species, the role of tools in social interaction and cognition, the bounds of cognition, and the many ways language can be used in social interaction and cognition from a linguistic perspective. Although some implications for AI and socially interactive technology are presented, this thesis does not intended to develop some general guidelines or methods for building more advanced artificial systems. Furthermore, I neither consider aspects such as one’s own view of the body.
as a social and cultural entity (cf. Shilling, 2005; Stam, 1998) nor address obesity and eating disorders such as bulimia and anorexia.

The previous sections of this introductory chapter present the research domain, describe the aim and motivations of the research, and address its theoretical significance to cognitive science as well as some practical relevance to artificial intelligence and socially interactive technology. Furthermore, the research process is described, and the relationship between body, embodiment, and embodied cognition, is discussed, disentangling some misconceptions of embodiment and describing how I view these concepts. What this thesis is not about has also been declared. A brief outline of the remainder of this thesis follows.

**Chapter 2** provides an overview of the history of the conceptions of the relation between mind and body. It begins with the ideas of Plato and Descartes, continuing via Darwin’s work to behaviorism, various (alternative) approaches of agent-environment interaction to the rise of the cognitive revolution in the mid-1950s. The inception of cognitive science resulted in computationalism and the main characteristics of the traditional approach and the criticism thereof are addressed. The final section of the chapter summarizes the main ideas concerning the relation between mind and body from the Ancient Greeks to computationalism. In further chapters the thesis illustrates that some of the earlier approaches of agent-environment interaction, in particular the work of Piaget, Merleau-Ponty, Vygotsky, and Mead to mention a few, are highly relevant to contemporary discussions of the embodied nature of socially interaction and cognition.

**Chapter 3** describes why and how embodied and situated approaches to cognition became relevant in the mid-1980s. It also offers an overview of basic ideas, characteristics, levels and concepts relevant to embodied cognitive science. Subsequently it portrays different approaches to, and views of, embodiment and embodied cognition in current embodied cognitive science, and also discusses several notions and aspects of embodiment as well as what kind of a body is required for natural cognition and AI. In the chapter, the significance of movement in embodied cognition is especially considered, and the concept of the **body in motion** is introduced. Finally, the social dimension of embodiment is briefly discussed.

**Chapter 4** discusses different aspects of embodiment in social interaction. Experimental findings of social embodiment effects in social psychology, emotions, and attitudes are reviewed. Furthermore, phenomenological issues as well as neurological underpinnings of embodiment in social interaction are discussed. In particular, the chapter addresses embodied simulation theories, and the action-perception linkage of mirror neurons. Subsequently the focus moves on to discuss embodied linguistics, exploring the role of embodiment in language and gesture, and in particular their interrelatedness. Finally, four fundamental functions of embodiment in social interaction are identified.
**Chapter 5** investigates issues of social interaction and cognition, analyzing and discussing in further detail the common characteristics, levels, kinds and methods of studying social interaction and cognition. In addition, two different metaphors concerning how to view social interaction and cognition are compared, and finally the nature and methodological study of social interaction are re-characterized, offering alternative explanations to the traditional view. It also investigates the role of the *social body in motion* in child development that subsequently leads to an alternative embodied explanation of how the social mind develops.

**Chapter 6** motivates and describes the framework from both historical and 'state-of-the-art' perspectives. The first section can be regarded as a short summary of the ideas portrayed in the previous chapters, which subsequently progresses to the final framework for the embodied nature of social interaction and cognition.

**Chapter 7** concerns the empirical part of the thesis. The first section describes and motivates the research design of the empirical work. The following sections and their subsections describe the conducted case study in different social situations in which different aspects of embodied actions of socially interactive cognition are the focus of analysis. The case study also offers unforeseen empirical findings of embodied actions within socially interacting and cognition. The chapter ends with a discussion and some conclusions concerning the synthesis of the theoretical and empirical work.

**Chapter 8** summarizes the main contributions of this thesis, compares it with related work, and considers what kind of a body is necessary for cognition as well as the uniqueness of human cognition. It also discusses some methodological issues, and presents some implications to AI and socially interactive technology. It also offers some ideas for future research, and the thesis concludes with some closing remarks.
Chapter 2

2. Body and Mind – A Historical Perspective

What is the relation between the body and the mind? Is the body just a mechanical container for the immaterial mind as proposed by dualists such as Descartes? Or is it the computational physical input–output interface to the world, as functionalist philosophers and many cognitive scientists today would argue? Or are our cognitive processes strongly dependent on the body as the proponents of the embodied approach? Are in fact ‘body’ and ‘mind’ just two sides of the same coin? These questions do not have any clear answers yet, but philosophers and scientists have been discussing these issues for a very long time with very different outcomes. As seen in the following chapters, this discussion is still a “hot topic” in contemporary cognitive science, but in order to discuss these questions accurately, it is important to review their historical context, which is the overall goal of this background chapter.

More precisely, this chapter has four interrelated aims. Firstly, the historical background for the ‘study of mind’ is outlined in a chronological way, beginning from its philosophical roots to the foundation of cognitive science and the computational paradigm. I believe that for any science, in particular an interdisciplinary one such as cognitive science, it is of great significance to know what has happened in the past in order to understand the present, and then hopefully avoid repeating mistakes and misunderstandings that have already been made. The second aim is to explain why the idea of computationalism became the dominant position within cognitive science at the expense of alternative approaches. However, it should be noted that there is neither enough space nor time to offer the whole story, and therefore I present a somewhat simplified view with a focus on the key concepts. Then a glimpse of the wide source of theories of agent-environment interactions from which modern cognitive science may still have much to gain is presented, re-evaluating and possibly incorporating some of these

---

3 Cf. e.g. Bechtel, Abrahamsen & Graham, 1998; Clancey, 1997; Clark, 2001; Dreyfus, 1972/1979, 1992; Gardner, 1987 or Pfeifer & Scheier, 1999, for more complete overviews of the history of cognitive science and AI.
‘old’ ideas. The final aim is to situate the resulting framework for the embodied nature of socially interactive cognition within its historical context, and to some extent, to motivate it. Thus follows my story of the ebb and flow of embodiment throughout the history of theories of the mind.

2.1 The Roots of Dualism - from Plato to Descartes

The philosophical roots of cognitive science date back to the old Greeks, and Gardner (1987), for example, pointed out that cognitive science has a long past but a short history. According to Gardner, the Greeks devised the mathematical inventions of logic and geometry and consequently, the notion that all human reasoning could be a sort of calculation appeared. The very concept of thinking of the human being as a reasoning device, calculating according to some rules has its origin in Plato’s ideas. Plato (428/427 BC – 348/347 BC) argued that all knowledge has to be described in a form of pure explicit definitions. As a consequence, he believed that phenomena that could not be formalized explicitly, for example, bodily skills and feelings, could not count as knowledge. He therefore made a distinction between the rational mind and the body with its emotions and skills. He admitted that there is some kind of connection between the two of them, but he never explained how it worked. This was the starting point of the Western philosophical tradition, assuming that reckoning is the ‘language of the mind’ (Dreyfus, 1972/1979). It was supposed that the matter of knowledge was embedded in the human mind from birth; the hard job was to bring forth this knowledge to consciousness, which is demonstrated in the Platonic dialog between Socrates and the young slave boy Meno (Gardner, 1987).

This conceptual separation of mind and body was confirmed by Aristotle (384 BC – 322 BC) who divorced the practical from the theoretical and defined the human being as ‘the rational animal’. Consequently, our sensorimotor skills and their relevance for coping with our surroundings were neglected and also the know-how in everyday activities (Dreyfus, 1972/1979). Aristotle made a distinction between form and its matter, and asked if the form can exist independently in the material world or ‘in form of the mind’ (‘information’ in today’s terms, according to Freeman & Núñez, 1999). In addition, Aristotle might be considered to have been the first biologist and ethologist, since he systematically studied animals and plants in his search for a “scheme of nature”. He argued that the soul was the form of the matter of the biological body and argued that this was true in all living organisms. His classification attempts of animals resulted in a “ladder of life”, which situated simple organisms at the bottom and the human being at the top. Aristotle’s work can be regarded as a forerunner to evolutionary theory (Ritvo, 1991; Sparks, 1982). The Greeks’ view of knowledge continued to dominate in the Western intellectual tradition, and Aristotle reached his pinnacle during the middle Ages. However, in the eras of Renaissance and Enlightenment, philosophers also started to pay attention to findings made in the empirical sciences that had emerged (Gardner, 1987).

The French philosopher René Descartes (1596-1650) also claimed that the mind is separated from the physical body, resulting in the concepts of two
different substances, and formulated the dualistic viewpoint. According to Dreyfus (1972/1979) and Cisek (1999), Descartes confirmed the ideas from Plato rather than adding anything radically new. Descartes was convinced that only humans possessed a rational mind, whereas animals and plants were mere machines, guided by their physical surroundings as puppets. Descartes’ assertion is primarily disembodied in the sense that he viewed the mind as a non-physical substance, with almost no connection to the body, arguing that the human soul actually ruled the mind and directed the body how to behave. The only way to study this rational mind was to use Descartes’ method of reason, which tried to ascertain how the mind (res cogitans) ruled the body (res extensa). Later philosophers, such as Hume and Kant, tried to analyze the function of the rational mind itself, and discussed what was possible or necessary to know about the world, and which possible constraints actually existed (Dreyfus, 1972/1979). According to Dreyfus, Descartes was the first person to imagine the possibility of building robots, but he also addressed some limitations in the machine itself, which he stated as follows:

Although such machines could do many things as well as, or perhaps even better than men, they would infallibly fail in certain others ... For while reason is a universal instrument which can be used in all sorts of situations, the organs of a machine have to be arranged in a particular way for each particular action. From this it follows that it is morally impossible that there should be enough different devices in a machine to make it behave in all the occurrences of life as our reason make it behave (Descartes, 1637, quoted in Dreyfus, 1972/1979, p. 235-236).

Consequently, Descartes demonstrated or believed he demonstrated the need of a uniquely human soul as the ‘pilot of the corporal boat’, since he was aware of the boundaries of this mechanistic view of nonhuman beings (Dreyfus, 1972/79; Freeman & Nuñez, 1999). On the other hand, vitalsists as Stahl (1660-1734), for instance, disagreed with the mechanistic view as proposed by Descartes. Instead, they argued that all living creatures were provided with some sort of ‘vital energy’ comparable to a spirit or soul. The debate between the proponents of these opinions was still ongoing when Darwin presented his theory of evolution (cf. Ziemke, 2000, 2001a).

2.2 Darwin’s Work and its Legacy

Charles Darwin’s book The Origin of Species (1859) was a serious attempt to explain the development of the different species which inhabit our world. However, Darwin’s theory was not the only one of its kind. Russell Wallace had developed a similar theory independently, which in fact forced Darwin to publish his own ideas (Cosmides et al., 1997). Although not totally original, Darwin’s theory was the first that actually could explain how change came about, which Darwin claimed was through the struggle for survival between individuals, the individual variations among conspecifics, and finally the proving that more off-spring are born in every generation than can actually survive to maturity. Darwin called it ‘natural selection’ which Spencer later labeled “survival of the fittest’. This line of argument was the cornerstone in
the theory of evolution (Cosmides et al., 1997; McFarland, 1993; Sparks, 1982).

Darwin also assumed that patterns of behavior were the product of evolution, meaning that human behaviors had roots in animal behavior (Gould & Gould, 1999). Although he was unwilling to talk discuss humans, he supposed that they are a part of the “tree of life”, arguing that humans to some degree have a common past with other living beings (McFarland, 1993). Nevertheless, inspired by the improvement in ‘hard sciences’ like physics, Darwin tried to identify ‘the laws of behavior’ in both animals and humans. This way of thinking resulted in the book *The Descent of Man* (1871, quoted from Gould & Gould, 1999, p. 3) in which he argued that:

> [t]he difference in mind between man and higher animals, great as it is, certainly is one of degree and not kind. We have seen that the senses and intuitions, the various emotions and faculties, such as love, memory, attention, curiosity, imitation, reason, and so on, of which man boasts, may be found in incipient even sometimes in well-developed conditions in lower animals.

Darwin himself certainly believed there was a mental linkage between animals and humans, although this was a standpoint that many people, not least in Victorian England, had difficulties in accepting. Many agreed that the theory was adequate for animals, but not for humans who where considered unique due to ‘their special minds’ (Gould & Gould, 1999; Ritvo, 1991). However, as described by Sheets-Johnstone (1999), Darwin strongly stressed the continuities of “mental powers”, and did not separate the mental abilities from the actual body. Nevertheless, he argued as follows:

> [E]xperience shows the problem of the mind cannot be solved by attacking the citadel itself – the mind is function of body – we must bring some stable foundation to argue from (Darwin, 1836-1844/1987, quoted in Sheets-Johnstone, 1999, p. 435).

It can be concluded that Darwin seriously stressed what today could be called an embodied approach to the study of the human mind. However, Sheets-Johnstone (1999) argued that these lines in Darwin’s work have been misinterpreted. Instead of studying the mind “as the function of the body”, researchers have reduced the body and its interactions with the environment to include only the brain itself.

Furthermore, according to Ekman (1998), Darwin was interested in ascertaining how our “habits had been gradually acquired”. In particular he wanted to find an evolutionary explanation for human facial expressions and emotions. Darwin held the belief that expressions were “emblems for the emotions” and tried to describe which “emotional states” resulted in different expressions. He also investigated human facial expressions in different parts of the world in an attempt to discover whether they were universal or not (Ekman, 1998). As a result, in 1872 Darwin published *The Expression of the Emotions in Man and Animals* (hereafter referred to as Expression), which included not only human expressions and emotions, but also these from
other animals, such as dogs, cats, horses, monkeys and apes. Darwin’s way of explaining the expressive and emotional similarities between species was to ascribe animals (both high-level and low-level animals) with human-like emotions and feelings. For example, the dog feels *pleasure* when working because it is its *duty*, and the cat is in a *loving mind* when rubbing against somebody’s legs, as well as ants are *despaired* when their ant-hill is destroyed (Darwin, 1872/1998; Sparks, 1982). However, Darwin focused primarily on facial expressions and paid little attention to gesture and body language, which he considered the product of socially learned conventions, only mentioning their communicative role in a few words, and did not elaborate on them further. According to Ekman (1998), Darwin’s reason for ignoring these communicative features could have been an overreaction to Bell’s claim that human expressions were God’s gift to mankind. For that reason, he might have decided the best way of confronting Bell’s opinion was to ignore it completely. Instead, Darwin (1872/1998) raised *what*, *how*, and *when* questions about emotions and expressions in general. Nevertheless, his main interest concerned *why* a certain expression is proper for a certain emotion. According to Ekman (1998), Darwin was very interested in this ‘why’ issue and its significance to his idea of the continuity of species, since Darwin wanted to demonstrate that human beings were not a distinctive species, created by some kind of God. On the contrary, he insisted that also humans were the product of evolution. However, Bell had earlier successfully proposed that “there were muscles in the human face without analog in the animal kingdom, designed by the Creator for the display of specifically human emotions” (quoted in Ekman, 1998, p. xxv). Instead, Darwin’s aim was to convince public opinion that the three principles of ‘what’, ‘how’ and ‘why’ could explain all primate expressions, which hopefully might result in an acceptance of his idea of the continuity of species. Creationists argued that if expressions were universal, it would be the result of the heritage from Adam. Instead, Darwin concluded that there were general significant similarities worldwide and claimed “the same state of mind is expressed throughout the world with remarkable uniformity” (Darwin, 1872/1998). Accordingly, human emotional expressions and facial gestures are innate, since their universality is a result of evolution, and therefore *not* a uniquely human capability (Ekman, 1998).

According to Cosmides et al. (1997), Darwin was the first to actually combine psychology and biology, but he was also responsible for the split between the two. Darwin’s work was in fact a serious attack against dualism, but the message was not recognized. This can be partly explained by his co-worker Wallace’s personal view, which did not support the anti-dualistic stance. Wallace was actually devoted to the Cartesian worldview, arguing that human mental abilities have to be explained in supernatural terms, as a kind of soul (in Cosmides et al., 1997). Hence, Darwin, in fact, tried to bridge this gulf between body and soul in his work (cf. Darwin, 1871), arguing that mental abilities could be explained in the same evolutionary way as the physiological factors. As a result of his fear of public opinion and lack of support from Wallace on this issue, Darwin elaborated two parallel lines of explanations for the development of the human mind. These were a
phylogenetic explanation that stressed descent, and an adaptationist explanation that emphasized selection (Cosmides et al., 1997). The adaptationist perspective has merely been factored out within psychology, and only the phylogenetic branch has been explored. This is in fact partly due to Darwin himself, since he mainly tried to explain the human mind from a phylogenetic perspective, making the case for the idea of a common human ancestor, originating from Africa (Cosmides et al., 1997). That is, Darwin believed in a single universal line of development, with roots from lower animals to fulfillment in the supposedly ‘highest’ human being, i.e. the contemporary European. In order to validate this claim, he had to run counter to the argument presented earlier in Descent (1851), in which he had stressed the role of environmental conditions and adaptations, resulting in several lines of evolutionary development (Ingold, 2002a). Consequently, Darwin had to argue that a kind of supremacy of reason to inherited endowment existed. Nevertheless, Ingold (2002a) pointed out that Wallace (cf. 1870) indeed had found some deficiencies in this line of argument. He was in fact more familiar with indigenous people than Darwin, arguing that so-called ‘savages’ were really endowed with the same intellectual powers as Europeans, and that these ‘primitive’ people had also realized sophisticated cultural achievements. However, Wallace was unable to explain his view according to the theory of ‘natural selection’, and therefore devoted the endowment of superior brains to creationism (Ingold, 2002a). In addition, the adaptationist perspective would not fully support Darwin’s claim of a single line of development view, because this line focused on the differences between separate species. Hence, Darwin suspected that the adaptationist perspective of his theory really should have convinced the dualists they were actually right, namely that the human being is the unique species. Consequently, they may interpret that different qualities in separate species really could lead to such matters as a soul in humans (Cosmides et al., 1997).

Despite the disruptive issues addressed above, Expression can be said to have been the first serious work on behavior in history (Cosmides et al., 1997; McFarland, 1993). According to Allen and Bekoff (1997), Darwin usually attributed cognitive abilities to animals on the basis of his own observations instead of using controlled experimental studies and this approach resulted in an ‘anecdotal cognitivism’ (Allen & Bekoff, 1997). Despite this anecdotal approach, he seems to have been the first person in history to apply a comparative and phylogenetic method to the study of behavior (Allen & Bekoff, 1997). Although his empirical methodology can be regarded as convenient in today’s terms, Darwin’s initial behavioral observations had a great impact, resulting in an increased interest in the study of behavior (Cosmides et al., 1997; McFarland, 1993). Darwin’s ambivalent view resulted in two contrary lines of behavioral research, the anthropomorphical line and the mechanistic line (Ziemke, 2000, 2001a). The anthropomorphical line overstressed the mental abilities of animals, for example, beavers were described by Lewis Morgan as having an extraordinary understanding of hydraulics and architecture, a conclusion he made by observing the beavers’ ability to build advanced constructions of
nests, channels and dams. However, he had actually been deceived by the beavers’ innate behavioral repertoires (Gould & Gould, 1999; Manning & Dawkins, 1998; Sparks, 1982). Moreover, the main proponent of this anecdotal line was Darwin’s disciple Romanes, who presented anecdotes of ‘clever’ animals in his book *Animal Intelligence* in 1882, in which he inferred mental phenomena quite subjectively, e.g. fish were reported to experience jealousy and spiders as enjoying music (Sparks, 1982). However, Romanes’ uncritical appraisal of animal mental abilities resulted in an attack by another Morgan, namely Lloyd Morgan (1894), who became well known for his declaration, nowadays referred to as ‘Morgan’s canon’, which states:

> In no case may we interpret an action as the outcome of the exercise of a higher psychological faculty, if it can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale (Lloyd Morgan, 1894, p. 53).

Lloyd Morgan was critical of anecdotal anthropomorphism, but did not claim that animals in fact lacked mental abilities, since ‘lower psychological states’ actually are mental states. Hence, Morgan insisted that scientists should rather try to find a simpler and more restrictive explanation of a behavior, instead of attributing higher mental capacities to animals (Allen & Bekoff, 1997; Manning & Dawkins, 1998).

This line of argument should be considered in the case of the horse *Clever Hans*, which was one of the most famous examples of ‘intelligent’ animals in the early 1900s (Allen & Bekoff, 1997; Gould & Gould, 1999; Sparks, 1982). His owner believed that animals were as clever as humans, if they were given the opportunity to become educated. He therefore trained his horse Hans, which apparently turned out to possess remarkable cognitive capabilities. Consequently, ‘Clever Hans’ was regarded as a prodigy, since the horse was able to solve mathematical problems and answer questions, by tapping the correct answer with his front leg. Moreover, the horse was able to return the correct answers with or without the presence of his trainer. Although many scientists, for instance the prominent Oskar Heinroth, were skeptical of the horse’s abilities, in their investigations of the case they were unable to discover any fraud or evidence that the horse was actually cued by its trainer. But in 1904 the mystery of ‘Clever Hans’ was revealed by Oskar Pfungst. He concluded that the horse did not really understand the questions, but instead was skilled in reading human body postures, e.g., a minor elevation of the eye-brows, holding the breath for a second or straightening one’s posture. These visual hints cued ‘Clever Hans’, since the questioner unconsciously altered body and/or head posture(s) when the horse reached the correct number of hoof taps, as a kind of involuntary relaxation of the “pressure” of the situation. As a result, ‘Clever Hans’ could not present the right answer when the questioner himself did not know the correct answer, since the horse was then not able to notice the slight change in body posture when the right answer was at hand. The story of ‘Clever Hans’ was in fact the first scientific case demonstrating that it could be hazardous to appeal to consciousness or any cognitive abilities in explanations of animal behavior. The horse had actually learned to associate
events, and the ‘Clever Hans debacle’ resulted in an ending of anecdotal
cognitivism, following Morgan’s canon (Allen & Bekoff, 1997; Gould & Gould,
1999). Although, the ‘clever’ horse failed to be good at arithmetic, he was in
fact very sensitive to perceptual cues, and excelled in reading human bodily
postures. According to Knapp and Hall (1997), it has been demonstrated that
horses are able to recognize changes in bodily postures as tiny as 0,2
millimeters, such as slightly dilated nostrils. Furthermore, the story of
‘Clever Hans’ might also serve as an illustrative example of the role of
embodiment in social interaction, since this case highlights two important
points. Firstly, humans continually show their feelings and attitudes, and
secondly, that other beings both humans and animals are able to sense and
interpret these aspects, although their interpretations might differ
significantly.

Returning to the discussion of the two separated lines of behavioral
research, Douglas Spalding argued that ‘clever’ behaviors were built into the
animals mind from birth. He proposed that behavior in a sense was
automatic in both animals and human beings, arguing that ‘voluntary acts’
were the result of automatic inborn nervous processes. Both humans and
animals were “conscious automata”, with no free-will or independence of the
laws of nature (Sparks, 1982). Hence, this was the mechanistic line of
research influenced by Darwin, given that he contrasted in his book The
Power of Movement in the Plants (1880) the downward digging behaviors of
earthworms and moles with the tendency of plant roots to grow downwards,
and then explaining both behaviors as a response to gravity. This concept
was further developed by von Sachs, who demonstrated tropism. Tropism is
the non-voluntary action driven by external forces, such as light or gravity in
flowers’ and plants’ tendency to turn towards the sun (von Sachs, 1882;
Ziemke, 2000). Ziemke (2000, 2001a) pointed out that Loeb (1859-1924) was
the most radical and influential of the mechanistic theorists, heavily inspired
by von Sachs’ work on tropism. Loeb identified similar behaviors in several
simple organisms and extended the notion of tropism to include animal
behavior. Hence, this led to the mechanistic assumption that both animals
and humans were just objects directed by the environment, lacking
subjective mental abilities. Loeb declared higher animals and humans were
controlled by environmental forces, arguing that human beings were
‘biological machines’, with a kind of robot-like quality in today’s terms.
However, Loeb pushed his idea to the extreme, and his explanations of both
animal and human behavior were too simple and were strongly criticized by,
for example, von Uexküll (see subsection 2.4.1 and Sparks, 1982; Ziemke,

Another proponent of mechanistic theories was Sherrington (1857-1952)
who studied and identified reflexes. He illustrated in The Integrative Action of
the Nervous System (1906) how the nervous system integrated reflexes into
adaptive behaviors and explained in what ways reflexes operated.
Sherrington declared that the reflex was the fundamental unit of behavior,
characterized by the reflex arc. He argued that this reflex arc was organized
in three distinct structures; an effector organ which was stimulus sensitive,
being wired via a nervous pathway to a receptor which initiated a response (Manning & Dawkins, 1998). Thus, the behavior of the body was predetermined by the wiring of the nervous system that linked stimulus to response. Furthermore, Sherrington claimed that this notion of reflexes was an abstraction, arguing that no part of the nervous system could react without affecting the whole nervous system, since all parts were intertwined together (see Ziemke, 2000, 2001a). Hence, Sherrington’s notion of the reflex arc would strongly influence the research of behavior, although it was biased and interpreted differently. Furthermore, Ziemke (2000) noted that Loeb (e.g. 1918) further developed von Sachs’ work on tropism, extending the notion of tropism to include animal behavior. However, Loeb’s theory of tropism fell short from empirical results, and he unfortunately confused the notions of tropism and taxes (nowadays the former concept refers to ‘immobile’ plants, whereas the latter concept refers to mobile organisms). Despite this mistake by Loeb, the importance of tropism and taxes has proved to be significant in the study of animal behavior. In 1940, Fraenkel and Gunn published the book _The Orientation of Animals: Kineses, Taxes and Compass Reactions_, which strongly agreed with Loeb’s view of the objective study of animal behavior, and the book contains many examples of different sorts of taxis. Fraenkel and Gunn (1940) argued that many kinds of animal behavior can be described by different taxes, either working together or against each other. As pointed out by Ziemke (2000) as well as Sharkey and Ziemke (1998), contemporary work on behavior-based robotics and AI has its roots in the early work on taxes.

Cisek (1999) identified two important issues in the research on behavior during the end of the 19th century and the beginning of the 20th century. Firstly, he pointed out the split in the study of living organisms. As a result, one line of thought concentrated on studying behavior as such, whereas the other focused merely on bodily physiology. However, Cisek (1999) argued that the motive for this split was for practical reasons, although it indirectly led to a gulf between the discoveries made within the field of behavioral studies on the one hand, and the field of physiology on the other, which resulted in a lack of communication between these two research fields. Secondly, Cisek (1999) argued that the behavioral sciences as a whole have been suffering because different approaches, such as logic, mathematics, and physics. They have attempted to explain and verify behaviors and mental phenomena in a mathematical way, in order to validate their scientific endeavor, arguing that it would otherwise not be recognized as ‘real’ science (Cisek, 1999). As a consequence, these fields have tried to discover the ‘laws of behavior’, which Darwin attempted as well. Hence, the legacy of Darwin’s work resulted in the split within the study of behavior and the unanswered question of dualism. One line of thought became experimental psychology which dealt with issues regarding learning and intelligence, and the other line became the scientific study of animal behavior and how they perceive the world (later known as ethology). As a result of these two branches, experimental psychology and ethology, scientists also developed different research methods (Cosmides et al., 1997; McFarland, 1998).
In the late 19th century, some researchers began to explicitly study human behavior, and in 1879 Wilhelm Wundt established the first psychological laboratory in Leipzig, Germany (Bechtel & Zawidzki, 1999; Shaughnessy & Zechmeister, 1997). In attempts to find the ‘atoms of the mind’, Wundt tried to explain behavior by referring to the thoughts, feelings and emotions which crossed the mind during conscious experience, This method of working was termed introspectionism and had a tremendous impact during the early days of experimental psychology (Gardner, 1987). Wundt also provided the first classification of gestures that was intended for scientific purposes, but this line of research did not continue. Moreover, in the United States, William James conducted studies on mental capabilities from another perspective (Bechtel & Zawidzki, 1998). James was critical of introspectionism, and instead followed the adaptationist line of Darwin, paying particular attention to brain functions and instincts. He claimed that the mind consists of functionally specialized mechanisms or instincts, and each of them had evolved via adaptation to handle different issues. James sought to investigate these procedures of mental activity in daily life, and this standpoint became known as functionalism (Gardner, 1987).

Indeed, the prevailing ‘units of analysis’ in experimental psychology were actually stimulus and response. Psychologists had been influenced and inspired by discoveries within physiology in general, and the function of the nervous system in particular. However, the notion of the reflex arc became misinterpreted by psychologists. The initial description of the reflex arc implied a neural circuitry of ‘coordinated action’ (Sherrington, 1906), but the psychologists considered it as two distinct units (MacFarland, 1998). They assumed that this process began with the stimulation of the organism, which initiated some sort of a mental act or awareness, resulting in a response. This means, they separated the stimulus from its response, and treated the response as a separable and independent event, which they claimed was the effect or result of a dependent stimulus. James (1890, p. 372), for example, stated that “[t]he whole organism, it will be remembered, is, physiologically considered, but a machine for converting stimuli into reaction”. However, Dewey (1896) was very critical of this restricted interpretation of the reflex arc, arguing that this psychological model actually was a legacy of mind-body dualism (Clancey, 1997).

Hence, Dewey (1896) proposed an alternative view, arguing that the organism interacts with the world through self-guided activity that coordinates and integrates sensory and motor responses. He addressed the need for a new description of the reflex arc, and he proposed the function of circular sensory-motor coordination, claiming “[w]hat we have is a circuit, not an arc or broken segment of a circle. This circuit is more truly termed organic than reflex, because the motor response determines the stimulus, just as truly as sensory stimulus determines movement” (Dewey, 1896, p. 363). For this reason, he rejected the linear model of stimulus and response, and argued that all activity is always an on-going ‘sensory-motor coordination’:
we find that we begin not with a sensory stimulus, but with a sensori-
motor coordination, the optical-ocular, and that in a certain sense it is
the movement which is primary, and the sensation which is
secondary, the movement of body, head and eye muscles determining
the quality of what is experienced. In other words, the real beginning
is with the act of seeing; it is looking, and not a sensation of light. The
sensory quale gives the value of the act, just as the movement
furnishes its mechanism and control, but both sensation and
movement lie inside, not outside the act. Now if this act, the seeing
stimulates another act, the reaching, it is because both of these acts
fall within a larger coordination; because seeing and grasping have
been so often bound together to reinforce each other, to help each
other out, that each may be considered practically a subordinate
member of a bigger coordination (Dewey, 1896, p. 359-359).

Hence, the ‘act of seeing’ is related to the circumstances of the situation in
which the looking process occurs. According to Dewey (1896) meaning and
response emerged together and there was no need for an analyzing or higher
descriptive function, it is sufficient with a lower level motor-sensory
coordination. Thus, this view proposed an alternative explanation of how the
mind works, without an intervening “consciousness” controller as a kind of
soul. In my opinion, Dewey actually stressed the situated and embodied
nature of mind, first by stating the circumstances in a particular situation
(situatedness), and then describing the underlying mechanisms in the form of
its sensory-motor coordination (embodiment) in that particular situation.
However, the aspects of introspectionism and functionalism (together with
Dewey’s view of sensory-motor coordination) should be swamped away and
replaced by behaviorism (Gardner, 1987). Behaviorists argued that it was
misleading to study either sensations or the operations of the mind or to
consider bodily contributions. Skinner (1938), for instance, argued that the
arbitrary associative links between stimuli and response were merely the
outcome of learning, and claimed that the way in which an animal or human
was embodied was entirely irrelevant.

2.3 Behaviorism
Following and expanding Lloyd Morgan’s argument, behaviorists argued that
only observations of overt behavior should be the object of study. The
initially false assumption about Clever Hans’ cognitive ability resulted in an
attempt to bring psychology more in line with the ‘hard sciences’, focusing
on replicable, controlled experimentation that recorded measurable behavior.
The reason for this methodology was that these methods would make it
possible to control and manipulate behavior and then provide explanations
that covered the behavioral patterns of both animal and humans during
variable conditions (Allen & Bekoff, 1997). Behaviorists followed Darwin’s
statement of a ‘mental continuum’ between animal and man, then arguing
that animals could be viewed as simple models of human learning. Thus,
behaviorism was the first rigorous attack against the dualistic assumption,
since it totally ignored any mental content. However, little, if any, interest
was directed towards in what ways an organism was embodied. Hence,
research on the role and relevance of bodily contributions to cognition was
overwhelmed by the tide of behaviorism (Allen & Bekoff, 1997; Chrisley & Ziemke, 2003; Clark, 2001; Cosmides et al., 1997; Sparks, 1982).

Behaviorism has some of its roots in the work of animal learning from the beginning of the 20th century, such as the work of Pavlov (Gould & Gould, 1999; McFarland, 1993). He studied the mechanisms of digestion in dogs and regarded the biological body as a machine (following the mechanical view), arguing the digestive system functions like a factory. He measured the quantity of saliva produced during feeding in the dog’s mouth and discovered the conditioned reflexes, arguing they were the “atoms of action”. Classical conditioning was identified around 1905 when Pavlov noticed that the dogs started to salivate before the food was at hand. The dogs learned to associate the sound of the dishing out of food with its delivery. As a result, Pavlov constructed the formula that a sign stimulus resulted in a physiological response. Moreover, Pavlov noticed that if a bell or a similar sound was repeatedly activated just before the unconditioned stimulus was presented, this conditioning (or training) stimulus would finally trigger the salivation response itself, which was called operant conditioning or learned conditioning. Pavlov supposed that any cue an animal could sense could be used as a conditioned stimulus for every response (Gould & Gould, 1999; McFarland, 1993; Sparks, 1982).

Another person who also investigated animal learning was Thorndike (1874-1949). Although he studied chickens running through different mazes, there was a significant difference between his own and Pavlov’s experiments (McFarland, 1993; Thorndike, 1913, in Sparks, 1982). While Pavlov’s dogs were always rewarded, Thorndike’s chickens were only rewarded when they performed the desired behavior, which also led to improved results. Consequently, Thorndike proposed the existence of a “law of effect”, i.e. responses to a certain situation that was followed by ‘satisfaction’ or success were strengthened, and vice versa (discomfort or failure resulted in weakness), and therefore there was no need for any mental content. However, Thorndike did not explicitly claim that humans actually behaved in the same way as animals (McFarland, 1993; Thorndike, 1913, in Sparks, 1982).

Hence, the mechanistic view proposed by Loeb and Sherrington was strongly in line with Pavlov’s work, which offered the behavioral sciences some physiological respectability. This happened during an era when the behavioral sciences were still envious of the formalistic or ‘hard’ sciences, and there was a need for an ‘objective’ science of behavior, especially after the embarrassing Clever Hans debacle (McFarland, 1993). As a result, behaviorism, which became the dominant approach from the beginning to the middle of the 20th century in the United States, emerged (Gould & Gould, 1999; McFarland, 1993; Sparks, 1982). In 1913 Watson published Psychology as the Behaviorist Views It, in which he argued that psychology should only study carefully defined stimuli and their overt responses, and then formulated rules that predicted which behavior should be the result of a given stimulus (Gould & Gould, 1999). Moreover, scientists had to neglect
every speculation about what might be happening in the ‘mind’. Watson and Loeb met, and agreed that animals were “mindless machines” (Sparks, 1982). However, Watson (1930) went beyond Pavlov’s assumptions of the power of classical conditioning and instead argued that all behavior is learned, even the physiological bodily processes, arguing:

[T]here are then for us no instincts- we no longer need the term in psychology. Everything we have been in the habit of calling an ‘instinct’ today is largely the result of training ... [Think] of each unlearned act as becoming conditioned shortly after birth. – even our respiration and circulation (Watson, 1930, quoted in Gould & Gould, 1999, p. 48).

Hence, behaviorists claimed, contrary to the views of Darwin and James, that instincts actually do not exist. Instead, all behavior is the result of learning and “all mental content comes from the external world”, via association processes that link stimulus and response (Cosmides et al., 1997). Moreover, Watson (1930) argued that all learned responses were chains of unconditioned responses in a form of ‘reflex chains’ and this learning process was supposed to be *equipotential*, meaning there was no inherent advantage of some stimulus-response association over another possible stimuli-response association in the organisms (Cosmides et al., 1997). In addition, the proponents of behaviorism reinterpreted Darwin’s work, but they over-stressed his idea of the mental continuum of species, which resulted in some misleading conclusions. Watson and his colleagues argued that humans and animals in fact learn and behave similarly, and as a result, scientists would gain knowledge of human behavior by studying rats, pigeons and other animals, as well as experimentally controlling their environments in a way that was not ethically acceptable for humans (Cosmides et al., 1997). Ironically, behaviorism, which had its roots in Darwin’s claim of phylogenetic continuity, resulted in an intensely anti-Darwinist view as behaviorists argued that Darwin’s theory of evolution would not shed any light on human behavior. On the other hand, they claimed that learning and environmental factors in one way or another ”insulate behavior from evolutionary shaping and analysis” (Skinner, 1953, in Cosmides et al., 1997).

Skinner was more radical than Watson, arguing that scientists should treat the mind as a ”black box” (Skinner, 1953). He asserted that no statements or investigations of the mind should be made, since there was no mental content, no instincts, no abilities and no emotional processes. The mission of the behaviorists was to explain human behavior in terms of classical and operant conditioning, stressed by the mental continuity of the species. The only inherited ability in animals and humans was the “genetic endowment” of classical and operant conditioning (Cosmides et al., 1997). Consequently, *antimentalism* and *equipotentiality* were the two pillars of behaviorism. However, these behavioristic assumptions were too optimistic and exaggerated, and several scientists criticized this behavioristic manifesto already during its heyday. As discussed in the following section, the main reasons for their criticism were either they had discovered phenomena which
could not be explained in behavioristic terms or they actually paid attention to mental content.

2.4 Beyond the Bounds of Behaviorism

There were researchers, mainly outside the United States, who were not devoted to behaviorism, and conducted work in other directions. This section presents some of these works, focusing on approaches and theories that can be viewed as early attempts of embodied and/or situated theories of cognition, as well as the social nature of cognition, although not explicitly expressed under these labels. The work of von Uexküll and the succeeding branch of ethology (subsection 2.4.1), the work of the Gestalt school and the philosopher Merleau-Ponty (section 2.4.2), Piaget’s genetical epistemology (subsection 2.4.3), and the socio-historical approach proposed by Vygotsky (subsection 2.4.4) are briefly outlined. Furthermore, Bartlett’s work on social cognition is included, and then we meet Dewey again, since his philosophical epistemology has strong similarities with the theories of embodiment and situatedness (subsection 2.4.5). Finally, Mead’s work on the relational nature of intelligence is portrayed (subsection 2.4.6).

2.4.1 The field of ethology and the work of von Uexküll

As a consequence of the major impact of behaviorism in the United States, ethology emerged primarily as a research field in Europe. According to Alcock (1999), it might be argued that the academic field of animal behavior actually emerged in 1937, when the first journal within the area appeared (in German), but it was in fact during the 1950s that the research field as such became widely accepted as a serious branch of science. Burghardt (1985a, p. xvii) for instance, declared that the “naturalistic study of animals, focusing on activities essential to their survival and informed by an evolutionary attitude, is the hallmark of ethology”. Hence, the underlying approach in ethology is biological rather than psychological, and its founders stressed the importance of studying animals in their natural habitats and not in artificial environments such as laboratories and Skinner boxes. While the proponents of behaviorism almost denied any forms of instincts, arguing for their equipotential assumption, ethologists instead followed Darwin’s adaptational line and stressed the importance of innate behaviors (McFarland, 1998). As described by McFarland, Darwin’s impact on ethology was strong and threefold. Firstly, the theory of natural selection laid the foundation for studying animal behavior from an evolutionary point of view. Secondly, his ideas on instincts can be considered a precursor to the inception of classical ethology. Finally, his work resulted in an increased interest in studying animal behavior as such.

The roots of ethology can generally be found in the early work of European zoologists, but the fundamental discoveries that founded the branch of ethology were in particular Whitman’s and Heinroth’s taxonomy of, and comparative work on, different behavioral patterns in birds (Lorenz, 1985). Although Heinroth formulated the “comparative study of gesture” and named it ethology, it was actually his disciple Lorenz, who became the main promoter of the new science. (Besides Lorenz, the founding fathers of
ethology are today usually considered to be Tinbergen and von Frisch (e.g. Alcock, 1999; Gardner, 1987; Sparkes, 1982; Walther, 1999). Lorenz, for instance, particularly emphasized the importance of sensitive observations of wild and ‘domesticated’ wild animals that he had raised in his home. Tinbergen and von Frisch conducted simple, though very significant experiments, in the natural field as well as on captive animals in ‘natural’ conditions (Bekoff, 2002; Bekoff, Allen & Burghardt, 2002).

Another influential person was Jakob von Uexküll (1864-1944) who had a tremendous impact on ethology in general, and Lorenz’s work in particular (Walther, 1999; Ziemke, 2000, 2001a). According to Walther (1999), although von Uexküll’s ideas strongly influenced Lorenz, he rarely credited von Uexküll’s work. This neglect was probably due to von Uexküll’s open and profoundly anti-Darwinian standpoint, which was contrary to the opinion of Lorenz, who was personally a great admirer of Darwin’s evolutionary theory (Walther, 1999). Probably also as a result of his anti-Darwinian stance, von Uexküll’s work became less known during his lifetime, though he was not against evolutionary explanations as such. Another reason might be the fact that he published only in German (Walther, 1999; Ziemke, 2000, 2001a). Von Uexküll coined the term *Umwelt* in order to describe the surrounding and unique subjective world of each species, including humans. That particular concept has recently received much attention in embodied cognitive science and robotics (cf. e.g. Brooks, 1991; Clark, 1997; Emmeche, 2001; Ziemke, 2000, 2001a), and von Uexküll’s work is probably the most influential from the field of ethology in contemporary discussions of embodied cognition.

Von Uexküll was initially active in physiology but has become better known for his later work in theoretical biology. According to Ziemke (2001a), von Uexküll was strongly inspired by Kant’s notion that all knowledge is determined by the knower’s subjective perception and consideration, arguing that biologists have to study “the phenomenal world or the self-world of the animal” (1934/1957, p. 319). Indeed, von Uexküll argued that the task of biology was to improve Kant’s work further, by more fully examining the role of the body, and the relation between subject and object. He strongly criticized the mechanistical theories in general, and the work of Loeb in particular, although he did acknowledge the role of chemical and physical ‘forces’ in the behavior of quite simple organisms (e.g., von Uexküll, 1909). Accordingly, Uexküll (1909) investigated which factors in its surroundings might affect the organism, and how this occurred.

In order to answer these questions, he began with organism itself, as opposed to an anthropocentric view, according to which the human worldview is completely separated from other species. As a result, nearly all human objects and artifacts might disappear and only the objects remains that exert an effect on the organism in correlation to its construction plan, and “the construction plan by itself creates the environment of the animal”

---

4 von Uexküll seems to have believed that God was the force behind the ‘divine’ plan (cf. von Uexküll, 1934/1957).
(von Uexküll, 1909, p. 223 original emphasis). Hence, von Uexküll (1909) differentiated between subjective (Umwelt) and objective (Innenwelt) experiences, arguing that it is essential to understand both in order to fully comprehend an organism’s behavior (Allen & Bekoff, 1997). The ‘inner world’ of an organism is made up by the construction plan, which filters out and regulates the crucial factors of the surroundings. Von Uexküll (1909) strongly argued that the basis for biology would be to study the construction plans of different species, bringing the fields of anatomy and physiology together again. This means, the research focus of biology, in von Uexküll’s view, should be the study of how the environment and the inner world are connected to each other. However, his own empirical work was limited to studies of invertebrates, whereas his theoretical work addressed both ‘lower’ and ‘higher’ animals (see von Uexküll, 1909). Broadly speaking, he proposed that a ‘special biology’ of all animal species could be realizing through studying the organization of the building plans in separate species and then comparing them. Hence, these comparative analyses should not only focus on the actual functions, but also concentrate on different materials or tissues, thus criticizing Loeb’s work on tropism for only addressing the former aspect.

Consequently, he was critical of behaviorism as well, arguing that their proponents ignored the organism’s subjective and embodied nature of being by treating “all living beings as mere machines” from an objective standpoint (von Uexküll, 1957, p. 319). He maintained that the proponents of these theories had disregarded the importance of subjective experiences, paying no attention to the subject’s (embodied) experience, for instance, how our sense and motor organs supply our perception and actions. In his own words:

The mechanists have pieced together the sensory and motor organs of animals, like so many parts of a machine, ignoring their real functions of perceiving and acting, and have even gone to mechanize man himself. According to behaviorists, man’s own sensations and will are mere appearance, to be considered, if at all, only as disturbing static. But we who hold that our sense organs serve our perceptions, and our motor organ our actions, see in animals as well not only the mechanical structure, but also the operator, who is built into their organs, as we are into our bodies. We no longer regard animals as mere machines, but as subjects whose essential activity consists of perceiving and acting. We thus unlock the gates that lead to other realms, for all that a subject perceives becomes his perceptual world, and all that he does, his effector world. Perceptual and effector world together forms a close unit, the Umwelt (von Uexküll, 1957, p. 320).

This means that different species experience the surrounding world in different ways due to their various body designs perception capabilities. It has been noticed that a bee perceives its surrounding world quite differently from a human, since they, for example, can see infrared light, which humans cannot do by nature (without an apparatus). Therefore, the bee’s subjective view of the world is separate from the human perception and interpretation of the ‘same’ world, since they inhabit their own ‘effective environments’ (cf. Clark, 1997).
In order to illustrate his idea of Umwelt, von Uexküll (1957) used the tick as an example, illustrating the gulf between the subjective experiences of different species. In fact, the mated female tick needs a blood meal before she can lay her eggs, and in order to complete that task, it has to be equipped with certain sensory and behavior repertoires so called *functional circles*. It is through several such functional circles that the tick becomes embedded in its own world. The tick is guided through those circles via particular perceptual and effector signs. The tick’s skin is sensitive to light, thus leading it up from the ground to a brighter position on a branch or grass blade. There it hangs until a mammal passes by which emanates butyric acid, and once that butyric acid reaches the ticks receptors, it drops onto the mammal. This means, the *perceptual sign* of butyric acid alters into a *perceptual cue* that in turn triggers an *effector sign*, releasing the tick’s legs which allows and the tick to drop onto the mammal. When the tactile cue of hitting the mammal’s *coat* is triggered, the tick begins to move around, searching for warmth, and when it finds the skin it will trigger a burrowing behavior, after which the tick starts to burrow in and suck blood. However, the tick has no sense of taste, and will drink any fluid of the right temperature once it has perforated the membrane. When the tick has finished her first and last meal, it will drop down, lay her eggs in the earth, and then die.

However, the mammal continues to generate butyric acid during the later stages, although the acid has lost its meaning for the tick, since it is actually sensitive to certain signs at different phases. Hence, the relation between the object and the subject is illustrated as a *functional circle*, which is not a set of random reflexes, but rather a collection of context dependent “well planned successions” which follow after each other (Uexküll, 1957, p. 324). According to Ziemke (2000), von Uexküll should neither be considered vitalistic nor mechanistic. Instead, he stressed that the organism’s parts are created in such a way that they together will form a *whole*, a kind of behavioral entity, functioning as an acting subject. Consequently, that behavioral entity emerges into a “systematic whole” through its functional embedding in its Umwelt, which he stated as follows:

> We are not concerned with the chemical stimulus of buturic acid, any more than with the mechanical stimulus (released by the hairs), or the temperature stimulus of the skin. We are concerned solely with the fact that, out of the hundreds of stimuli radiating from the qualities of the mammal’s body, only three become the bearers of receptor cues for the tick. Why just these three and no others?
>
> What we are dealing with is not an exchange of forces between two objects, but the relations between a living subject and its object. These occur on an altogether different plane, namely between the receptor sign of the subject and the stimulus of the object … her Umwelt (von Uexküll, 1957, p. 325).

As described by Ziemke (2000, 2001a), von Uexküll (1928) distinguished between living organisms and machines, and the major differences lie in the
construction and the autonomy of the system itself, and the former is described as follows:

Every machine, a pocket watch for example is always constructed centripetally. In other words, the individual parts of the watch, such as its hands, springs, wheels, and cogs, must always be produced first, so that they may be added to a common centerpiece.

In contrast, the construction of an animal, for example, a triton, always starts centrifugally, from a single cell, which first develops into a gastrula, and then into more and more organ buds.

In both cases, the transformation underlies a plan: the ‘watch-plan’ proceeds centripetally and the ‘triton-plan’ centrifugally. Two completely opposite principles govern the joining of the parts of the two objects (von Uexküll, 1982, quoted from Ziemke, 2000, p. 17).

This means, von Uexküll (1928) particularly stressed that living organisms have an innate “meaning-quality” in their organs, contrary to the parts of a machine which can never develop centripetally, a line of argument closely related to his notion of the autonomy of living systems versus machines. As described by Ziemke (2000), von Uexküll (1928, p. 34) argued that “each living tissue differs from all machines in that it possesses a ‘specific’ life-energy in addition to physical energy”. However, this ‘life-energy’ is a kind of quality in the chemical/physical system, and not some ‘vitalistic’ immaterial energy (Ziemke, 2000). This ‘life-energy’ makes it possible for cells to respond to certain stimuli in a particular way, since living cells are able to perceive and act according to their specific receptor and effector signs. As a result, the behavior of the organism is “meaningfully organized” and not “mechanically regulated” according to von Uexküll (1982). He argued that machines operate mechanically, simply following the physical and chemical laws of cause and effect and not able to either grow or change. He furthermore claimed that machines have fixed structures, created and ‘built in’ to the machine by the human designer. Hence, machines are heteronomous, since they cannot regenerate or repair themselves if they break down (Ziemke, 2000, 2001a).

Living organisms, on the contrary, can both grow and regenerate on their own, since they have their functional role themselves, allowing the tissue to repair and grow autonomously. In sum, von Uexküll (1928) claimed that “machines act according to plans (the human designers), whereas living organisms are acting plans” (von Uexküll, 1928, cf. Ziemke, 2000, 2001a).

According to Burghardt (1985), von Uexküll’s work provided the inspiration for several classical ethological concepts such as sign stimulus, releasers, and innate release mechanisms. During this initial era of ethology the belief was that animals at times responded in an instinctive manner to specific stimuli so called sign stimuli. Alcock (1999, p. 25), for example, defined “an instinct as a behavior that appears in fully functional form the first time it is performed, typically such behaviors are mechanically triggered by a simple cue of some sort”. For instance, Tinbergen (1951) conducted some experiments with male sticklebacks that are sensitive to red spots, and his results suggested that it might be some innate release mechanisms (IRM) in the fish. In addition, Lorenz (1937) argued that much of animal behavior was
constructed by some “fixed-action patterns” (FAP) or species dependent instincts, and mainly genetically determined (Alcock, 1999).

Alcock (1999) pointed out that early ethologists studied many behavioral cases, which were not easily altered by specific environmental influences. Tinbergen and Perdeck (1950), for instance, discovered that newly hatched herring gulls somehow ‘know’ how to be fed. They pecked at a red dot on the bills of their parents, which responded by regurgitating some food. Another example, also described by Alcock (1999), concerns the rolling behavior in the incubating greylag goose. If an egg was taken and placed at a distance, the goose retrieved the egg to the nest using a standardized behavior pattern to roll the egg back (Alcock, 1999). Hence, from an observer’s point of view, it actually seemed as if the goose knew what to do. However, this behavior does not end even when the egg is taken away, since the goose will actually complete the behavior as such, with or without the egg (cf. Tinbergen, 1951). Moreover, Tinbergen (1958) discovered that gull chicks would rather peck at a long thin striped red stick than on a realistic model of a parent gull, since the red striped stick in fact functions as a ‘better’ releaser of begging behavior when the chick is quite young. Lorenz and Tinbergen called this instinctive response fixed action pattern (FAP), and supposed that the fixed action pattern was the response which plays itself out to completion, once activated by a simple sensory cue. They called this key component of the object that triggers the FAP a sign stimulus or a releaser, in the case when the sign stimulus was a social response between individuals (Alcock, 1999). This means, the red dot on the stick functions as a releaser for the begging behavior of the gull’s offspring.

In sum, the field of ethology offered a comparative framework for the study of behavioral functions, evolution (phylogeny), physiology and behavior development (ontogeny) in animals (cf. Tinbergen, 1951). These different levels of analysis are sometimes presented as Tinbergen’s four questions’ (e.g. Segerståle & Molnar, 1997), and provides a stable basis for the study of behavior (Manning & Dawkins, 1998). This framework has also been applied on human behavior, in particular Eibl-Eibesfeldt’s (1989) work of human ethology. However, his work focuses on finding the ‘pure nature’ of human beings, although it can be questioned if the hunting and gathering people he studied actually represent the ‘plain and true’ nature of human beings, since these tribes are also embedded in a social and cultural surrounding. Thus, it can be argued that, in fact, there are no such ‘natural’ human beings.

2.4.2 From the Gestalt school to Gibson and Merleau-Ponty
Although the predecessor of the Gestalt school was Ehrenfels’ work in the 1890s, on melody perception, the real starting point was Wertheimer’s book concerning the visual perception of movement in 1912 and later studies conducted together with his assistants Köhler and Koffka (Gardner, 1987). They noticed that in human perception, the overall characteristics of an object, as its main shape or contour, were of greater importance than the partial elements (Bechtel, Abrahamsen & Graham, 1998). As a result, proponents of the Gestalt school took a holistic approach, and argued that a
solution to a problem may occur in one step, and not in several series of stimuli and responses as argued by behaviorists. The answer to a problem resulted in a form of a gestalt, explained as a ‘picture in the mind’, an ‘insight’ or a kind of ‘aha-experience’. Consequently, Gestaltists concluded that ‘the whole is more than the sum of its parts’. Besides the studies conducted on humans, Köhler (1925) investigated the problem-solving abilities of chimpanzees, and concluded that this insight-ability or ‘gestalt’ was present in some apes as well.

However, the Gestalt approach lacked a theoretical foundation for the perceptual phenomena observed, and as a result the research field failed to maintain its influential and central position. The reason for this is twofold. Firstly, new findings in neuroscience and initial models of information-processing provided seemingly plausible alternative explanations of the same evidence (Dreyfus, 1972). Secondly, advocates of the Gestalt school relocated to other places around the world, due to the political climate in Europe during the 1930s (Bechtel, Abrahamsen & Graham, 1998; Gardner, 1987). However, some of the ideas of the Gestalt school were used as a foundation for the work of Gibson, who developed an ecological approach of perception, in which he introduced the concept of direct perception in an attempt to bring the environment back into psychology, stressing the dependency of the context for our perceptual ability (Gibson, 1966, 1979). He claimed that objects in the world afforded certain actions, allowing organisms to pick up certain ‘invariants’ and dynamical features of the environment. The organisms could then immediately transform this environmental information into action, without the need for any stepwise processes (Mace, 1977). Hence, Gibson’s general ideas can be viewed as situated/embodied in the sense that he emphasized the importance of the interaction between the agent and the environment for intelligent behavior.

The French philosopher Merleau-Ponty (1908-1961) was inspired by the Gestalt school and tried to find support for some of their ideas in his work. He claimed that the mind was essentially embodied and interacting with the surrounding world (Dreyfus, 1972/1979). Merleau-Ponty then further claimed that it is actually the body which provides meaning for the mind, thus taking a stance separate from both mentalism and materialism (Priest, 1998). According to Priest (1998), Merleau-Ponty was critical of behaviorism, arguing in *The Structure of Behaviour* (1942/1963) that neither is psychology ‘reducible’ to biology nor biology ‘reducible’ to physics. Instead, Merleau-Ponty suggested the human mind really functions in a similar way as the Gestaltists explained perceptual phenomena. This means, the interpreted perception of an object (or a person’s behavior) in itself is

---

5 According to Dreyfus (1972/1979) the Gestaltists would never have accepted the ‘computer metaphor’ for mind, since they did not believe in stepwise instructions, they rather viewed the mind as an emerging whole; a kind of ‘gestalt’.

6 Initially published in French in 1942, but the English version was published in 1963. The French original of the title word ‘Structure’ was the French translation of the German ‘Gestalt’ (Priest, 1998).
ambiguous and dependent on the perceiver’s (or observer’s) own conscious or unconscious preconceptions, resulting in the emergence of different, alternative ‘gestalts’ or interpretations (Priest, 1998). Hence, studying behavior from an observer’s point of view, by ‘objectively’ noticing the physiology of the actions was insufficient, since no component of behavior could be reduced to its assumed sub-parts. Consequently, the whole behavior as such is always interpretable from different angles, since we can never eliminate alternative explanations of a particular action, and behavior is not the result of external causes alone (Priest, 1998). Similarly, Merleau-Ponty (1942/1963) claimed that on the other hand, behavior could not be explained internally in terms of conscious intentionality (Loren & Dietrich, 1997). Loren and Dietrich (1997) pointed out that Merleau-Ponty was frustrated with the dualistic view in philosophy, and the “step back”7 (epoché) from the real lived world, which was the common view in phenomenology, as proposed by Husserl and Heidegger for instance. Merleau-Ponty maintained that these beliefs essentially divorced embodiment from consciousness and cognition8.

Merleau-Ponty was actually inspired by Heidegger’s work9 and the idea of ‘being-in-the-world’, but he went an essential step further, arguing that being embodied is the core ‘essence’ of human subjectivity of being in the world, which he stated as follows in his Phenomenology of perception (1945/1962):

> when I reflect on the essence of subjectivity, I find it bound up with that of the body and that of the world, this is because my existence as subjectivity is merely one with my existence as a body and with the existence of the world, and because the subject I am, when taken concretely, is inseparable from this body and world (Merleau-Ponty, 1962, p. 408, quoted in Mingers, 2001, p. 111, emphasis added).

Hence, Merleau-Ponty’s core concept was the idea that ‘I am my body’ as a kind of “embodied cogito”, i.e. it is not the brain that does the thinking, instead it is done by the body (Loren & Dietrich, 1997). According to Merleau-Ponty’s (1962) ideas, the brain is merely a part of a larger system, the central nervous system and the whole body, and in order to understand intelligent behavior we must study the whole system. He therefore claimed that human behavior must be seen structurally, through the interplay between the body and the nervous system in relation to the environment. In order to do so, he pinned down Heidegger’s more abstract level of description, and addressed both biological and physiological matters in his work, discussing different aspects of animal and human behavior (Loren &

---

7 This “step back” or epoché is the divorced study of consciousness from a kind of meta-level perspective, by putting the everyday experience and the outer world in brackets. It has its roots in Descartes’ method of reason.

8 According to Loren and Dietrich (1997) as well as Priest (1998), Merleau-Ponty did not separate cognition from consciousness. The analysis of cognition actually is one aspect in the analysis of consciousness. Hence, what counts as cognition counts as consciousness as well.

Dietrich, 1997; Mingers, 2001). Hence, in his view, the behavior of an organism is the result of the interaction between the organism’s subjective perception of the environment conducted by its body and nervous system, which he characterized as follows:

The properties of the object and the intentions of the subject ... are not only intermingled; they constitute a new whole. When the eye and the ear follow an animal in flight, it is impossible to say “which started first” in the exchange of stimuli and responses. Since all the movements of an organism are always conditioned by external forces, one can, if one wishes, readily treat behaviour as an effect of the milieu. But, in the same way, since all the stimulations that the organism receives have in turn been possible only by its preceding movements which have culminated in exposing the receptor organ to the external influences, one could also say that behaviour is the first cause of all the stimulation ... But in the organism itself ... which chooses the stimuli in the physical world to which it will be sensitive. The environment (Umwelt) emerges from the world through the actualization or the being of the organism – an organism can exist only if it succeeds in finding in the world an adequate environment (Merleau-Ponty, 1963, p. 13, quoted in Mingers, 2001, p. 13).

Hence, this structural relation can be viewed as a kind of ‘circular loop’, with no start- or end-points. This means, the world neither determines our perception, nor does our perception generate the world (cf. Varela et al., 1991). As a result, Merleau-Ponty disregarded dualities such as perceptive and motoric abilities, as well as object and subject. Instead, he argued that perception and action are entirely intertwined, because perception constantly entails motor (muscular) actions, and these actions in turn generate novel perceptions and so on. Consequently, there is no dominance of one side over the other; instead there are “mutual affordances” between the organism and its external world (Mingers, 2001). Moreover, Merleau-Ponty claimed it is actually the body which has the necessary ‘knowledge’ to perform the task at hand, since the body ‘knows how to act’ and ‘how to perceive’ through the history of its phylogenetic and ontogenetic interactions with the environment. He maintained that “[a] movement is learned when the body has understood it, that is, when it has incorporated it into its ‘world’” (Merleau-Ponty, 1962, p.139).

Loren and Dietrich (1997) argued that Merleau-Ponty viewed this relation between the embodied organism and certain objects in the environment as intentional, i.e. the way an object appears to the organism is meaningful, although this does not necessarily mean cognitively meaningful, but rather that it is biologically meaningful. They further pointed out that Merleau-Ponty supposed this ‘meaning’ was the outcome of the body as such, claiming that bodies actually are conscious, which means bodies are ‘cognitive’ in themselves. Hence, he viewed cognition as a sort of biological phenomenon instead of a mental one, in view of the fact that our cognition is bound to the world through our embodiment, arguing that all organisms are “condemned to meaning” prior to any cognitive states. This bodily ‘meaning’ or ‘intentionality’ is the result of the organism’s orientations to the world, depending on and affected by its past history, which he stated as follows:
The gestures of behaviour, the intentions which it traces in the space around the animal, are not directed to the true world or pure being, but to the being-for-the-animal, that is, to a certain milieu characteristic of the species: they do not allow the showing through of a consciousness, that is a being whose whole essence is to know, but rather a certain manner of treating the world, of ‘being-in-the-world’ or of ‘existing’ (Merleau-Ponty, 1963, quoted in Loren & Dietrich, 1997, p. 353).

The relation between the external world and the organism is represented in the states of the animal’s nervous system, and these states mediate the animal’s behavior in the environment. Additionally, bodily intentionality offers the necessary basis for cognitive intentionality, since it has its roots and origin in bodily intentionality. This means, Merleau-Ponty did not separate intentionality from the body as is usually done in more analytic formulations of intentionality (Loren & Dietrich, 1997).

According to Dreyfus and Dreyfus (1999), as well as Dreyfus (2002), Merleau-Ponty (1962) developed two significant concepts that actually address these ‘bodily intentions’, namely the intentional arc and the maximal grip. The ‘intentional arc’ is characterized by the close connection between the embodied organism and its world. It functions like a feed-back loop between the skills learnt from past experiences and the demands of the present situation at hand. The notion behind the intentional arc is that all past experiences are projected back on the world. The organism acts from its own point of view, and the outcome is due to the organism’s history of past experience with similar situations or things. Hence, the organism is ‘afforded’ a certain response, which is guided by the organism’s previous ‘knowledge’ or experience. Consequently, there is no need for representations in the mind, since the best representation of the outside world is the world itself, which is presented in this feed-back loop or intentional arc itself (Dreyfus, 2002). The ‘maximal grip’ then, is the body’s tendency to respond to these demands in the most appropriate way, namely to bring forth a more optimal ‘gestalt’, from the organism’s point of view. This is illustrated in the following statement:

For each object, as for each picture in a gallery, there is an optimum distance from which it requires to be seen, a direction viewed from which it vouchsafes most of itself: at a shorter or greater distance we have merely a perception blurred through excess or deficiency. We therefore tend towards the maximum of visibility, and seek a better focus as with a microscope (Merleau-Ponty, 1962, p. 302, quoted in Dreyfus, 2002, p. 378).

Furthermore, Merleau-Ponty (1962) argued that all higher animals including humans always strive to get a maximal grip on a situation. However, this is not a goal directed activity, instead it is purposive in the sense of being a kind of ‘basic’ motivation, since organisms always try to find a more optimal ‘gestalt’. Hence, this basic motivation is the driving force, not any stated request to achieve a certain goal. The body is tuned by the situation to reach a flow of balance and equilibrium (Dreyfus & Dreyfus, 1999). Hence,
according to Dreyfus (2002), the intentional arc is the result of the tendency toward a maximal grip.

According to Mingers (2001), Merleau-Ponty tried to find in his last, but unfinished book *The Visible and the Invisible* (1964/1969), the fundamental primitive of the mutual independence between the organism and its world, being eager to bridge the gap between object and subject. In order to do so, he went beneath his earlier notion of the *structural* relation between the organism and its world. Hence, he concluded that both sides rather are different aspects of the same primary whole 'brute being', namely the *flesh*. This new notion can be viewed as two sides of the same coin. This means, the lived body has a twofold character, being able to see or touch on the one hand, and being seen or touched on the other hand. Mingers (2001, p. 116) exemplified this by saying “when we touch an object we are also touched by it and, even more reflexively, when we touch ourselves we are both toucher and touched in a dual sense. It is like a measuring instrument that measures its own internal states”. Hence, this double nature has its roots in the flesh, and in order to describe the very nature of the flesh, Merleau-Ponty argued the flesh is “not matter, is not mind, is not substance. To designate it, we should need the old term “element”, in the sense it was used to speak of water, air, earth, and fire, that is: in the sense of a *general thing*, midway between the spatio-temporal individual and the idea” (Merleau-Ponty, 1969, p. 139, quoted in Mingers, 2001, p. 116).

Thus, so far only questions concerning low-level perception and bodily intentions in Merleau-Ponty’s work are addressed, but he actually tried, contrary to other phenomenologists, to situate his theory in a more cultural context (Neuman, 2001). The main reason for this cultural inclusion may have been his alliance with the French anthropologist Claude Levi-Strauss (Priest, 1998). As a consequence, he argued that each aspect of being in the world was important, and therefore ‘being-in-the-cultural-matrix-of-the-world’ was an “indispensable moment of the lived dialectic” (Neuman, 2001).

Merleau-Ponty (1962) pointed out that these more ‘basic’ bodily intentions could be elaborated further by moving from a “literal meaning to a figurative meaning”, maintaining that their new meaning obviously had a new significance. In order to explain this shift of meaning, he mentioned the skill acquisition of dancing and driving a car, since both behaviors have roots in motor activities. This means, the embodied organism has to practice its muscular movements and bodily actions until it has developed the accurate patterns of behaving in a certain situation, and the very process is the same for cultural practices (Dreyfus, 1972/1979). In a similar way he illustrated how abstract concepts derive their meaning from embodied experience. The notion of a triangle can be stated in an abstract mathematical way, but its underlying understanding can only be grasped through our embodied nature of being in the world (Mingers, 2001, but see the work of Lakoff and Núñez, 2000 mentioned in the introduction).
With respect to language, Merleau-Ponty (1964/1969) argued that human language rests on the same intentionality as its bodily counterpart, which has its roots in intentional relations such as desires, wants, and so on. He claimed that these intentions actually existed in the child before it had the ability to express them linguistically (Loren & Dietrich, 1997). Moreover, he argued that the most basic form of communication is bodily behavior and these acts are the basic foundation for cooperation among organisms. Human language then, is an extension of these bodily acts. Children acquire language in the same way as they acquire other bodily or cultural skills, and the meanings of the words come through their practical use, since “[I] learn it as I learn to use a tool, by seeing it used in the context of a certain situation” (Merleau-Ponty, 1962, p. 402, quoted in Mingers, 2001, p. 115). However, when humans actually have acquired a language, it will control human cognition and communication (Loren & Dietrich, 1997). Merleau-Ponty (1964) treated the relation between language and thought as an intertwined process, arguing that the use of language actually is the process of thinking; either we speak to ourselves or to other persons. Hence, it is through verbalization that consciousness arises (Mingers, 2001).

In sum, the main characteristics of Merleau-Ponty’s work were his non-dualistic and anti-behavioristic explanation of embodied experience. He convincingly argued that bodies are intentional in themselves, as they know ‘how to act’, making cognition a biological phenomena rather than a mental one. Hence, he stressed the connection between action and perception, claiming that perception is really a motor act. Although he emphasized cultural aspects of embodiment, he stressed psychological explanations of perceptual phenomena more, probably as a legacy from his interest in the Gestalt school.

2.4.3 The genetic epistemology of Piaget

The Swiss scholar Jean Piaget (1896-1980) stressed the importance of sensorimotor activity for the emergence of intelligent behavior (e.g. Piaget 1952, 1954), and he is well known for his cognitive development theory. In order to explain the process of intellectual development he proposed a genetic epistemology, meaning the study of the development (genesis) of different kinds of knowledge (epistemology). He did not primarily view his research as limited to psychology, but rather as a combination of biology and logic (cf. e.g. Boden, 1994; Fischer & Kaplan, 2003). According to Sinha and Jensen (2000), the basic motivation of his genetic epistemology was to offer a developmental and biologically based reformulation of the philosopher Kant’s synthetic theory of knowledge. Piaget, like von Uexküll, was deeply inspired by Kant’s idea that the subject’s knowledge is dependent on its own perceptual and conceptual abilities (cf. Ziemke, 2001a). Although, Kant did not address the development process behind the forms and categories that are central in subjective experience, this developmental aspect became the major theme in Piaget’s work (Boden, 1994). During his life he was a prolific producer of literature, and this section can only very briefly address the central tenets of his work.
Cosmides et al. (1997) pointed out that Piaget searched for a middle way between Lamarck’s and Darwin’s work, on the one hand, and the Gestalt school and behaviorism on the other, in order to realize his scientific endeavor (cf. Boden, 1994). Initially, Piaget was a biologist, studying how simple organisms (molluscs) adapted to their environment, arguing that organisms adapt to their environment by different processes. These processes are ‘innate’ (or evolved adaptations) that guide learning and development, being present more or less in all animals, including humans. Piaget’s interest in children’s cognitive development arose when he created tests for investigating intelligence. He discovered that children made systematic kinds of errors, and recognized that these patterns appeared within quite fixed age spans. Piaget argued that children’s intellect was the outcome of their own logic, which was not the same as the logical rules of adults (Fischer & Kaplan, 2003). In 1937, he published the famous book, The Construction of Reality in the Child, in which he offers a model of how infants and children develop different concepts, such as space, time, objects and causality.

A central concept in Piaget’s work is schemata (e.g. Boden 1994; Wadsworth, 1996) or action schemes (e.g. von Glaserfeld, 1995). These schemata/action schemes are the cognitive structures by which humans and other cognizers organize and adapt to the environment, which is similar to von Uexküll’s functional circles (cf. Ziemke, 2001a). However, it should be noted that these schemata are not physical or mental entities; they should rather be viewed as processes within the nervous system, which correspond to different concepts and categories. Moreover, schemata are reflexive by nature, and they can be inferred from mere reflex motor actions in newborns, like sucking and grasping behaviors (Wadsworth, 1996). Piaget supposed that an infant initially possesses very few schemata, but as the child grows and develops, these schemata progressively become more generalized and discriminated. This means, schemata are not fixed, since they change and become refined during ontogeny. For instance, a newborn infant sucks on nearly anything that is put into his/her mouth, but after a while the infant learns to differentiate between stimuli that can be associated with milk and stimuli that cannot. Hence, the infant has to develop various distinctive sucking schemata, one for milk producing stimuli, and another for non-milk producing stimuli (Wadsworth, 1996). However, these basic schemata should not be characterized as ‘mental’ or ‘cognitive’ in the traditional sense, rather they can be viewed as perception-action loops, functioning as a precursor of forthcoming ‘mental’ activities (Boden, 1994). During the course of development, the basic reflexive schemata of the child improve, and become more differentiated and numerous, gradually resulting in a more complex network. All the schemata in an adult are derived from the basic and reflexive sensorimotoric ones. Thus, cognitive development is the process of construction and reconstruction of these dynamical and integrated schemata, and Piaget stated that “every schema is...coordinated with all other schemata

10 Originally published as La Construction du Reel Chez l'Infant in 1937, the English version appeared in 1954.

According to von Glasersfeld (1995), the concept of schema or action scheme is derived from the biological description of reflexes, but not limited to a ‘simple’ stimulus-response mechanism. Instead, this particular notion actually contains three crucial elements. Firstly, the identification of a certain situation, subsequently, a specific activity associated with the situation at hand, and finally the infant’s subjective expectation that the performed activity actually produces the desired result. This means, the process of an infant’s searching for the mother’s breast, in order to find milk, should be viewed as a process of meaningful organization of agent-environment interaction (cf. Ziemke, 2001a). Von Glasersfeld (1995, p. 14) mentioned that Piaget considered “cognition as an instrument of adaptation, as a tool for fitting ourselves into the world of our experiences”. Hence, cognition in Piaget’s sense concerns the organization of the agent’s sensorimotor experiences and interactions with the environment.

The processes of change and refinement of schemata are characterized as assimilation and accommodation (cf. e.g. Boden, 1994; Fischer & Kaplan, 2003; Wadsworth, 1996). On the one hand, assimilation is the mental process of integrating and adding new “perception-action loops” or conceptual substance into an existing schema, allowing growth of the actual schema. It should be noted that this means no radical change to the schema itself. Accommodation, on the other hand, is the process of constructing new schemata or modifying an earlier existing one in order to fit the new stimuli. It is important to notice that the resulting schemata are the child’s individual constructions, and not any integrated ‘copies’ of the external reality. The form and content of a certain schema is determined by the individual’s unique experiences of perception-action interactions with the environment. This means, the overall behavior of an agent reflects both assimilation and accommodation. The state of balance between assimilation and accommodation is called equilibrium, and equilibration is the process of moving from imbalance to balance. Equilibration is a self-regulatory internal mechanism, using the mechanisms of assimilation and accommodation, which allow external experiences to be incorporated into schemata. However, it should be noted that equilibrium is a temporary state, since new experiences might result in a state of disequilibrium, then triggering the processes of either assimilation or accommodation (cf. e.g. Boden, 1994; Fischer & Kaplan, 2003; Wadsworth, 1996).

Wadsworth (1996) pointed out that the relationship between action and knowledge is central in Piaget’s genetic epistemology. In addition to the processes of accommodation and assimilation, physical and mental actions are also necessary for the development of intelligence. Moreover, Piaget viewed knowledge as a construction emerging from the child’s active interactions. Fischer and Kaplan (2003, p. 681) paraphrased Piaget as follows:
I am a constructivist. I think that knowledge is a matter of constant, new construction, by its interaction with reality, and that it is not pre-formed. There is a continuous creativity.

Piaget identified three kinds of knowledge, each of them resulting from the child’s interactions with the environment, namely physical, logical-mathematical, and social knowledge (Wadsworth, 1996). Physical knowledge is characterized as the subject’s knowledge of events and concrete objects, which is derived from the child’s sensorimotoric manipulations of these objects. Through these actions the child discovers and constructs physical knowledge about the shape, size, weight, texture, of a certain object such as a pot, for example. Logical-mathematical knowledge is the result of mental actions (thinking) about experience with objects. It should be noted that the child invents this form of knowledge rather than discovers it as with physical knowledge. However, it is important to note that the child’s inventions are derived from his/her earlier sensorimotor actions with objects. An example of logical-mathematical knowledge is number concepts, which the child develops by manipulating different arrangements of objects, materials and sets. In this form of knowledge, the objects themselves serve as a medium for the child’s inventions, rather than being the focus of interest. The last form of knowledge is social knowledge, which is constructed from the child’s interactions with other people. Examples of social knowledge include social and cultural conventions of how to behave in certain situations, e.g. rules, laws, values, morals, ethics and language systems. As a consequence of different cultural settings, various forms of social knowledge exist. From the beginning, the infant is more dependent on sensorimotoric experience, before he/she is able to possess the ‘power’ of symbolic representation in the form of language. The initial and basic schemata are a kind of sensorimotor representations of objects, and later on these sensorimotor actions become more conceptual in their form (Wadsworth, 1996). Broadly speaking, the child can only construct knowledge by directly acting on objects or interacting with others, not by viewing other people manipulating these objects or telling the child about their characteristics. Hence, constructed knowledge is the result of the child’s subjective firsthand experiences of interactions with objects and other people.

Piaget identified four major cognitive developmental stages that correspond to qualitative differences in intellectual development (cf. e.g. Boden, 1994; Fischer & Kaplan, 2003; Gredler, 1992; Miller, 1983; Wadsworth, 1996). These stages should not be viewed as distinct steps on a stairway, but rather as a successive and cumulative process. Piaget viewed cognitive development as a continuum, wherein the latter stages are derived from previous ones, being incorporated and transformed to a ‘higher’ level, and each stage in turn consists of several sub-stages. Piaget characterized the main four stages as follows (cf. e.g. Gredler, 1992; Miller, 1983; Wadsworth, 1996):

1. The stage of sensorimotoric intelligence (0–2 years of age).
2. The stage of preoperational thought (2–7 years of age).
3. The stage of concrete operations (7–11 years)
4. The stage of formal operations (11–15 years).
During the first stage ‘cognition’ is primarily characterized as sensorimotor intelligence, and it addresses no conceptual thinking, but the child’s behavior begins to be goal-directed. The cognitive development is mainly focused on constructing and discriminating different schemata, building concepts about how reality works. In the next stage, preoperational thought, the child is able to develop different representational systems like language and other conceptual systems. Piaget regarded language as a demonstration of the symbolic function, displaying a certain stage of mental development. In fact, Piaget rather viewed language as an aid for intellectual development, but not as a necessary means for a producer of intelligence (Wadsworth, 1996). The child’s reasoning and logical ability is limited to tasks directly at hand, being dominated by perception, and viewed as pre-logical. In the concrete operations stage, the child begins to apply more formal logical thought to real problems that he or she encounters directly. It is in the final stage that the child has developed the ‘highest’ form of reasoning and can apply logical reasoning to all kinds of problems, and is capable of hypothetical reason and abstract thought. Hence, the child has developed the most advanced form of cognitive functioning. It should be noted that the age spans are not inflexible, but should rather be viewed as normative, although Piaget believed that the order is fixed and regression from one stage to another would not occur. Piaget also claimed his theory of cognitive development is universal, as a consequence of the logical formalization of his theory. The basic organizational force of intellectual development is logic, and this is highlighted in the four stages, characterizing different forms of logical thinking (Fischer & Kaplan, 2003; Gredler, 1992; Sinha & Jensen, 2000; Wadsworth, 1996).

In order to move within and between these stages, Piaget identified four important factors for cognitive development, namely maturation, active experience, social interaction, and equilibration (cf. e.g. Wadsworth, 1996). Maturation and heredity play a role in intellectual development, setting a wide constraint for the developmental process. The importance of active experience is already addressed in this section, and it is central for assimilation and accommodation. Social interaction is characterized as the interchange of ideas among individuals, and it is of major importance for the development of social knowledge. According to Sinha and Jensen (2000), it is often argued that Piaget neglected the social dimension, which they, however, consider a common misunderstanding. They point out that he actually stressed the social dimension as an essential factor for the development of cognition, as stated, for example, in the quote below:

Whether we study children in Geneva, Paris, new York or Moscow, in the mountains of Iran or the heart of Africa, or on an island in the Pacific, we observe everywhere certain ways of conducting social exchange between children, or between children and adults, which act through their functioning alone, regardless of the context of information handed down through education. In all environments, individuals ask questions, work together, discuss, oppose things and so on; and this constant exchange between individuals takes place throughout the whole of development according to a process of socialization which involves the social life of children among
themselves as much as their relationships with other children or adults of all ages (Piaget, 1972, p. 35, cited from Sinha & Jensen, 2000).

According to Sinha and Jensen (2000), Piaget supposed that socio-cultural factors could either accelerate or retard the developmental process. He actually admitted that in some cultural settings the final stage of formal operational thinking might not emerge in the individuals. This means, Piaget viewed the social-cultural factors as a necessary and sometimes limiting condition, which could only influence the speed, not the direction, to the terminal point of (logical) cognitive development. The final important factor of intellectual development is the previously mentioned mechanism of equilibration, which functions as a coordinator and self-regulator between the three other factors during cognitive development.

Piaget’s ideas initially did not fall on receptive ground outside the French-speaking community. Possible reasons for this were that he was only published in French, and the dominance of the behavioristic era in the United States (Fischer & Kaplan, 2003). However, as Vauclair and Perret (2003) point out, it is amazing to note that Piaget already in 1955 founded the International Center for Genetic Epistemology (ICGE) in Geneva, financed by the Rockefeller Foundation. Moreover, Piaget was positive towards multidisciplinary research, and can be regarded as a forerunner to the interdisciplinary approach in modern cognitive science. His most significant contribution is the claim that cognitive changes are the outcome of a developmental process. Sinha & Jensen (2000) point out that while his stages are not central in today’s developmental theories, his constructivistic, biological, and epigenetical ideas really are.

However, it should also be mentioned that Piaget’s theory has been criticized heavily. The main criticisms have focused on his extensive research on his own children, his use of methodologically flawed experiments, his inaccurately corresponding age-spans, and his failure to explain the mechanisms behind the process of equilibration (cf. e.g. Miller, 1983; Wadsworth, 1996). Boden (1994) noted that Piaget actually underestimated children’s abilities and innate capabilities, both initially and later on during development. Another main issue is that his theory is based on intellectual ‘ideals’ of the Western tradition, not paying enough attention to cultural differences in cognitive development, as illustrated by his claim that his theory was universal (Cole & Wertsch, 1996). Another issue is his neglect of the relational aspect of cognition, an issue that Vygotsky and Mead much more strongly emphasized.

2.4.4 The cultural and historical approach of Vygotsky

The Russian scholar Lev Vygotsky (1896-1934) viewed individual cognition and intelligence as culturally based, grounding his theory in the cultural history of the human species and the child’s interactions with other people in its particular culture. Unfortunately, he died at the age of 37 in 1934, and his work did not reach the Western world until 1962, when the first public

---

11 Most of this section has been published in Lindblom & Ziemke (2002, 2003).
translation in English appeared. One cause for this delay is probably the fact that Vygotsky’s work was banned in the Soviet Union from the mid-1930s to the mid-1950s. According to Kozulin (1986), Vygotsky was initially active during an era when Russian psychology was dominated by behavioristic reflex theories, proposed by, e.g., Pavlov and Bekhterev. Besides contemporary Russian psychology, Vygotsky was inspired by the work of Hegel and Darwin, and was familiar with Gestalt psychologists such as Koffka, Buhler, and Köhler, as well as the early work of Piaget. However, Vygotsky himself was critical of both behaviorism and Gestalt psychology; he argued that these studies, in their “zoological models” removed the essential differences between human and animal intelligence. He therefore claimed that human intelligence was “more than a leather sack filled with reflexes” (1979, p. 9), arguing that the existing psychological theories had failed, since they were not capable of explaining all the structures of human behavior. Instead, Vygotsky (1979) searched for a psychological theory that would describe the development of the abilities that are exclusively human. He claimed this could only succeed if all dimensions of the human mind were analyzed, but not in the form of introspectionism. Vygotsky (1979) was critical of this ‘mentalistic’ convention, since it, in his opinion, confined itself through circular reasoning into which states of consciousness were ‘explained’ by the term of consciousness. As an alternative, he argued that if consciousness is taken as the subject of study, then its explanation must be sought in some other dimension of reality. Vygotsky proposed that socially meaningful activities play this role as a ‘producer’ of consciousness, arguing that the individual mind is constructed from the outside, through interactions with other people. In his own words:

The mechanism of social behavior and the mechanism of consciousness are the same ... We are aware of ourselves in that we are aware of others; and in an analogous manner, we are aware of others because in our relationship to ourselves we are the same as others in their relationship to us (Vygotsky, 1979, p. 29).

This means, the nature of individual human intelligence is, according to Vygotsky, developed through interactions with the environment in general, and more precisely it is the result of social interactions with other human beings.

Vygotsky (1978) distinguished between elementary and higher mental functions. He asserted that our elementary mental functions had to be those functions that were genetically innate and existed both in humans and (other) animals. These elementary or natural mental functions include for example, simple memory, perception and attention. These mental functions are controlled by the recognition of co-occurring stimuli in the environment, which Vygotsky (1978) referred to as signalization. According to Vygotsky, the higher or cultural mental functions are exclusively human and emerge dynamically through radical transformations of the lower ones. In elementary functions there is a direct link between a stimulus in the environment and a response from the creature, which Vygotsky (1978) expressed by a stimuli → response formula. However, for a higher mental
function the structure differs significantly, since it involves an *intermediate link* between the stimulus and the response, as illustrated in figure 1.

![Stimuli Response Mediated act](image)

**Figure 1.** The organization of higher behaviour via a mediated act. (Adapted from Vygotsky 1934/1978, p. 40).

Vygotsky (1978) claimed that this type of organization is fundamental to all higher cognitive processes, although typically in a much more complicated structure than illustrated above. The intermediate link involves the psychological tool which is ‘drawn into’ the cognitive operation to fulfill a special function, namely creating an altered relation between stimulus and response. The higher mental functions lie *outside* the individual, in the form of psychological tools and interpersonal relations. The lower functions do not disappear in the ‘developed’ or ‘enculturated’ mind, but undergo some re-organizations according to particular forms of human cultural activity (Kozulin, 1986).

Vygotsky particularly focused on the factors that distinguish between elementary and higher mental functions. Primarily he mentioned the shift of control from the environment (*signalization*) to the individual’s voluntary regulation of his/her behavior. He subsequently claimed that social origins and nature are the driving forces of higher mental abilities, as well as the use of *psychological tools* that mediate higher mental functions. For example, Vygotsky (1977; 1978) argued that such a simple operation as tying a knot in a handkerchief to function as a memory cue altered the psychological construction of remembering. As a result, the memory process was extended beyond the biological inherited factors; the incorporation of artificial or self-generated stimuli in the form of *psychological tools* was the key difference between animal and human behavior. He maintained that previously in human evolution, humanlike ancestors developed simple tools, and this invention led to a shift of behavior, resulting in an important change in the pattern of thinking. Vygotsky (1977, 1978) called this process of conveying meaning to arbitrary stimuli as *signification*. He argued that (other) animals were not capable of performing such operations, which demarcate the starting point of human intelligence.

The invention, and use of arbitrary stimuli as psychological tools to perform advanced cognitive ‘tasks’ such as remembering, decision-making, etcetera is according to Vygotsky (1977; 1978), analogous to the human invention and use of technical tools such as hammers, saws, spades and ploughs. However, this analogy has significant differences, because the two separate activities have crucial distinctions according to Vygotsky. The basic foundation in the analogy between a psychological tool and a technical one lies in their *mediating function*, which characterizes both of them.
Consequently, they can be included in the same category, from a psychological standpoint. Vygotsky (1978) argued that the essence of the use of psychological tools for mediated activity is they influence and have an effect on human behavior, since actions conducted with these psychological tools create thoughts. In 1933 he stated “the central fact about our psychology is the fact of mediation” (Vygotsky, 1982, p. 166, quoted from Wertsch, 1985, p. 15).

The most important distinction between a technical tool and a psychological one lies in how they affect human behavior. The technical tool is externally oriented, towards changing objects, whereas psychological tools are internally oriented, by changing ways of thinking, controlling, regulating, and organizing behavior. As a consequence, both technical and psychological tools transform cognition. The psychological tools bridge the gap between elementary and higher mental functions, and they include “various systems for counting; mnemonic techniques; algebraic symbol systems; works of art; writing; diagrams; maps, and technical drawings; all sort of conventional signs, and so on” (Vygotsky, 1981, p.137, quoted in Cole & Wertsch, 1996). Of the psychological tools, mediating our thoughts, feelings and behavior, he considered language the most significant. Vygotsky (1978, 1986) stressed that the primary function of language, in the form of speech, is a device for social contact, and interpersonal communication, influencing other people. Later, this social speech transforms and becomes egocentric speech, which internalizes social speech for the child’s own ends. Vygotsky (1986) argued that this egocentric speech is a shift from social speech (between people) to inner speech, which ‘goes’ inward into the mind, by directing our own thinking. Consequently, the interpersonal becomes intrapersonal, and ‘actions’ with this special psychological tool create thought. Thus language liberates us from our immediate perceptual experience and allows us to also represent the past, the future and the un-present. Thinking and language are dynamically related, since understanding and producing language are processes that transform the process of thinking.

As a result of his analysis of the differences between animal and human behavior, resulting in elementary and higher functions, Vygotsky identified two different influences on psychological development, namely biological principles and sociohistorical factors (Vygotsky, 1978). According to Vygotsky (ibid.), biological factors are part of our ontogenetic development, and incorporate the development of the physiological body. These biological factors control the initial months of life in infants, responsible for the development of perception, basic memory and spontaneous attention. Vygotsky called the emergence of these elementary mental functions natural (or primitive) development. The second line of development is sociohistorical, and it appeared with the invention and use of culturally based psychological tools (signification) in primitive humans. These tools function as ‘regulators’ of human social behavior, and especially language is an important ‘organizer’, both in the form of speech and written text. The line of sociohistorical development separate human behavior from animal behavior, as well as also having a significant role in the cognitive development of the
individual child, since the child literally is born into the psychological tool systems of its particular culture. Vygotsky characterized the importance of these two lines of development for individual intelligence as follows:

The cultural development of the child is characterized first by the fact that it transpires under conditions of dynamic organic changes. Cultural development is superimposed on the process of growth, maturation, and the organic development of the child: It forms a single whole with these processes. It is only through abstraction that we can separate one set of processes from another.

The growth of the normal child into civilization usually involves a fusion with the processes of organic maturation. Both planes of development - the natural and the cultural - coincide and mingle with each other. The two lines of change interpenetrate one another and essentially form a single line of sociobiological formation of the child’s personality (Vygotsky 1960, p. 17, quoted from Wertsch, 1985, p. 41).

Hence, the cognitive abilities of an ‘enculturated’ adult human are the product of these processes of cognitive development, in which ‘primitive’ and ‘immature’ humans are transformed into cultural ones. In simple terms, the child initially has to learn the particular psychological tools of its culture, and then learns how to use them to master and control its own behavior. This transformation process, from elementary (or natural) mental functions to more complex higher functions is described (not explained) by two key principles, namely, the process of signification (using psychological tools), and a principle referred to as the general law of cultural development (Wertsch, 1985). The essence of the latter is as follows:

Every function in the child’s development appears twice: first, on the social level, and later, on the individual level; first, between people (interpsychological), and then inside the child (intrapsychological) ... All the higher functions originate as actual relations between human individuals ... The transformation of an interpersonal process into an intrapersonal one is the result of a long series of developmental events. ... The internalisation of socially rooted and historically developed activities is the distinguishing feature of human activity, the basis of the qualitative leap from animal to human psychology (Vygotsky, 1978, p. 56-57, original emphases).

Vygotsky (1978) called this process of transforming an interpersonal process (human-to-human interaction) into an intrapersonal one internalization. To illustrate the essential role of social interactions during this transformation process Vygotsky (1978) used the example of the development of pointing in the child. He claimed that initially it is only a simple and incomplete grasping movement directed towards a desired object, only represented by the child's reaching and grasping movement, and nothing more. When the caretaker comes to help the child, the meaning of the gesture situation itself changes, by obtaining another meaning, as the child’s failed reaching attempt provokes a reaction, not from the desired object, but from another person. The individual movement ‘in itself’ becomes a gesture ‘for-others’. The caretaker in this case interprets the child’s grasping/reaching movement as a kind of pointing gesture, resulting in a socially meaningful
communicative act, whereas the child itself at the moment is not aware of its communication ability. However, after a while the child becomes aware of the communicative function of its movements, and then begins addressing its gestures towards other people, rather than the object of interest that was its primary focus initially. Thus, “[t]he grasping movement changes to the act of pointing” (ibid, p. 56). As Kozulin (1986) pointed out, it is essential to note that the child itself is the last person who ‘consciously’ grasps the ‘new’ meaning of its own pointing gesture.

Another central concept in Vygotsky’s theory is the so-called zone of proximal development, which is related to the process of internalization in the child, transforming interpersonal functions into intrapersonal ones. It is in the zone of proximal development that the child learns, through social interactions, how to use the tools available, especially the psychological ones. Vygotsky (1978, 1986) noticed that when a caretaker gives meaning to the child’s interaction, when the child is unable to do so for itself, the child is working in the zone of proximal development, which Vygotsky characterized as follows:

> It is the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers (Vygotsky, 1978, p. 86).

The caretaker realizes the child’s achievement by means of hints, explanations, encouragements, regulating and controlling the child’s focus of attention and so on. Vygotsky (1978) also related imitation and learning to the zone of proximal development. He argued that a child can merely imitate what is within its zone of proximal development, explaining if a caregiver presents a too advanced solution to a problem, the child did not grasp the solution, even when the solution was presented repeatedly. The child can therefore only ‘imitate’ and adopt a solution to a problem or an activity if it is within the boundaries of the child’s particular zone of proximal development. Moreover, Vygotsky (1978) argued that only humans possess a zone of proximal development:

> A primate can learn a great deal through training by using its mechanical and mental skills, but it cannot be made more intelligent, that is, it cannot be taught to solve a variety of more advanced problems independently. For this reason animals are incapable of learning in the human sense of the term; human learning presupposes a specific social nature and a process by which children grow into the intellectual life of those around them (Vygotsky, 1978, p. 88, original emphasis).

Thus, according to Vygotsky, the ‘mind’ of the chimpanzee, for example, can never be developed and extended further than their biological heritage, since they lack a zone of proximal development. In other words, Vygotsky followed an evolutionary course stressing that both biological and social factors are responsible for higher mental functions. They are not controlled by conditioned reflexes or genetically determined, rather ‘mediated actions’. However, the continuity ends here, as Vygotsky claimed that the process of mediated action was a uniquely human ability. However, it can be pointed
out that Vygotsky misjudged animals because in his time relatively little was known about the richness of their social interactions and, perhaps even more importantly, the similarities of the biological mechanisms underlying human behavior and that of other animals (for a more detailed discussion, see Lindblom & Ziemke, 2003). According to Davydov and Radzikhovskii (1985), there is a major gulf between ‘Vygotsky the psychologist’ and ‘Vygotsky the methodologist’. They point out that Vygotsky almost exclusively focused on the socio-cultural forces in his empirical studies, and that he neglected the biological line of development, especially the physical maturation in the child during its first years of life. They further argue that Vygotsky tended to view the biological factors as ‘raw materials’, which were then transformed by the socio-cultural forces, whereas he mentioned almost nothing about how changes in the biological factors may influence the socio-cultural ones. Wertsch (1985), on the other hand, argued that Vygotsky himself was aware of the necessary, but not sufficient conditions provided by the biological factors, since he assumed that the natural factors play the major role in early ontogeny, and that the cultural forces take the leading role later on. Hence, Wertsch argues that Vygotsky did not view advanced cognition and thinking as the outcome of social factors alone, since he stated that “culture creates nothing; it only alters natural data in conformity with human goals” (Vygotsky, 1960, p. 200, cited in Wertsch, 1985).

In sum, Vygotsky’s theory of cognitive development particularly stresses that individual intelligence emerges as a result of biological factors (embodiment, one might say in today’s terms) that interact with a physical and especially a social environment (in today’s terms: situatedness) through a developmental process. Interestingly, whereas behaviorism had a great impact within psychology in the United States by ignoring any mental content, the Vygotskyan approach and its successors took a totally different approach already during the 1930s and onwards, stressing the role of mediation, embodiment and situatedness for the development of higher mental functions.

2.4.5 Dewey and Bartlett

Two scholars who are usually not explicitly associated with theories of embodied and situated cognition (for an exception see Clancey, 1997), but who in my opinion deserve more attention than they so far have received are presented here. The reasons for mentioning them are that John Dewey’s and, in particular, Mead’s ideas (presented in the following section) have some interesting similarities with Vygotsky, since they both stressed the biological and social nature of intelligence. Moreover, Sir Frederic Bartlett emphasized the constructive aspect of cognition, and his schema theory has had a great impact on the role of internal representations in the forthcoming field of cognitive science. However, his ideas of schema theory focused on cultural and social aspects, which were largely ignored in the theories proposed by classical cognitive science.

The American philosopher John Dewey (1859-1952) tried to elaborate a non-dualistic view between knowledge and the world, arguing that it is not
possible to go beyond ourselves in order to see the ‘real’ world (cf. e.g. Field, 2001). Hence, it is not possible to think of the world without describing it from the subjective point of view, and this line of argument is the basic foundation of all his work. According to Marsh (2006), Dewey can be considered as the first “ghost-buster”, in his efforts to provide an anti-dualistic/non-Cartesian philosophy of mind. Furthermore, Field (2001) points out that Dewey tried to elaborate a new ‘theory of knowledge’, arguing that the traditional epistemologies, such as rationalists and empiricists, had drawn a too sharp distinction between thought and knowledge, since thought was supposed to exist apart from the world (Field, 2001). Dewey’s theory rejected these traditional assumptions, and he argued that knowledge should be viewed in a practical way, instead of striving after the ‘figment of your imagination’ of objective knowledge (Gardner, 1987). According to Field (2001), Dewey was influenced by psychologist William James and German philosopher Hegel, but his primary source of inspiration was Darwin’s theory of natural selection. Dewey supposed that Darwin’s idea had swept away ‘supernatural’ explanations of the origins of species, since the theory of natural selection stressed that living organisms are the product of a natural and sequential process of adaptation to the environment (Field, 2001). Consequently, Dewey proposed a naturalistic approach to a ‘theory of knowledge’, which considered the development of knowledge as a process of active human adaptations to changing environmental conditions. Contrary to traditional approaches of ‘knowledge’, Dewey described thought genetically, as the product of the interaction between organism and environment, and knowledge was supposed to have a ‘practical instrumentality’ in the guidance and control of that interaction (Field, 2001). As a result, Dewey used the term “instrumentalism” for his new approach. The first significant application of this ‘naturalistic understanding’ was offered in his seminal article ‘The reflex arc concept in psychology’ in 1896, previously presented in more detail in section 2.2. In that article, he attacked the basic units of analysis within behavioral psychology, namely ‘stimulus’ and ‘response’.

Broadly speaking, behavior was at that time described as a conversion of a response to a certain stimulus, but Dewey argued that this description is partial, since it only explained what happens between the reception of a stimulus and the generation of an action. Instead, Dewey supposed that behavior should rather be viewed as a control process, which coordinates ongoing activity (cf. e.g. Clancey, 1997; Pfeifer & Scheier, 1999). He maintained that perception and action are closely coupled, and he called this coupling “sensory-motor coordination”. This close coupling is especially present in contemporary work in ‘active vision’ (cf. Pfeifer & Scheier, 1999). Furthermore, there is evidence that vision should not be viewed as passive information processing, instead it is an active and integrated sensorimotor outcome. Current research on active vision is an instantiation of the general idea of Dewey’s claim that perception and action are closely linked (cf. e.g. Clancey, 1997; Pfeifer & Scheier, 1999).

According to Field (2001), Dewey proposed in Experience and Nature (1925) that the mind has a social origin, and that the human individual is initially a
social being. Hence, he viewed the mind as an emergent function which has its foundation in natural processes, shaped through interactions between human beings and the world in which they live (Field, 2001). Dewey’s ideas can be interpreted in today’s terms as being situated and/or embodied, and Pfeifer and Scheier (1999) point out that the approach of ‘sensory-motor coordination’ is of fundamental importance for an embodied cognitive science. Instead of viewing different forms of intelligent behavior (e.g. memory, perception, and learning) as ‘sense-model-act’ series, the principle of ‘sensory-motor coordination’ offers an alternative explanation. However, despite the fact that the work of Dewey is well-known in many scientific areas in general, his work is to a large extent ignored in the cognitive science literature, although his name is mentioned quite often within the areas of learning and education. Similarly, Rockwell (2005) argues that Dewey’s work is grossly disregarded in the situated/embodied approaches to mind, albeit he is a fore runner to these ideas. Dewey’s work, for instance, is only typically mentioned in a few sporadic lines or not at all (cf. e.g. Bechtel, Abrahamsen & Graham, 1998; Clark, 1997; Gardner; 1987; Hendriks-Jansen, 1996; Varela et al., 1991). Although his work has strong similarities with today’s embodied and situated theories of cognition, only Pfeifer and Scheier (1999), Clancey (1997) as well as Rockwell (2005) strongly address Dewey’s ideas in depth from an embodied/situated perspective of cognition.

Sir Frederic Bartlett (1886-1969) was an experimental psychologist active in Great Britain, and he is most well known for his work on memory which he presented in his landmark book, *Remembering - a Study in Experimental Psychology and Social Psychology* (1932). In that book he described his studies on memory and presented the foundation for his theory of schema, which later influenced the fields of cognitive science and psychology (Roediger, 2003). Roediger (2003) points out that the main theme in Bartlett’s (1932) book is to emphasize the constructive nature of cognition. In one of Bartlett’s most well-known experiments, the subjects read a native American folktale called *The war of the ghosts*, and then are encouraged to retell the story at different time intervals. Bartlett recognized that the subjects changed details in the story that they did not understand, since the cultural context of the story was quite different from their own socio-cultural background. Bartlett coined the term “effort after meaning” in order to refer to the subject’s actual behavior when they were confronted with elements that they did not understand in the foreign folktale.

As a result of these observations, Bartlett suggested that memory is a subjective process of reconstruction, which in fact is a highly social act, and the idea for schema theory was born (Sinha & Jensen, 2000). Roediger (2003, p. 320) paraphrases Bartlett’s conclusion that “the most general characteristic of the whole group of experiments was the persistence, for any single subject, of the form of his first reproduction” and the utilization “of a general form, order and arrangement of material seems to be dominant, both in initial reception and in subsequent remembering”. Bartlett denoted the ‘general form’ that humans encoded and remembered their experiences as “schema”. According to Sinha & Jensen (2000), Bartlett considered the
concept of schema as both a common principle of cognitive organization, and as a unit of analysis for revealing social and cultural variations in different models of the world. Moreover, Saito (1996) mentions that Bartlett suggested that more advanced cognitive processes, like remembering and thinking for instance, are active, selective and constructive processes, which are formed and influenced to a great extent by social factors.

Clancey (1997) points out that Bartlett is usually credited for introducing the concept of schema as traditionally used in the descriptive models of cognitive science and AI. However, he notes that Bartlett himself strongly argued against prototypical descriptions of object and the term schema as such, since it could be viewed as something persistent, and not stressing its constructive nature. As Clancey (1997) reveals, the strong emphasis on representations during the 1970s and 1980s generally missed Bartlett’s point of view, and the fact that “Bartlett was striving for a psychological theory that viewed the organism as a whole, within its everyday experience of perceiving and acting in some context” (Clancey, 1997, p. 48).

Saito (1996) points out that Bartlett’s phylogenetic explanation of cognition has similarities with the theories of Piaget and Vygotsky, since Bartlett also stressed a “tripartite” origin of cognition, namely biological, psychological and social factors. Saito (1996) notes that including the social perspective of cognition does not automatically rule out the individual aspect, the real problem actually starts when taking an ‘either-or’ position. Instead, Saito proposes an approach that encompasses all these significant factors, taking a multi-level perspective. As a departing point for such a hypothetical approach, Saito (1996) reviews Bartlett’s (1923, 1932) work on the social foundation of cognition. His conclusion is that previous reviews of Bartlett’s work concerning social origins do not pay enough attention to Bartlett’s multi-level analysis which, in his opinion, emphasizes an embodied mind approach.

2.4.6 Mead’s theory of the socially interactive mind

Finally, the American philosopher George Herbert Mead (1863-1931) is considered one of the founders of Pragmatism as well as having great influence in the area of social psychology (Cronk, 2005). However, his most significant contribution, from a cognitive viewpoint, is the work on the role of social interaction for the emergence of mind and consciousness. Similarly, Lewis (1999) stressed that of utmost importance in Mead’s theory of the interactive mind is the opinion that knowledge of the self and knowledge of others are interdependent.

Mead described in *Mind, Self and Society* (1934) how the individual mind and self emerges from social interaction, in particular, he stressed that individual cognition is exceptionally and deeply social. In other words, according to Mead (1934), social practice is prior to the structures and processes of individual cognition. Given that mind arises from social interaction and communication, it cannot be explained independently of these processes. Like the mind, the self is also socially emergent, which means that the
individual self is the *outcome* of social interaction and *not* the precondition for social interaction and cognition. This means that Mead’s theory of the self is in contrast to individualistic approaches of the self, which assume priority of selves to the social process (Cronk, 2005; Gillespie, 2005). For instance, Mead pointed out that:

The self is something which has a development; it is not initially there, at birth, but arises in the process of social experience and activity, that is, develops in the given individual as a result of his relations to that process as a whole and to other individuals within that process (1934, p. 135).

Mead (1934) considered the process of communication as involving two stages, namely *conversation of gestures* and *conversation of significant gestures*. It should be strongly emphasized that Mead, in both cases, took for granted the social context within which two or more individuals interact with one another. In general, *conversations of gestures* are manual gestures, bodily postures, facial expressions and so on. According to Cronk (2005), a famous example of this kind of communication is the dog-fight, in which Mead described how two dogs interact with each other through bodily expressions.

The act of each dog becomes the stimulus to the other dog for his response. There is then a relationship between these two; and as the act is responded to by the other dog, it, in turn, undergoes change. The very fact that the dog is ready to attack another becomes a stimulus to the other dog to change his own position or his own attitude. He has no sooner done this than the change of attitude in the second dog in turn causes the first dog to change his attitude. We have here a conversation of gestures. They are not, however, gestures in the sense that they are significant. We do not assume that the dog says to himself, “If the animal comes from this direction he is going to spring at my throat and I will turn in such a way.” What does take place is an actual change in his own position due to the direction of the approach of the other dog (ibid., p. 42-43).

Taken together, communication occurs between the dogs without any awareness of the responses that the gestures generate in the other dog because neither one nor the other individual dog is aware of the reactions of the other to its gestures. This means, the dog is unable to respond to its own gestures from the standpoint of others. Despite the fact that the individual participants in the conversation of gestures interact socially, they obviously do not know that they are communicating (Cronk, 2005). In other words, the phase of conversation of gestures is considered *unconscious communication* in Mead’s view. *Conversation of significant gestures*, however, is communication through human language, which Mead mostly referred to as vocal gestures. He maintained language is the means of social interaction by *significant symbols*. The crucial characteristic of a significant symbol is its ability to function correspondingly to both *i*) the individual making the *vocal gesture*, and to *ii*) the response called out in others to whom the *vocal gesture* is directed. Furthermore, Mead’s approach is developmental, because it is from the conversation of gestures that human language arises in the form of
conscious communication. Thus, more advanced forms of communication emerge from more or less basic forms of social interaction. In humans, language exceeds but does not eliminate the conversation of gestures and characterizes the transition from non-significant to significant interaction (Cronk, 2005; Gillespie, 2005).

Gillespie (2005) argues that ‘taking the attitude of the other’ is the crucial characteristic of the significant symbol. This means, the significant symbol is not only a shared singular perspective, but needs the self to take the attitude of both the self and the other at the same time. In other words, it functions as a kind of social duality beyond mere sharing between individuals. Gillespie (2005) proposes that this ‘social duality’ is accomplished by Mead through the social act. In order to develop the argument, we return to the previously presented reflex arc by Dewey. Broadly speaking, Mead’s social act is a reply to the dualistic conception of interaction that dates back to Descartes (see subsection 2.2) as well as an elaboration of the Deweyan act. Mead claimed that the Deweyan act is incomplete since it only deals with subject-object interaction, and is therefore unable to shed any light on mind and consciousness. Instead, Mead focused on subject-subject interaction, which then became the breakthrough in his work. Gillespie (2005) denotes Mead’s approach to the feedback theory of consciousness, because it is the feedback we receive from others that establishes the meaning of our actions, which in turn generates consciousness. Citing Mead, Gillespie writes:

> We are conscious of our attitudes because they are responsible for the changes in the conduct of other individuals. A man’s reaction toward weather conditions has no influence upon the weather itself. It is of importance for the success of his conduct that he should be conscious not of his own habits of response, but the signs of rain or fair weather. Successful social conduct brings one into a field within which consciousness of one’s own attitudes helps toward the control of the conduct of others” (Mead, 1910a, p. 403 in Gillespie, 2005, p. 25).

In accord with this remark, we return to Vygotsky’s pointing example, in which the child’s reaching attempt is goal directed, but notably the child initially has no awareness of the act itself, since only the desired object is meaningful, and not the act of reaching. However, Vygotsky then introduced the mother, and Gillespie emphasizes that from her perspective, the reaching act is meaningful. Hence, through social feedback, the mother changes the grasping attempt into pointing, but it should be stressed that the social feedback does not imply the child is yet able to take the perspective of the other, in this case – the mother. Accordingly, the mother’s attitude toward the child, engaged in the Deweyan act with some object, provides the necessary social feedback. Therefore, there is a distinct shift from a dyadic to a triadic model of interaction, and meaning emerges from these subject-subject interactions. Thus, according to Mead (1934), meaning arises within

---

12 Although this stance has similarities with Vygostky’s ideas (see subsection 2.4.4), Gillespie (2005) mentions that this social ‘mirror’ approach dates back to the work of Adam Smith (1759, in Gillespie 2005) by which Mead might have been inspired.
social interaction between individuals, and is not pre-given *per se*. Furthermore, conversation of significant (vocal) gestures functions both externally, i.e., between different individuals, and internally, i.e., between the individual and oneself. Most generally, social cognition is always *relational*, and Mead’s act is a social act. Hence, the relational nature of interaction proposed by Mead (as well as Vygotsky), is the significant difference between the Deweyan and the Meadian acts.

However, according to Gillespie (2005), Mead himself came to reject this view, i.e. subject-subject interaction for two reasons. On the one hand, he was discontent with the stance that the living and animate world offers more feedback than the physical and inanimate world. On the other hand, non-human animals, even social insects, receive feedback from each other, and yet do not appear to have either consciousness or mind. For these reasons, he concluded that feedback *per se* does not lead to consciousness or mind, arguing that a dog can be taught to ‘point’ with its paw toward food without the dog being aware of the meaning of pointing. However, the core problem with the social feedback theory is its disguise of the subject-object model of interaction, because the other is regarded as a ‘pseudo other’, since it nevertheless can be an object or mechanical device (Gillespie, 2005). Accordingly, the major focus is still on the actor, and the other is not endowed with any independent perspective. The crucial point is the incorporation of the ‘perspective of the other’ into the analysis of the subject-subject interaction, which is then referred to as *subject-subject relation* (Gillespie, 2005). It is here that Mead’s concept of ‘taking the attitude/rôle of the other’ becomes relevant in order to portray the significance of Mead’s social act.

Most generally, Gillespie points out there are two related insights in Mead’s work on the emergence of consciousness and self-mediation. Firstly, Mead defined consciousness as becoming other to oneself, and it is only through becoming ‘other to self’ that this reflexive self-mediation of consciousness is able to occur. Secondly, Mead recognized, as Gillespie puts it, “self is already other from the perspective of other within social interaction, and thus if self could take the attitude of the other (toward self) then we would have an explanation of consciousness” (ibid., p. 27). The critical point, however, as Gillespie stresses, is how the self can take the ‘attitude of the other’.

The underlying problem here is the divergence of perspectives within social interaction, which is commonly mistakenly overlooked. Gillespie emphasizes that “we automatically conceive of a situation from the perspectives of diverse participants simultaneously” (ibid., p. 27). Returning to Vygotsky’s pointing example, the issue can be reformulated in the following way. Despite the fact that the actions of both child and mother partly comprise the other’s situation, they are clearly engaged in different situations. How can the child then go beyond its own perspective and subsequently be able to experience the mother’s situation, in which the child is ‘the other’? The Meadian answer is the *social act*, which is a more specific kind of social interaction than the mere intertwining of two Deweyan acts. A Meadian social act refers to an interaction that has become an institution with
established roles or positions (e.g., teacher/student, seller/buyer, cop/thief etc.). In this case the established positions are the mother and child, and these interactions are quite stable over time. According to Gillespie, the introduction of social structure and stability is fundamental for the shift from a socially interactive Deweyan act to the relational Median social act. It is generally through the use of language that children begin to take the 'attitudes of others' toward themselves as a kind of perspective-taking. Given that human language functions as a crucial mechanism for mind, it also functions as the elementary foundation of the self. Mead (1934) claimed:

I know of no other form of behavior than the linguistic in which the individual is an object to himself, and, so far as I can see, the individual is not a self in the reflexive sense unless he is an object to himself. It is this fact that gives a critical importance to communication, since this is a type of behavior in which the individual does so respond to himself (Mead, 1934, p.142).

According to Cronk (2005) and Gillespie (2005), Mead proposed that besides language, the social emergence of the self is further developed through two certain kinds of social interaction, namely **play**, and **game**.

Crucial to the generation of self-consciousness in the play and games of children, as in linguistic interaction, is the process of **role-playing**. Cronk writes that through playing, children take the role of another and act as though they actually were the other (e.g., doctor, mother, house-keeper, Cinderella, and so on). Accordingly, in this kind of role-playing a child puts itself into a **single role** at a time, and a child’s play experience of the other is a "specific other" (Cronk, 2005). The game, however, entails a more complex form of role-playing, since children must internalize the roles of all the others involved within the game. For example, in soccer, children must be aware of the roles of the referee as well as the different types of players, such as forward, back and goal keeper. Furthermore, children must also follow the rules of soccer which determine how the various roles in the game are determined to act. This configuration of “roles-organized-according-to- rules” results into Mead’s “generalized other”, in which the attitudes of all the participants form a symbolized unity. Cronk argues that self-consciousness in the full sense of the concept is achieved only when the individual child can perceive itself from the standpoint of the ‘generalized other’. Taken together, these three kinds of symbolic interaction are regarded as the major pillars in Mead’s socialization approach.

Cronk (2005) points out that even if the self is considered to emerge from symbolic interactions, it should not be regarded as a passive reflection. Instead, the individual’s reaction to the social realm is an **active** process in which one decides what to do next from the attitudes of others. This does not imply that the behavior is in any way determined, i.e. there actually are two sides of the self (Cronk, 2005; Mead, 1934).
Mead (1934) distinguished between ‘me’ and ‘I’ in order to explain how these aspects of self are related in conduct (as illustrated in Figure 2), namely the two phases that i) reflect the attitude of the generalized other ('me'), and ii) respond to the attitude of the generalized other ('I'). Most generally, 'I' is the momental awareness of the social 'me', because the 'I' of this moment is present in the 'me' in the next moment. Therefore, the 'I' functions as the novel and situational reply of the individual's 'me'. In addition, Cronk (2005) indicates there is a dialectical relationship between society and the individual, and this dialectic is enacted in terms of the polarity of the 'me' and the 'I'. In other words, the 'I' and the 'me' have a dynamic relation to one another.

According to Varela (1994), Harré elaborated Mead’s theory of the social nature of mind and self, and introduced a discursive turn to Mead’s ‘taking the role of the other’. Harré (in Varela, 1994) considered the beginning of ‘taking the role of the other’ as the consequence of learning the linguistic practices of using the pronouns of one’s culture, i.e., first, second, and third person. From this discursive turn, the second-person standpoint refers to Mead’s the 'I' whereas the third-person standpoints refer to the 'me'. Furthermore, Harré stated that the interplay of the elements of self, the 'I', and the “me”, or first person pronominal practices, systematically vary across different cultures and languages, which implies that both the 'I' and the 'me' must be socially constructed. Consequently, there is cross-cultural evidence for the social emergence of self in favor of Mead’s approach (Varela, 1994).

Thus, Mead considered the human mind as emerging out of the interaction of biological individuals in a social sphere. In particular, he stressed that mind should neither be considered as a ‘substance’ located in some transcendent realm nor existing as some event that takes place within the human physiology. According to Cronk (2005), Mead rejected the traditional view of mind/body dualism as well as the attempt to explain mind merely in terms of physiological and neurological responses as in behaviorism. Indeed, Mead can be regarded as behavioristic in the sense that he claimed mind could be completely explained by behavioral terms, but only through the denial of the existence of a ‘substantial’ entity. Instead, he considered mind the natural function of human organisms, in which it was neither possible nor desirable to deny the existence of mind entirely. The human physiological organism is a necessary but not sufficient condition of mental processes, since the emergence of mind is dependent upon interaction.
between the human organism and its social environment. In particular, linguistic interaction is of crucial importance since there is no advanced mind or thought without language. Thus, mind is not reducible to the neurophysiology of the organic individual, but emerges in "the dynamic, ongoing social process" that constitutes human experience (Mead, 1934, p. 7).

Taken together, Mead’s theory of the socially interactive mind and consciousness is highly compatible with today’s embodied and situated approach of cognitive science. However, since Mead usually focused on vocal gestures and disregarded the role of manual ones, he failed to explicitly address other bodily aspects of social interactions, despite the fact that he considered humans as embodied biological persons in a social context (cf. Farnell, 1999).

2.5 The Fall of Behaviorism and the Rise of a Science of Mind

Despite the serious criticisms raised against behaviorism from several of the researchers addressed above, behaviorism managed to be the leading approach in psychology and animal behavior, particularly in the United States, through the first half of the 20th century. However, its great impact started to decrease during the mid-1950s, with the following two major reasons responsible for that shift in interest. Firstly, ethologists challenged the equipotential assumption, and secondly, the advent of the computer confronted the anti-mentalistic assumption and paved the way for the ‘cognitive revolution’.

However, some animal behaviorists had earlier obtained results that could not be explained merely in behaviorist terms. According to Gould and Gould (1999), classical and operant conditionings do not suggest any evidence of higher mental processes or cognition in animals, although Tolman (1948) discovered a kind of latent learning in rats. His findings illustrated that animals are capable of performing novel behaviors by ‘planning’ and not only by ‘goal-seeking’ and trial-and-error learning. Tolman conducted the following experiment, already in 1948. A rat was allowed to walk around in a two-armed maze which had two ends; one black area end, and one white area, each containing food. The next day the rat was given an electric shock in a black box and on its release back into the maze the following day, it unmistakably ran towards the white area. The rat demonstrated it was able to link its previous, but unrelated experiences separated in time, place and context, by making a plan of a ‘safe’ run in the maze. This behavior could not be predicted or explained by behavioristic theories, and Tolman’s finding was ignored and commonly scorned in his own research field. This neglect could be due to the fact that Tolman argued for some ‘mental processing’ underlying the rat’s planned novel behavior, which he termed as a kind of ‘cognitive map’ (Gould & Gould, 1999).

Although earlier scholars, like Tolman, for example, had assumed there was some mental content, animal researchers’ greatest contribution to the fall of behaviorism was their demonstration of the incorrectness of the assumption
of *equipotentiality*. One of the strongest pieces of evidence was Breland and Breland’s (1961) discovery of biases in classical conditioning, revealing that ‘innate’ behaviors or tendencies overtook the conditioned or learnt behaviors, which they stated in their article *The misbehavior of organisms* (1961) as follows - obviously a reply to Skinner’s article *The behavior of organisms* (1938).

In our attempt to extend a behavioristically oriented approach to the engineering control of animal behavior by operant conditioning techniques, we have fought a running battle with the seditious notion of instinct ... After 14 years of continuous conditioning and observation of thousands of animals, it is our reluctant conclusion that the behavior of any species cannot be adequately understood, predicted or controlled without knowledge of its instinctive patterns, evolutionary history, and ecological niche (Breland & Breland, 1961, cited in Gould & Gould, 1999, p. 57-58).

Another study of crucial importance was performed by Garcia and Koelling’s (1966). This study revealed that it was difficult and perhaps impossible to teach a rat, using classical conditioning, to avoid flavored water, even if the rat received an electric shock. Furthermore, according to Gould and Gould (1999), other animal studies demonstrated the impossibility of training rats to avoid electrical shocks by pressing a lever. The learning task was only successfully performed by the rats’ natural flight behaviors of jumping or running-away. For example, rats were unable to associate feeding with flight behaviors and pigeons seemed to function in a similar way. Pigeons were able to learn to peck to obtain food, but could not learn to peck to avoid electric shocks. Hence, there seems to be some inherited constraints in the learning and behavioral repertoires of animals, since some kinds of tasks were easier to learn than others. Broadly speaking, these studies demonstrated that classical and operant conditioning are foiled by ‘instincts’, in the sense of functionally specialized adaptations in the animal’s behavioral repertoire.

The role and relevance of social interactions became evident in Harlow’s (1958, 1959) horrid experiments on infant rhesus monkeys. He separated the infants from their mother’s shortly after birth and raised them in separate cages, in which artificial ‘mothers’ were placed, constructed of either wired or soft frames. Half of the rhesus monkeys were bottle fed from the wired dummy mother and the other half by the ‘soft’ dummy mother. The result revealed that the infant apes preferred the ‘soft’ dummy mother over the ‘wired’ mother, whether it offered food or not. Moreover, another study also indicated that growing up isolated, resulted in permanently socially impaired monkeys even as adults (Masson & McCarthy, 1995).

Despite the fact that a number of researchers criticized the mechanistic view of behavior, several machines were built, and they were mainly inspired by Loeb’s work on tropism and Sherrington’s work on reflexes (Sharkey & Ziemke, 1998; Ziemke, 2000), in attempts to test the mechanical perspective. Additionally, advances in computer technology and influences from cybernetic theory resulted in efforts to build machines that illustrated some
aspects of ‘intelligent behavior’. As described by Bechtel, Abrahamsen & Graham (1998), cybernetics is characterized as the theory of ‘control and communication in the animal and the machine’, which is realized through a feedback mechanism. The feedback mechanism allows living organisms to maintain and steer themselves in dynamical environments, adjusting their internal components to external conditions and changes. According to Ziemke (2000), Grey Walter’s (1953) work on constructing two artificial tortoises ‘Elmer’ and ‘Elsie’ around the 1950s, has had a strong impact on today’s embodied AI research. Walter’s work which was influenced by Loeb and Sherrington’s notions of tropism and reflexes, demonstrated that the ‘tortoises’ were able to display phototaxis and ‘hunger’, i.e. they were able to recharge their batteries when necessary. Additionally, ‘Elmer’ and ‘Elsie’ had a small electronic ‘nervous system’, two ‘sense reflexes’ and two ‘receptors’, enabling them to be sensitive to light and ‘touch’ (being able to re-charge their batteries). The artificial tortoises were able to avoid obstacles, attracted to light, and constantly moved around except when re-charging their batteries, which was the only time when they were attracted to the bright light of the ‘hutch’. As a result, these artificial creatures were able to display a number of different behaviors, and can be considered the “archetypes of biologically inspired robotics” (Ziemke, 2000). Moreover, Grey Walter (1953), who was also inspired by Pavlov’s work on conditional reflexes, was the first person to developed and use reinforcement techniques for robot training (Ziemke, 2000). The experiments conducted by Grey Walter demonstrated that a rather simple learning mechanism made it possible for a robot to interact with its environment by adapting its behavioral repertoire (Ziemke, 2000).

Ziemke (2000) points out that the early cybernetic and mechanistic attempts of building artificial creatures demonstrate that intelligent behavior was considered as the result of the interaction between an agent (artificial or natural) and its environment. Despite the fact that von Uexküll was very critical of the mechanistic view, the technical realizations of bringing forth successful behavior had general similarities with von Uexküll’s work on functional circles and signs, which seemed to have a central role in that interaction. However, these early attempts of modeling intelligent behavior was gradually replaced with an approach that significantly separated internal representations from the external world - the computational paradigm. Consequently, the field of cybernetics largely died down as an approach for studying mental phenomena, but it can be viewed as a primary multidisciplinary effort in that direction (Bechtel, Abrahamsen & Graham, 1998; Varela et al., 1991).

The death blow for behaviorism was the advent of the computer. An influential person in computer science was Alan Turing, who realized the idea of computational machines or a so-called ‘Turing machine’. A Turing machine is an abstract automaton, and neither its physical realization is significant nor how it is physically executed. Its only significance is the process the machine performs. As a result, the concept of simulating human information processing in electronic computers emerged (Pfeifer & Scheier,
In Turing’s (1950) classical paper *Computing machinery and intelligence*, he provided the basis for viewing the mind as computational, discrete and disembodied (Wilson, 1998). In the famous Turing test, the method for evaluating if a machine can think, the experimental set-up was totally disembodied. Furthermore, in order to investigate intelligence, Turing (1950) argued for the importance of drawing “a fairly sharp line” between human’s physical and intellectual capacities claiming that his test should only address intellectual aspects, stated as follows:

No engineer or chemist claims to be able to produce a material which is indistinguishable from the human skin. It is possible that at some time this might be done, but even supposing this invention available we should feel there was little point in trying to make a ‘thinking machine’ more human by dressing it up in such artificial flesh. The form in which we have set the problem reflects this fact in the condition which prevents the interrogator from seeing or touching the competitors, or hearing their voices (Turing, 1950, quoted in Wilson, 1998, p. 76).

Hence, the intellectual capacities addressed by Turing were highly disembodied, since it was obviously unnecessary for intelligence to be ‘dressed up in flesh’, which followed the Cartesian view of splitting the mind from the body (Wilson, 1998). However, in the conclusion of the same paper, Turing discussed the ability to build ‘learning machines’ (cf. Lindblom & Ziemke, 2003). He realized the difficulties of attempting to build 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”:

Instead of trying to produce a program to simulate the adult mind why not rather try to produce one which simulates the child’s? If this were then subjected to an appropriate course of education one would obtain the adult brain. Presumably the child-brain is something like a notebook as one buys from the stationers. Rather little mechanism, and lots of blank sheets. ... Our hope is that there is so little mechanism in the child-brain that something like it can easily be programmed. The amount of work in the education we can assume, as a first approximation, to be much the same as for the human child (Turing, 1950).

Turing himself carried out some experiments with simple ‘child-machines’, but after his death in 1954, most AI researchers focused on other approaches. Turing (1950) actually seemed to have considered neither human-like embodiment nor the full range of human social interactions as particularly crucial for the child-machine’s mental development. With reference to the famous case of Helen Keller (cf. e.g. Leiber, 1996), Turing argued that “[w]e need not be too concerned about the legs, eyes, and so on”, as long as “communication in both directions between teacher and pupil can take place by some means or other”. Furthermore, he envisioned the machine being tutored by humans, but also claimed “one could not send the creature to school without the other children making excessive fun of it”. However, since the mid-1990s a number of researchers in situated/embodied AI and robotics have begun to take seriously the idea that the creation of artificially intelligent systems might require not only
physical situatedness and embodiment, but also some form of child-like development in interaction with some social environment (cf. Lindblom & Ziemke, 2003).

The dawn of the computer age and progress within information theory (e.g. Shannon & Weaver, 1948) took place at the same time, leading the way for the information-processing metaphor of mind. The computer was interpreted as evidence that internal representations and mental processes can occur in a physical device. As a result, the antimentalistic assumption in behaviorism was considered to be proven false, and the ‘new science of mind’ - cognitive science was founded in 1956; the start of the so-called ‘cognitive revolution’. AI was brought to light during the same time through its attempts to build ‘intelligent machines’ (Gardner, 1987). The inception of the so-called ‘cognitive revolution’ can be traced back to the Symposium on Information Theory, which took place at the Massachusetts Institute of Technology, in September 1956. The program during the symposium included, for example, the demonstration of Newell and Simon’s ‘Logic Theory Machine’, which was an information processing system able to prove a complex logical theorem (Gardner, 1987; Pfeifer & Scheier, 1999). According to Pfeifer & Scheier (1999), Newell and Simon claimed they had built the first ‘thinking-machine’.

Moreover, the linguist Chomsky submitted a transformational grammar based on information theory, and Miller presented the ‘magical number seven’, which illustrated how the human short-term memory might function. According to Miller (1979, in Bechtel, Abrahamsen & Graham, 1998, p. 37) September, 1956 was when “cognitive science burst from the womb of cybernetics and became a recognizable, interdisciplinary adventure in its own right”. Another important meeting in 1956 was the so-called ‘Dartmouth conference,’ the aim of which was to discuss ‘thinking machines’, and many of the founding fathers of artificial intelligence were among the participants (Gardner, 1987; Pfeifer & Scheier, 1999). Of particular importance to the fall of behaviorism was the Chomsky’s (1959) review of Skinner’s book Verbal Behavior (1957). Skinner suggested that language was also controlled by external stimuli, but was unable to fully explain this statement by behavioristic concepts. Chomsky presented a new theory of language based on linguistic transformation, and was able to illustrate how humans could produce totally ‘new’ utterances (Pfeifer & Scheier, 1999). Broadly speaking, these new thinkers, together with several others, paid attention to what happened inside the ‘skull’, i.e. our internal processes, and reopened closed doors by modeling or conducting investigations of ‘mentalistic’ phenomenon that behaviorism had neglected.

As a result of the advent of the computer and the ‘cognitive revolution’, little attention, if any, in the following years was directed to studies in biology and animal behavior, because the new science of mind considered that those earlier attempts were wrong and misleading in the study of cognition (Gardner, 1987). Cognitive science, as a research field, studied ‘higher’ mental processes, which animals were assumed to lack, and researchers tended to relate studies on animal behavior in general and animal learning in particular with the ‘horror’ of behaviorism. Nevertheless, even if the
behaviorists treated the mind as a ‘black box’, this box was in any case situated to some degree. It was placed in the surrounding world (although in an artificial environment) and reacted to it through real-time interactions. The early mechanistic and cybernetic models, for instance Grey Walter’s artificial tortoises, which can be regarded as one of the precursors of the AI field, would be more or less completely ignored during the era of computationalism, and the same fate awaited theories that stressed agent-environment interaction (Ziemke, 2001a). In sum, the behaviorists can be regarded to having over-emphasized the role of the environment, while the forthcoming ‘cognitivists’ instead would under-emphasize its importance for intelligent behavior (cf. Lloyd, 1989). In chapter 3 current theories and models of embodied cognition, to some degree, attempt to re-unit and/or integrate the two.

2.6 Computationalism: The Disembodied Paradigm

The computational paradigm was founded on two major hallmarks, namely representations and the computer metaphor for mind. This section presents the main theoretical characteristics of this paradigm, and illustrates how these major features became implemented and modeled in artificial systems. As a final point, some reasons are presented that try to explain why the computational paradigm became the dominant approach in the scientific study of mind on behalf of alternative theories.

As a consequence of the development of computer technology it became possible to realize Leibniz’s notion that numbers could be designated to concepts, which could then be manipulated by formal rules, such as Boolean algebra. As a result, the concepts themselves could subsequently be manipulated, laying the foundation for the law of thoughts (Bechtel, Abrahamsen & Graham, 1998). Accordingly, the role of representations was emphasized, and it was proposed that the ‘mind’ itself had mental (internal) models of the surrounding environment. A representation can be considered the mapping between the elements in the external world and the internal symbolic representations, and functions as an internal ‘mirror’ of the external environment. The concept of mental models goes back at least to Craik (1943, in Ziemke, 2000), who discussed the idea of having a ‘small-scale model’ of explicit knowledge about the world in the mind. Thus, thinking was viewed as the processing of those internal representations or symbols in the brain. As a result, researchers became influenced by formalistic knowledge representations and therefore various formal representational languages were created such as ‘schemata’ (e.g. Posner, 1973), ‘semantic nets’ (Collins & Quillian, 1969), ‘frames’ (Minsky, 1975), ‘production systems’ (e.g. Andersson, 1983) and ‘scripts’ (Schank & Abelson, 1977). Symbolic thought was assumed to be analogous to language processing in many ways, since both include sequential syntactical processing of words, as in the proposed representational models (e.g. Fodor & Pylyshyn, 1988). Hence, the only form of situatedness in such systems is to have a mental model of the surrounding world, whereas embodiment did not matter at all (Ziemke, 2001a).
The dawn of the computer age and progress within information theory (e.g. Shannon & Weaver, 1948) occurred simultaneously, resulting in another cornerstone of classical cognitive science and traditional AI, the computer metaphor for mind (cf. e.g. Cisek, 1999; Ziemke, 2001a). Neisser (1967), who can be viewed as the founding father of cognitive psychology, stated in his influential book *Cognitive Psychology* (1967):

> the term ‘cognition’ refers to all the processes by which sensory information is transformed, reduced, elaborated, stored, recovered and used. It is concerned with these processes even when they operate in the absence of relevant stimulation, as in images and hallucinations. Such terms as sensation, perception, imagery, retention, recall, problem-solving, and thinking, among many others, refer to hypothetical stages or aspects of cognition (Neisser, 1967, p. 4).

Taking such a view as the general framework for the study of mind, Neisser (1967, 1976) supposed there was a one-to-one mapping between the computations carried out by a digital computer and human thinking, since both the computer and the human mind “accept information, manipulate symbols, store items in ‘memory’ and retrieve them again, classify inputs, recognize patterns and so on” (Neisser, 1976, p. 5). Thus, cognitive psychology in the form of ‘the computer metaphor for mind’ became equivalent with human cognition. Neisser (1967) stressed that the actual task for cognitive psychologists was to understand the ‘program’, and not the ‘hardware’. In order to perform that task, he made the analogy between investigating human cognition and trying to understand how a computer was programmed. This analogy was motivated by the belief that the human brain processed information in a similar way to the computer, which Pfeifer & Scheier (1999) illustrate as follows:

> It became natural to think of human beings as information processing systems that received input from the environment (perception), process that information (thinking), and act upon the decision reached (behavior). This corresponds to the so-called sense-think-act cycle. Psychologists could now talk about “encoding”, “search”, “retrieval”, “matching” and other information processing operations. The hope was to establish a strong theoretical and formal ground for conceptualizing human behavior that would replace behaviorist psychology. It seemed that anything humans do could be viewed in information processing terms: reading, remembering facts, recognizing objects, drawing logical conclusions, solving difficult problems, playing chess, conducting a conversation, and so forth. Moreover, models couched in information processing terms were easy to formalize in terms of computer programs (Pfeifer & Scheier, 1999, p. 37).

As a result of equating cognition with the function of a computer, which is an artificial system that has the ability to process information, any machinery that could carry out a similar process independently of its hardware (brain tissue or mechanical device) was said to be, at least potentially intelligent. This view which is called *functionalism* (Putnam, 1960) distinguishes between hardware and software levels and claims ‘it is not the meat, it is the motion’ that matters. This means, cognition or intelligence could be analyzed at an algorithmical level or as computer processes without
paying attention to the underlying structure of the device, on which the processes are performed. Consequently, the challenge for cognitive science was to discover the inner processes (or software) underlying intelligent behavior, and ‘intelligence as computation’ became the cognitivists’ manifesto (Pfeifer & Scheier, 1999).

Accordingly, many computer programs were constructed in a functionalistic way that simulated higher human cognitive abilities, e.g., Newell and Simon’s General Problem Solver (1961) and more specific expert systems like MYCIN (Shortliffe, 1976). MYCIN was an expert system that proposed diagnoses of diseases caused by infections, with laboratory tests and symptoms as input, and MYCIN then presented some suitable antibiotics to cure the illness as output. An additional example of a program that used computational representations was SHRDLU (Winograd, 1972), which mastered conversations about a small, simulated world of blocks. This program was able to respond to various issues about the block world and could also handle some manipulations with the blocks as outputs to stated questions.

In 1976, Newell and Simon presented the Physical Symbol System Hypothesis (PSSH), a kind of hallmark for the computational paradigm, claiming that thinking and cognition is symbol manipulation. The hypothesis states that symbol systems implemented in any device (e.g., brain, hardware, paper) have both the necessary and sufficient resources for general intelligence. By ‘necessary’ they meant that any system that does not have these properties cannot be intelligent, and ‘sufficient’ implied that a system having these characteristics had the potential to be intelligent (Pfeifer & Scheier, 1999). Accordingly, the computer could be interpreted as evidence that ‘mental’ activities can occur in a physical device, without any need to speculate about an intervening ‘spirit’ or ‘soul’ in the system. Consequently, higher cognitive abilities were viewed as the formal rule-based, processing of internal symbols inside the brain analogous to a computer program, while body and environment were reduced to some kind of input and output devices.

According to Cisek (1999), the computer metaphor for mind became successful because it offered convincing answers to major questions that confronted the field of psychology during the initial decades of the 20th century. He claims that the previous immaterial soul and mechanical body dualism was replaced with a mind/body dualism within the computational approach. Cisek (1999) points out that the explicit split between the immaterial mind and body resulted in two separate interfaces between the mind and the external reality, namely perception and action. However, humans are equipped with an organ, which has the ability to link these two processes together, allowing humans to have free will, rational thought and consciousness. Despite the fact that this kind of dualism was rejected, computationalism replaced the immaterial mind with a mechanical concept of cognition (Cisek, 1999). Consequently, the same structure of the design remained, laying the foundation for the sense/perception-think/cognition-
**action model of intelligent behavior. Accordingly, every psychological behavior was categorized either as perceptual, cognitive or motor control. The major focus for cognitive scientists was to investigate how cognition operates by converting perception into actions. However, even if the mind was initially considered distinct from the physical world, it could actually be studied scientifically by concepts of internal states. An attempt to solve the mind/body problem was functionalism, arguing that mental states correspond to functional states in the brain (or in the computer). However, this approach had to explain the qualitative difference between brain and mental phenomena (Cisek, 1999).

Thus, computationalism offered the following explanations of the issues addressed above (Cisek, 1999). Firstly, the computer can be considered a model for how the mind works, namely by abstract symbol manipulation of internal representations. Secondly, it allowed a foundation for the discussion of non-dualistic internal states like memory, for instance. Thirdly, it offered the exciting metaphor of functionalism, which allowed mental states to be viewed as the software, running on the hardware. In addition, this functionalistic view provided an elegant answer about how to solve the gulf between biology and psychology, namely that cognitive phenomena are the software running on the (biological) hardware. Finally, computationalism provided a mathematical formalism in the languages of predicate logics, syntax, and information theory. All of these answers to the issues addressed earlier emerged within a short time span, and therefore it was not surprising that the computer metaphor became the viable and only model of how the mind actually works. However, not everyone was swayed by the ‘power’ of computationalism, and there were some which survived like “mammals under the dinosaurs” (Cisek, 1999).

### 2.7 Criticisms of Computationalism and the Inception of Connectionism

In the late 1970s several lines of criticism emerged against the ‘narrow’ computational approach. Dreyfus (1972/1979) and Searle (1980) made the most significant attacks against the strong faith in computationalism. The common theme in their attacks was the lack of connections between the external world and the internal representations. Dreyfus (1972/1979), in his book *What Computers Can’t do: a Critique of Artificial Reasons*, argued that the quandary for the traditional approach is that the knowledge in a computer is represented from the ‘outside’. He stated:

> Even if it represents all human knowledge in its stereotypes, including all possible human situations, it represents it from the outside like a Martian or a god. It isn’t situated in any of them, and it may be impossible to program it to behave as if it were (Dreyfus, 1972/1979, p. 52).

This means, someone had designed and declared the ‘knowledge’, since it is not present or situated in the program itself. The only relation of the ‘knowledge’ to the external environment is through the creator of the program, who decided *how* to conceptualize elements in the surrounding
world. The designer also made the selection what is relevant or not to be represented. The program did not have a direct link or mapping between the external world and the internal representations of it. The mapping went via the designer and the direct link was an ‘illusion’ by the observer, who shared the linking between the external world and the internal representations with the designer. For this reason the system itself lacked a common sense understanding, because it only dealt with predetermined facts. As a consequence, the program’s creator/observer in some sense was its only context (Ziemke, 2000, 2001a).

Furthermore, Dreyfus (1972/1979) maintained the traditional AI-programs only mastered micro-worlds, slices of limited problem areas and/or isolated parts of human cognitive abilities. Examples of micro-worlds are game playing computers, language ‘understanding’ using a form of scripts, and expert systems like MYCIN. The common factor is all of them could not manage tasks in a natural environment. Dreyfus (1972/1979) strongly attacked the lack of situatedness in these formal representations. The motivation for his criticism was that the programs did not have the necessary background information, independent of the amount of explicit knowledge and rules they enclosed, which was the consequence of the lack of ‘first hand semantics’ in traditional AI programs (Ziemke, 2000). According to Dreyfus (1972/1979), it is not sufficient or possible to represent ‘everything’. Instead he argued that we had to go beyond the formal representations and take the body and the surrounding world into account “since intelligence must be situated it cannot be separated from the rest of the human life” (ibid., p. 62). This ‘rest of human life’ is the body’s influence on cognition, cultural factors, and common sense knowledge, which may be impossible to define explicitly. Therefore it would not be possible to represent intelligence within a traditional computer program. According to Ziemke (2000), Dreyfus might have been the first person to use the concept of ‘situatedness’ within the area of AI, and Dreyfus (1972/1979) motivated his argument that situatedness and embodiment might be the core properties for intelligence as follows:

Human beings … they are, as Heidegger puts it, always in a situation, which they constantly revise … We can see that human beings are gradually trained into their cultural situation on the basis of their embodied precultural situation in a way no programmer using KRL [Knowledge Representing Language] is trying to capture. But for this very reason a program is not always-already-in-a-situation … it seems that our sense of our situation is determined by our changing moods, by our current concerns and projects, by our long-range self-interpretation and probably also by our sensory-motor skills for coping with objects and people – skills we develop by practice without ever having to represent to ourselves, our body as an object, our culture as a set of beliefs, and our propensities as situation → action rules. All these uniquely human capacities provide a “richness” or a “thickness” to our way of being-in-the-world and thus seem to play an essential role in situatedness, which in turn underlies all intelligent behavior (Dreyfus, 1972/1979, p. 52-53, emphases added):
This means, Dreyfus (1972/1979) argued that intelligence requires a ‘background of common sense’ which humans are equipped with by virtue of being embodied, and situated in their physical, social, and cultural world. However, computers are not embodied or situated in that sense, and therefore computers do lack ‘practical intelligence’. Consequently, Dreyfus (1972/1979) claimed that computers can be considered ‘existentially’ stupid, despite the fact that they can successfully deal with formal languages and logical relations.

Dreyfus (1972/79) particularly emphasized the role of embodiment in intelligent behavior, arguing that in order to make embodiment a tentative approach in cognitive science and AI, there was a need to demonstrate how it would be possible to perform a task like waving a hand in the air without appealing to any principles of physics and/or geometry. He proposed an alternative, embodied explanation of intelligence, in which the body itself plays a significant role by means of three functions that might be lacking in (digital) computer programs. Firstly, he mentioned the ‘inner horizon’, which can be illustrated as a kind of partial unspecified anticipation of perception. For instance, when perceiving a certain object, the surrounding context, previous experiences, and certain anticipations affect the possible interpretations of the object. Hence, the possible interpretations are neither made at random nor are they predetermined. Secondly, the overall character of this ‘inner horizon’ decides the meaning of the detailed elements it integrates and becomes determined by them. Thirdly, the body’s capacity to transfer this perceived anticipation to a motoric action, and as stated by Dreyfus (1972/1979, p. 255), “all these are included in the general human ability to acquire bodily skills. Thanks to this fundamental ability an embodied agent can dwell in the world in such a way as to avoid the infinite task of formalizing everything”. Consequently, representations may not be necessary, since it appears probable that humans can, for instance, learn to swim by developing the necessary movements through practice, without any need to represent the bodily (and muscular) movements in a kind of formalistic structure. Moreover, he noted that studies in developmental psychology have demonstrated that “learning of specific details takes place on a background of shared practices which seem to be picked up in everyday interactions not as facts and beliefs but as bodily skills for coping with the world” (ibid., p. 47). In sum, these sensory motor skills appear to underlie all the “higher” mental functions, and these kinds of pattern recognitions are not achievable for the digital computer that obviously lacks a body, and therefore cannot utilize these functions (Dreyfus, 1992). Dreyfus (1972/1979) finally addressed the questions if it would be possible to make an artificial embodied agent and if it is possible to build such a robot by using digital computers. Dreyfus believed that, hypothetically, it would be possible to construct such an artificial embodied agent, but it would require that the agent be equipped with sensorimotor mechanisms that resembled the ones humans and other living organisms have.

---

13 The term was coined by Husserl.
Searle’s (1980) much debated Chinese Room Argument addressed the lack of understanding in the computer program itself, and Searle (1980) made the distinction between ‘strong AI’ and ‘weak AI’. He described ‘strong AI’ as the view that a computer program really could understand and think, i.e. has a real mind, while ‘weak AI’ is the view that a computer program could be used as a tool for studying intelligence. A brief description of his thought experiment, the Chinese Room Argument, follows. A person is located into a room and given a narrative in Chinese; but the person does not understand a word of Chinese. There is also a rule book in the room that explains how to link Chinese signs to each other. The person in the room receives from the outside questions written in Chinese that he/she must reply to in Chinese. In order to manage this task, the person uses the rules and the given set of Chinese signs, and as a result, he/she sends out correct answers (in Chinese), even though the person in fact knows no word of Chinese. To an outside observer it seems as if the person in the room actually understands Chinese. According to Searle (1980), the person in the room is behaving exactly the same way as a running computer program, manipulating symbols (Chinese signs) by following some rules. The same happens in traditional AI programs, they behave as if they were intelligent, but something important is missing. The program itself has no understanding of what it is doing; there is a lack of intentionality, since the person or program never interprets what the symbols stand for. The person does not understand what the story in Chinese is about, and neither does a computer system running a program, that is, understanding what it is doing in the sense of others’ interpretation of the program’s operation. Thus, Searle (1980) reached the conclusion that there is no relation between the internal representations (symbols) and the external represented objects in the world within traditional AI systems. The interpretation of the symbols is made by the designer (or the observer) but not by the system itself. This problem is nowadays called the ‘symbol-grounding problem’ (Harnad, 1990). However, as pointed out by Ziemke (2000), some misinterpretations of Searle’s conclusion exists (Searle, 2003). He did not actually claim that attempts to build intelligent machines were doomed to fail, rather he argued that the mistake conducted by contemporary AI was that it dealt with computer programs but “has nothing to tell us about machines” (Searle, 1980, i.e. embodied systems, in today’s terms). This means, he actually accused AI of being dualistic, since disembodied computer programs were characterized as intelligent. As an alternative, Searle argued that AI systems should be equipped with some of the “causal powers” of living organisms, such as perception, and action, allowing these systems to link their internal structures to their ‘own’ real world (Searle, 2003; Ziemke, 2000).

Another challenge to cognitive science was to extend the functionalistic view by taking the neurological aspects of cognition into account, and in the 1980s a new form of modeling cognitive phenomena emerged, namely connectionism (e.g. Rumelhart & McLelland, 1986). However, connectionism was not a totally new approach; rather there was a re-emergence of neural network research. Already in 1943, McCulloch and Pitts demonstrated that networks of neuron-like units were able to compute complex logical
functions, and in 1958, Rosenblatt presented his seminal paper on the perceptron. Despite having all the fundamental ideas, neural network research did not progress well at that time (Bechtel, Abrahamsen & Graham, 1998; Pfeifer & Scheier, 1999).

Connectionists use artificial neural networks (ANN), which are, to some degree, able to imitate and resemble the brain’s distributed functions at a general level. Accordingly, connectionism regards cognition as brain activity, consisting of a huge collection of interconnected simple units (neurons), which process information in parallel, contrary to symbolic and formalistic models in which the processing is discrete. Broadly speaking, the two main characteristics of neural networks are their ability to learn and their emergent properties (for a more detailed description see e.g. Pfeifer & Scheier, 1999; Ziemke, 2000). Each individual neuron in a network receives ‘activity’ from their closest neighbors, and every neural connection has a weight, which can be positive (exhibitory) or negative (inhibitory). The forwarded ‘activity’ is determined by the product of the numerical weight and the strength of the signal from the sending neuron. The strengths of the connections between the neurons can be altered by adding new input to the network and applying some learning techniques. Consequently, neural networks are capable of making mappings from input to output by themselves, as a kind of learning process. While formalistic symbolic systems were only able to simulate and handle well-defined tasks, neural networks are flexible and can manage fuzzy input. Therefore, they are useable for assignments that are not formalizable, e.g. pattern-recognition tasks. Hence, neural networks are able to carry out tasks that the designer has not directly programmed into them. This means, a neural network functions as a self-organizing system, similar to the brain, and through this massive parallel processing the ability of cognition emerges. Accordingly, cognitive processes should not be regarded as symbol manipulation, but as activity pattern (Clark, 1997; Pfeifer & Scheier, 1999).

Broadly speaking, the same major criticisms that were directed at representationalism and functionalism have also been turned against connectionism (cf. Ziemke, 2000, 2001a). For instance, artificial neural networks do not solve the ‘symbol grounding’ problem themselves, since most networks begin from the designer’s specification of the task. This means, networks do represent properties in the world, although at another level, but it is still the designer that decides from the ‘outside’ which properties should be represented. Moreover, the biological plausibility of neural networks is very limited, and therefore do not serve as realistic models of biological neurons, given the trade-off between biological authenticity and algorithmic simplicity. Consequently, there is a tentative risk in making this simplification, since it is plausible one might miss the very characteristics of intelligence in natural organisms (Pfeifer & Scheier, 1999). Another criticism was the request for evolutionary explanations, since it has been argued that only simulating human intelligence is not satisfying, instead we have to ask how and why these abilities come to be there (Hendriks-Jansen, 1996).
In conclusion, the role of embodiment, situatedness and environment was still neglected within both traditional cognitive science (computationalism) and connectionism (see Clark, 1997). However, as Pfeifer & Scheier (1999) point out, neural networks are viewed as information processors, but if they become embedded in physical robots they might, hypothetically, transform robots into *intelligent autonomous agents*.

### 2.8 In Summary

From the brief historical overview, it is clear that the idea of mind or cognition being embodied is by no means new. At different times and for different reasons, theorists have sought to understand the mind from alternative perspectives. Thus, the idea of the mind as a kind of ‘rational calculating device’ that was earlier proposed by Plato and Descartes, was realized through the use of the digital computer as the metaphor for mind. However, for several reasons the rational and formalized perspective became the dominating approach in cognitive science and symbolic AI, and consequently the role of the body and the environment was disregarded in efforts to find the key mechanisms for intelligence. Clancey (1997) mentions that Lakoff (1987) offered an appropriate review of the computational paradigm, which is as follows:

The traditional view is a philosophical one. It has come out of two thousand years of philosophizing about the nature of reason. It is still widely believed despite overwhelming empirical evidence against it ...

We have all been educated to think in those terms ...

Modern attempts to make it work assume that rational thought consists of the manipulation of abstract symbols and that these symbols get their meaning via a correspondence with the world, objectively constructed, that is independent of the understanding of any organism...

Though such views are by no means shared by all cognitive scientists, they are nevertheless widespread, and in fact so common that many of them are often assumed to be true without question or comment (Lakoff, 1987, cited in Clancey, 1997, pp. 51-52).

Dreyfus (1972/1979) also emphasized that the computer in fact was the product of over 2,500 years of traditional thinking of Plato’s legacy. Additionally, Cisek (1999) noted that the so-called cognitive revolution was just “old wine in new bottles” in a philosophical sense. However, the emerging viewpoint of embodied cognition claims that our cognitive processes are deeply rooted within our brain and body’s interactions with the surrounding world, which can be viewed as a Copernican revolution against computationalism. This is discussed in detail in the following chapters.
Chapter 3

Perhaps the most obvious, and most overlooked, aspect of human intelligence is that it is embodied.
Brooks et al., 1998

We need to find a cure for the Cartesian sickness.
Michael H. Agar, 2005

There is now enough empirical evidence to reject a disembodied theory of the mind as biologically implausible.
Gallese, 2004

3. Embodied Cognitive Science

Having reviewed the history of the relationship between body and mind in the previous chapter, ending with the criticism of the traditional computationalist paradigm in cognitive science, I now turn our attention to the alternative approach of embodied cognition. However, it should be noted that the relation between situated and embodied cognition is far from well-defined. Sometimes these terms are used interchangeably, and in other cases as separate approaches. The aim of this thesis is not to evaluate and discuss the relation between situated and embodied cognition. Henceforth we take the embodied approach as the primary focus of interest, although these concepts are closely intertwined with each other. On the one hand, Matarić (2001, p. 82) argues that “[s]ituatedness refers to having one’s behaviour strongly affected by the environment ... Embodiment is a type of situatedness; it refers to having a physical body and thus interacting with the environment through the constraints of that body”. On the other hand, as Gibbs (2006, p. 1) puts it, “embodiment in the field of cognitive science refers to understanding the role of an agent’s own body in its everyday, situated activity”. Generally speaking, embodiment can be viewed as the mechanisms that make it possible to be situated, and consequently it can be argued that embodiment might be a ‘stronger’ position than situatedness, which implies that both concepts are of crucial importance in the alternative embodied, situated, and distributed approaches to the study of mind.

14 The connection between situated and embodied cognition, however, is far from being definite or well-defined, but they have similar historical roots and can be considered as alternatives against computationalism (cf. e.g. Anderson, 2003; Clancey, 1997; Matarić, 2001, Wilson, 2003).
The following section 3.1 offers an overview of the inception of the embodied and situated approach(es), and presents basic ideas and concepts. In section 3.2 I discuss what particularly characterizes embodied cognition, addressing different notions and levels of embodiment, which subsequently propel what kind of body is required for cognition. In section 3.3, the role of relevance of body in motion, that is, movement and kinesthetic experience are discussed - rather neglected topics in current embodied approaches of mind. Section 3.4 takes a closer look at the social dimension of embodiment, which is another area that has not yet received necessary attention in contemporary embodied cognitive science. Furthermore, it describes how I will tackle the issue of the social aspects of embodiment and interaction as we continue on the journey through this thesis.

3.1 Putting Brain, Body and World Together Again

The embodied approach in cognitive science arose during the 1980s as a reaction against formalism and computationalism. Broadly speaking, while cognitivists claimed that cognition takes place inside the skull in the form of abstract symbol manipulation, arguing that the body only served as an input and output device, proponents of embodied cognition offered a radical shift in explanations of cognition. Literally speaking, the emerging viewpoint of embodied cognition claimed that cognitive processes depend on experiences that come from having a body with particular sensorimotor capabilities interacting with the surrounding world. Accordingly, a number of overlapping, but not necessarily identical embodied views emerged under various names, such as embodied action (Varela, Thompson & Rosch, 1991), embodied cognition (Clark, 1997), embodied intelligence (Brooks, 1991), embodied cognitive science (cf. Clark, 1997, 1998, 1999a, 1999b, 2001; Pfeifer & Scheier, 1999), embodied mind (Varela, Thompson & Rosch, 1991), and embodied AI (Franklin, 1997). As Dorffner (1999, p. 23) points out, the common theme in these approaches of embodied cognition is that “they put the body back into the focus of cognitive science, emphasize the interaction between the agent and the environment, view experience and individual differences as at least as important as detached logical principles, and apply a more holistic stance toward research on human cognition”. Thus, instead of viewing the mind as a ‘mirror’ of the world as in the computational paradigm, Clark (1997), for instance, argues that the brain has the role of a controller for embodied activity, and therefore we should not divorce thought from embodied action.

However, this stance is obviously not totally new, rather it can be viewed as a re-emergence of several previously addressed ideas, advocated by, for example, von Uexküll, Merleau-Ponty, Piaget, Vygotsky, Bartlett and Dewey, who in today’s terms, particularly stressed situatedness, embodiment and

---

15 It should be noted, however, that in order to present and discuss the different concepts, each researcher’s own vocabulary is used, and there is no explicit interpretation made regarding the relation between the terms ‘views’, ‘notions’ and ‘levels’.

16 This line is borrowed from the title of Clark’s (1997) book.
biological explanations of cognition (see the respective subsections in section 2.4). Their theories fell on fertile soil when the criticisms raised against computationalism (see section 2.7) were taken seriously and when empirical evidence required alternative explanations. According to Lakoff and Johnson (1999), the view of mind in computationalism came from an a priori analytic philosophy (cf. section 2.8), rather than being based on empirical results. The situation in contemporary embodied cognitive science is rather the opposite, since the view of the mind as embodied instead comes from empirical evidence than merely from philosophical assumptions, although the influential roots to earlier scholars clearly demonstrate that theories of embodiment are not created ‘out of nothing’. As Lakoff and Johnson summarize it,

...in the emergence of second-generation cognitive science, there was no a priori commitments to the existence of a certain theory or model, rather it was a commitment to make sense of a vast range of phenomena from diverse empirical domains that resulted in an embodied situated cognitive science. The concrete results were empirical discoveries not anticipated in advance. Indeed, they were quite surprising (1999, p. 80).

The linguist George Lakoff and philosopher Mark Johnson (1980), for example, suggested that abstract concepts may be metaphors of physical and bodily concepts. Mark Johnson (1987) later argued that human meaning, understanding, and rationality are the results of embodied experience of being in the world. Johnson (1987), for instance, emphasized that metaphors emerge from the embodied experience of the structures of the individual human body such as gravity being the persistent significance for the concepts of ‘Up’ and ‘Down’ in language and human cognition. In the same year, ethnomethodologist Lucy Suchman (1987) published her influential book Plans and Situated Actions, in which she stressed the impact of the momentary circumstances in a situation more than the importance of internal representations, claiming "that all activity, even the most analytic, is fundamentally concrete and embodied" (ibid. p. viii).

The embodied era within AI began when researchers started to model intelligence from a bottom-up perspective (cf. e.g. Brooks, 1986, 1991, 1999). This new approach also referred to as Nouvelle AI or New AI, offered an opportunity to model and implement earlier theories and ideas in a way that was not previously possible. One of the most influential proponents of New AI is Rodney Brooks (1991) who particularly focused on the challenges of getting robots to act in the real world, and he claimed that situatedness and embodiment were the two cornerstones of the new approach to synthesize intelligence, which he characterized as follows:
[Situatedness] The robots are situated in the world – they do not deal with abstract descriptions, but with the here and now of the world directly influencing the behavior of the system.

[Embodiment] The robots have bodies and experience the world directly – their actions are part of a dynamic with the world and have an immediate feedback on their own sensations (Brooks, 1991b, p. 571, original emphases).

The shift towards an embodied approach within AI resulted in a crucial role for mobile robots that interacted with the environment. Clark (1997, p. 6) for instance, characterized an autonomous agent as “a creature capable of survival, action, and motion in real time in a complex and somewhat realistic environment”. Thus, the aim of New AI was to simplify the task of modeling human intelligence, by starting with some easier and simpler agents than humans, focusing on survival-related foundations such as real-time interaction and integration of sensory-motor functions. Another reason for building autonomous agents was addressed from an engineering point of view, when, for example, NASA needed self-regulating mobile robots for gathering and conveying information on other planets. The traditional way of modeling AI was not suitable for space adventures, since the robots had to act on their own and could not be controlled from earth. Accordingly, the challenges of building such mobile robots resulted in second thoughts about the previous approach to modeling adaptive behavior and intelligence (Clark, 1997). Brooks’ so-called behavior-based robots were designed through activity-based decomposition, which means the robot was equipped with independent activity-based subsystems instead of functional decompositions, and there was no central planner controlling the robot’s behavior. The different subsystems did not explicitly communicate with each other; instead, they competed among themselves in order to be in charge, via interactions with the external environment through its sensors. As a result, an adaptive behavior emerged. Examples of such behavior-based robotic agents are, e.g., the artificial cockroach Periplaneta Computatrix (Beer & Chiel, 1993), and the soda can collecting mobile robot Herbert (Connell, 1989, in Clark, 1997).

Today, many cognitive scientists as well as AI researchers believe that being embodied is a core property for intelligence both in humans and artificial systems. However, if we take the stance that human cognition actually is embodied, what exactly does that mean? Are humans embodied due to their physical nature, their phenomenological nature, biological/evolutionary history, or their individual developmental and social history?

### 3.2 The Nature of Embodied Cognition

One way to find a contemporary definition of the concept of embodiment is to look in *The Encyclopedia of Cognitive Science* (2003), in which embodiment is characterized as follows:
An understanding of how cognition is realized or instantiated in a physical system, especially a body, may require or be required by an account of a system’s embedding in its environment, its dynamical properties, its (especially phylogenetic) history and (especially biological) function, and its nonrepresentational or noncomputational properties (Chrisley & Ziemke, 2003, p. 1102).

Although the above definition is rather broad, it is quite understandable since such a definition serves as a general introductory explanation of the concept in the wider field of contemporary cognitive science. However, Chrisley and Ziemke (2003) further elaborate the definition, stating that embodied approaches usually entail some or all of the following aspects to various degrees:

- Acknowledgement of the role that the body and its sensorimotor processes can and do play in cognition. Some aspects of the system that would, on the traditional view, be considered mere matters of implementation, are instead taken to be crucial components.
- Understanding of cognition in the context of its (especially evolutionary) biological function: to support the activities of the body.
- A view of cognition as real-time, situated activity, typically inseparable from and often fully interwoven with perception and action (Chrisley & Ziemke, 2003, p. 1102).

Their definition strongly considers modeling aspects of embodiment, implemented in artificial systems and explicitly lacks some fundamental aspects of human nature, such as language and culture. On the other hand, Gibbs (2006) more strongly emphasizes these aspects of embodiment in his cognitive premise, in which human activity and language are of major importance for thought and cognition.

People's subjective experiences of their bodies in action provide part of the fundamental grounding for language and thought. Cognition is what occurs when the body engages in the physical, cultural world and must be studied in terms of the dynamical interactions between people and environment. Human language and thought emerge from recurring patterns of embodied activity that constrain ongoing intelligent behavior. We must not assume cognition to be purely internal, symbolic, computational and disembodied, but seek out the gross and detailed ways that language and thought are inextricably shaped by embodied action (Gibbs, 2006, p. 9).

Taken together, these two main characterizations of embodiment strongly illustrate the difference between the previous (dis-embodied) approach of computationalism and the embodied view(s) of cognition. Furthermore, they offer some general claims about the nature of embodied cognition, but do not provide a more detailed picture of the embodied mind.

A tentative way to further ‘pin down’ the essence of embodiment is to characterize different levels of embodiment, given that the body as a whole is not the one and only as well as the appropriate level for analyzing and understanding cognition (Gibbs, 2006). Lakoff and Johnson (1999), for example, portray three levels of embodiment, which they label the neural level, the phenomenological level, and the cognitive unconscious level.
• **Neural embodiment** concerns structures that characterize concepts and cognitive operations at the neural level...

• **Phenomenological level** is conscious or accessible to consciousness. It consists of everything we can be aware of, especially our own mental states, our bodies, our environment, and our physical and social interactions...

• **The cognitive unconscious** is the massive portion of the iceberg that lies below the surface, below the visible tip that is consciousness. It consists of all those mental operations that structure and make possible all conscious experience, including the understanding and use of language (Lakoff & Johnson, 1999, p. 102-103).

According to Lakoff and Johnson, **neural embodiment** basically relates to neurophysiological circuits such as color and spatio-relation concepts, and operates together with the rest of the body. The **phenomenological level** refers to the “feel” of experience and qualia, such as the taste of an apple or the sound of a Beethoven symphony. Phenomenological reflection also belongs to this level. The level of the **cognitive unconscious** includes all the unconscious knowledge and thought processes which are guided by the sensorimotor aspects of embodiment, in particular those cognitive processes that form basic-level and spatial-relation concepts. Given that the body is fundamental at this level, all of the cognitive processes and structures, including language, are grounded in patterns of bodily activity and experience. In short, this level is proposed to clarify conscious activity and experience that cannot exactly be understood on its own terms. In sum, these three levels operate jointly, given that the detailed characteristics of the phenomenological and cognitive unconscious levels arise from the nature of the neural level, together with the experience of the external environment. Because of that tight interconnection between these levels, the topographic maps and orientation-sensitive cells of the body bring forth our spatio-relational concepts as well as the neural circuits resulting in human color categories, for example. It should be noted, however, that this is not the same as saying the neural level is the ‘hardware’ on which some independent ‘software’ runs. To describe all aspects of the human mind, *all* three levels need to be though they may not necessarily be sufficient considered, because ”people are not just brains, not just neural circuits. Neither are they mere bundles of qualitative experiences and patterns of bodily interaction” as Lakoff and Johnson (p. 104) maintain.

Furthermore, Núñez (1999) proposes another classification of embodiment, in which there are also three levels, but he addresses the *extent* to which cognition truly is considered to be embodied.

• **Trivial embodiment** – cognition and the mind are directly related to the biological structures and processes that sustain them...This view holds not only that in order to think, speak, perceive, and feel, we need a brain – properly functioning brain in a body – but also that in order to genuinely understand cognition and the mind, one can’t ignore how the nervous system works.

• **Material embodiment** – First, it sees cognition as a decentralized phenomenon, and second it takes into account the constraints imposed...
by the complexity of real-time bodily interactions performed by an agent in a real environment."

- **Full embodiment** – Full embodiment explicitly develops a paradigm to explain the objects created by the human mind themselves (i.e., concepts, ideas, explanations, forms of logic, theories) in terms of the non-arbitrary bodily-experiences sustained by the peculiarities of brains and bodies. An important feature of this view is that the very objects created by human conceptual structures and understanding (including scientific understanding) are not seen as existing in an transcendental realm, but as being forth through specific human bodily grounded processes (Núñez 1999, p. 55-56)

Briefly stated, **trivial embodiment** is according to Núñez the view that cognition rests on physiological and neurological processes, and in order to study cognition this neurological foundation has to be considered. This stance is by no means controversial (except for dualism), and can be viewed as the ‘lowest’ degree of embodiment. The next level, **material embodiment**, focuses on low-level processes such as action and perception, taking place in a real-time situation. Furthermore, this stance illustrates one of the most basic claims of embodied approaches to cognition. The final level, which is the ‘strongest’ **full embodiment**, sharing the basic views of trivial and material embodiment, is the ‘heaviest’ commitment of embodiment, as it claims that all cognition, even abstract thinking and reasoning, are based on being embodied.

Clark (1999a), on the other hand, addresses another point of view concerning embodiment; focusing on what methodological and theoretical effects an embodied approach may have on cognitive science in general. He speculates how far embodiment could go as well as to what extent it is a genuine alternative to the traditional view of information processing and computationalism rather than just a more or less compatible complement. In order to answer this question, he distinguishes between simple versus radical embodiment. In simple embodiment, the traditional foundation of cognitive science is preserved, and the nature of embodiment is merely considered a constraint of the ‘inner’ organization and processing. Radical embodiment, on the other hand, goes much further and treats the facts of embodiment as “profoundly altering the subject matter and theoretical framework of cognitive science” (ibid. p. 348). However, the distinction between the two stances is not fixed; instead they might overlap. According to Clark (1999a), there are numerous examples of simple embodiment, still relying on internal representations and inner models. He argues that Lakoff & Johnson’s (1999) work on bodily metaphors belongs to this category, stressing the inner representational realm in the form of putting old ideas in a new ‘embodied’ outfit. According to Clark (1999a), radical embodiment instead involves one or more of the following claims:

- In order to understand the complex interactions between brain, body and world, there is a need for new analytic methods and tools, e.g., dynamical systems theory (cf. Beer, 1995; Kelso, 1995; Port & van Gelder, 1995; Thelen & Smith, 1994).
• Addressing the inadequateness and unnecessarily of the traditional notions of internal representations and computationalism.
• The typical decompositions of the cognitive system into inner and functional subsystems are misleading. Instead, there is a need for alternative views that do not cut brain, body and world apart.
• The subject matter of cognitive science needs to be reassessed.

This means, a “mature science of mind” goes beyond the individually bounded scope, taking the bodily and environmental aspects as crucial parts of human intelligence. Following Vygotsky (see subsection 2.4.4) and Hutchins (1995), Clark (1999a, 1999b) particularly stresses the use of scaffolding, in the form of artifacts and culture as crucial components for enhancing our biological cognitive abilities. He argues “that unraveling the ways in which biological brains couple themselves with these very special kinds of ecological objects, is surely one of the most exciting tasks confronting the science of embodied cognition – and the one that might shed great light on the role of embodiment in more abstract domains” (ibid. p. 349). However, it can be argued that Núñez’s levels of embodiment are a mix between Clark’s two stances. On the one hand, Núñez wants to develop a new ‘paradigm’ of embodiment. On the other, however, he does not explicitly address the role of the external environment. Instead, he is rather bound by the ‘brain and body’- viewpoint, which is also the major impression of Lakoff and Johnson’s three levels of embodiment.

Hirose (2002) discusses embodiment from an ecological standpoint, making a distinction between the act of embodying, and the state of being embodied, inspired by the work of Gibson (e.g. 1966, 1979, see also subsection 2.4.2). Hirose argued that the majority of research conducted in embodied cognition so far, has focused on being embodied, demonstrating how cognitive abilities such as language (e.g. Lakoff & Johnson, 1999), for instance, is embodied in a physical body. Instead, Hirose (2002) pays attention to the act of embodying, which is illustrated in Polanyi (1964) and Bateson’s (1972) example of the blind man using a stick (see subsection 3.2.1). When using the stick to probe the surroundings, it ceased to be an added tool, instead becoming a part of the body for an accustomed user and, as a result, the bounds of the actual body can be extended beyond the skin. Hence, Hirose points out that an “external object can be regarded as a part of the perceptual system”, arguing that “the boundary is, in fact, changeable and can be extended by tools. Therefore, to explore the relation between the body and cognition, therefore, research should address how tools become part of the body, focusing on the act of embodying, in addition to the status of being embodied” (ibid. p. 291). Consequently, Hirose views the body as a dynamical entity, arguing that future research should pay more attention on the dynamic nature of embodying. In a similar vein, Clark (2003) argues that humans are uniquely suited for integrating ‘meat and metal’, which makes humans ‘naturally born cyborgs’. He further argues that our minds are predisposed to finding and integrating non-biological resources. While we are good at emotions, senso-motoric information, and pattern recognition, our memory and logical reasoning can be supported by technology. Since
humans are experts in integrating technology and biology, such integration enables us to think and feel beyond what is biologically given. This, according to Clark, is the strength of the human intellect.

Another distinction of embodied approaches to cognition which is suggested by Chrisley and Ziemke (2003), distinguishes between epistemological and metaphysical approaches to an embodied cognitive science. The epistemological approach states that bodily concepts are required in order to explain and understand cognition. In contrast, a metaphysical approach does not stress that cognitive processes must be realized in a body or that research within AI should proceed by constructing embodied robots. This means, the metaphysical approach does not make any claim about embodiment in explanations of cognitive activity, even if an artificial system as such would have a bodily shape.

Summing up, after nearly two decades of research under the banners of embodiment and embodied cognition, there is, so far, no common understanding of what actually constitutes embodied cognition. However, during recent years, several authors have actually tried to characterize embodied cognition (e.g. Anderson, 2003; Chrisley & Ziemke, 2003; Wilson, 2002; Ziemke, 2001b, 2003). Nevertheless, there are also many levels of definitions and notions of embodiment which range from rather less clear cut, to more precise and narrow ones (cf. e.g. Anderson, 2003; Brooks, 1991; Brooks et al., 1998; Chrisley & Ziemke, 2003; Dautenhahn, Ogden & Quick, 2002; Lakoff & Johnson, 1999; Núñez, 1999; Riegler, 2002; Sharkey & Ziemke, 2001; Ziemke, 2001a, 2001b, 2003). Hence the following subsections present and discuss a number of these different views of embodied cognition. In particular, both Wilson (2002) as well as Anderson (2003) have made serious attempts to distinguish these different views, which are presented in more detail in the subsequent two subsections. Subsection 3.2.3, then portrays different notions of embodiment.

3.2.1 Six views of embodied cognition
Wilson (2002) points out that the notion of embodied cognition actually holds a number of diverse stances and she argues that if the notion as such is to remain useful, it is necessary to disentangle and evaluate these views. Consequently, according to Wilson (2002), the six most important claims of embodied cognition are the following:

1. Cognition is situated. Cognition takes place in the context of a real world environment, involving perception and action.
2. Cognition is time-pressured. Cognition is constrained by the limitations of real-time interactions with the surroundings.
3. We off-load cognitive work to the environment. In order to handle these constraints, we take advantage of the external environment in order to reduce the cognitive workload.
4. The environment is a part of the cognitive system. The ‘skull’ alone is not the proper unit of analysis, since the interaction between the environment and the mind are so closely coupled.
5. **Cognition is for action.** The function of the mind is to control and guide action, and therefore cognition should be viewed from a situation-appropriate perspective.

6. **Offline cognition is body-based.** Cognition that takes place ‘offline’ is also based on sensorimotor mechanisms.

The first stance, *cognition is situated* is not particularly controversial, and it is commonly argued that situated activity is one of the cornerstones of embodied cognition (cf. e.g. Brooks, 1991; Clancey, 1997; Clark, 1997, 1998, 2005; Hendriks-Jansen, 1996; Kirshner & Whitson, 1997; Pfeifer & Scheier, 1999; Thelen & Smith, 1994; Ziemke, 2003). However, Wilson points out that it is of great importance to determine what situated cognition actually means, since she claims that falsely the existence of non-situated cognition has been implied (see Greeno & More, 1993). Broadly speaking, Wilson argues situated cognition is “cognition that takes place in the context of task-relevant inputs and outputs. That is, while a cognitive process is carried out, perceptual information continues to come in that affects processing, and motor activity is executed that affects the environment in task-relevant ways” (Wilson, 2002, p. 626). As examples of situated cognition she includes activities such as driving a car or a holding a conversation. However, she criticizes that this definition in fact excludes many cognitive activities, such as planning and remembering, since they take place ‘offline’, meaning the particular cognitive activity is not dependent on the actual environmental interaction at hand. Accordingly, Wilson argues that the ability of offline cognition might be one of the driving forces of human intelligence, from an evolutionary point of view, while overstating the role of ‘real-time’ situated cognition in the evolutionary past of the human species, may instead hinder the understanding of this kind of cognition. More recently, Clark (2005) also discusses the issue regarding the situated nature of human cognition, arguing that “human reason is, we may say, disengaged but not disembodied” (Clark, 2005, p. 236). According to him, there is no sharp line between so-called online versus offline cognition, given that both processes are running in parallel. He introduces the concept of *surrogate situatedness* in order to explain how we as humans create and use human-built structures in order to transform the space of higher-level cognition. Hence, he stresses that “we actively create restricted artificial environments that allow us to deploy basic perception-action-reason routines in the absence of their proper objects” (p. 233), which he dubs *surrogate situations*. To be more precise, Clark characterizes the concepts in this way.

> By a surrogate situation I mean any kind of real-world structure that is used to stand in for, or take the place of, some aspect of some target situation. By a target situation, I mean an actual, possible, or at least superficially possible real-world event or structure that is the ultimate object of my cognitive endeavor (Clark, 2005, p. 236).

Some examples provided by Clark are the use of a dotted line, a drawing, and when the target situation is the as yet non-existing door on a sketch, in which the surrogate situation is the context provided by the drawing. This
means, according to Clark, that these sketches and etc, both allow human cognition to be disengaged while at the same time offering a concrete place in which to organize action-perception couplings of an essentially real world-like kind of interaction.

Similarly to Wilson, Clark (2005) also discusses evolutionary and developmental issues. According to him, surrogate situations provide a midpoint in the history of development, between completely offline imaginative modes of cognition and the more time-constrained real-world interactions here and now. However, he points out that there is a clear qualitative difference between surrogate situations and normal real-time situations, as in catching a ball or playing tennis. The difference concerns both the content, given that surrogate situations benefit from idealization and abstraction, and therefore omit much concrete details. Also the timing differs, since surrogate situations help to slow down some of the temporal constrains present in real-world interaction. Thus, the ability to create surrogate situations blurs the ‘strict’ distinction between online versus offline cognition. Similarly, Svensson (2007) points out that the forms of offline and online cognitive processes should not be considered as two distinct classes of cognitive phenomena. According to him, so-called “offline processes” can both be involved in activities that are characterized by responsiveness to environmental alteration, but also in activities that are more or less independent of the immediate happenings in the environment (Svensson, 2007). In addition, it should be noted that the use of the terms online and offline cognition is not the most appropriate choice of terms, given that they implicitly refer to the computer metaphor of mind (see section 2.6). Although I make this objection to the naming and use of these terms, they are used here and throughout the thesis because they are the commonly well-known terms in the literature. To introduce other concepts may lead to some confusion of what I mean or refer to. To move on, the following two claims by Wilson actually evaluate ‘real-time’ situated cognition in more detail.

The stance cognition is time-pressured, stresses the fact that a living organism has to cope with changing situations ‘here and now’, and consequently situated cognition must deal with time-pressure. Research conducted on autonomous agents has emphasized the significance of time-pressure as a constraint for shaping behavior. Despite the fact that these real-time activities in themselves cannot be considered particularly ‘intelligent’, Wilson suggests that more advanced structures of cognition can be grounded in successive layers of environmental real-time interactions (e.g. Brooks et al., 1998; Clark, 1997; Pfeifer & Scheier, 1999). Brooks (1991) stated that the computer metaphor for mind (see section 2.6) results in a so-called representational bottleneck. He claims the time it takes for the sensory

---

17 Some tentative alternative terms for online cognition could be ‘first-order situatedness’ or ‘primary situatedness’. Clark’s surrogate situatedness can then be considered as ‘second-order situatedness’ or ‘secondary situatedness’. For offline cognition I suggest ‘third-order situatedness’ or ‘tertiary situatedness’. I am thankful to Henrik Svensson for discussing these concepts.
inputs to be translated to a ‘symbolic code’ in which the planning (or cognition in general) takes place and then decoding the outcome back into another format for motor response, takes a vast amount of time and requires too many resources. Consequently, the computer metaphor for mind hinders rapid and demanding real-time interactions, according to Brooks, since the agent has not enough time to update the internal representation\textsuperscript{18} as a result of the time-pressure. Therefore, the agent must generate cheap and efficient tricks to handle the problem, in order to avoid a cognitive ‘break down’ or a decreased cognitive performance. Furthermore, Wilson points out that cognitive activity in general is not performed in the state of “being in a hurry”, but did indeed emphasize the great importance of sensorimotor coordinations in time-locked activities such as walking, object manipulation, and sport performance.

In the third view, the use of \textit{off-loading cognitive work to the environment} is stressed, as a way of handling the demands of real-time cognition. Wilson notes that although humans mostly run their cognitive processes ‘offline’, there are still situations that must be carried out \textit{in situ}. Consequently, we have to consider our cognitive limitations when acting in real time and one strategy of reducing the cognitive burden is, according to Wilson, the use of \textit{epistemic actions} (Kirsh & Maglio, 1994). These actions are used to alter the surrounding environment in order to ease the cognitive work to be done. An example of using epistemic actions can be found in Kirsh and Maglio’s (1994) study of the computer game Tetris. The goal when playing Tetris is to fit geometrical falling blocks as neatly as possible into the bottom row, which contains previously dropped blocks. Furthermore, the player can rotate the blocks during the fall in order to obtain an optimal fit. Kirsh and Maglio (1994) pointed out that the players actually used real rotation movements to simplify the task, rather than mentally representing a solution and then executing it. Hence, as pointed out by Clark (1997), we may ‘think’ faster with the hands, than the brain.

Another way of offloading cognitive work is the use of \textit{effective environments} in autonomous agent research, constructing an environmental \textit{niche} for the agent. A \textit{niche} world is constructed by letting certain characteristics of the environment bear some of the information needed in order to simplify for the agent. As a result, the information-processing load may be reduced (Brooks, 1991; Clark, 1997, 2005). Hence, the idea of \textit{niche-dependent sensing} can be traced back to the work of von Uexküll (see subsection 2.4.1), who explained how different species may experience the surrounding world in different ways due to the various designs of their body as well as their perception and action capabilities. With regard to Herbert, his niche is the offices at MIT’s artificial intelligence lab, and the features in the environment that Herbert is

\textsuperscript{18} It should be noted that there is an ongoing debate concerning the role of internal representations by proponents of embodied cognition, who usually stress the use of situative resources in a situation more than the possible occurrence of internal representations see e.g. Agre, 1993; Beer, 2000; Brooks, 1991; Clark, 1997; Markman & Dietrich, 2000; Riegler, von Stein & Peschl, 1999; Suchman, 1993; Svensson, 2007; Vera & Simon, 1993.
‘sensitive’ to are chosen to stimulate his task, namely collecting empty soda cans on tables. The selected features include, for example, shape of tables, soda cans, and holding a soda can in his ‘hand’. Other objects such as carpets, people, and walls are just obstacles to avoid, and consequently Herbert neither experiences the colors of the walls nor makes a difference between various obstacles. Accordingly, Herbert’s Umwelt differs considerably from the Umwelt of the humans working in the same lab. However, Wilson argues our ability to off-load cognition is much broader than manipulating the world itself, and stresses the potential of symbolic off-loading. This concerns the way of performing and manipulating symbolic or representative cognitive actions directly in the world such as counting on one’s fingers, drawing diagrams, doing arithmetical with a paper and pen, instead of mentally representing them. Wilson maintains this kind of situational decoupled strategy is of great importance in cognition. She suggests this strategy can be seen in gesturing while speaking, eventually serving a cognitive function for the speaker by facilitating the shaping of the thoughts to be expressed (cf. Iverson & Goldin-Meadow, 1998, and sections 5.3 and 5.3.2). Hence, noticing this is particularly important for the research topic addressed in this thesis, while Wilson concludes the use of bodily resources in off-loaded cognition may have far-reaching consequences for a general understanding of embodied cognition.

The claim that the environment is a part of the cognitive system, is “deeply problematic” according to Wilson (2002), since it has been argued cognition is not an activity bounded by the ‘skin and skull’, but rather supposed to emerge from the interaction between brain, body and environment. As a result, the cognitive activity is distributed across the agent and the interacting situation, and in order to understand cognition the agent’s and the situation at hand should be the proper unit of analysis (cf. e.g. Clark, 1997, 1998; 1999b, 2005; Hutchins, 1995; Thelen & Smith, 1994; Wertsch, 1998). However, Wilson is only critical to the second part of this claim, namely the method of analyzing the individual and the environment as one unified system and the use of the term system as such. She maintains studying the distributed causality within the system will not lead to an understanding of cognitive organization and functioning. However, Hutchins’ distributed approach to cognition (1995) for instance, is concerned with cognition at a ‘higher’ level, such as team performance in ship navigation, i.e. inter-individual rather than intra-individual cognition. Analyzing ship navigation, Hutchins (1995) illustrates how multiple embodied biological brains combine with tools (sextants, alidades, etc.), and media (maps, charts, etc.) during performance. These external resources allow the human users “to do the tasks that need to be done while doing the kinds of things people are good at: recognizing patterns, modeling simple dynamics of the world, and manipulating objects in the environment” (ibid., p. 155) (section 4.4 discusses Hutchins’ distributed cognition approach in more detail). Accordingly, Clark (1999b) claims the environment can be viewed as a “source of cognition”, since it complements biological computation and processing, which he states as follows:
The external environment, actively structured by us, becomes a source of cognition —enhancing ‘wideware’— external items (devices, media, notations) that scaffold and complement (but usually do not replicate) biological modes of computation and processing, creating extended cognitive systems whose computational profiles are quite different from those of the isolated brain (Clark, 1999b, p. 349, original emphasis).

However, it should be noted that in this quote Clark mentions the use of scaffolding (Wood, Bruner & Ross, 1976), which Wilson does not refer to at all, although the term is commonly present in the literature of situated and embodied cognition (cf. e.g. Clark, 1997, 1998, 1999a; Hendriks-Jansen, 1996; Rogoff, 1990, 2003; Wertsch, 1998). Broadly speaking, scaffolds are external structures that function as a kind of supportive framework, simplifying the cognitive activity for an individual agent. Clark (1997) claims that the notion of scaffolding has its roots in the work of Vygotsky (see section 2.4.4) since he stressed “the way in which experience with external structures might alter and inform an individual’s intrinsic modes of processing and understanding” (cited in Clark, 1997, p. 45). Accordingly, the previously mentioned strategies for offloading cognitive work to the environment actually function as scaffolds as well as Clark’s (2005) idea of surrogate situations. Additionally, Hendriks-Jansen (1996) stresses that one important role of scaffolds is to bootstrap and launch the infant into a social and cultural environment, allowing it to cognitively develop and think in intentional terms, communicate through language and manipulate tools and artifacts. Similarly, Clark (2005) argues that language, in form of words and symbols (whether spoken or written), is a special kind of human-made surrogate situation that alters and amplifies our thoughts into higher-level cognition.

However, together with colleagues (Susi, Lindblom & Ziemke, 2003) I discuss a related question, namely where exactly to draw the boundary between a cognitive system and its environment. This was illustrated by Polanyi (1964) and Bateson (1972) with the classical example of a blind man using a stick: What are the bounds of the blind man’s system - does it or does it not include the stick? The recent discussion of the appropriate unit of analysis in the study of cognition has led to a debate concerning the distinction between cognitive agents and the non-cognitive scaffolds they use, viewing the whole unit as one cognitive system (e.g. Clark & Chalmers, 1998; Hutchins, 1995). Adams and Aizawa (2001) have formulated detailed criticisms of these theories, which they consider guilty of blurring that distinction, since they consider cognition as “restricted to the confines of our brains” (ibid. p. 44). However, as pointed out by Bateson (1972) himself, questions about whether a mental system is bounded by skin or skull, are in

---

It should be noted that Vygotsky himself never used the term scaffolding, instead it was coined by Wood, Bruner & Ross (1976). According to Mauro Goulder (personal communication, 2002), Wood and Bruner were actually the first Westerners to access Vygotsky’s follower Galperin’s work, but they did not acknowledge Galperin’s work in their well-known article. Galperin investigated the impact of different kinds of instruction on problem-solving processes (see Susi, 2006). However, the spreading of Galperin’s idea to the Western world was delayed due to the political climate.
fact “nonsense questions”, since the boundaries of cognition, as well as the appropriate units of analysis, depend on what we want to explain. Hence it is only natural for cognitive scientists interested in situated, embodied and distributed cognition to choose as their units of analysis cognitive agents in situ, i.e. embedded in their environments. Whether or not such units of analysis should be referred to as cognitive systems is only a secondary question, and it could be concluded that debating where to draw the bounds of cognition is to some extent simply ‘much ado about nothing’ (Susi, Lindblom & Ziemke, 2003).

The fifth claim, **cognition is for action**, stresses the role of the mind to guide action. She argues cognitive mechanisms such as perception and memory have to be considered from a situation-appropriate perspective. For instance, Wilson notes that empirical work in visual perception (e.g. Ballard, 1996; Ballard et al., 1997) suggests the purpose of the visual system is not to construct an internal representation of the outside world, which then should be manipulated by the ‘higher’ cognitive architecture. Instead, it has been argued that some sorts of visual input directly prime motor activity, without the need for an intervening ‘higher’ cognitive process to guide action. In the case of memory, Wilson addresses Glenberg’s (1997) claim that the traditional approach to memory as a ‘storage’, should be replaced by the view of memory as “perceptuomotor patterns”, characterizing short-term memory as the undertaking of specific action skills. However, Wilson does not address the work conducted in New AI such as action-oriented representations (cf. Clark, 1997), and sensory-motor coordinations (e.g. Nolfi & Floreano, 2000; Pfeifer & Scheier, 1999). Pfeifer and Scheier (1999) mention that work conducted in active vision suggests biological vision systems do not passively process the visual input, but instead actively interacts with the environment. Accordingly, Bajcsy (1988, in Pfeifer & Scheier, 1999, p. 431) has stated that “we do not just see, we look”, stressing the active process of linking perception and action. It should be noted that the use of sensory-motor coordination within active vision has roots in Dewey’s (1896) earlier criticism of the reflex-arc in psychology (see section 2.2), and that contemporary work on active vision in AI is influenced by Dewey’s ideas (see Clancey, 1997; Pfeifer & Scheier, 1999).

The use of action-oriented representations can be illustrated by the mobile robot Herbert, in which there exists no obvious distinction between perception, planning (cognitive process), and action. Instead, there is a direct coupling between perception and action, accomplished by so-called subsumption architecture (Brooks, 1991). The robot does not rely on any internal model of the environment, and there is no internal communication between the different behavioral layers of the system. Hence, the controlling of Herbert's behavior is accomplished via “action-loops” between the environment and the architecture. As previously mentioned, Herbert’s task was to collect empty soda cans, and the robot's prevailing behavior was to use a laser scanner to locate and then move towards them. When Herbert arrived at a can, its arm reached out, but it is important to note that there was not actually any internally generated command to reach. Instead the reaching behavior was activated when the moving behavior ended. Similarly,
there was no grasping command, but in the moment when the infrared ray between the 'fingers' was broken, Herbert’s 'hand' grasped (Connell, 1989). Hence, it could be argued that action-based representations, as addressed above, function in situ, but it might be an overstatement to claim that advanced cognition functions in a similar way.

Although his final claim, offline cognition is body-based, has received little attention, it may be the best documented and the most powerful of all (Wilson, 2002). Moreover, as pointed out by Ziemke (2003), this is actually the only claim that explicitly addresses the role of the body in cognition. In order to explain her stance, Wilson returns to her example of symbolic off-loading; using the finger tips as a way of performing and manipulating symbolic or representative cognitive actions directly in the world, instead of mentally representing them. This action can be considered as a surrogate situation, using Clark’s (2005) vocabulary, but Wilson then advances a step further by pushing this counting activity inward, allowing only the priming of the motor action, but no overt motion actually takes place, that is, offline cognition. Hence, this sort of mental strategy in the form of a simulation process, she argues, may open up for new explanations of how 'offline' embodied cognition functions, since it has been claimed many cognitive processes indeed utilize sensorimotor functions in this 'hidden' way. This means, structures that were initially evolved for perception and action 'online', seem to be able to run 'offline', supporting higher-level cognitive abilities. Wilson mentions that such simulation models have, for instance, been proposed by Dennett (1995), and Glenberg (1997), but see also section 4.2 for more details.

Broadly speaking, these authors propose that the same neural mechanisms underlie both 'online' and 'offline' embodied cognition, simulating some aspects of the physical world by using the sensorimotor resources as a kind of representational device. Moreover, Wilson refers to experimental evidence for this kind of phenomena in mental imagery, memory, reasoning and problem-solving, which are all well-established cognitive domains showing illustrative examples of 'offline' embodied cognition. For example, she argues working memory seems to be functioning as a kind of symbolic off-loading, but instead of off-loading the workload to the outside, working memory off-loads information 'inside', on perceptual- and motor control systems (see Wilson, 2001), which is further addressed and elaborated on in section 3.4. However, Wilson does not mention the work conducted in AI, modeling sensorimotor simulations in autonomous agents (cf. e.g. Clark & Grush, 1999; Ziemke, Jirenhed & Hesslow, 2005). Matarić (1997) for instance, addresses the need to scale up research in embodied AI to deal with high-level cognition, and Clark and Grush (1999) suggest that research within situated robotics and autonomous agents needs to address the role of representations to consider higher-level cognitive abilities. They argue this can actually be done by internally simulating perception and action, using this emulation as a sort of internal representation (for details cf. e.g. Svensson, 2007; Svensson, Lindblom & Ziemke, 2007, and section 5.2).
In sum, Wilson states that there is a general trend in progress that stresses the whole embodied nature of human cognition, arguing that the body actually serves the mind, and that the ability of ‘offline’ cognition may have been an important factor for the development of human intelligence. However, she frequently focuses on the individual mind, neglecting the social nature of cognition and the broader notion of scaffolding as a strategy for offloading cognitive workload on the environment. Consequently, Wilson can be regarded as more of a ‘Piagetian’ than a ‘Vygotskian’, since she pays little attention to the role of the environment in enhancing and supporting cognitive functions. However, there are similarities between her proposal of symbolic offloading and the use of scaffolding, although she still focuses more on the individual than taking a broader perspective of cognition. The main contributions of her paper are the identification and disentangling of the different claims, in particular the introduction of the terms ‘online’ and ‘offline’ cognition and the explanations thereof, although the terms offer an implicit connection to the assumption that the mind functions as a computer.

3.2.2 Four complementary aspects of embodied cognition

Anderson (2003) has made another attempt to characterize embodied cognition. As a starting point for his description, he states that the main difference between situated and embodied cognition concerns their focus on the symbol-grounding problem (see Harnad, 1990, and subsection 2.7.2), with the latter approach considered to a greater extent. However, as he points out, it is obvious that these two approaches are closely related and complementary. Anderson has outlined four different aspects of embodiment, each of which playing an important role in shaping, limiting and grounding advanced cognition, namely physiology, evolutionary history, practical activity, and socio-cultural situatedness.

The first aspect, physiology, explains that the mind is not only embodied in the sense that our cognitive processes are instantiated in neural mechanisms, but suggests the functional design of the sensorimotor systems have an important role in other areas, such as concept formation and reasoning. For example, Anderson reveals that our color concepts are the result of the body’s physiology, since it is argued that focal hues, i.e. the ‘most’ red of red for example, is more easily processed by the sensorimotor system than brighter or darker nuances of red, and peripheral colors like pink and orange. Hence, he maintains our sense organs are dynamic instruments of exploration, having a central part in our understanding of the world. In the second aspect, evolutionary history, Anderson emphasizes the phylogenetical roots for our human intellect. In view of the fact that our bodily physiology has great similarities with our closest relatives, it has been argued that our intellectual capacities might also be similar, stressing there is a mental continuum between different species. Hence, in order to study cognitive processes, we have to ask, paraphrasing Hendriks-Jansen (1996), “[h]ow did this ability come to be there?”
Anderson’s next aspect, **practical activity**, highlights that practical and embodied activity can have cognitive and/or epistemic meaning, since it can be a part of a problem solving process, functioning as an external scaffold. The final aspect, **socio-cultural situatedness**, is according to Anderson, the most complex aspect of all, arguing that the distinction between situatedness and embodiment does not make sense at this level. Hence, he explains that humans and higher mammals are situated in a social and cultural context, implying that actions may have social and cultural meaning requiring background knowledge and ‘know how’ in a certain situation. However, he does not address how this background knowledge could be gained from an embodied point of view. On the other hand, he stresses the need for social and cultural scaffolding as a way of bootstrapping the accurate behavior in a cultural setting.

At first glance, Anderson’s four aspects of embodiment have some similarities with Wilson’s (2002) six claims, but a closer look also reveals major differences. For instance, Anderson’s aspects are not explained as elaborately as Wilson’s. However, while Anderson’s final claim, socio-cultural situatedness, is almost absent in Wilson’s article, as already mentioned, Wilson offers a more precise explanation of higher level cognition, stressing ‘offline’ simulated sensorimotor processes, which is not addressed in Anderson’s aspects at all. In addition, she points out that focusing on real-time interactions, which humans have in common with our closest relatives, could be misleading, since the ability of offline cognition might be unique for the human species. Nevertheless, they both stress the impact of the human ability of symbolic off-loading and scaffolding. However, Anderson further elaborates this final aspect later in his article, and inspired by Clark (1997), stressing the central role of socio-cultural situatedness as a scaffold for higher level cognition, in which the environment will be a part of an extended cognitive system. Nevertheless, this is actually the only stance of embodied cognition Wilson has some difficulty accepting.

Anderson’s four aspects of embodiment were derived from Lakoff and Johnson’s (1999) book *Philosophy in the Flesh*, which is argued (cf. Anderson, 2003; Neuman, 2001), to be one of the strongest and most complete statement so far, concerning the claim that all cognition in the end is ultimately grounded in embodiment. However, as described by Anderson (2003), Lakoff and Johnson (1999) do not explicitly state what the body actually is, and other criticisms of their work have been raised, for instance, by Neuman (2001). Neuman maintains there are three major difficulties in their work: they do not address cultural differences of the embodied mind, their conception of the body as a pre-given entity, and finally that the mind seems to be determined by the body in a casual and mechanical sense, ignoring the feedback between body and mind. Hence, Neuman concludes that Lakoff and Johnson (1999) actually make some serious epistemical errors by ignoring both the bodily and cultural dimension of cognition. However, Lakoff and Johnson (1999) in fact declare “what we understand the world to be like is determined by many things: our sensory organs, our ability to move and to manipulate objects, the detailed structure of the brain,
our culture, and our interactions in our environment, at the very least: *What we take to be true in a situation depends on our embodied understanding of the situation*, which is in turn shaped by all these factors. Truth for us...depends on such embodied understanding” (ibid. p. 102). However, while the content of their book does not provide more detail with regard to these views, as one might assume, from the above quote, they do consider linguistic expressions from different languages. It should be noted, however, that the major aim of their book is to re-interpret the Western philosophical tradition from an embodied point of view, which actually is one cultural dimension of embodiment.

### 3.2.3 Some additional notions of embodiment

Having addressed some attempts of characterizing embodied cognition, we move on to present and discuss different notions of embodiment or the kind of body required for cognition. Hence, if we adopt the stance that human cognition is embodied, to what degree are humans, and other natural and artificial systems actually embodied?

This perspective of embodiment is particularly derived from the field of AI, speculating if the embodied approach actually will lead to a ‘strong AI’, namely a robot with a mind. In order to address this question, Sharkey and Ziemke (2001) distinguish between two poles of embodiment, namely mechanistic and phenomenological embodiment. From an anti-mentalist stance, they characterize mechanistic embodiment as the view focusing on robots rather than computer programs. Accordingly, taking a mechanistic view of embodiment means that cognition is embodied in the control system of a sensing and acting machine, and nothing more. Hence, it is the behavior that is grounded in the interaction between agent and environment. This kind of robot control is inspired by the aforementioned work on reflexes by Sherrington and Loeb’s work on tropisms (see section 2.2). Other typical examples of mechanistic embodiment are Grey Walter’s work on the artificial tortoises Elmer and Elsie (see section 2.5), and the soda can-collecting robot Herbert (see section 3.1 and subsection 3.2.1). Phenomenological embodiment, on the other hand, refers to a mental or a subjective world, originating from the work of von Uexküll (see subsection 2.4.1). Von Uexküll was, as mentioned before, critical of mechanistic theories, since they neglected the subjective nature of the living organism. Instead, von Uexküll argued that the parts of the organism are integrated into a purposeful whole system, allowing it to construct its own phenomenological world. Hence, the notion of phenomenological embodiment stresses the essential part of having a *living* body, and therefore being able to experience a subjective world. Despite the fact that the robot Herbert might be both ‘situated’ and ‘embodied’, Sharkey and Ziemke (2001) argue the robot does not experience its world in the same way as living organisms, since it actually has no living body and “does not experiences the world directly”. Thus, they conclude strong phenomenological embodiment is impossible with current technology, since robots lack a living body. However, it is possible to create simulations and models of embodied cognition, in a kind of a *weak* embodiment. In addition, they point out the question of *strong mechanical* embodiment is

93
more complicated, arguing the behavior of a robot is only meaningful from the observer’s point of view, not in the ‘robot’s world’, and making other interpretations is to recapitulate the Clever Hans error (see section 2.2). In sum, Sharkey and Ziemke (2001) conclude that *strong embodiment* is (so far) impossible from both a mechanical and a phenomenal point of view, but *weak embodiment* is possible for both aspects.

An obviously closely related question is what kind of a body is required for (embodied) cognition, and what kind of an artificial system can be considered embodied. Ziemke (2003) argues that many discussions concerning embodiment do not stress what kind of a body is required for cognition. As a consequence, this leaves open the question whether a humanoid robot (a robot with a humanlike design) for instance, could have the same sort of cognition as its living counterpart. In order to discuss these questions, Ziemke (2003) identifies and contrasts six different notions of embodiment found in the literature, but as he explains, they should not be viewed as distinct positions, but rather an attempt to group related ideas together. The six notions are as follows:

- **Structural coupling** between agent and environment
- **Historical embodiment** as the result of a history of structural coupling
- **Physical embodiment**
- ‘**Organismoid’ embodiment**, i.e. organism-like bodily form, such as humanoid robots.
- **Organismic embodiment** of autopoietic, living systems
- **Social embodiment**, the role of the body in social interaction

The first, and broadest notion of embodiment is *structural coupling*, meaning that any system can be considered embodied if it is “structurally coupled” to its environment, but this notion does not necessarily require having a body. Franklin (1997), for instance, argues software systems can be intelligent without a body in a physical sense, but he further claims “they must be embodied in the situated sense of being autonomous agents structurally coupled with the environment” (ibid. p. 500). Ziemke mentions that the term structural coupling has its roots in the work of Maturana and Varela (1980, 1987), who view cognition as a biological phenomenon. Quick *et al.* (1999), as well as Dautenhahn, Ogden and Quick (2002) have been inspired by the idea of structural coupling when they made an attempt to formulate a “precise definition” of embodiment. Additionally, their work is also an attempt to clarify weaknesses in Pfeifer and Scheier’s (1999, p. 649) definition of embodiment, which is as follows:

**Embodiment**: A term used to refer to the fact that intelligence cannot merely exist in the form of an abstract algorithm but requires: a physical instantiation, a body. In artificial systems, the term refers to the fact that a particular agent is realized as a physical robot or as a simulated agent.
Dautenhahn et al. (2002) claim this definition actually does not say anything of importance for the notion of embodiment. For instance, they state according to this definition, a living fish in the water and a dead fish in the supermarket are embodied in the same way, since the definition as such lacks differences in degrees concerning the level of embodiment. Additionally, the definition takes no notice of the environmental circumstances within which the system is embodied. Hence, following Pfeifer and Scheier’s (1999) definition of embodiment, every system having a body, either biological, metallic, live or dead, is nevertheless embodied according to Dautenhahn, Ogden and Quick (2002). They explain that embodiment, particularly in robotic research, is considered the same as being physically instantiated, which they argue is not the same as being embodied. Hence, they declare no explicit analysis of “the nature of embodiment” exists, but there are rather a number of considerations regarding the consequences of being embodied. They further noted that it may be impossible to reach a unified definition of embodiment within the areas of robotics, cognitive science and biology, within the near future.

Thus, after analyzing which specific properties an agent (natural or artificial) needs to be equipped with to make them embodied, they conclude that considering only agent-oriented characters in isolation does not suffice, but rather there is a need to focus on the relationship between the agent and its surrounding. As a result, they want to quantify embodiment, measuring the relationship that exists between the agent and the environment, in which it is embedded. As a result, they propose a minimal definition of embodiment, which states:

\[ \text{A system } S \text{ is embodied in an environment } E \text{ if perturbatory channels exist between the two. That is, } S \text{ is embodied in } E \text{ for every time } t \text{ at which both } S \text{ and } T \text{ exist, some subset of } E \text{’s possible states with respect to } S \text{ have the capacity to perturb } S \text{’s state, and some subset of } S \text{’s possible states with respect to } E \text{ have the capacity to perturb } E \text{’s state} \] (Dautenhahn, Ogden & Quick, 2002, p. 400).

Dautenhahn et al. (2002) actually claim that their minimal definition does not rule out any aspect of being embodied. Despite the fact that this definition is one of very few attempts to define embodiment, it has been argued that the above definition is of limited use. Riegler (2002), for example, explains although “this attempt to clarify the notion of embodiment is an important first step, it is at the same time an insufficient characterization” since “every system is in one sense or another structurally coupled with its environment. This applies to living creatures as well as to computer programs, since both are exposed to perturbations” (Riegler, 2002, p. 341). Thus, as Ziemke claims (2003), this minimal definition of embodiment does not make any distinction between cognitive and non-cognitive systems, and it could be argued that this definition has little to tell us about embodied cognition. This claim is in fact illustrated in Quick et al.’s (1999) own example of a granite rock (S), on the Antarctic tundra (E). The rock is persistently perturbed by the forces of the arctic wind which in turn perturbs the flow of the air. Thus, according to the above minimal definition of
embodiment, the rock is actually an embodied system, which indeed not many cognitive scientists will insist on.

The notion of **historical embodiment**, stresses that the embodiment of a system is the outcome of previous agent-environment interactions. This means, besides being structurally coupled for the moment, the historical aspect may in many cases result in co-adaptations between, for example, the system and its environment. Varela, Thompson and Ross (1991), for example, argue that “knowledge depends on being in a world that is inseparable from our bodies, our language, and our social history – in short, from our embodiment” (ibid. p. 149). Likewise, Ziemke (1999) notes “[n]atural embodiment [of living systems] ... reflects/embodies the history of structural coupling and mutual specification between an agent and environment in the course of which the body has been constructed”. Moreover, Riegler (2002) describes this historical coupling in a similar vein, arguing that “a system is embodied if it has gained competence within the environment in which it has developed” (ibid. p. 347). Despite the fact that this notion is a little more restrictive than the previous one, stressing the evolutionary and/or developmental roots of embodiment, Ziemke concludes the notion of historical coupling obviously is not particularly restrictive either. The notion actually seems to be applicable to nearly all living systems, and it does not exclude systems without a physical body.

The notion of **physical embodiment** omits systems lacking a “physical instantiation”, which was aforementioned in Pfeifer and Scheier’s (1999) definition of embodiment. Hence, this notion excludes Franklin’s (1997) software agents, but not Quick et al.’s (1999) granite outcrop or any other object of little importance for cognitive science. Moreover, Ziemke argues although the notions of historical and physical embodiment could be considered special instances of structural coupling, they are not equal. Riegler (2002) remarks that his own definition of embodiment (cf. above) “does not exclude domains other than the physical domain” especially “[c]omputer programs may also become embodied”, but only if they are the result of their own self-organization, and not the outcome of a designer. Similarly, Ziemke argues living systems are cases of physically embodied systems that are also historically embodied, while many other physical systems are not historically embodied ones.

The notion of **‘organismoid’ embodiment** limits embodiment to organism-like bodies, i.e. physical bodies which to some extent have a similar form of living organisms, and are equipped with some kind of sensorimotor capabilities. Hence, this notion is proposed to cover both living organisms and their artificial counterparts, although only the former could be considered historically embodied. However, although embodiment has been regarded as one of cornerstones of new AI since the early 1990s, only recently have researchers begun to build humanoid robots as a way of modeling truly human-like intelligence in artifacts (see section 8.2). However, Ziemke points out that although humanoid robots and real human beings could be viewed as a case of ‘organismoid’ embodiment, this still leaves open
the question concerning what the supposed significant bodily differences between humans and other living organisms are from a cognitive viewpoint. Thus, he argues, it might be the case that having arms, legs or other bodily parts may be of crucial importance for human-like embodied cognition, although the notion of ‘organismoid’ embodiment does not in fact reveal why a ‘bodily design’ is necessary.

The notion organismic embodiment includes only living bodies, and is inspired by the work of von Uexküll (see subsection 2.4.1), along with Maturana and Varela’s (1980, 1987) biological explanation of cognition. Ziemke explains that these approaches, broadly speaking, claim that cognition in fact is what living systems do in interaction with the environment. Maturana and Varela (1987, p. 29) characterize cognition as the “effective action of a living being in its environment”, and in their view, ‘affective action’ is the same as ‘knowing’. Knowing, in turn, is defined as “operating effectively in the domain of existence of living beings”. Consequently, they claim knowledge is the result of an ongoing interpretation that emerges from our biologically embodied structures of understanding, by living in and experiencing a social and cultural environment. Hence, these actions make it possible for us to “make sense” of our being in ‘our’ world. They argue the ‘mind’ functions as an autonomous system (having operational closure) and there is a mutual relation between the mind and the world, through structural coupling. It is through structural coupling that a complex system enacts a world that is brought forth through a history of structural couplings. However, this structural coupling should not be viewed as similar to the input/output mapping in the computer metaphor for mind. In that model, meaning is created from the outside by the designer, while in an enactive system, meaning is the result of the organization and history of the system itself. Hence, the organism and the environment have been developed and bound together, through the process of natural drift.

Varela et al. (1991) point out that Darwin never stated the “law of natural selection”, but instead used the term as an analogy for explaining his idea (Mitchell et al., 1997). The evolutionary process is usually viewed to function as a ‘force’ of natural selection, but he never made the claim this force actually existed. Obviously, there is no optimal design or solution, but instead ‘evolution’ can be viewed as a satisfier or/and a tinkerer, taking what is present at hand and not being an inventor. This means, evolution should not be viewed as a “field of forces” (selective pressures, genetic variety, optimizing the fitness), since this explanation is too simple, instead, evolution can be regarded as a natural drift (Varela et al., 1991). Accordingly, different visual systems have evolved in various animal species, and they should not be considered as optimal adaptations to different regularities in the world. Instead, they should be explained as the result of different histories of natural drift. They are enacted or brought forth through a specific history of structural coupling. The environment and the organism cannot be separated since they in fact are codetermined in evolution, which is dependent on the perceptually guided activity of the organism and the environmental regularities. For
example, the ultraviolet light patterns in flowers have coevolved with the
ability to ‘see’ ultraviolet light in honey bees, since both the flowers and the
bees have a history of structural coupling with each other.

Following this line of argument, von Uexküll (1928), as well as Maturana
and Varela (1980, 1987) argued that there is an apparent difference between
living organisms and human-made machines. However, Ziemke notes that
these two approaches use different terms in explaining their view.
Nevertheless, the general idea is basically the same, although for obvious
reasons elaborated in more biological detail in their own theory. Hence, they
view all living systems as autonomous or autopoietic, whereas machines are
regarded heteronomous and allopoietic (cf. Ziemke, 2000, 2001a). As
previously mentioned (see subsection 2.4.1), von Uexküll (1928)
characterized autonomous systems as ones capable of growth, able to repair
themselves and containing their own functional ‘self-steering’ rule. In
addition, these systems have a past of recurrent interactions with the
environment. Heteronomous systems, on the other hand, are described as
ones with no growth ability, unable to repair themselves, having no historical
foundation of reaction, and following rules designed by humans. As a result,
von Uexküll argued that the construction of living systems and machines
diffs, since living systems are constructed centrifugally, whereas machines
are constructed centripetally. Broadly speaking, living systems begin with
one single cell, from which the different organs then develops, according to
its own rules. Machines, on the contrary, consist of different parts
assembled to form a ‘unified’ system, following rules designed externally of
the system.

While Maturana and Varela’s (1987) characterization of different systems is
similarly directed, they distinguished between the organization and structure
of a system. Beginning with autopoietic (living) systems, they denote
organization as the relations that must exist among the components of a
system, allowing it to be a member of a certain class, for instance, belonging
to the class of vertebrates or the parts and connections that exist in a watch,
independent of the certain material. The particular organization that
characterizes living systems is their ability of continuous self-production,
namely an autopoietic organization. Hence, there is no separation between
producer and product, since the components of the system are produced by
the system itself, and the fundamental variable to maintain a constant in an
autopoietic organization is the own organization as such. Structure, on the
other hand, is characterized as the realization of the actual components, and
the relations between them. These components and relations constitute a
particular unity, making its organization real. In vertebrates this is realized
in a living animal, consisting of different kinds of cells, which make up the
body’s physiology. However, machines, on the other hand, have a structure
designed and assembled by a human designer from the ‘outside’, (cf. von
Uexküll’s notion of centripetally), and its organization is specified by the
concatenation of processes not produced by the system itself and they are
independent of the machine as such.
As Ziemke explains, with current technology, organismic embodiment is limited to natural systems. However, this notion of embodiment does not address the uniqueness of human embodied cognition, since all living systems according to this notion are actually embodied. To conclude, he argues that in fact, this notion of embodiment is perhaps also of limited value to cognitive science, since it really avoids the question of what constitutes human and higher-animal cognition. However, Maturana and Varela (1987) in fact distinguish between different kinds and levels of autopoietic systems. Starting with different kinds of living systems, they differentiate between unicellulars (living organisms consisting of one single cell) and metacellulars (living systems consisting of numerous numbers of cells, being reciprocally coupled). These metacellulars are so-called second-order autopoietic systems, since they include cells that are components of their structure, and having operational closure. Additionally, they address third-order structural coupling, which arises when metacellulars enter into interaction with another organism. When such a third-order structural coupling occurs, these organisms will emerge to a new phenomenological dimension, characterized as particularly complex behaviors via interactions between these organisms. Thus, social phenomena and communication are the result of this spontaneous organization of third-order couplings, and they generate a certain kind of internal phenomenology (We return to this topic in chapter 5, when we discuss the nature of social interaction).

Finally, the notion of social embodiment is addressed, and exemplified with Barsalou et al. (2003) notion of social embodiment, which is as follows: “states of the body, such as postures, arm movements, and facial expressions, arise during social interaction and play central roles in social information processing” (ibid. p. 43.). However, Ziemke points out that their notion of social embodiment addresses the role of embodiment in social interactions rather than discussing what sort of body is required for social embodiment. This means, the what question needs to be addressed in future research concerning embodiment and social interaction.

Thus far, many aspects of the embodied nature of cognition have been portrayed. We now highlight some issues that need clarification and more detailed addressing in the current theories of embodied cognition, in order to develop a "mature science of the mind" as Clark stated in section 3.2. Taken together, the lack of agreement concerning embodiment has resulted in some oversimplifications of the role of the body in cognition. For example, the discussion has mainly focused on a "static" body, i.e. what kind of physical realization of the body is necessary for cognition. The notions of embodiment presented in the above subsection (3.2.3), for instance, range from quite basic shapes that interact with the environment to more organism-like bodily forms as well as humanlike robots. The point I want to make is that the crucial aspect of the “body in motion” has received little interest. Therefore, current theories of embodied cognition need to pay more attention to the role and relevance of “body-in-motion” for cognition, which is the topic in the following section 3.3.
3.3 Body in Motion

As previously stated, while appearance so far has received much attention in discussions of embodied cognition, the crucial role of the body in motion has received very little, although research in various domains has shown the relevance of locomotion experience to human cognition (cf. e.g. Adolph & Berger, 2006; Farnell, 1999; Gallagher, 2005; Hurley 1998; Sheets-Johnstone, 1999; Thelen & Smith, 1994; Van Gelder & Port, 1995). Furthermore, Trevarthen (1977, in Hendriks-Jansen 1996) pointed out that a practical motive is another reason for the neglect of dynamical aspects in psychological research. Since bodily movement patterns of humans were difficult to observe with the technology of the time, cognitive science consequently became more of a static science of perception, cognition and action than a science of embodied dynamic interactions. Moreover, Varela (1992, in Farnell, 1995, p. 23), for example, states that “[s]ocial scientists are body-dead because they are conceptually brain-dead to signifying acts within the semiotics of body movements”. However, when researchers actually paid attention to embodied movement, it often appeared that, as Farnell (1995) formulates it, the moving body has lost its mind.

Furthermore, Farnell also claims that besides treating cognizers as embodied agents, we have to recognize the embodied agent moves. Similarly, Sheets-Johnstone (1999, p. 4), argues “bodies are hardly at the forefront of thought about thought; neither for that matter is movement”, and subsequently, bodily movements are consistently taken for granted or ignored. As she puts it, while much research has focused on mind, not much has focused on bodies ‘as if minds have nothing to do with bodies’ (p. 6). According to Sheets-Johnstone (1999), the subjective tactile-kinesthetic experience of one’s own moving body is the bedrock of thinking. This means, the body is not a bridge connecting the subjective and the objective body; the real issue of embodiment is not the ‘packaging.’ Instead, the self-experienced bodily understanding is the elemental and unsurpassable unity of embodied actions. She notes that although Merleau-Ponty (see subsection 2.4.2) is commonly viewed as the ‘Knight of the Body,’ he actually overlooked the deeply ingrained role of self-experienced movement in embodied beings. She also claims that the core of being is the relation between the body and movement, stressing ‘consciousness does not arise in matter; it arises in organic forms, forms that are animate’ (p. 43). The human experience of movement is stunning. Human infants are not born inanimate, but already moving, as fetuses begin to produce self-induced movements somewhere between five and six weeks after conception, shortly after developing body parts with which to move. Already at 12 weeks, a majority of the fetuses’ arm movements are directed toward something, such as their own faces and bodies, and the environmental features of the womb (Adolph & Berger, 2006). There is hence a need to discover how we ‘put ourselves together’, in order to catch ourselves in the tactile-kinesthetic ‘apprenticeship’ of our own bodies (Sheets-Johnstone, 1999).

Accordingly, Adolph and Berger (2006) emphasize “movement is perhaps the most ubiquitous, pervasive, and fundamental of all psychological activities. It
is the hallmark of animacy and the essence of agency. Across development, self-initiated movements of the eyes, head, limbs, and body provide the largest source of infants’ perceptual experiences” (p. 167). In accordance with this remark, Gallagher (2005) emphasizes that intelligent behavior which is demonstrated in the human infant should not only be measured by their physical manifestations, given that they actually are those processes.

In order to clarify the role played by the ‘body in action’ for cognition and consciousness, Gallagher (2005, 2007a) proposes that a conceptual distinction between body image and body schema is needed. He states that these terms have been used ambiguously in the literature, resulting in much confusion. According to him, they refer to two different although closely related systems. Phenomenological reflection implies the difference between taking an intentional attitude towards one’s own body as well as having a capacity to move or to exist in the action of one’s own body. The concepts of body image and body schema relate to this phenomenological difference (Gallagher, 2005).

- **Body image** consists of a system of perceptions, attitudes, and beliefs pertaining to one’s body.
- **Body schema** is a system of sensorimotor capacities that function without awareness or the necessity of perceptual monitoring (p. 24).

The distinction between body image and body schema, however, is not an easy one to make, since the two systems interact and are highly coordinated; in the context of intentional action, as well as in pragmatic and socio-cultural contexts (Gallagher, 2005). He states that **body image**, besides involving on-going perceptions, can also include mental representations, beliefs, attitudes when the object of such intentions concerns one’s own body. In other words, the image a person has of his/her own body. The body image is also able to contribute to the control of movement, given that our attention to our bodies’ visual, tactile, and pro-prioceptive experiences may support us when developing new movements and actions. When learning a new movement, such as a dance step, improving the backhand in tennis, or imitating the novel movements of others, we may consciously monitor and correct our own movements. Briefly stated, the body image is the perception of movements.

In contrast, the **body schema** involves certain motor capacities, abilities, and habits that both enable and constrain movements and the maintenance of posture, referring to what extent, and in what precise way one’s body constrains or shapes the perceptual field (Gallagher, 2005). The body schema usually operates best when the intentional object of perception is something other than our own bodies, which means that the body schema is a sensory-motor system that functions below the level of self-referential intentionality (Gallagher, 2005). Most generally, movement and the maintenance of posture are accomplished by the almost automatic performances of the body schema, which explains why adults neither need nor have to maintain a constant body percept in order to move around the world. But this does not imply that the body schema operates as merely a
matter of reflex, because body schema controlled movements can be accurately shaped by the goal-directed behavior of our intentions. For example, when we reach for a glass of water with the intention of drinking from it, our hands ‘shape themselves’ in order to pick up the glass, and this happens without our conscious awareness. In other words, body schema supports intentional activity, without being a form of consciousness or cognitive function, given that the attention and awareness in those cases are focused on external object(s), and not on the particular accomplishment of movement and locomotion (Gallagher, 2005). Briefly stated, the body schema is the accomplishment of movement. All in all, the perception of one’s own or somebody else’s movement is interrelated is complex ways. For instance, when the awareness of one’s body becomes explicit in terms of monitoring or directing perceptual attention to aspects such as bodily position, movement, or posture, then such awareness supports the perceptual aspect of the body image. This awareness may then interact with the body schema in complex ways, without being equivalent to a body schema itself. Furthermore, the contribution made to the control of movement by body image will always find its complement in capacities that are defined by the operations of a body schema that continue to function to maintain balance and enable movement. Such operations are always behind our awareness. This means, a body schema is not reducible to a perception of the body; it is never equivalent to a body image (Gallagher, 2007a).

With regard to the role of movement in cognitive development, research in motor development has usually been considered with the mere study of physical growth, motoric milestones, and so on, but there is emerging empirical evidence revealing interesting and important links between cognitive and motoric functions (cf. e.g. Adolph & Berger, 2006; Thelen & Smith, 1994). Traditionally, the systems of perception and action have been considered as having no essential contact with each other, as well as constituting separate research areas where perception is the dominant object of study, and action has been neglected (Hurley, 1998). Hurley argues against the traditional “Input-Output picture “of perception and action, following in the footsteps of Dewey’s work on the reflex arc (see section 2.2 and subsection 2.4.5). Most generally, she portrays the mainstream view of the mind as a kind of sandwich, with perception the bottom slice, actions the top slice, and cognition the filling in between, thus implying that perception and action operate separately. However, she emphasizes that perception, action, and environment are deeply intertwined, stressing perception and action are interdependent of each other, given that action shapes perception as much as perception shapes action in a continuously evolving loop. However, Adolph and Berger, for instance, emphasize that much of their work focuses on perception-action coupling “with the aim of understanding developmental changes in the perceptual, cognitive, social, and emotional processes that contribute to the adaptive control of motor actions” (2006, p. 162). They point out that the traditional textbook image of a motionless eyeball waiting to receive perceptual information is overly simplistic and artificial. On the contrary, the eye, head, limb and body movements bring the perceptual system to the available information. That is, perception-action
loops guide future actions rather than merely bring them forth. Thus, the interrelations between perception and action are fundamental for the adaptive control of motor action and cognition (Adolph & Berger, 2006; Hurley, 1998). Furthermore, perceptual experience is response-dependent, since certain kinds of response actions may interfere with perception, whereas others instead enhance perception. Moreover, action is not a mere effect of consciousness, but instead helps to constitute conscious experience, and intentional action or agency is essential for consciousness (Hurley, 1998). Thus, this view is highly compatible with an embodied approach of cognition.

In addition, research conducted by Thelen and Smith (1994) has, for instance, explored the changing shape and performance of infants’ reaching, kicking, and self-induced locomotion movements, to illustrate the close coupling of developmental trajectories, proposing a dynamical systems approach to cognition and action. As Adolph and Berger noted, maybe the most far-reaching implication of research on action-perception coupling to motor development is that motor skills have become significant to developmental psychologists who would not normally consider motor control. The fact that bodily actions and movements are directly observable, in contrast to psychological entities, offers a unique approach for studying developmental processes. Adolph and Berger argue that there is no need to infer the content of children’s thoughts, perceptions, emotions, intentions, linguistic representations, and so on, from overt behaviors (such as speech, gestures, facial expressions or eye movements) or images of brain activity, since, unlike covert mental events, motor behaviors are ‘out in the open’ and also ‘have shape’. This means, there is no inferential leap separating an embodied action from a description of its form (Johnson, 2001). Accordingly, this has implications other than mere visibility, because movements and motor actions are embodied and embedded in a material and social environment, in which the individual’s embodiment is constantly performed by a creature with certain bodily capacities and deficiencies (Adolph & Berger, 2006). That is, the functional significance of embodiment and its possibilities for action depend on the fit between the actor’s body and the properties of the surrounding environment. Briefly stated, the central and inescapable feature of motor action is that the opportunities and constraints of the body as well as the environment are continually changing.

Moreover, Adolph and Berger state that researchers of action-perception do not normally consider social interaction and cognition as central topics of study in their work. However, children’s possibilities for action are not merely constrained by their immature bodies and limited skills, since in their being part of the social world, caregivers provide social scaffolding as well as various cultural tools that are then incorporated into children’s action-perception couplings. Hence, the development of action is not a lonely enterprise, since infants typically acquire new motor skills in supportive social contexts. When providing toddlers a finger to hold when learning to walk exemplifies caregivers providing various kinds of social and material scaffolds (see Clark, 1997). Cultural differences in child rearing practices
emphasize the caregivers’ role in providing opportunities for action. Consequently, infants acquire, for instance, their sitting, and self-induced locomotion actions at different timescales (cf. Greenfield, 2002; Rogoff, 2003; Rogoff & Lave, 1984).

Summing up, in this section I have described and highlighted that there is an essential relation between the ‘body in motion’ and cognition, and as Gallagher (2005) puts it, movements are the part and parcel of cognition, and thus embodiment actually ‘shapes the mind’. We return to the particular issue of body in motion in section 5.5, while the role and relevance of the social body in motion in social interaction and cognition is discussed. In more detail, the relationship between self-induced locomotion and the development of self and other social cognitive processes in human children are discussed and analyzed further.

3.4 Toward the Social Dimension of Embodiment

Given that the neglect of body in motion in current theories of embodied cognition has been addressed in the previous section, we now turn our attention to another oversimplification in current embodied cognitive science, namely the social dimension of embodiment. The main focus in most theories of embodied cognition has until now been on the relation between the individual body and its cognitive processes, in interaction with the physical environment (cf. e.g. Chrisley & Ziemke, 2003; Clark, 1999; Núñez, 1999; Riegler, 2002; Wilson, 2002; Ziemke, 2003). This is quite surprising since the ability to engage in social interaction is one of the building blocks of human culture, as well as the foundation for the complexity of social life and social cognition. However, Johnson and Rohrer (2007, p. 19), for instance, point out that “embodied cognition is often social and carried out cooperatively by more than one individual”. Consequently, there is a need to move beyond the current emphasis on the interactions between the individual and the physical environment, to interactions between agents and their social environment. Others have also recognized this, and argue that the theoretical knowledge of embodiment needs to be extended beyond current notions and/or levels of ‘individual’ embodiment (cf. e.g. Anderson, in press; Riegler, 2002; Sinha & Jensen, 2000; Ziemke, 2003).

Sinha and Jensen (2000), for instance, address the appeal for a social dimension of embodiment in cognition. They stress that although the social sphere has been mentioned in earlier writings (e.g. Lakoff, 1987; Varela et al., 1991) which neither further explain how the relation between the physical and the social world is realized, nor in what ways the organism’s social and physical environments relate to each other. In other words, although the relevance of the social experience is in no way denied, it is not further explored either. Therefore, the ‘Embodiment thesis’ needs to continue “beyond the body”. This means, they call for an extended notion of

---

20 Larger parts of this section have earlier been published in Lindblom & Ziemke (2003; 2005, 2007).
embodiment that is not restricted to the “humanly corporeal”, but incorporates the social dimension of embodiment as well (Sinha & Jensen, 2000). In particular, they argue the embodiment approach, as currently formulated, suffers from the same limitations as Piagetian developmental theory (cf. subsection 2.4.3), namely not paying sufficient attention to the role and relevance of culture and society in human cognition. Moreover, they note that although the embodied approach has challenged the mind/body dualism at the individual level, the dualistic stance between the individual and society is still present. Hence, they state the “embodiment thesis” so far, does not explore its specific role in cognition and language. Similarly, Riegler (2002) notes since other individuals are a part of one’s world the social dimension is more a sophistication than a contradiction of embodiment. In accordance with this remark, Anderson (in press) states that since at least some organisms are coupled to a physical as well as social environment, various “perturbatory channels” between the organism and the social world that also matter to the character of its cognitive processes consequently exist. He suggests nearly everything that has been proposed in Wilson’s article *Six views of embodied cognition* (see subsection 3.2.1) can be recapitulated through emphasizing the interaction between the organism and its social environment. All in all, there seems to be a need to focus more on the issue of social embodiment, which is the main focus of the following chapters.

The lack of understanding of the knowledge of social embodiment has thus far resulted in some oversimplifications of the role of the body in social interaction and cognition. The two perspectives from which this can be considered constitute the starting-points for the following two chapters.

One perspective concerns the poorly understood role of the body in social interaction and social cognition despite the interest in social interaction that has been presented. Empirical evidence from social psychology, for instance, has revealed how social thought and judgments can be affected by bodily states, actions and motivations. Although these findings, among others, indicate that the body might be used as a mediator or resonance mechanism in the process of perceiving others, its function is still poorly understood. Consequently; I shall in the next chapter describe and discuss how embodiment contributes to social interaction, beyond the general terrain of embodiment presented in this chapter. This issue constitutes the major topic of *chapter 4*, and the consequences of this chapter can be considered way stations on the road to the final framework of embodied social interaction in section 6.2.

The other perspective describes that mainstream theories of cultural and social cognition tend to overlook the role of the body in social interaction by treating embodied social interactions such as body posture, gaze and gesture as nothing but the visible outcomes of mental intentions. Since one of the reasons for the absence of the body in the social sciences is the anxiety of slipping into *biological determinism*, social sciences, therefore commonly regard mind as superior to and independent of the body. These issues and
alternative characteristics of the relational nature of cognition and development are, among other issues, explored in more detail in chapter 5.
Chapter 4

Our body is not just the executants of the goals we frame or just the locus of the causal factors which shape our representations. Our understanding itself is embodied. That is, our bodily know-how and the way we act and move can encode components of our understanding of self and world.
Taylor, 1993

The body, as represented in the brain, may constitute the indispensable frame of reference for neural processes that we experience as the mind; that our very organism rather than some absolute external reality is used as the ground reference for the constructions we make of the world around us and for the construction of the ever-present sense of subjectivity that is part and parcel of experiences; that our most refined thoughts and best actions, our greatest joys and deepest sorrows, use the body as a yardstick.
Damasio, 1995

4. Embodiment and Social Interaction

This chapter presents and reviews different perspectives on social aspects of embodiment, since recent work in cognitive science and related disciplines indicates that the body and its sensorimotor processes play important roles in social interaction, cognition and language. The theoretical considerations and empirical findings range from disciplines such as social psychology, phenomenology, neuroscience, and gesture to linguistics. In general, section 4.1 portrays empirical findings, mostly from social psychology, which presents several compatibilities between bodily and cognitive states in various sorts of on-line and offline social interactions, such as mimicry, emotions, attitudes, and social perception.

Section 4.2 begins with phenomenological aspects of intersubjectivity and proceeds to the possible neurological underpinnings of such social understanding, especially addressing theories of embodied simulations/re-enactments and mirror neurons. Generally speaking, the different theories concerning ‘theory of mind’ which have emerged, can be broadly divided into ‘theory theories’ and ‘simulation theories’ (Carruthers & Smith, 1996; Gallagher, 2005). Broadly speaking, theory theorists argue that the ability to predict and interpret the behavior of others is supported by a folk psychological theory, based on the structure and function of the human mind (cf. the previous chapter). The theory as such could be innate and
modularized, learned or acquired. Followers of the simulation theory, on the contrary, claim that the mind-reading ability is not any sort of theory. Instead, it is an ability to project oneself into another person’s point of view; simulating what it is such as to be in the other person’s situation (Carruthers & Smith, 1996). In order to simulate, there is necessary for an ability to imitate the ‘inner states’ of another person and it has been supposed that the body and its sensorimotor processes can be used as a linking device when perceiving others. It should be noted, however, that this ‘radical’ view should not be misinterpreted as claiming there is a direct correlation between so-called ‘objective’ (third-hand perspective) neurological states in the brain and ‘subjective’ (first-hand perspective), phenomenological experience, which might be the impression at first glance. On the contrary, as pointed out by Gallagher (2005), bridging the troubled water of social cognitive neuroscience and phenomenology through a direct mapping is no viable approach, because “there is no short cut that can bypass the effects of embodiment” (p. 244). This means, he suggests that the Cartesian divide between body and mind might be blurred, and perhaps erased. According to Gallagher (2005), the standard models of mind such as representational, functional or computational are still oversimplified, despite their complexity, since they altogether neglect to consider the effects of embodiment.

Then I subsequently tune up the microscope, and in section 4.3, present linguistics from an embodied perspective, describing how language might be grounded in embodiment and movement. In particular, focusing on the ways gesture and speech are interrelated at both the neurological as well as at operative levels. Finally, in Section 4.4 the findings from the presented disciplines are analyzed, and generalized into four fundamental functions of embodiment in social interaction.

4.1 Embodied Social Psychology

Semin and Smith (2002) point out that findings in social psychology and current research in embodied cognition have a lot in common. While on the one hand, the relevant ‘situation’ of interest is actually a social situation, on the other, several interesting phenomena in social psychology could be explained from an embodied perspective (Semin & Smith, 2002). Concerning embodiment, social psychologists report how social thought and judgments can be affected by bodily states, emotions, actions and motivations (cf. e.g. Barsalou et al., 2003; Niedenthal, et al., 2005a; Niedenthal, et al., 2005b; Semin & Smith, 2002).

4.1.1 Four general social embodiment effects

Barsalou et al. (2003) explain that there are plenty of empirical findings implicating connections between embodiment and social cognition. Furthermore, with social embodiment they mean that “states of the body, such as postures, arm movements, and facial expressions, arise during social interaction and play central roles in social information processing” (ibid., p. 43). Moreover, social psychologists have described four forms of (social) embodiment effects, which have been unexplained from a unified
theoretical perspective. However, Barsalou et al. (2003) argue that theories of embodied cognition actually offer a framework for explaining and unifying these social embodiment effects, which depending on the situation, range from simulations to actual full-blown executions. Accordingly, they offer a theory of social embodiment, and describe how this theory will explain these social embodiment effects, which are the following ones.

- **Perceived social stimuli do not only produce cognitive states, but also bodily states**
- **Perceiving bodily states in other people results in bodily mimicry in oneself**
- **Own bodily states produce affective states**
- **The compatibility of bodily states and cognitive states adjusts performance effectiveness**

Firstly, perceived social stimuli do not only produce cognitive states, but they also produce bodily states. For instance, Weisfeld and Beresford (1982, in Barsalou et al., 2003) reported that high school students who received good exam results adopted a more erect posture than those that received poor grades, who instead adopted a less erect position. Barsalou et al. (2003) speculate if a related question is whether the social event as such triggered the bodily reaction directly, or if mediating mechanisms exist. It could be the case that an emotional state was triggered by the receiving of the grade, which in turns resulted in a bodily state. Another example is Bargh, Chen and Burrows’s (1996, in Barsalou et al., 2003) experiment which primed subjects with concepts related to elderly (‘gray’, ‘bingo’, and ‘wrinkles’). The result demonstrated that the group primed with these critical words primed embodiment effects such as moving more slowly when leaving the experiment laboratory, while this embodied effect was lacking in the control group primed with neutral words. Other experiments have demonstrated similar effects (cf. e.g. Aarts & Dijksterhuis, 2002; Dijksterhuis & Bargh, 2001; in Barsalou et al., 2003). Additionally, there is evidence that perceived social stimuli produce facial responses; viewing pleasant visual scenes produces positive facial responses, while negative ones seem to produce negative expressions. Hence, is can be argued that the perceived social stimuli automatically adjust expressive facial reactions, and this effect is reported when subjects view pictures imaging another human or reading about fictional characters (cf. e.g. Andersen, Reznik & Manzella, 1996; Cacioppo et al., 1986; Vanman & Miller, 1993, Vanman et al., 1997; in Barsalou et al., 2003). Moreover, social stimuli affect embodied aspects of communication, since subjects primed with words related to a certain stereotype affected their behavior according to the particular priming (cf. Dijksterhuis & van Knippenberg, 2000; in Barsalou et al., 2003). Broadly speaking, theses studies show that various social stimuli produce embodied responses in the perceiving person himself/herself.

Secondly, perceiving bodily states in other people actually results in bodily mimicry in oneself, namely that the embodied responses mimic the perceived social stimuli. For instance, viewing a smiling person produces, to
some degree, the same facial expression in the observer. They identified three forms of mimicry, namely bodily, facial and communicative mimicry.

With regard to bodily mimicry, it is reported that bodily actions between interacting people often become synchronized (cf. e.g. Bernieri, 1988; Bernieri, Reznik & Rosenthal, 1988; in Barsalou et al., 2003). For instance, Chartrand and Bargh (1999, in Barsalou et al., 2003) point out that subjects mimicked the experimenter’s actual behavior, such as rubbing the nose or shaking a foot pretty often. Evidence of facial mimicry is widely documented in the literature (cf. e.g. Dimberg, 1982; Provine, 1986; in Barsalou et al., 2003). O’Toole and Dubin (1968, in Barsalou et al., 2003), for example, point out that mothers tend to open their mouths after their infants have opened their own during feeding. Embodied mimicry is also present during communication, matching the speech rate, utterance duration, and emotional tone. For instance, Bavelas et al. (1998, in Barsalou et al., 2003) report that listeners often mimic speakers’ manual gestures, and it has been argued that different kinds of synchronization do assist interacting partners during conversation by establishing understanding, cooperation and empathy (e.g. Bernieri, 1988; laFrance, 1985; Neuman & Strack, 2000; Semin, 2000; in Barsalou et al., 2003).

Thirdly, one’s own bodily states produce and trigger affective states, means embodiment not only functions as a response to a social stimulus but also constitutes a tentative stimulus in itself. Hence, this is the reverse of the first identified social embodiment effect. Barsalou et al. (2003) distinguish between bodily and facial elicitation. Evidence of bodily elicitation presented by Stepper and Strack (1993, in Barsalou et al., 2003), for instance, shows how bodily positions or postures actually influence the subjects’ affective state. Subjects were induced into either an upright or slump position during a cover story about measuring ‘ergonomic positions’. During the experiment all the subjects performed an achievement test, and received false feedback that they had done well on the assignment. Afterwards the subjects were asked to rate their feelings of ‘pride’ at the time, and those that had been in an upright position experienced more pride than those who had been in a slump position. Hence, the subjects’ posture actually influenced their affective states. Riskind and Gotay (1982, in Barsalou et al., 2003) have reported similar results. As Barsalou et al., point out, facial elicitation, on the other hand, is well documented in the literature and is often referred to as facial feedback (cf. e.g. Adelmann & Zajonc, 1989; in Barsalou et al., 2003). Despite the fact that the accounts of these effects differ, it can be noted that many studies demonstrate that the shaping of the face into an emotional expression, for instance a smile, tends to produce the corresponding affective state, namely ‘happiness’. Strack, Martin and Stepper (1988, in Barsalou et al., 2003) demonstrated that subjects holding a pen in their lips or teeth, similar to smoking a cigar, rate cartoons differently in conjunction with their holding of the pen. Holding the pen by the teeth seems to trigger the musculature of smiling, whereas holding the pen between the lips tends to activate the frowning musculature. The result indicates that the subjects who held the pen with the teeth rated the
cartoons funnier than the others who held the pen between their lips. Barsalou et al. argue these results are consistent with the social embodiment hypothesis, since the expressions associated with the musculature affected the evaluation of the cartoon, despite the fact that the subjects were not aware their musculature had been manipulated into a facial expression.

Finally, the compatibility of bodily states and cognitive states adjusts performance effectiveness. This final social embodiment effect concerns more complex relationships in cognitive behavior. Barsalou et al. note that when embodied and cognitive states are compatible, the cognitive work proceeds more easily. On the other hand, when they are incompatible, the cognitive performance is less efficient. They argue that these relations demonstrate the importance of the interaction between the body and higher cognition, further suggesting that advanced cognition utilizes embodied representations, as a consequence of the interference between the two of them. For instance, several motor performance compatibility effects are reported, and generally embodiment interacted with the task that the subjects performed. For example, Chen and Bargh (1999, in Barsalou et al., 2003) demonstrated that subjects actually respond faster to ‘positive’ words than ‘negative’ ones, when pulling a level towards themselves instead of pushing it away. This means, motor performance is actually more optimal when compatible with cognitive processes. According to Barsalou et al. (2003) similar results are reported for memory tasks (see e.g. Förster & Strack, 1997, 1998; Laird et al., 1982; in Barsalou et al., 2003), face recognition (see Zajonc, Pietromonaco & Bargh, 1982; in Barsalou et al., 2003), facial categorization (see Niedenthal et al., 2001; Wallbott, 1991; in Barsalou et al., 2003), word recognition (see Förster & Strack, 1996; in Barsalou et al., 2003), reasoning (see Riskind, 1984; in Barsalou et al., 2003), and secondary task performance (see Förster & Strack, 1996; in Barsalou et al., 2003). Broadly speaking, Barsalou et al. stress that the compatibility between embodiment and cognition is also present for non-social stimuli, supporting a broader pattern of embodiment-cognition compatibility. They conclude that common mechanisms which produce this compatibility effect across different domains exist. Additionally, embodiment seems to be central to cognitive processing, in view of the fact that bodily states interact extensively with cognitive states. Thus, they conclude that embodied interaction is ubiquitous in human cognition.

4.1.2 Social embodiment effects in attitudes, social perception and emotions

In more recent work, some of the abovementioned researchers focus explicitly on traditional conceptions in social psychology, such as attitudes, social perception, and emotions (Niedenthal et al., 2005a). Generally speaking, they suggest that social-information processing involves embodiment, with which they refer to both “actual bodily states and to simulations of experience in the brain’s modality-specific systems for perception, action, and introspection” (Niedenthal et al., 2005a, p. 184). Moreover, Niedenthal et al. (2005a) differentiate between peripheral (body-based) and central (modality-based) senses of embodiment, and they particularly focus on the latter, in which they include the sensory-,
and introspective systems that respectively underlie perception of a current situation, action execution, and conscious experiences of emotions, motivation, and cognitive operations. Additionally, following Wilson (2002; see subsection 3.2.1), they address these topics from online (i.e., perceiver interacts with actual social objects, e.g., mimicking a happy facial expression) and offline (i.e., perceiver represents social objects in their absence, e.g., understanding the concept happiness or recalling a happy experience) perspectives on embodied cognition. Niedenthal et al., (2005) argue that the distinction between online vs. offline is helpful in systematizing the findings within social psychology, and besides, it can function as a way to conceptualize the acquisition and the use of knowledge, as well as hopefully recognizing similarities between their underlying embodied mechanisms. It should be noted that the review by Niedenthal et al. (2005a) that follows is by no means exhaustive, but emphasizes representative empirical findings for each category.

- **Embodiment of attitudes**
- **Embodiment of social perception**
- **Embodiment of emotions**

Firstly, **Embodiment of attitudes**, as Niedenthal et al. (2005a) pointed out, Darwin (see section 2.2) already addressed attitudes from an ‘embodied’ perspective, claiming that posture expresses an organism’s affective response towards an observed object. Nevertheless, for several reasons, the bodily aspects of attitudes have not been considered in mainstream research within social psychology. Concerning online embodiment in the acquisition and processing of attitudes, Niedenthal et al. (2005a) emphasize that empirical studies show that bodily postures and motoric activities, such as nodding heads (in agreement) or shaking heads (in disagreement) are related with positive or negative preferences and action predispositions toward objects (cf. Wells & Petty, 1980, in Niedenthal et al., 2005a). Similarly, Tom et al. (1991, in Niedenthal et al., 2005a) studied how such head movements influence the attitudes towards a pen placed on the table in front of the participants during the testing of head phones. Afterwards, a naïve experimenter offered the ‘old’ pen that had been placed on the table during the experiment or a ‘new’ pen the subjects had not seen before. Depending on the performed head movements, i.e., nodding in agreement or shaking in disagreement, the participants favored the pen that correlated to the developed attitude. In other words, the nodding participants chose the ‘old’ pen, whereas the head-shaking participant preferred the ‘new’ one.

Regarding offline aspects of attitudes, as Barsalou et al. (2003) argue, conceptual processing is maximally efficient when the conceptual information is consistent with embodiment effects (see the fourth social embodiment effect in subsection 4.1.1, i.e., compatibility of bodily states and cognitive states adjusts performance effectiveness). It is reported that Förster and Strack (1997, 1998, in Niedenthal et al., 2005a) revealed a similar effect in a long-term memory task. In their study, participants had to generate the names of famous persons, and then classify the celebrities according to
whether they liked, disliked or were neutral about them. During the generating names phase (e.g. ‘Jane Fonda’ or ‘Clint Eastwood’), participants were instructed to either place their hands beneath the table and pushed upward (inclining an approach behavior) or on top of it and pushed downward (inclining an avoidance behavior). As a result, the participants directed to conduct an approach behavior named more celebrities they liked, whereas those that performed an avoidance behavior named more they disliked. This means, attitudes appear to be partly influenced by embodied states, whether the performance is carried out online or offline.

Secondly, with regard to embodiment of social perception, the concept social perception refers to impression formation in a social situation (Niedenthal et al., 2005a), and subsequently, the previously addressed findings of mimicry and imitation in the literature can be viewed as online embodiments of social perception (see subsection 4.1.1). One example is the previously reported finding by O’Toole and Dubin (1968), which demonstrated that mothers open their mouths in response to their infants’ mouth opening during feeding. It is important to consider, however, that imitation goes beyond mere facial expressions, for example individuals engaged in conversations tend to synchronize their speech rate, utterance duration, prosody, postures, manual gestures, and other characteristics of social interaction (cf. Niedenthal et al., 2005a). Moreover, it is argued that such conversational synchrony facilitates cooperation and empathy between interacting partners, which is confirmed by empirical findings (see Chartrand & Bargh, 1999, in Niedenthal et al., 2005a).

Proceeding to offline aspects of social perception, Niedenthal et al. (2005a) refer to the previously presented study by Bargh, Chen and Burrow (1996), who demonstrated that participants primed with words related to stereotypes of elderly (e.g. ‘bingo’, ‘gray’) walked more slowly from the laboratory to the elevator than the control group primed with neutral words. It should be noted that the words used to prime the elderly stereotype were not associated with motoric functions. Apparently, these abstract concepts referring to elderly stereotypes activated the participants’ actual walking behavior, resulting in a slow walking embodiment effect. Furthermore, Andersen, Reznik, and Manzella (1996, in Niedenthal et al., 2005a) created descriptions of fictional characters, based on personality descriptions of significant others the participants liked in their ordinary lives. Later, in the experimental situation, while the participants read the descriptions of the fictional characters, they tended to display positive facial expressions. When the participants instead read descriptions of the fictional characters based on persons they disliked, they were inclined to show negative facial expressions. In short, these studies demonstrate that such an abstract activity as merely reading about social stimuli influences the subjects’ actual facial expressions.

Thirdly, regarding embodiment of emotions, Niedenthal et al. (2005a) highlighted that the linking between embodiment and emotions can actually be traced back to William James (see section 2.2), who, generally speaking,
argued that the basis of emotion is the bodily activity that occurs in response to an emotional stimulus (see also Damasio, 1999). Returning to Niedenthal’s et al. (2005a) position, embodiment is essentially engaged in information processing about emotion, online, and most notably, also offline, in which humans represent the emotional meaning in abstract entities such as words.

Niedenthal et al. (2005a) stress that besides the ubiquity of embodied responses to what they call non-emotional actions and movements, there is accumulating evidence that humans also mimic others’ emotional facial expressions (see Busch et al., 1989; Dimberg, 1982; in Niedenthal et al., 2005a). Bavelas et al. (1986, in Niedenthal et al., 2005a), for instance, reported that when a co-researcher of theirs actually participated within their own experimental situation and deliberately faked an injury, grimacing in pain, the observing participants then also grimaced. The extent of the participants’ grimaces correlated with how clearly they could see the confederate’s face. This means, emotion imitation seems to occur automatically without any higher mediating conscious awareness. Moreover, it has been demonstrated that participants react to subliminal happy and angry expressions with minor smiles and frowns (see Dimberg, Thunberg & Elmedal, 2000, in Niedenthal et al., 2005a). Additional findings also suggest embodied effects of subliminal facial expressions beyond facial mimicry, and participants subliminally exposed to happy or angry faces were subsequently asked to try a novel beverage. The result demonstrated that there was a correlation between the like vs. dislike of the beverage (drinking much beverage vs. drinking less beverage) and the subliminal exposure of happy vs. angry faces. Similarly, Wallbott (1991, in Niedenthal et al., 2005a) asked participants to categorize emotional facial expressions in photographs of humans, and as the participants conducted their task, their own faces were videotaped. The recording demonstrates that the participants tended to mimic the facial expressions during classification, and there was a correlation between the accuracy of their classification and the magnitude of mimicry. In other words, bodily feedback from facial mimicry plays an important role in the ability to process emotional facial expression.

Proceeding to offline embodiment in emotion, Riskind (1984, in Niedenthal et al., 2005a), for instance, demonstrated that participants’ retrieval of pleasant or unpleasant autobiographical memories was influenced by the manipulation of facial expressions and postures. As a result, adopting an erect posture and also smiling hastened the retrieval of pleasant autobiographical memories, compared to the speed of retrieving unpleasant memories. Similarly, the previously described study by Stepper and Strack (1993) in subsection 4.1.1, concerning emotionally responding to self-evaluation of performance, demonstrates a similar correlation effect between offline embodiment of emotion and bodily posture.

Furthermore, Niedenthal et al. (2005a) present the work of Zajonc and Markus, in which already in 1984, they argue that social psychologists should consider embodiment in the interplay between emotion and
cognition. According to Zajonc and Markus’ hard interface theory (HIT), bodily states already constitute “hard representations” of social knowledge, and they suggest that the bodily movement itself has a representational content and does not require a more cognitive or “mental” representation in order to possess representational value (cf. Zajonc & Markus, 1984, in Niedenthal et al., 2005a). Niedenthal et al. (2005a) used the ‘chewing gum’ experiment in order to illustrate HIT, in which subjects were asked to study 78 photographs of human faces. One group was instructed to mimic, and imitate the facial expressions, as well as the head and gaze orientation of the persons depicted in the photographs. While another group was blocked to mimicry through chewing gum, a third was required to squeeze a sponge with their non-preferred hand, as a way to manage motor control. Finally, a fourth group had to judge the head orientations and the facial expressions of the persons in the photographs. Following this first part, the subject received a face-recognition test. In accordance with the prediction of the embodiment view, memory performance was best in the mimicry group (73%), whereas the chewing gum group’s score was worst (59%). The other two groups’ score were in between. The results indicate that the subjects’ imitation of the depicted facial expressions, function as ‘muscular’ representations in themselves, which enhanced memory through “consistent elaboration”. As pointed out by Niedenthal et al. (2005a), this might explain why many people have such a good memory for faces, because humans unconsciously construct hard representations of others’ faces that complement more ‘mental’ representations. In addition, in HIT, the concept ‘hard representation’ does not only entail the muscular system, but also the gastrointestinal, glandular, and cardiovascular systems. In other words, the ‘gut’ feeling of emotion and affect and the muscular system function ‘by themselves’ without any transition to a higher and ‘soft’ representational level (cf. Zajonc & Markus, 1984, in Niedenthal et al., 2005a).

In Niedenthal et al. (2005b) the focus is particularly on emotion knowledge and embodiment, and it is stated that research on emotion has been conducted from several different perspectives during the years. However, they suggest that emotion should be regarded from an embodied perspective, and the argument follows the same line as in their earlier writings presented above. They also differentiate between bodily states which are often unconscious are embodiments of emotion and feeling states that are conscious (cf. Damasio, 2003). As Niedenthal et al. (2005b) explain, the general debate about whether emotion is conscious or not becomes more reasonable if embodiments of emotion can be considered “unconscious until consciously attended to and manifested as feelings” (ibid., p. 23). They emphasize that this distinction is useful in the interpretation of a number of findings that would at first glance appear to be inconsistent with the embodied perspective. They present empirical evidence which demonstrates that people are to varying extent consciously aware of their embodied emotional experiences, since humans tend to account for their beliefs of embodied states of emotion rather than an accurate readout of these states. By separating the notions of bodily and feeling states, they suggest that such biases are the norm and consistent with other empirical findings of
emotional experiences. In short, empirical studies show that bodily posture, facial and vocal expressions have “emotion-specific, facilitative effects on self-reported emotional feelings, as well as effects on other measures of emotional experience” (ibid., p. 30). Moreover, they argue that these embodiments “not only modulate ongoing emotional experience, but also facilitate the generation of the corresponding emotions” (ibid., p. 30). Summing up, the empirical findings presented in these subsections (4.1.1-4.1.2), as well as other studies, provide significant evidence for embodiment effects in social-emotional information processing that support crucial cognitive and social functions.

Despite the fact that the above presented social phenomena involve embodiment in various degrees, no unified theory linking them together exists yet according to Barsalou et al. (2003) and Niedenthal et al. (2005a, 2005b). They note that traditionally, embodiment is regarded as an appendage to these phenomena, but argue that embodiment as such is the core property of social representations. Furthermore, they offer a framework of modality-specific reenactments of perception, action, and introspection. This process has two phases, namely storage of modality specific states, and a partial reenactment of these states at a later occasion, through modal simulation processes in different simulators (cf. Barsalou et al. 2003). A simulator integrates information across a category’s instance, faces for example, and a simulation denotes a specific conceptualization of a certain category, e.g., the face of the king of Sweden. Additionally, they argue that humans develop “deep-rooted knowledge” about recurrently experienced situations, which they call ‘situated conceptualizations’. They suggest that this particular ‘knowledge’ is represented as “modality-specific simulations” of ‘situational components’ for the appropriate human beings, objects, and actions, for example. When some situational component in turn activates a ‘situated conceptualization’, an inferential process of the current situation arises via pattern completion. The cues that initiate those situated conceptualizations might be embodied states, which also may function as an inferential mechanism, triggering a previously experienced conceptualization. If there is a matching of current and previous embodied states, the processing of information will be facilitated (cf. Barsalou, 1999). In Niedenthal et al. (2005a), however, there are some elaborations of the nature of simulation. For instance, in the original account of Barsalou’s (1999) theory of perceptual symbols system (PSS), it was required that the bodily states initially be encoded into neurological representations in modality-specific systems that subsequently allowed bodily states to be involved in cognitive processing. In following elaborations of PSS, as in Barsalou et al. (2003), bodily states are themselves central to the situated conceptualizations underlying ‘higher’ cognitive functions – and not only the simulations of these particular bodily states (Niedenthal et al., 2005a).

However, it seems that they altogether (Barsalou et al., 2003; Niedenthal et al., 2005a, 2005b) fail to clearly explain how this simulator actually works, and it can be argued that their theory is a mix between ‘traditional’ representations and mental imagery, since the ‘situated conceptualization’
can be viewed as a kind of symbol, having similarities with the classical representation of frames. In their cases, it could be said that their theoretical framework adds an embodied icing to the traditional information-processing cake in the sense that it acknowledges embodiment by taking as its starting point perceptual/modal representations. Yet they continue to explain cognition largely in terms of internal representations and the computational processes manipulating them. Therefore, it can be argued that their framework of simulation is a ‘weaker’ version of classical representational theories, in view of the fact that they put them into an “embodied outfit” - which is not grounded in human embodiment, but rather places embodied constraints on our information processes, to use Clark’s (1999a) notion of weak embodiment (see section 3.2). Nevertheless, their attempts can be regarded as an important move in order to highlight the importance of perceptual aspects in cognitive processing, but their simulation theory needs to be refined and elaborated further in order to account for the underlying mechanisms of social embodiment. Moreover, Niedenthal et al., (2005b) strongly suggest that the embodiment of simulation of others’ emotions actually provides the meaning of the perceived event, but this claim presupposes that the person in question actually knows and understands the underlying meaning of the emotional embodied experience. This significant issue is not further elaborated, but I return to it once more in chapter 5, when discussing how this emotional meaning is supposed to emerge and develop through social interaction with others (see subsections 5.5.1-5.5.3).

4.2 Phenomenological Experience and Neurological Underpinnings of Embodiment

Humans and many other animals are essentially social beings, and Blakemore, Winston and Frith (2004), for example, argue that much of the brain must have evolved to handle social communication and interaction. Current findings in social and cognitive neuroscience provide strong evidence for this embodied interpretation of social phenomena. The interest in social cognitive neuroscience, the empirical study of the neural mechanisms underlying social cognitive processes, has increased rapidly in recent years. For instance, simulation theories of mind-reading, as well as work on so-called mirror-neurons as the neurobiological underpinnings of social interaction are good examples of more ‘radically’ (see Clark, 1999a, in section 3.2) embodied views of social cognition.

Already in the 1940s, Merleau-Ponty (see subsection 2.4.2) argued that intercorporeality constitutes the basis of intersubjectivity and social interaction. Intercorporeality means that one’s own lived body is the locus of intersubjective experience, which is based on the close intermodal links between visual perception, kinaesthetic-proprioception, and motor behavior (e.g. Gallagher, 2007a). In recent times, Dautenhahn (1997), for instance, emphasize that humans have the ability to read ‘social signs’, emotions and

---

21 Parts of this section have been published in Svensson, Lindblom & Ziemke (2007), but here they are elaborated further.
states of another agent’s mind, and particularly stresses that the agent’s own understanding of such interaction is more significant than the exchange of communication signals. She hypothesizes that a phenomenological dimension of social understanding might be founded in embodied mechanisms that allow biological agents (in particular humans) to read “social signs” and another agent’s mind, by simulating the other agent’s emotional stance. Hence, this can be accomplished through interactive feedback actions between the emphasizing person, who is attending another human that expresses her own experience, and the other way around. In order to receive and apprehend empathy, several conditions are involved. Firstly, there is a ‘willingness’ to put oneself in the other person’s emotional state, and then resonate with that other person, which results in an immediate recognition of the other person’s felt experience. The recognition phase is characterized as the consequence of ‘inner resonance’, which means that the empathic person actually dynamically reconstructs the experience of the observed person by simulating the experience herself. As a result, the empathic person offers a communicative expression in the form of an empathic response, which the other person then feels and experiences as being understood.

Hence, Dautenhahn actually addresses a simulation process of social understanding, stressing that the attribution of mental states takes place in an active embodied system, resulting from simulation processes in an individual body with its own ‘biographical’ past, using the agent’s own body and experiences as the point of reference. Dautenhahn also notes that embodied simulations provide a way to overcome the gap between computationalism and phenomenology, by identifying mechanisms which could mediate between the ‘inside’ and the ‘outside’ dimension of embodied cognition. In sum, she assumes that a social, cognitive mechanism of “experiential bodily understanding” which makes it possible for an agent to engage in joint communication activity exists. Broadly speaking, this implies that one’s own understanding of social interaction and another agents’ mind is more than the exchange of communication signals.

Furthermore, recent neuro-scientific findings imply that such a simulation mechanism may rely on special kinds of visio-motor neurons called mirror neurons, which become activated both when performing specific goal-directed hand (and mouth) movements and when observing or hearing about the same actions (e.g. Gallese & Goldman, 1998; Gallese, Keysers & Rizzolatti, 2004; Kohler et al., 2002; Rizzolatti, 2005; Rizzolatti et al., 2002). Since mirror neurons respond to both conditions, it has been argued that the mirror system functions as a kind of action representation, since it links ‘action’ and ‘action-perception’. Consequently, this mirroring mechanism enables the agent to understand the meaning of the observed action by embodied simulation. Moreover, it has been speculated that the mirror system might be a basic mechanism necessary for imitation and ‘mind-reading’, i.e., attributing mental states to others (e.g. Rizzolatti et al., 2002; Svensson, Lindblom & Ziemke, 2007). Taken together, the consideration of the mirror neuron system as the neurobiological underpinning of
‘intcorporeality’ and simulation theories as the basis of social interaction and mind-reading, provide significant examples of more embodied views of social cognition.

4.2.1 Embodied simulations

In recent years, more detailed accounts of how the sensorimotor structures of the brain are involved in cognition have been developed in several disciplines, often taking into account data from neurophysiological and neuroimaging studies. As Hesslow (2002) points out, this notion of bodily simulation is not entirely new. He mentions that Alexander Bain, for example, in 1896 suggested that thinking is basically a covert form of overt behavior that does not activate the body and therefore remains invisible and to external observers. Today’s simulation theories, based partly on data from neuroscience, can further clarify the possible role of simulation in cognition, and thus explain in a more concrete way than before the embodiment of cognition. These accounts show that the traditional strong division between perception and action (cf. Hurley, 1998, in section 3.3 as well as Dewey’s work in subsection 2.4.5), as well as between sensorimotor and cognitive processes, needs to be revised. A particular kind of ‘embodiment’ theories that has emerged in different contexts is the so-called emulation or simulation theories (e.g. Barsalou, Solomon & Wu, 2003; Decety, 1996; Frith & Dolan, 1996; Grush 2003, 2004; Gallese, 2004, 2005; Hesslow, 1994, 2002; Jeannerod, 1994, 2001). The basic concept is that neural structures responsible for action and/or perception are also used in the performance of various cognitive tasks.

Hesslow (2002), for instance, proposes that simulated chains of covert behavior as a kind ‘offline’ cognition can to a large extent explain cognitive behavior. More precisely, at least human brains, from a certain age, have the ability to reactivate previous perceptions and actions in the absence of any sensory input or overt movement (cf. Clark’s surrogate situations in subsection 3.2.1). For example, these simulations of actions or perceptions are through, conditioning coupled to achieve internal simulations of organism-environment interactions (Hesslow, 2002). A similar, technically more detailed account of the general idea is formulated by Grush in his emulation theory of representation (Grush 2003, 2004; see also Clark & Grush, 1999). Based on the control-theoretic concept of forward models (emulators), previously used to account for motor control (e.g. Wolpert & Kawato, 1998), Grush developed an emulation theory for several types of cognitive processes, including perception, imagery, reasoning, and language. Briefly stated, he argues that emulation circuits are able to calculate a forward mapping from control signals to the (anticipated) consequences of executing the control command. For example, in goal-directed hand movements the brain has to plan parts of the movement before it starts. To achieve a smooth and accurate movement proprioceptive and/or kinesthetic (and sometimes visual) feedback is necessary, but sensory feedback per se is

22 To which degree animals are capable of so-called “mental time travel”, i.e., recollection of specific past events or anticipation of the future, is still an open question. For a detailed discussion see Clayton, Bussey & Dickinson (2003).
too slow to affect control appropriately (Desmurget & Grafton, 2000). The “solution” is an emulator/forward model that can predict the sensory feedback resulting from executing a particular motor command. A further prediction is that the emulator circuits are achieved by the reactivation of the same sensorimotor processes used in overt action and perception (e.g. Grush, 2004; Hesslow, 2002; Jeannerod, 2001).

In short, the simulation account argues that cognitive processes are achieved by the reactivation of the same neural structures used for physically sensing, moving and manipulating the environment, but also the conceptualization and understanding of intersubjectivity and language (cf. Subsection 4.3) can be viewed as a kind of embodied simulation. Contemporary theories which emphasize simulation of social experiences in modality specific systems are, for instance, Damasio’s (1995, 1999) and Nielsen’s (2002) approaches of emotion, Barsalou’s (1999) theory of perceptual symbol systems and its followers (previously addressed in subsection 4.1.2), and Gallese’s approach (2004) of intersubjectivity. In line with this remark, it has also been suggested that simulation mechanisms to play a vital role for action recognition (e.g. Gallese, 2004, empathy (e.g. Decety & Chaminade, 2003), and even language understanding (e.g. Glenberg & Kaschak, 2002, 2003).

Thus far, this subsection portrays the general idea of embodied simulation, but the focus is now explicitly on specific approaches that address the social dimension. There are, however, differences between them, and they vary to what extent the individual needs to simulate in order to contribute to the embodied understanding of others, leading to a continuum of various degrees of embodied simulations.

On the one hand, Gallese’s (2004) theory of the shared manifold of intersubjectivity (see section 5.2 regarding intersubjectivity) propose that all kinds of interpersonal relations, such as imitation, mind-reading and empathy, depend, at a basic level, on the foundation of a shared manifold space, which then is characterized by routines of embodied simulations. Gallese (2004) addresses this issue from both an evolutionary perspective as well as from current findings in cognitive neuroscience, arguing that the ‘mental Rubicon’ between behavior-readers (non-human primates) and mind-readers (humans) could not happen by the flicking of a magic switch. Furthermore, based on findings in current neurological studies, he stresses “there is now enough empirical evidence to reject a disembodied theory of the mind as biologically implausible” (p. 166).

Gallese (2004) emphasizes that in order to simulate, the task depends on what we want to understand, given that the behavior of others is neither objectively given nor expressed in social situations. In other words, depending on our social situations, there are different kinds of interactions.

---

23 According to Blakemore, Frith and Wolpert (1999), this is also why for most people it is not so easy to tickle themselves: the forward model produces predicted sensory feedback that “prepares” the agent.
between the embodied person and its surrounding, resulting in different ways that humans can act upon the world, or simulate to do so. Take together, all these different kinds of social interaction, both online and offline, so to speak, rely on the same basic functional mechanism: embodied simulation, arguing “[o]ur brain models the behavior of others, much the same as it models our own behavior” (p. 167). He characterizes simulation as an “implicit mechanism meant to model the objects and events that the mechanism itself is supposed to control while interacting with them” (ibid., p. 168). This means, as Gallese explains, simulation is a control functional mechanism, and its function is to model the objects to be controlled.

Gallese (2004) suggests that the pretend state, i.e., the simulation that is used by the interpreter in order to understand the behavior of the agent, is not the result of a deliberate and voluntary act on the interpreter’s side. On the contrary, he advocates that it is automatic, unconscious and pre-reflexive. Furthermore, simulation is not limited to the motor system, since it actually uses various parts of the brain (but see Jacob & Jeannerod, 2005 for a critique of motor theories of social cognition). Instead, he proposes that simulation is how we always model reality, which means that what is usually denoted a representation of reality is not a copy of what is objectively given, but an interactive model of the reality that cannot be known by itself. The conceptual framework of the ‘shared manifold’ stresses that the organization of self-other identity is the major force for the cognitive development of more articulated and advanced forms of interpersonal relationships, allowing us to understand others. This shared manifold comprises all the dimensions of social cognition, and is described at three different levels (Gallese, 2004).

i) phenomenological level: being a part of our society, and ‘empathy’ in its broad sense;
ii) functional level: ‘as if’ modes of interactions enable models of self/other; and
iii) sub-personal level: the level of activity in the neurons.

It should be noted that the shared manifold of intersubjectivity does not entail our experiencing others as we experience ourselves, since it only enables and bootstraps mutual simplicity. In other words, the role of embodied simulation is to model the interaction between an individual and its environment, both online and offline, and then interpersonal relations are then automatically established by means of simulations.

On the other hand, Gallagher (2005) stresses that the ‘interactive practice of mind’, as he formulates it, neither needs any theory as in theory-theory (TT) nor simulation theory (ST) as in Gallese’s simulationist approach. Instead, he argues that the understanding of the other person is primarily neither theoretical nor based on an internal simulation, since it is a kind of embodied practice. It should be noted, however, that Gallagher does not deny the cases when we use the ability of theoretical interpretations or/and simulation, since these occasions are, according to him, rather rare in
proportion to the majority of our interactions. This means, TT and ST at best explain some narrow and specialized set of cognitive processes, which only sometimes are used in social interactions. Indeed, he advocates that in the cases when we lean more on advanced strategies, they are already shaped by primary embodied practices (Gallagher, 2005). The major problem, according to him, is the assumption that communicative interaction between two people is a process that takes place between two ‘Cartesian minds’. By ‘Cartesian minds’, Gallagher refers to the view which requires that one’s understanding usually involves a retreat to a realm of ‘theoria’ or ‘simulacra’ into a set of internal operations that becomes decoded and externalized in another modality such as speech, gesture, or action. That is, there is always some kind of higher level processing being carried out in cognition. In contrast, he argues that communicative interaction is accomplished in the very action of pragmatic embodied interaction, through the expressive movement of speech, gesture, and the environmental and contextual factors of interaction itself. Therefore, the idea that the understanding of another person involves an attempt to theorize an unseen belief or simulate in mind-reading is challenged. Instead, he proposes that only when our ‘second-person pragmatic interactions’ or our evaluative attempts to understand others break down do we choose to use more specialized practices of third-person explanation and prediction, i.e., simulation as such is mostly carried out offline. However, it is of major importance to be aware of the different perspectives in these situations. On the one hand, human interaction usually unfolds fluently, with ourselves an integrated part of that actual situation (first-hand perspective), or on the other hand, when we try to evaluate what happens through embedded reflections on possible actions rather than leaning on detached considerations of mental state (third-hand perspective). This means, in order to interpret and understand other people in real-time interaction, Gallagher (2005) suggests that humans seldom need to move beyond the present embodied and expressive actions at hand in order to grasp and gain an understanding of the other person.

In order to support this claim, Gallagher (2005, 2007b) reinterprets the very idea of simulation, by addressing the problem of the separation of the processes of perceiving and simulating in the common accounts of simulation. The common view of simulation stresses that the simulation is secondary to and distinguishable from the perception. Instead, he proposes an alternative interpretation of this neural activity, namely as part of the neural processes that underlie intersubjective perception (Gallagher, 2007b). According to him, this claim requires that perception is conceived of as a temporal enactive sensory-motor phenomenon (cf. Hurley, 1998, discussed in section 3.3). In this regard, there is not any discrete process that involves perception plus simulation, but rather a direct intersubjective perception of what the other is doing (Gallagher, 2007b). He argues that phenomenologically, when one sees another person’s action or gesture, one directly perceives or immediately ‘sees’ the meaning in the action/ gesture, without the need to simulate it. He presents (2005) brain-imaging studies, in which subjects were asked to simulate their own movements (first-person perspective) or another person’s (third-person perspective) movement. The
result shows that there is no additional brain activity in favor of an extra level or effort as a kind of simulation, meaning there is no evidence for viewing simulation as an ‘extra’ step (cf. e.g. Barsalou, 1999; Dautenhahn, 1997; Gallese, 2004) over and above the perception. Indeed, Gallagher’s point is “that there is no evidence that perception and simulation are two separate systems. Rather, one could say, in effect, perception of action is already an understanding of the action; there is no extra step involved that could count as a separate simulation routine” (ibid., p. 223, and note the close connection to Dewey’s work on the reflex arc in section 2.2 as well as ‘Body in motion’ in section 3.3). In other words, the neurological underpinnings of what could count as simulation are part and parcel of the (re-)activations that correspond to the original perception from an embodied pragmatic perspective (Gallagher, 2005). This poses another problem, however, namely where to draw the line between perception and other (cognitive) processes. Subsequently, the need of an internal model is questioned, and as Gallagher (2005) explains, “[t]he required model is the action of the other, and it is already being perceived. Why would one need to ‘read off’ the meaning of an action on an internal ‘as if’ model, indirectly, when one is observing that very action performed by the other?” (ibid., p. 224). Furthermore, according to Gallagher (2007b), proponents of simulation theory stress that simulation involves the instrumental use of a first-person model to form third-person “as if” or “pretend” mental states, but he argues that this is not a possible explanation. He argues that we cannot control these re-active sensorimotor processes at a personal level, and for that reason we cannot use them as a model. Similarly, another proposed idea that the brain itself, at a subpersonal level, is using these reactivations as a model (cf. Damasio, 1999, 2003), which does not make sense either according to Gallagher (2007b). Thus, his major point is that “the neural systems neither activate themselves nor take the initiative, but are activated by the other person’s action”. Thus, “the other person has an effect on us. The other elicits this activation... It is not us (or our brain) doing it, but the other who does it to us” (ibid., p. 8-9).

However, the question is whether the proponents of embodied simulation actually stress that embodied simulation really involves a ‘as if’ model or if they only are sloppy with words? A possible, and perhaps better, account of how to consider the term embodied simulation, as a way of not indicating an ‘as if’ model (as in the cases of Barsalou (1999), Dautenhahn (1997), and Gallese, (2004)), can be found in Svensson (2007), namely as a re-activation mechanism. He states “a central mechanism of cognition is the simulation of bodily states or, more specifically the re-activation of neural circuitry also active in bodily perception, action, and emotion” (Svensson, 2007, p. 1). This is in line with Gallagher’s claim that what many proponents of simulation theory denote as ‘simulation’ is not simulation in any genuine sense of the word. However, Gallagher’s (2007b) criticism mostly concerns simulation as an implicit and/or explicit process going on during online cognition, but he does not further elaborate the issue concerning simulation as representations or models in offline cognition. According to Svensson (2007), one of the important factors to understand the embodiment of higher-level
cognition, is to consider embodied simulations as offline representations, which presumably are accomplished through the “sharing” of neural mechanisms between sensorimotor processes and higher-level processes. In other words, embodied simulations can be considered as representations, but in a restricted sense, which Svensson refers to as offline representations (but see also Clark’s (2005) surrogate situations in subsection 3.2.1). This means, embodied simulation processes can function as offline representations, given that something must be “standing-in-for” the issue not present at hand. In the case of offline cognition, through embodied simulation, the stand-in relation serves “as reactivations of sensorimotor processes, that is, roughly speaking, the internal replication of agent-environment interactions” (Svensson, 2007, p. 99).

Generally speaking, when comparing Gallese’s and Gallagher’s approaches concerning embodied simulation in social interaction, it can be argued that Gallagher is more radical than Gallese. According to Gallagher, there is no need for an extra step or a implicit simulation as a kind of ‘as if’ model in online social interaction, as proposed by Gallese, since the embodied practice in itself provides the primary foundation of the social understanding. This means, the body serves as the basic medium for understanding others from a first-hand perspective, and only in more complex situations (e.g. breakdowns, evaluation, offline interaction and cognition), do humans lean on embodied re-active ‘simulation, which in turn are based on prior online embodied practices. Gallese, in contrast, stresses that we do implicitly simulate in all instances of social interaction, both online and offline. Indeed, Gallagher’s account is more parsimonious from an embodied perspective, since he strongly emphasizes the role of online embodied practice in the process of understanding others, and not relying on any model, ‘stand in’ or surrogate situatedness completely all the time.

In addition, Niedenthal et al., (2005a) also address the problem of the representational capacity of the body, since several critics argue that firstly, bodily responses and feedback are too slow and undifferentiated to serve as the underpinning of experience. A second problematic situation is that the same bodily state may be associated with several different cognitive and emotional ‘representations’. However, they present some possible responses to the “body-is-to-crude-too-slow-and-too-varied”, advocating that the motor system actually is able to support very subtle distinctions, such as the sensorimotor processing for generating spoken utterances in human language. Furthermore, although there is a limited number of bodily states, their combinations can support a vast variety of ‘representational’ differences. Finally, by focusing on the modality-specific systems represented in the brain instead of actual muscular- and sensory processes, they avoid the major criticism, arguing that the circuits in the brain are fast and flexible enough to handle a large amount of bodily states at the same time (Niedenthal et al., 2005a).

In summary, with regard to whether these bodily-based processes are carried out online or offline, or to the extent of embodied simulations and/or
re-activations, these approaches imply that bodily states are involved in social interaction and cognition, and may constitute the very foundations of the particular social cognitive phenomena in question. Furthermore, such a mechanism may rely on, or be a part of special kinds of visio-motor neurons in the premotor cortex in macaque monkeys, namely mirror neurons, which exemplify how perception, action, social cognition, and even language come together at the level of single neurons.

4.2.2 Mirror neurons and the action-perception linkage

Mirror neurons provide a key example of sensorimotor brain structures also involved in (social) cognitive processes. Although different hypotheses exist, many theories regarding the function of mirror neurons emphasize their role in social cognition (cf. e.g. Gallagher, 2005; Gallese & Goldman, 1998; Gallese, Keysers & Rizzolatti, 2004; Rizzolatti & Arbib, 1998; Rizzolatti et al., 2002). Mirror neurons are certain kinds of neurons that become activated both when performing a specific action and when observing the same goal-directed movements of an experimenter (cf. e.g. di Pellegrino et al., 1992; Rizzolatti et al., 1996; Rizzolatti et al., 2002). Originally, mirror neurons were described in the premotor cortex (area F5) of the monkey brain, but later observations demonstrate that they are also represented in the inferior parietal lobule (Rizzolatti, 2005). Although the existence of mirror neurons has so far been experimentally confirmed in single-neuron studies of monkeys, there is strong empirical evidence from functional imaging studies of the human brain that such a system is actually present in human beings as well, since these studies revealed activation of the likely human homologue to area F5 during action observation (e.g. Arbib 2005; Gallese et al., 2002; Rizzolatti, 2005). Moreover, EEG and magnetoencephalography studies have demonstrated activation of motor cortex while observing finger movements (cf. e.g. Rizzolatti, 2005).

There are, however, various interpretations of the functional role of the mirror neurons. On the one hand, it is argued that the mirror neuron system can be interpreted as a kind of observation-execution mechanism or resonance mechanism, which links the observed actions to actual actions of the subject’s own behavioral repertoire. That is, it enables the agent to understand the meaning of the observed action by either activating (cf. Gallagher, 2005) or simulating (e.g. Gallese & Goldman, 2002; Rizzolatti et al., 1996; Rizzolatti & Arbib, 1998) the observed action through its own sensorimotor processes. This means, mirror neurons can be interpreted as representations of actions, used both for performing and understanding actions. On the other hand, it is also suggested that the function of this matching system might be a part of, or a precursor to, a general mind-reading capability that allows one to adopt the point of view of other conspecifics by matching or simulating their mental states with a resonant state of one’s own, i.e. putting oneself in another’s ‘shoes’ in order to understand or predict mental states and behavior. In addition, it is proposed

24 Single neuron recordings of the type used in these experiments eventually destroy the neurons recorded from, and thus for ethical reasons cannot be used in humans.
that the mirror neuron system is the neurological underpinning from which language developed (e.g. Arbib, 2005; Rizzolatti & Arbib, 1998).

According to Rizzolatti (2005), the question what is the function of the mirror neurons or the mirror neuron system is ill posed, since they may not have any particular functional role at all. Instead, the properties of mirror neurons indicate that the primate brain, and perhaps also other mammal brains, is endowed with a mechanism able to map illustrative descriptions of actions, performed in the higher visual areas onto their corresponding motor part. In other words, as Rizzolatti explains, this matching mechanism may underlie a variety of functions, which depend on what aspects are considered, such as particular circuits of mirror neurons, connections to other systems, aspects of the observed action, and the species in question.

As Rizzolatti (2005) points out, it might sound odd that in order to recognize an action, one should activate the motor system, but there are empirical findings which support this general idea. For example, it has been found that observers undertake motor facilitation in the same muscles as the one used by the observed individual (cf. Fadiga et al., 1995). This means, even while only observing the actions of another individual, a neural ‘triggering’ event in fact takes place in the observer. This means, the linking between action and perception offers an ‘intuitive’ understanding of the observed action, i.e., what it means to do it and what the action really is about, compared to when presented in mere visual aspects. Thus, this linkage, which is present in both humans and monkeys, enables ‘real’ action understanding of the observed action. With this action understanding as the basic foundation, other functions can further develop, some of which are found only in humans, such as imitation (Rizzolatti et al., 2002). Despite the fact that monkeys have mirror neurons, they lack the ability of ‘true’ imitation or have this capacity in a very limited form (Visalberghi & Fragaszy, 2002). Consequently, a related question is whether the mirror neuron system is involved in imitation or not, and if the answer is yes, then why do not monkeys imitate?

According to Rizzolatti (2005), there is evidence that the mirror neuron system is involved in direct repetition of actions (e.g. Iacoboni, et al., 1999) as well as in imitation learning (e.g. Buccino et al., 2004). However, there is a significant difference between the characteristics of the mirror neuron system in monkeys and humans. In monkeys it responds during the observation of goal directed actions, whereas in humans, it is also activated by intransitive, meaningless actions (cf. Fadiga et al., 1995). According to Rizzolatti (2005) the monkey mirror system seems to be tuned to portray the goal of actions, i.e., is able to carry out the goal of the action. However, it is unable to grasp the details in the way in which the goal is accomplished, and subsequently cannot replicate the observed action. In other words, there is a continuum of action-understanding in primates, which ranges from response facilitation, emulation and ‘true’ imitation (see Tomasello, 1999 for an overview of different kinds of imitation). Furthermore, current brain imaging experiments demonstrate that the prefrontal lobe plays an important role in
imitation learning and this lobe area appears to have the function of combining elementary motor acts, e.g., specific movements of the fingers, into more advanced motor patterns (Buccino et al., 2004). In view of the fact that humans have a large expansion of the frontal lobe, it might be the case that the frontal lobe of monkeys is too limited to produce the more advanced activity pattern.

The major advantages that can be gained from observing the actions of another individual concern both ‘what’ the actor is doing as well as ‘why’ the actor is doing it, where the latter aspect offers an opportunity to grasp the intention of the observed action (Rizzolatti, 2005). Iacoboni et al. (2005), for instance, investigated the understanding of intentions of others while watching their actions in different conditions. Human subjects were exposed to three different stimuli; grasping hand actions without a context, context only, and in two different contexts (either ‘drinking – to have tea’ or ‘cleaning - after having tea’). The fMRI data shows that actions embedded in contexts generated a significant increase of activity in the pre-motor mirror neuron areas of the brain, indicating that the mirror neuron system is also involved in grasping the intention of others automatically. This means, the role of the mirror neuron system seems to be more complex than mere action-recognition, otherwise a similar response should have been displayed while watching grasping actions regardless of whether the context of the observed action was present or not. Furthermore, there were different activations between the ‘drinking’ and ‘cleaning’ contexts, which imply there are certain neurons in the human inferior frontal cortex that particularly ‘code’ the why aspect of the action. Consequently, this kind of mirror neurons is suggested to “discharge in response to the motor acts that are most likely to follow the observed one” (Iacoboni et al., 2005, p. 533). They suggest, in addition to ‘classical’ mirror neurons, there are neurons which are visually ‘triggered’ by a certain motor act, such as grasping, but do not ‘discharge’ during the execution of the same motor act. Instead, they discharge from another act functionally related to the observed act, such as bringing a glass to the mouth as in ‘drinking’. These neurons are referred to as logically related neurons, and, although they have been reported previously their functional role has not been discussed extensively. Iacobani and colleagues suggest that these neurons are part of a neurological chain that codes others’ intentions. The triggering factor of these neurons seems to be a specific context in which an action is usually followed by a particular successive motor act. As they explain, “observing an action carried out in a specific context recalls the chain of motor acts that typically is carried out in that context to actively achieve a goal” (p. 533). In other words, these neurons might be ‘a step ahead’ of the present moment, and may therefore be a more parsimonious account to Gallese’s simulation theory, but more in accordance with Gallagher’s embodied practice.

These logically related neurons could be the basic foundation of the proposal that the mirror neuron system might be a part of, or a precursor for “mind-reading”, i.e., attributing mental states to others. In particular, Gallese and Goldman (1998) argue that such mechanisms can explain how an agent
determines what mental states of another agent have already occurred. When mirror neurons are externally activated by observing a target agent executing an action (allowing the subject to evaluate the meaning of the other’s action), the subject knows (visually) that the observed target is currently performing this very action and thereby “tags” the “experienced” action as belonging to the target. This has been verified in an fMRI experiment, in which mirror areas, besides action understanding mediate the understanding of others’ intentions (cf. Iacoboni et al., 2005). In addition, recent evidence shows that the mirror mechanism is involved in empathy as well as emotions. Wicker et al. (2003), for instance, demonstrated that during the observation of a facial expression of disgust, there was neural activation in the same area of the brain that is activated when exposed to disgusting odors. This means, there are neural activations in the insula both when a person experiences disgust, as well as when the emotion of disgust is generated by the facial expression of another person. In other words, feeling/experiencing emotions function analogously with action understanding, based on the same resonance mechanism.

In accordance with this remark, we return to Gallagher’s (2005) non-simulationist view (see subsection 4.2.1), in which he stresses that mirror neurons and shared representations are not primarily the mediators of simulation but the enactment of direct intersubjective perception. He exemplifies this view in the imitation of facial expression, emphasizing that infants have no need to simulate the facial gesture internally, as an extra-step, since through actually seeing it, they already simulate it on their own faces. This means, one’s own body is already communicating with the other’s body at unconscious and perceptual levels that are sufficient for intersubjective interaction to emerge. Furthermore, brain imaging experiments with human subjects sitting still observing others moving have indicated that the mirror system seems to distinguish between biological and non-biological actions (Blakemore, Winston & Frith, 2004). It is commonly argued that another person’s action can influence one’s own actions, and Sebanz et al. (2003) revealed that the presence of another person altered the timing of the response time of a subject carrying out a spatial compatibility task. Moreover, the observation of another person’s actions has an impact on one’s own actions, and interference effects occur when there is a mismatch between one’s own actions and the observed ones (Blakemore, Winston & Frith, 2004). However, these interference effects seem to occur only for observed human actions and not while, for instance, observing a robot performing interfering actions (Kilner, Paulignan & Blakemore, 2003). Blakemore, Winston and Frith (2004) ask what is special about human biological actions and why mirror systems require biological action to be activated. Furthermore, little is known about how the subject can distinguish its own actions from those of others, given that to a large extent the same neural mechanisms underlie both action observation and own action (cf. Blakemore, Wolpert & Frith 2002). However, it might be possible to resolve this question in the future using techniques for the dual scanning of two brains, which would thus facilitate the recording of simultaneous responses of two interacting humans (Blakemore, Winston & Frith, 2004).
addition to action-recognition, mirror neurons are also considered to be involved in more complex social actions, such as gesture and language. Rizzolatti and Arbib (1998), for instance, suggest that the human communicative and linguistic capacity is a natural extension of action-recognition based on mirror neuron mechanisms. This could provide a tentative explanation of why and how the human Broca's area, involved in gesture and language processes, emerged from area F5. Arbib (2005), for instance, suggests that the mirror system provides the causal mechanism for basic intentional interaction and thus might constitute the foundation of human language. As Rizzolatti (2005) points out, however, it is obvious that the mirror neuron mechanism itself is unable to explain the whole complexity of speech and human language, but actually clarifies one of the fundamental aspects of social interaction and communication, namely how the interacting partners are able to share the communicated meaning of a dialogue. Accordingly, from discussing mirror neurons and the mirror neuron system, I now tune up the microscope and portray the role of embodiment in human language, particularly the role of gesture in communication and thinking.

4.3 Embodied Linguistics

In order to address the issue of embodiment in human language, it is necessary to consider the traditional ‘cognitive’ divide between speech and gesture in the study of linguistics. This divide has several reasons, which are both methodological and epistemological. The search for the evolutionary origin of language, for instance, encouraged interest in body movements; according to Farnell (1999), the anthropologist Mallery compared the gestures of Native Americans’ signing systems and deaf sign languages in the 1880s. The purpose was to find the original sign-making ability of humans that led to the emergence of spoken language in human ancestors (Farnell, 1995). However, this line of research waned when Darwin presented his theory of evolution. In contrast, Boas actually stressed the cultural specific nature of human movement, and studied the accompanying body movements of Native American dances, as well as in the performance of oral literature (cf. Boas, 1890, 1972). However, Boas chose to omit the relation between gesture and language in his influential introduction to the Handbook of American Indian Languages (1911). As a result of this exclusion, Boas set the stage for the future of (American) linguistic research that led to a rather narrow view of language (Farnell, 1995). Thus, it could be argued that Boas’ view was one major cause of the split between gestures and speech in language studies.

According to Knapp and Hall (1997), Boas’s protégé Edward Sapir also recognized the interplay between gestures and speech, stating that “we respond to gestures with an extreme alertness and, one might almost say, in accordance with an elaborate and secret code that is written nowhere, known to none and understood by all” (Sapir, 1928, p. 556, in Knapp & Hall, 1997). Nevertheless, for some reason, Sapir did not address this issue further in his research (Farnell, 1999). Efрон, another student of Boas, actually conducted the pioneering research on gesture and environment, and
in his classical Gesture and Environment (1942), he made three important contributions to the study of gestures. Firstly, he developed detailed and innovative methods for studying body language and gesture. Secondly, he elaborated a framework for classifying nonverbal behavior, and finally, he stressed the importance of cultural influences in shaping gesture and bodily movements. Other classical approaches to studying nonverbal behavior include Birdwhistell’s (1970) work on ‘kinesics’ (body movements) and Hall’s (1959, 1966) work on ‘proxemis’ (space) (cf. e.g. Knapp & Hall, 1997). The study of nonverbal communication increased after the Second World War, when the interest in human communication was strongly inspired by information theory and cybernetics. McNeill, citing Kendon, pointed out that “once human action was conceived of as if it were a code in an information transmission system, the question of the nature of the coding system come under scrutiny” (Kendon, 1982, in McNeill, 1992, p. 4). It should be noted that in those days, researchers distinguished sharply between digital and analog codes. While the digital code was illustrated by the linguistic system, the analog code was supposed to be present in prosody, posture, facial expression and gesture. As a result, analog codes were regarded as paralinguistic, – besides language (McNeill, 1992). Therefore, nonverbal communication was supposed to exist beside thought and cognition, following the same line of argument that the cognitivists used concerning the body. As a consequence, the study of non-verbal communication has not dealt with cognition and thinking (Farnell, 1999). However, in the beginning of the 1980s, Adam Kendon challenged this view, and was among the first who seriously considered how speech and gesture interact and convey meaning in communication. David McNeill (1992) also produced some groundbreaking work in speech and gesture, demonstrating that the hand movements while we talk are closely intertwined with speech in timing, function, and meaning (Goldin-Meadow, 2003).

4.3.1 Language as grounded in embodiment
In more recent times, the epistemological divide (i.e., verbal versus non-verbal interaction) in linguistics may be bridged from an embodied perspective. Iverson and Thelen (1999), for instance, particularly stressed that the hand and the mouth are tightly coupled in the mutual cognitive activity of language production. They advocate that “speech and gesture have their developmental origins in early hand-mouth linkages, such that as oral activities become gradually used for meaningful speech, these linkages are maintained and strengthened. In short it is the initial sensorimotor linkages of these systems that form the bases for the later cognitive interdependence “(ibid., p. 20). They propose that language and movements are very closely related in the brain, which is accomplished through a dynamic coordination between these two systems, influencing and supporting each other. Iverson and Thelen (1999) present the following empirical evidence in support of this view. Firstly, it seems that some language and motor functions share the same underlying brain mechanisms. Ojemann (1984), for example, demonstrated that there seems to be a common brain mechanism for sequential movement and speech production, which are located in the same area of the brain. Moreover, Fried et al. (1991) point out there are some
indications that the vocal tract, as well as the hands and arms are represented in closely related sites in certain brain areas. Secondly, particular brain regions normally associated with motor functions are involved in language tasks (cf. e.g. Grabowski et al., 1998; Petersen et al., 1989; Pulvermüller, et al., 1996). Classical 'language areas' (e.g. Broca’s area) become activated during motor tasks (cf. e.g. Bonda et al., 1994; Krams et al., 1998). Finally, there seems to be a close link between patterns of loss and recovery in certain motor and language functions of some types of patients. For instance, language losses in patients suffering from aphasia show a parallel dysfunction in gesturing (cf. e.g. Hill, 1998; Kimura & Archibald, 1974).

Moreover, they point out there are close connections between the oral and manual systems in the infant at birth, such as the Babkin reflex; i.e., newborn babies open their mouths when applying pressure to their palms, and gesturing has positive effects on language development in infants (cf. Goodwyn & Acredolo, 1993, 1998). To conclude, Iverson and Thelen argue that converging empirical evidence exist revealing that the systems of hand and mouth movements are not separate systems; but should rather be viewed as intimately linked in language production. From an embodied point of view, this integrated communicative “speech-language-gesture” system is a convincing proposition for a sensorimotor origin of thought and cognition, as well as of continued embodiment during life. In addition, they propose that this linkage is represented in commensurate codes, allowing a seamless meshing between different modes of information processing though not in the form of abstract symbolic manipulation.

However, some researchers have argued that conceptualization and language understanding cannot be achieved through the manipulation of amodal, arbitrary symbols alone but have to be grounded in the body as well as in bodily interaction with the environment. On the one hand, Gibbs (2006) points outs that the majority of research on brain functions and language does not accurately consider the importance of kinesthetic experiences in communication and language, although these issues mostly require actual body movement. As he explains, “this neglect has seriously undermined scientific understanding of the relations between mind and body, and, more specifically, linguistic meaning, communication, and embodiment” (p. 159). Similarly, Lakoff and Johnson (1999) suggest that moving the body might be the most common form of action, which subsequently indicates that the general structure of ‘control schemas’ for bodily movements is used to characterize aspectual structure, that is, the general structure in action and events. According to them, given that concepts of bodily movements that are represented by verbs such as ‘grasp’, ‘pull’, ‘lift’, tap’ and so on concern motor actions, it should not be surprising that the motor schemas of bodily movement structure such concepts. Consequently, their major point is that there is no genuine perceptual/conceptual distinction, since the conceptual system makes use of significant parts of the sensorimotor system. Farnell (1999) however, criticizes Lakoff and Johnson’s approach to cognitive linguistics, arguing that it continues the dualistic misconception when the
body is still disguised as an objective ‘natural organism’ which in one way or another is capable of direct correspondence with the surrounding world through experience. In order to exemplify her claim, she disapproves of the standpoint that body movements and their orientations provide experimental grounding for ‘kinesthetic image schema’. She agrees it is certainly true that the dynamics of corporeal sensations are to be found in the anchored domains of human experience. Such an approach, however, maintains dualism since it implies the moving body only provides the basic origins of abstract reason and cognition, but nothing more than the experiential grounding for spoken language and cognition. This means, instead of asking how metaphors “instantiate” image schemas, it is more appropriate to explore the lived experience from which image schemas are derived as abstract products of cognitive reflection. Briefly stated, the ordering of cause and effect should be questioned. When bodily movements are reduced to mere physical experience, the physical being and its bodily actions are denied the status of embodied forms of knowledge. In contrast, Farnell (1999), for instance, emphasizes that the numerous ‘metaphors we move by’, such as ‘grasping an idea’ or ‘life is a journey’ make use of body movements for metaphoric and metonymic purposes. Accordingly, the imaginative capacity is not simply indirectly embodied, given that linguistic metaphors, metonyms, and images are based on bodily experience’ as Lakoff (1987) claimed. Instead, as Farnell explains, “our imaginative capacity is directly embodied because action signs themselves can be imaginative tropes, some of which integrate with or are taken up in spoken language forms”.

On the other hand, Glenberg and Kaschak (2002) outline an explanation of language in accordance with the idea of cognition as body-based simulation, suggesting that language is partly achieved through the same neural structures used to plan and guide action. Under the heading of the indexical hypothesis they developed an account of language comprehension partly based on simulation of action. They argue that the meaning of a sentence is achieved by a process that indexes words to perceptual symbols, i.e., a kind of modal symbols based on records of the neural states that underlie perception (cf. Barsalou, 1999), which in turn retrieves the available affordances in the situation and determines their relevance through the particular sentence construction. Thus, the understanding of a sentence is essentially achieved through a simulation of action using the same neural systems active in overt behavior. An empirical result that supports the close coupling between language and action is the “action-sentence compatibility effect” (Glenberg & Kaschak, 2002). It was found that the sensibility of a sentence is modified by physical actions. Reaction times increased when subjects read “toward sentences” that implied action toward the reader, such as ‘open the drawer’ and had to provide the answer through an incongruent action, i.e., moving the hand away from the body. Conversely, when subjects answered through an action congruent with the sentence, reaction times decreased. It might be worth noting that Glenberg and Kaschak not only included sentences describing concrete, physical transfers, but also sentences describing cases of abstract transfer, such as "Liz told you the story" (2002, p. 560). The action-sentence compatibility effect was also
present when reading these more abstract sentences. Further evidence is derived from experiments on language comprehension and construction that are only explainable by implicating perception and action systems as predicted by the indexical hypothesis (Glenberg & Kaschak, 2003). To give but one example, Barsalou, Solomon and Wu (1998) describe an experiment which showed that presenting a modifier that potentially reveals internal features has an effect on feature listing not predicted by standard amodal approaches. The standard models predict that listing the features of half a watermelon as opposed to a whole watermelon would only differ with regard to amount, i.e., half a watermelon is smaller than a whole watermelon. The experiment illustrated, however, that subjects listed more internal features such as seeds, which can be explained if the concepts are based on perceptual symbols. Moreover, Markman and Brendl (2005) reveal that the four general social embodiment effects presented by Barsalou et al. (2003) in subsection 4.1.1 are not always tied to the subject’s body, but sometimes the actions and corresponding effects are performed in relation to a non-physical instantiation of the self (i.e., moved away from the subject’s physical body). In such cases, the mere simulation of actions, according to Markman and Brendl, is not sufficient to explain the phenomena, since actions are usually tied to the subject’s body and egocentric perspective.

Similarly, Gallese (2005) presents connections between action-understanding and the use of predicates in language. In recent studies of mirror neurons, these neurons were investigated under two conditions, namely hidden and full visual scenes. In the visual condition, the monkey was able to see the entire action, e.g., a hand-grasping movement. In the hidden condition, the same action was carried out, but its crucial and final part, i.e., the interaction with the actual object, was invisible, and the monkey merely ‘knew’ that the target object was present. The result, however, demonstrated that more than half of the mirror neurons responded in the hidden condition (e.g. Umiltà et al., 2001). This implies that the intention behind the action actually was mediated, despite the fact that the monkey did not see the actual hand-object interaction, which means that “out of sight is not out of mind” (Gallese, 2005, p. 33). He suggests that the goal of the action was still hinted at, arguing that the gap of missing visual information is filled by simulating the complete action. Furthermore, transitive actions that are accompanied by specific sounds were also studied. Certain ‘audio-visual’ mirror neurons become activated not only when executing or seeing a particular auditory action, e.g., the breaking of a peanut, but also when only hearing the sound of cracking nuts. The action of tearing paper gives similar results. This means, the audio-visual mirror neurons are able to compensate for the missing ‘information’, and still seem to interpret the actual goal of the action. In order to carry out the action goals, the monkeys’ multimodal-driven (activations) and/or simulations are conducted in neurons situated in the ventral premotor cortex of the monkey brain, which instantiates properties strikingly similar to the symbolic properties characterizing human thought (Gallese, 2005). Indeed, the action (activation)/simulation of audio-visual mirror neurons is, according to Gallese (2004, 2005) reminiscent and similar to the use of predicates in
human language. The verb ‘to break’ conveys its meaning in different contexts, such as seeing somebody breaking a peanut, hearing somebody breaking a peanut, or breaking a peanut myself. The point Gallese makes, is that the verb ‘to break’ is the same, and the predicate is neither dependent on the context to which it applies nor the subject/agent that performs the action. This means, understanding the action does not require a declarative understanding, because it is meaningful in itself constituting a fundamental level of action-understanding.

Thus, from an embodied perspective, the activation and/or simulation of the mirror neuron system might function as the glue that binds hand, mouth and language together. Although Barsalou et al. (2003) point out that gestures actually constitute another important aspect of social embodiment, they do not address this issue further, arguing that gesture and language lies beyond their scope. However, gesture is the topic of the following subsection, in which we portray the particular ways gesture remains embodied in language.

4.3.2 Gesture in language

It has been recognized that gesture\(^{25}\) is a pan-human ability tightly connected and synchronized with speech (cf. e.g. Farnell, 1995, 1999; Gallagher, 2005; Goldin-Meadow, 1999, 2003; Goodwin, 2000; Iverson & Thelen, 1999; McNeill, 1992, 2005; Roth, 2002, to appear). McNeill (1992) for example, proposed that speech and gesture form a single system of communication, grounded in a common underlying thought process, emphasizing that “[g]estures do not just reflect thought but have an impact on thought. Gestures, together with language, help constitute thought…Gestures occur, according to this way of thinking, because they are a part of the speaker’s ongoing thought process” (ibid., p. 245, original emphases). This means, language is more than spoken words, since gesture seems to have an important role in language by helping speakers convey and express ideas. In this subsection, the nature of gesture as well as its functional role in communication and thinking is further elaborated on.

The criteria for a ‘gesture’ require that (i) the generation of hand movement takes place during the communicative process of speaking, and (ii) that it is not a functional act on an object or a person (Goldin-Meadow, 2003). It should be noted, however, that emblems, such as the ‘ok’ emblem, differ from gesture because they do not depend upon speech, in view of the fact their meaning is fully conveyed in themselves without speech, as well as having a constant form within a social community, and are not generated ad

\(^{25}\) It should be noted, however, that the sign languages of deaf people are different from gesturing in accompanying speech, since they beat all the burden of communication, and therefore sign language operates like spoken language. For instance, signers do not use iconicity to a greater extent, but use arbitrary signs. Likewise, sign language is segmented at the syntactic structure, morphological structure and phonological structure, in the same way as in spoken language. Signers, however, do gesture but usually with their mouths, but not much research is conducted on this topic (Goldin-Meadow, 2003).
hoc. This means, emblems resemble words, since their established and culturally shared meaning is understood without context and explanation (Goldin-Meadow, 2003). According to Goldin-Meadow (1999, 2003), various classifications of gestures exist, and McNeill (1992), for instance, identifies four different kinds of gestures that supplement speech.

- **Iconic gestures** obviously capture characteristics of the semantic content of speech. For example, making a twisting motion with one’s hand while saying ‘I can’t open this jar’.

- **Metaphoric gestures** are similar to iconic gestures since they are pictorial, but their content is rather more abstract than concrete. For example, when presenting different ideas these might be illustrated by treating them as bounded objects supported by hand movements, i.e., making gestures that discretely represent the proposed ideas. In addition, the gestures may spatially reflect whether the ideas are closely related to each other or not.

- **Deictic gestures** are ones that indicate objects, locations and other persons in the real world, such as pointing to a person and saying ‘she is my brother’s girlfriend’. It should be noted that deictic gestures could be used although the entity is invisible or not present. For instance, pointing at a chair to represent the person that earlier sat on it.

- **Beat gestures** catch the rhythmical pulsation of speech, and tend to bear the same form regardless of the content, for instance, moving one’s fingers or hands up and down, or back and fourth. They are usually generated with short, quick movements, and in order to put stress on a word, beat gestures demonstrate the significant role of that word in the discourse (Goldin-Meadow, 1999, 2003; McNeill, 1992).

Despite the close connection between gesture and speech in language, they generally differ in how they carry meaning (Gallagher, 2005; Goldin-Meadow, 1999, 2003). Indeed, gesture offers alternative ways of expressing ideas that are hard to articulate in speech, as well as when there is no proper word at hand for the actual meaning to be conveyed (Goldin-Meadow, 2003). Moreover, the temporal structure of gesture is different from the linear and segmented forms of speech (or written text), since gesture is more immediate and spatial, being able to express information through imagery. For instance, gesture varies in dimensions of time, space, form, trajectory, and so on, and is able to express meaning complexes without undergoing segmentation. Besides, gesture is able to present information simultaneously, which in speech needs to instead be expressed sequentially (Goldin-Meadow, 2003). For example, consider counting how a rabbit jumped off the road. Through gesture, the movement of the object of interest, its characteristics, and the manner of movement of the object’s path and location, are portrayed all together, by making an imaginary gesture of the path and location of the energetic jumps of the rabbit. In other words, it can be argued that spoken language is uni-dimensional in time, but gesture is multidimensional. Accordingly, gesture entails a progression of following stages, because it involves the accomplishments of movement and not just
semantics. However, the same argument can be used for speech, since at some level speech is in itself a kind of embodied gesticulation\footnote{According to Gallagher (2005, p.125) speech can be regarded as phonetic gesticulation.} (Gallagher, 2005). This notion is further developed in current work by McNeill (2005), in which the microgenesis of gesture and speech originates from a primary stage, or as McNeill denotes it, the ‘growth point’, which is then identified and analyzed. In addition, Niedenthal \textit{et al.} (2005a) note that work conducted by Rauscher, Krauss and Chen (1996, in Niedenthal \textit{et al.}, 2005a), for instance, demonstrate that participants who first watched an animated action cartoon, were significantly slower in re-describing spatial elements of the cartoon after a break, when their hands were prevented from gesturing (a cover story of recording skin conductance from their palms hindered hand movements). Apparently, blocking gesture impaired the access to our cognitive understanding.

Beginning with how children used gestures in solving Piagetian conservation tasks (cf. subsection 2.4.3), Goldin-Meadow and colleagues discovered that speech and gesture convey different information, but not necessarily conflicting meaning (cf. e.g. Goldin-Meadow, 1999, 2003). For example, they noticed that some children who offered an erroneous verbal explanation of the task nevertheless indicated in their gestures that they actually understood the present task and its solution. Hence, the children are generating a gesture-speech mismatch, which Goldin-Meadow (1999) characterized as a “communicative act in which the information conveyed in gesture is different from the information conveyed in the accompanying speech” (ibid., p. 423). This means, in a mismatch, the speaker’s speech and gesture convey different information, and the ‘extra’ ideas found in mismatches are only conveyed in gesture, which sometimes has a positive impact on learning and instruction. Through mismatches, children communicate unique information through their hands offering their teachers a ‘window’ into their minds. In other words, gesture reflects thoughts that cannot yet be articulated in speech. It should be noted, however, that some mismatches can also lead the listener astray. In addition, ordinary listeners can effortlessly take advantage of the meaning conveyed in gesture without any special training, being able to read ‘gesture on the fly’. Moreover, the information noticed from gestures by listeners can instead be expressed in speech, i.e., not necessarily being ‘tagged’ in gesture. This means, the expressed information, in either speech or gesture, can switch modality when the information is re-expressed (Goldin-Meadow, 2003).

Thus far, gesture seems to be a part of communication since gesture reflects thoughts, but Goldin-Meadow (2003) subsequently addressed whether the need to communicate is the force that drives humans to gesture. She presents empirical studies, which demonstrate that people generally gesture more when they can see each other, but we also gesture when we cannot see others, for example, while talking on the phone. Moreover, even blind people gesture as they speak, producing gestures at the same rate, and in the same manner as the sighted. Gesture also functions for speakers as well, since
gesture may provide speakers with a channel for expressing thoughts that are difficult to articulate in speech, using another medium (cf. Goldin-Meadow, 2003). In other words, besides conveying information to others, there is evidence that gesturing is a ‘boon to the gesturer’. As Goldin-Meadow explains, gesture may not only reflect thoughts it might also play a role in shaping thoughts, similarly as externalizing one’s problem by writing it down to solve it easier. Empirical work indicates that besides reflecting cognitive load, gesturing also has some impact on cognitive processing itself, and “whatever the mechanism, it’s clear that gesturing can help free up cognitive resources that then can be used elsewhere. At the very least, we ought to stop telling people not to move their hands when they talk” (Goldin-Meadow, 2003, p. 166). This means, according Goldin-Meadow, gesture may be a kind of prototype of ideas, shadowed but not hidden, which functions as an experimental vehicle of thought, in which linguistic meaning is grounded in bodily action (see subsections 4.2.2-4.2.3 and section 4.3). She speculated that gesture may be an explicit depiction of the action-meaning embodied in speech, suggesting that hand movements are physical instantiations of the speaker’s mental model. All in all, Goldin-Meadow hypothesizes that the division of labour between speech and gesture is based on a singular ‘integrated model’, in which the accessibility of the two modalities differs, given that some information can be accessible to both speech and gesture, some is reachable only for gestures and then some only for speech. In addition, it should be noted that gesture also has kinaesthetic dimensions, and this phenomenological understanding may be crucial for how the ‘body knows the world through the hand’ (Goodwin, 2000).

Despite the fact that Goldin-Meadow mentions aspects of embodiment and movement in gesture, she does not elaborate these issues further. Gallagher (2005), on the contrary, emphasizes that “[l]anguage is a modality of the human body. It is generated out of movement” (p. 107). Although gesture involves a direct mapping of meaning onto space and motion, it is not a kind of instrumental actions because normally gestures are neither controlled by the body schema nor by the body image (see section 3.3). Instead, he advocates that primarily and essentially gesture is coupled with linguistic and communicative processes, which means that gesture is irreducible to pure movements. It should be noted, however, that this stance on gesture is an alternative to motor theories of gesture (cf. Kita, 2000; Streeck, 1996 in Gallagher, 2005), in which gesture is viewed as the origin of language, and that spoken language gradually developed from embodied movements, a special kind of ‘oral motility’ as Gallagher explains. Similarly, speech can also be regarded as a sophisticated movement of the body. However, as Gallagher points out, this might have some truth, but not the complete truth. Indeed, there is a need to understand how gesture is produced. Empirical evidence from patients suffering from pathological limitations and defects in either their body image or body schema, suggest that the communicative and linguistic functions seem to transcend their pathological suffering. The person IW, suffering from some neurological limitations, and therefore must volitionally and visually guide and control his every day movements, e.g., reaching for a glas of water, is able to gesture in the dark
without visual feedback. This means, his bodily motor difficulties do not carry over to gesture, and indeed, his gesturing appears quite normal in launching and timing (Gallagher, 2005).

Gallagher suggests, based on findings from his work on IW, that while gesture is a certain kind of expressive movement, it may use or build upon the basic systems that control movement. However, as Gallagher explains, there is a need to make a distinction between instrumental (along with locomotive action), as a kind of motor action (with which IW has problems) and expressive or communicative actions. The fundamental distinction between these kinds of movements lies in the assumption that gesture is “an action that helps to create the narrative space that is shared in the communicative situation” (ibid., p.117), instead of being a mere movement in the service of communicative acts within human discourse. In other words, gesture is tied to a linguistic and communicative system, which also controls its manifestation. Therefore, Gallagher offers a communicative theory of gesture, in which gesture constantly entails the mapping of meaning into a linguistic space, whether it is purposefully produced or not. Additional evidence for the interpretation that gesture is a certain kind of expressive movement comes from persons who gesture with phantom limbs and those who are congenitally blind people (see Iverson & Goldin-Meadow, 1998). All in all, Gallagher (2005, p. 122) characterizes embodiment in relation to different kinds of movement as follows in Table 3.

**Table 1.** The relations between embodiment and different kinds of movements (from Gallagher, 2005, p. 122).

<table>
<thead>
<tr>
<th>Type of movement</th>
<th>Examples</th>
<th>Primary control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflex</td>
<td>Babkin-reflex</td>
<td>Automatic motor programs</td>
</tr>
<tr>
<td>Locomotive</td>
<td>Walking, crawling</td>
<td>Body schema</td>
</tr>
<tr>
<td>Instrumental</td>
<td>Reaching, grasping</td>
<td>Body image</td>
</tr>
</tbody>
</table>

According to Gallagher’s integrative theory of gesture, gesture is supposed to be i) embodied, i.e., constrained and enabled by motoric possibilities; ii), communicative; and iii), cognitive, i.e., contributing to the accomplishment of thought, and the shaping of mind. Gallagher, however, stresses that a total disregard of the movement nature of gesture would be a mistake, and he identifies four different aspects of movements in gesture:

- **Initiation of gesture** is generally unconscious and tied to cognitive and communicative requirements, but in some cases conscious control exercises veto power over gestural movement.
- **Launching and timing** are integrated with communicative and cognitive factors; which are tied to semantics and the pragmatics of the situation.
• *Morphokinesis* [shape] is integrated with communicative and cognitive factors tied to semantics.
• *Topokinesis* [location of one hand relative to the other and relative to other parts of the body] is dependent primarily on body-schematic processes and proprioception in the normal case (p. 123).

Taken together, Gallagher's major claim is despite the fact that gesture is a cognitive-communicative phenomenon, it is embodied in a very basic way since it requires motor control, but at the same time, gesture is irreducible to pure movement. Moreover, he points out that the connotation 'cognitive' factor does not explicitly mean conscious control, and, as he explains “gesture and language shape cognition in a pre-cognitive manner”. In other words, gesture enables people to embody and communicate their thoughts both through action and in interaction.

Furthermore, it is argued that gestures also have representational properties, and Goldin-Meadow (2003), for instance, emphasizes that the gestures accompanying speech are *symbolic acts that convey meaning*. In addition, Gallagher mentions that McNeill, (1992, p. 105) also emphasizes that gestures “are not arms waving in the air, but *symbols that exhibit meaning* in their own right”. As Goldin-Meadow points out, however, it is easy to overlook the symbolic nature of gesture since its encoding is usually iconic, i.e., a gesture simply appears such as what it represents. For instance, a twisting hand movement reminds you of the action used to open a jar, but it should be noted that the twisting action is neither equal to the actual act of twisting nor does is mean the word “open”. In line with this remark, Gallagher emphasizes the fundamental difference between instrumental acts (e.g., opening a jar or reaching out to pick up a glass), and the generation of a gesture *signifying* the very action of opening a jar or picking up the glass. In other words, the act of gesture achieves an entirely different function than the actual grasping or opening, because those actions have representational content, which is a cognitive and possibly a communicative function that requires the generation and expression of meaning (Gallagher, 2005). According to Gallagher, gesturing may for just that particular reason, differ significantly in its mechanism in comparison to the instrumental action, since the meaning and the communicative situation calls forth the cognitive and linguistic nature of gesture. Furthermore, Roth’s (2002) *naturalistic* study of the role of gestures in human communication, in different groups of physics students who have to develop a shared scientific language of electrostatics, demonstrated that present physical materials and gestures functioned as *representational* forms, and made it possible for the students to externalize thought. For instance, iconic gestures emerged via object manipulation and epistemic movements during the students’ lab investigation and those gestures seemed to precede speech production of the scientific concepts, since students expressed through gesture what they only later were able to articulate in speech or written text.

In addition, King (2004) points out that although both Goldin-Meadow (1999, 2003) and McNeill (1992) use gesture to investigate and unlock interior and
individual cognitive phenomena, a related question is how gestures facilitate or hinder social coordination in dyadic, triadic, and collaborative interactions (see 4.2 concerning various kinds of social interactions). According to King, this kind of gesture research is conducted to a lesser extent than studies of individual cognitive processes, but the work of Goodwin (cf. e.g. 2000, 2003), focuses on gesture within social interactions. Goodwin (2003), however, uses the concept of *symbiotic gesture*, in contrast to the more common classifications of gestures, and a symbiotic gesture refers not only to the particular movements of the individual’s hand and speech in isolation, but also considers the indispensable contextual surrounding in which the gesture is carried out that structures and scaffolds the gestural activity. Taken together, ‘symbiotic gesture’ goes beyond the bounds of the individual’s body, given that symbiotic gestures include environmental aspects that capture the whole meaning of the situation, which is greater than, and different from, its separate parts (see the close connection with the Dance Metaphor in 4.3.2). Although Goodwin (2003) argues that symbiotic gestures are rather common phenomena within situated human interaction, little research has been conducted on how structures in the environment play a significant role in the organization of gesture.

Recently, it has been suggested *bodily mimesis* (Donald, 1991) bridges the gap between language and embodiment (Zlatev, 2007). Generally speaking, bodily mimesis is the volitional use of the body for constructing and communicating representations in the form of imitation and gesture. It is worth highlighting that the emergence of these bodily representations for the sake of communication might serve as the foundation for the phylogenetic development of a “human language” in the form of a conventional symbolic system via triadic mimesis. It implies the understanding of communicative intentions, and is similar to Tomasello’s (1999, see section 5.2) suggestion that it is the understanding of others as intentional agents that distinguishes human beings from other primates. Triadic mimesis is, according to Zlatev, the crucial factor for the leap in cultural evolution. Furthermore, the ability to understand that a gesture corresponds to something presumably came naturally to our predecessors through the use of manual gestures to ‘describe’, for instance, that an animal went up that tree, in which the pointing gesture denotes both ‘animal’ and ‘tree’, and the upward movement of the arm ‘climbing’. According to Zlatev, the essence of triadic mimesis, serves as the basis for language and symbolic interaction. This means, Zlatev’s proposal is similar to ‘gesture-first’ theories of language origins, but differs from them because mimesis lacks full conventionality and systematicity, which were likely to have appeared when vocal calls became recruited for the purpose of disambiguating gestures (cf. Arbib, 2003). Thus, Zlatev argues that mimesis is the ‘missing’ link in the evolution of language, which subsequently explains the roots of symbolic interaction. Finally, Zlatev claims that the ways the mirror neuron systems operate as well as the fact that it is located in the human homologue to Broca’s area partly provide a ‘neural correlate’ for mimesis. However, there are some possible objections to Zlatev’s proposal. Firstly, his account misses the close interrelatedness of speech and gesture in language as discussed in subsection 4.3.2, given that
Zlatev suggests gesture proceeds speech, not viewing them as a united speech-gesture language system. Secondly, as also previously stated and shown, communicative gesture is not necessarily under conscious volitional control, as Zlatev claims, since IW is able to gesture without visible control. Thus, Zlatev misses that gesture in itself has communicative intentions and therefore cannot be regarded as instrumental movement reducible to pure movement (cf. Gallagher, 2005), which means the know how of gesture is not the same as the know-how of instrumental movement. Finally, Zlatev’s account is not embodied in the radical sense, since he strongly emphasizes mental concepts, intentions and so on, that are addressed from an alternative perspective in section 5.4. However, we return to this topic in section 5.5, when discussing the ontogeny of socially interactive cognition. Thus, once again, Zlatev’s proposal instead has an embodied outfit rather than being genuinely embodied in Clark’s (1999a) radical sense.

Concerning the connection between gesture, embodied simulations and mirror neurons, Rizzolatti and Arbib (1998) suggest that a series of preventative mechanisms are usually activated in order to hinder the actual (re-) production of the observed action. Occasionally, however, the premotor system will allow a tiny aspect of the simulated movement to be executed, and this short glimpse is (unconsciously) recognized by the other person, affecting both the actor and the observer. This means, the actor recognizes it as an intention in the observer, and the observer notices that her (involuntary) response, in turn, affects the behavior of the actor. During the course of evolution, the capacity to voluntarily control one’s own mirror system to emit signals, instead of the mere automatic leaking of parts of the mirrored actions, was essential for the emergence of a (basic) dialogue between two individuals which forms the core of language. Rizzolatti and Arbib further speculate that this new capacity of the mirror system was initially based on oro-facial movements, which is how all primates mainly communicate. Manual gestures were added later, as a way of complementing the oro-facial ones, since gestures increased the sender’s expressive power. The combination of oro-facial movements and gestures, according to Rizzolatti and Arbib, strongly implies the importance of controlled vocalization as an extension of oro-facial movements and gestures. The evolutionary pressure for more complex sound emission, together with the anatomical possibilities, resulted in a move of intentional interaction from its oro-facial and gestural origins to sound emission (cf. Corballis 1999, but consider Gallagher’s criticism against motor theories of gesture).

However, their hypothesis overlooks the integrated nature of speech and gesture in human language. McNeill (2005), on the contrary, emphasizes the thought-language-hand link in language origin between area 44 and 45 in Broca’s area of the human brain, highlighting the double characteristic of language, i.e., speech and gesture is the course of joint action in evolution. Area 44 is mainly responsible for the organization of action sequences, whereas area 45 is the part of Broca’s area that contains many mirror neurons, which McNeill suggests became self-responding to one’s own actions subsequently imbuing them to contain meaning. During the course
of phylogeny, these two systems became co-opted in order to unite gesture and vocalization. This means, speech and gesture evolved together to embody meaning, given that “language is an orchestration of actions by the hand and the vocal tract articulators...the flow of oral-‘gestural’ movements ... speech itself, which we usually think of as the primary linguistic medium, are in fact not the only medium. But language is also patterns for organizing actions. In part, what evolved was a new way of acting. How brain mechanisms orchestrate such motor actions in the oral and manual spheres” (McNeill, 2005, p. 247).

According to McNeill, the crucial shift in the function of mirror neurons occurred when they began to respond to significances other than the actions themselves, as a way of co-opting areas 44 and 45 in Broca’s area, providing the basis for recognizing the actions of others. In other terms, this co-opted system seems to be a part of a circuit for recognizing intentional goal-directed actions from one’s own actions or from others. It should be stressed, that McNeill emphasizes the relational nature of the mirror neuron system which, in my opinion, Rizzolatti and Arbib (1998) overlook in their hypothesis. They sidestep a crucial aspect in the new way of organizing movements by gesture (as communicative movements instead of instrumental movements, cf. Gallagher, 2005), namely the fact that gesture signifies things other than the actions themselves. As McNeill explains, the action of upward movement is not only a movement upward, but also the material carrier of a concept of ‘Upness’, metaphorically speaking, and the answer to how this ability of significant symbols emerged is present in Mead’s work (see subsection 2.4.6), since he explained the relational aspect of human language. Citing Mead, McNeill wrote “[g]estures become significant symbols when they implicitly arouse in an individual making them the same response which they explicitly arouse in other individuals” (Mead, in McNeill, p. 250). Consequently, meaningfulness emerges from the ability to simulate a social reaction of another in yourself, a way of reacting in your own actions similarly to the actions of others, which he denotes Mead’s loop. Furthermore, the mirror neurons system functions as the mechanism for this loop, and it provides a plausible explanation to why gesture is used so frequently in human social interaction. Gesture also has the important role of activating our own mirror neuron system, as well as offering oneself the ability to take the role/perspective of the other simultaneously (McNeill, 2005). This means, the consequences of linking areas 44 and 45 in Broca’s area is the generation of action sequences with meanings other than the meaning of the action itself, i.e., Mead’s loop, which is possibly a unique circuit in the human brain, subsequently resulting in the development of the self as an independent agent through social scaffolding in Vygotsky’s General law of cultural development (see subsection 2.4.4). Thus, the shift from non-significant to significant interaction, as in Vygotsky’s previously described pointing example can be explained neurologically by Mead’s loop. Finally, Mead’s loop creates a connection of gesture to discourse, given that this relational characteristic is also present in speech. Clark (1999a), for instance, stressed this dual aspects of language – “words and text are both real, external objects that we can encounter and
manipulate and key instruments of inner, abstract, environmentally
decoupled reason” (p 350). However, in my opinion, Clark focuses far too
much on the aspects of speech and written text as the only symbolic means
in human language, and thus totally neglects gesture (see also Clark, 2005,
in which he does not discuss gesture at all in regard to surrogate
situatedness).

4.4 Four Fundamental Functions of the Body in Social Interactions

In this section, I identify some fundamental roles of embodiment in social
interaction, based on the issues discussed in this chapter which offer highly
complementary rather than alternative views on the role of embodiment in
social interactions. By integrating these issues, we can obtain a deeper
understanding of socially embodied actions in social interaction.

In order to grasp the general functions of embodiment in social (cognitive)
interaction; it is important to consider that one cannot focus on just one
modality or aspect, such as posture, emotion, speech, facial expression or
gesture. On the contrary, it is necessary to consider the cross-modal
interaction among multiple modalities. I choose to use the term cross-modal interaction instead of multi-modal interaction, as a means of emphasizing the
interrelatedness between these modalities and experiences, rather than
viewing them as somewhat distinct, parallel and independent of each other.
Moreover, these fundamental functions of embodiment in social interaction
are not fixed, and are, to some degree, overlapping. Additionally, these
embodied functions are not ranked in-between. Furthermore, the different
kinds of social interaction and communication are not considered, since I
quite simply just address social interaction in general, from both the so-
called inside (first-hand perspective) as well as the so-called outside
perspective (third-hand perspective). Based on the previous ideas and
empirical findings, I identify the following four fundamental functions of the
body and its sensorimotor processes in social interaction:

- **The body functions as a social resonance mechanism**
- **The body functions as a means and end in social interaction**
- **Embodied actions and experiences function as a helping hand in shaping, expressing and sharing thoughts**
- **The body functions as a representational device**

All in all, these four fundamental functions which suggest that our socially
embodied actions do not simply serve to express our internal cognitive
processes, but instead are actually fundamental elements of doing and
experiencing cognition, are described as follows.

4.4.1 The body as a social resonance mechanism

From an embodied perspective, there is no need to decode or represent
embodied social stimuli to more “advanced” or cognitive states since the

These four fundamental functions of embodiment in social interaction have been
presented earlier, in shorter versions, in Lindblom (2005b, 2006) and Susi & Lindblom
(2005).
bodily states in themselves are actually cognitive or affective states, as Barsalou’s *et al.* (2003) and related work portrays (subsections 4.1.1 - 4.1.2). They point out, for instance, that bodily postures influence the subjects’ affective state, e.g., subjects in an upright position experienced more pride than subjects in a slump position. Additionally, it should be noted that this social resonance mechanism also functions reversely, i.e., cognitive and affective states influence bodily states; for example, the high school students who received good grades in an exam adopted a more erect posture than those who received poor grades. Furthermore, the common human ability of synchronizing, tuning into and mimicking others’ facial expressions and bodily behaviors, which are widely documented in the literature and experimental findings, for instance, is shown by subjects often mimicking an experimenter’s actual behavior such as rubbing the nose or shaking a foot.

The social resonance mechanism not only functions in direct or online social interaction, but also in offline interaction and when the social stimuli are more abstract, e.g., concepts. For example, subjects primed with concepts related to *elderly* people (e.g. ‘gray’, ‘bingo’, ‘wrinkles’) exhibited slower movements when leaving the experimental laboratory, as compared to the control group primed with neutral words. Moreover, subjects responded faster to ‘positive’ words (e.g. ‘love’) than ‘negative’ ones (e.g. ‘hate’) when asked to pull a lever towards themselves, which demonstrates a compatibility effect between cognitive and bodily states. In line with this remark, the field of phenomenology (cf. section 4.2) strongly emphasizes that one’s own embodied experiences are primarily the locus of intersubjectivity in order to grasp and gain the understanding of other people, meaning there is no need for an ‘as if’ model of the present social situation. In other words, it can be argued that the body and its sensorimotor processes function as a social resonance mechanism. The examples presented in this chapter, as well as other studies, demonstrate there is a strong relation between embodied and cognitive states in social interaction, since the bi-directional exchange between these states as well as between the interacting partners, occur automatically without any higher knowledge structure.

### 4.4.2 The body as means and end in social interaction

As mentioned in subsection 4.2.2, it has been supposed that this social resonance mechanism is accomplished by mirror neurons. It is also suggested that these mirror neurons and the mirror neuron system provide a link between ‘action’ and ‘action-perception’ in view of the fact they respond to both conditions. Consequently, mirror neurons enable the agent to perceive and understand the meaning of the observed action, and its system is subsequently considered the neurobiological and embodied underpinning of social experience, social interaction as well as even mind-reading and ToM. Apparently, the close linkage between first-hand and third-hand aspects of social interaction implies that the body and its sensorimotor processes are ‘cognitive’ in themselves, and not only bounded by the brain. The central issue here is ‘action-understanding’, which means that in order to *recognize* an action, besides activating the perceptual (e.g., visual or auditory) system, the motor system is also activated. This is in sharp
contrast to the still prevailing view of cognitive processes, in which the perceptually visible stimulus is encoded to a more ‘advanced’ and cognitive processing level, then decoded back to a motoric movement that executes the command as output. From an embodied perspective, this means perception and action should be regarded as closely interrelated when intuitively grasping the idea of the observed action. The significant benefit of the action-understanding linkage, despite its parsimonious account, is the inbuilt dual ability of grasping both the ‘what’ and ‘why’ aspects of the present action, i.e., what the action is about as well as catching the intention behind the movement (as described in subsection 4.2.2).

Most approaches that describe the social dimension of embodiment are generally based on the notion of social (cognitive) interaction results from embodied simulation or embodied practice (see subsection 4.2.1). This is exemplified by a matching system that allows one to adopt other persons’ point of view by simulating their mental states with a resonant state of one’s own, i.e., one is able to put oneself in another person’s ‘shoes’ through embodied simulation. Broadly speaking, this implies that one’s own understanding of social interaction and other agents’ mind is more than the exchange of communication signals. It should be noted, however, that the extent of using embodied simulations in social interaction differs. On the one hand, it is argued that we always simulate (both online and offline) in order to understand our social surrounding (Gallese, 2004). On the other hand, Gallagher (2005, 2007b) emphasizes that we only simulate when our embodied practices in on-line interaction break down or when we try to construct third-person explanations and/or predictions, in which the simulations are mostly carried out offline. There is not possible to draw a sharp line between online and offline cognition, but given the more radical nature of embodiment in this thesis, it goes in favor of Gallagher’s account. Moreover, empirical results support a close coupling between language and action, with bodily action perhaps being the roots of linguistic meaning (see subsection 4.3.1). In other words, the mirror neuron system makes it possible to establish various kinds of online and offline social interactions, which are fundamental for people’s ability to interact and understand each other.

Furthermore, there are different degrees of activation of and couplings with the mirror neuron system and other parts of the brain, which form a continuum between basic levels of action-understanding to more advanced ones, such as gestures and language (see subsections 4.3.1 – 4.3.2). Although the current knowledge of the underlying mirror neuron mechanism is unable to explain in detail the whole complexity of human social interaction, it does illuminate how interacting partners are able to share the communicated meaning of a dialogue, which is one of the fundamental aspects of social communication and human language, in which Mead’s loop plays a crucial role.

The functions of the body as a ‘resonance mechanism’ and also ‘a means and end’ might seem quite comparable. However, whereas the function of the
body ‘as a resonance mechanism’ simply means cognitive and bodily states of the interacting partners are reflected in both themselves and in-between them, it does not explain the relationship between their first-hand and third-hand experiences in social interaction. Instead, viewing the function of the body ‘as a means and end’ offers a tentative explanation of that particular linkage, thereby unifying the ‘inside and ‘outside’ perspectives of socially embodied interaction. In other words, the previously portrayed function of embodiment in social interaction mostly stresses ‘that’ the body and its sensorimotor processes function as a social resonance mechanism, whereas the second function rather focuses on ‘how’ this is accomplished.

4.4.3 Embodied actions as a helping hand in shaping and sharing thoughts

Besides speech, gesture is a significant (embodied) aspect of social interaction, which may provide important information to the listener, since gesture offers speakers the means of expressing thoughts difficult to articulate in speech. Accordingly, gesture is a crucial part of communication, and there is evidence demonstrating that gesturing is a ‘boon to the gesturer’, and besides reflecting thoughts, it might shape them. Gesturing may be a form of processing initial ideas that are shadowed but not hidden. As a result, gesturing functions as a vehicle of thought. Hence, gesture and speech complement, but do not compete, with each other (cf. subsections 4.3.1 – 4.3.2).

Furthermore, it is important to notice that gesture has not only visual, but also kinaesthetic dimensions, and this phenomenological understanding may be crucial for how the “body knows the world through the hand”. Despite the close connection between gesture and speech in language, they generally differ in how they carry meaning, which can be broadly characterized as when spoken language is uni-dimensional and discrete in time, and gesture is instead more multidimensional in time. Through gesturing, we are able to generate and embody dynamical associations between different matters, which can offer new insights to the present situation or problem at hand. In addition, gesture sometimes serves as an explicit instance of the action-meaning embodied in speech, implying that hand movements are physical externalizations of the speaker’s ideas. Broadly speaking, gesture enables people to embody and communicate their thoughts in action. Besides the ways gesture is part of the processes of shaping and expressing thoughts, it is also an embodied action that ‘lends a hand’ in the emergence of the shared social discourse in social interaction.

In other words, bodily actions, particularly gesture, and their experiences, function as a helping hand in shaping, expressing and sharing thoughts, since they operate both outwardly (inter-subjectively) as well as inwardly (intra-subjectively) in social interaction and cognition, in which our embodiment (paradoxically) both constrains and enables the possibilities of expressive movements.
4.4.4 The body as a representational device

It is widely acknowledged that one of the hallmarks of human (spoken) language is its representational capacity, which is, so to speak, the ability to correspond to something else. In addition to speech, there is the more controversial claim that the body also has representational properties, where certain kinds of gesture, portraying representational aspects, are the least provocative and most obvious examples of the body as an (external) representational device. Hence, it is argued that through gesturing we shape symbols that carry meaning by themselves (e.g., the iconic and metaphoric gestures described in subsection 4.3.2), despite their iconic character, and for instance, a twisting hand movement reminds us of the action used to open a jar. The essential difference, however, lies in the communicative act of gesture, which signifies the very action of opening a jar. This means, it is an action at another functional level, and subsequently it has representational aspects, since the twisting motion of the hand is not really the same as the actual twisting motion when opening the jar since the generation and manifestation of the twisting (representational) gesture involves communicative intent and perhaps (cognitive) meaning.

Proceeding from the representational aspects of gesture, the notion of the body as a representational device is echoed in more provocative proposals. For instance, Zajonc and Markus’ Hard Interface Theory (HIT) stresses that bodily states already constitute “hard representations”, suggesting these in-depth rooted bodily representations do not require encoding to more cognitive or “mental” representations to have representational value (cf. subsection 4.1.2). In line with this remark, although not as radical as in the HIT, is Barsalou et al.'s. (2003) suggestion that bodily states of situated conceptualizations underlie ‘higher’ cognitive functions. The same argument can be applied to the various kinds of embodied simulation (cf. subsection 4.2.1), in which re-activations of sensorimotor processes, instead of manipulating mental symbolic concepts, serve as the foundation for elementary and more ‘higher’ cognitive processes. More controversial, with regard to embodied simulation, is the opinion that simulation actually (sometimes) is an ‘extra step’. Indeed, the pragmatic embodied practices are usually necessary in order to grasp and gain an understanding of the other person, since the perception of another person’s actions is enough in online cognition (see Gallagher, 2005, 2007b in subsection 4.2.1).

The neurological roots of this ability could be the activity of the mirror neurons, since their linkage between ‘action’ and ‘action-perception’ might propose a kind of ‘action representations’ that are directly enacted in social interaction. Furthermore, since mirror neurons are able to compensate for missing ‘information’, and yet seem to ‘understand’ the goal of the action, it can be argued that the grasping of the action does not require a declarative understanding, since it is meaningful in itself (cf. subsection 4.3.1). In other words, the cognitive processes of social (cognitive) interaction are, to various extents, grounded in embodied representations that have representational content. It should be noted, however, that we are not denying mental concepts as such, but questioning the organization, since they may be the
result of and grounded in embodied interactions and not the underlying requirement for cognitive processes (cf. subsection 4.3.1). Furthermore, Mead’s loop considers the relational aspect of social interaction, and explains in what ways gesture functions both inwardly and outwardly, so to speak.

In summary these identified four fundamental functions of embodiment in social interaction obviously emphasize that embodiment really matters in social interaction, which goes far beyond the bounds of internal mental processes. We return to these functions when I motivate and present the framework in chapter 6.
Chapter 5

Information is created in the interface between perception and action
...the salience of the body... is missing in many theories of meaning
Fogel, 1993

...social scientists are body-dead because they are conceptually brain-dead
to signifying acts within the semiotics of body movements.
Varela, 1992

Nothing is more wrong-headed than calling meaning a mental activity
Wittgenstein, 1958

5. The Nature of Social Interaction and Cognition

In chapters 3 and 4 I mapped the terrain of embodiment theories in general, and addressed the social dimension of embodiment to some degree. However, before beginning to investigate and analyze social embodiment more deeply, I must consider the issue of social interaction and cognition. In this chapter, I portray and analyze in further detail different characteristics, levels, kinds and metaphors of, ways of studying social interaction and cognition, as well as its ontogeny in the following way.

Firstly, section 5.1 characterizes the basic foundation of social interaction and presents various definitions in mainstream social cognition. In subsections 5.1.1 and 5.1.2, I particularly present social interaction in general, making a ‘cut’ between social interactions in social insects and other living creatures like mammals and humans. Subsequently section 5.2 focuses mainly on different levels of social interaction relevant for human interaction, since the ability to engage in social interaction is a central building block of social life and cognition, and thus one of the foundations of human culture. Section 5.3 portrays and compares two different metaphors of the explanation of the nature of social interaction. The following section 5.4, revisits the characterizations of social interaction and its methodological study, and offers alternative explanations to the traditional view. Section 5.5 explains how different scaffolds and embodied actions contribute in the development of different kinds of social interaction and cognition during ontogeny, and then re-describes them from an embodied perspective. Finally,
section 5.6 summarizes the major issues discussed, and outlines the way ahead.

5.1 Characterizing and Studying Social Interaction and Cognition

In order to address the topic of social interaction from an embodied perspective, we return to Maturana and Varela’s (1987) notion of structural coupling (cf. subsection 3.2.3) and use their approach as the underlying foundation. Besides being coupled with a physical environment, an individual can also be coupled with a social one. Hence, as Maturana and Varela (1987) stressed, while an organism can be structurally coupled with the physical environment, as well as structurally coupled with another organism, the ‘perturbations’ between them need not to be distinguished from the ones that derive from the physical, and non-living environment. Adding recurrent interactions between the organisms through the course of ontogeny, results in co-drifting organisms, since “co-ontogenies with mutual involvement through their reciprocal structural coupling, each one conserving its adaptation and organization” (Maturana & Varela, 1987 p. 180). When this co-ontogenesis occurs, the co-drifting organisms generate a new phenomenological domain, called third-order (structural) couplings (TOCs), that may be rather complex when there is a nervous system involved. That is, social interaction. Figure 3 illustrates structural couplings with the environment as well as third-order structural couplings with another organism.

![Figure 3](image_url)

**Figure 3.** Schematic illustration of a) structural coupling with the environment respectively b) another organism (TOC) (redrawn from Maturana & Varela, 1987, p. 180).

Furthermore, Smith and Grasser (2005) emphasize that the mutual coupling with other conspecifics is similar to physical couplings, but with an ‘added twist’. They suggest that the reciprocal interactions, for instance, between a baby and its caretaker are significantly different from the infant’s interaction with a physical object. Instead of the more ‘fixed’ interactions with a physical object, social interactions are broader, more active, and more diverse. That is, the social partners of a social interaction offer much more than physical objects, and this fact has crucial consequences (see Mead’s view in subsection 2.4.6). Thus, social aspects of interaction add structure and support to the development and transformation of more basic interactions to more advanced ones through the interactive patterns that are mutually produced between the interactants. In other words, social interactions generally offer more diversity and complexity than physical interactions.
(Smith & Grasser, 2005; see also the work of Mead in subsection 2.4.6). Similarly, Maturana and Varela (1987) emphasized that while third-order structural couplings are a rather common phenomena, they differ both in i) how they occur, that is, in their underlying mechanisms and kinds of interactions they provide; and ii) the additional phenomena, that is, the outcome or product that emerges from the different kinds of interactions. Indeed, the variety of social interactions rests on behavioral couplings afforded by the organism’s nervous system (Maturana & Varela, 1987). This means, in order to understand social dynamics as a biological phenomenon, it is important to consider the different kinds of social interactions resulting from the huge diversity of behavioral patterns and means of interactions afforded by the organism’s embodiment in general.

In order to interact with other individuals different means of interaction with the social world are thus needed, which in turn affect the characteristics of the organism’s cognitive processes (cf. Anderson, in press; Maturana & Varela, 1987; Sinha & Jensen, 2000). It is well-known that animals and humans socially communicate and interact in a diversity of ways, ranging from quite simple notions of “social phenomena” in social insects to the ‘expressiveness’ of different communication abilities in animals and humans to high-level social cognition and mind-reading capabilities in humans (and to some degree in other primates). As the range of social interaction perspectives briefly described here indicates, there is quite a wide spectrum, extending from the social interactions of social insects, such as termites and ants, to mind-reading abilities in humans (and perhaps) in primates.

Generally speaking, two major kinds of social interactions can be distinguished. They concern the social interaction of insects, and that of mammals and birds, both of which are based on the individual’s morphological embodiment that subsequently impacts on the various kinds of social interactions of different species. As a result of this morphological division I identify two different kinds of embodiment. The following two subsections (5.2.1-5.2.2) portray this major division in more detail, and provide a more extensive discussion of exo-organismic embodiment and endo-organismic embodiment respectively.

5.1.1 Social interaction in social insects
Maturana and Varela (1987) presented the well-known example of social interaction of social insects such as bees, ants and termites that involves the entire ontogeny of participating organisms. Although social insects indeed are ‘social’, they usually interact indirectly, through the environment via an interchange of chemical substances, a mechanism denoted throphallaxis, and, in addition, through certain ‘communicative’ dances (Maturana & Varela, 1987). Figure 4 below illustrates indirect social interaction, since the relation between the social insects proceeds via the use of throphallaxis (solid arrows) and not directly between them (dotted arrows).
Furthermore, when they do interact directly through bodily interactions, the mechanism of structural coupling is also carried out via an interchange of chemical substances like pheromones and hormones. Depending on the content of the substances social insects distribute between themselves, the result is a variety of different and specialized roles diverging in their morphology. This divergent morphology is manifested in their activities and roles in the colony, such as the various castes of worker ants, the male, and the queen. Despite their different roles and morphology the means for social interaction is similar, via throphallaxis in one way or another, resulting in the ontogenies of individual participants being bound together in a co-ontogenetic structural drift of the whole colony. If the queen is removed, for instance, the chemical distribution and the overall chemical balance within the colony is disrupted, and as a consequence the feeding of the larvae changes so they will develop into queens. This means, the ontogeny of the participating others in the colony as a whole is dependent on the ontogenesis of certain individuals (Maturana & Varela, 1987). It should be said, however, that this process is different from von Uexküll’s (see subsection 2.4.1 and 3.2.3) notion that living bodies are constructed centrifugally. This most generally means living systems begin with one single cell, from which the different organs then develop, according to the cell’s own rules. In social insects, however, these ‘rules’ are not limited to being a part of the organism’s plan itself, given that external factors have crucial implications for the development of the individual’s embodiment since they might change body plans as they grow. However, if we return to the different notions of embodiment presented previously in subsection 3.2.3, it should be noted that Ziemke (2003) does not explicitly address this specific kind of organismoid embodiment in his discussion of various notions of embodiment.

The use of throphallaxis, or other objects, as a mediator in social interaction, is also referred to as stigmergy. Stigmergy functions as the missing link in the ‘coordination paradox’ of how to connect the individuals’ seemingly isolated behaviors with the overall collective and cooperative behavior of the colony from a holistic perspective (Theraulaz & Bonabeau, 1999; Susi & Ziemke, 2001; Susi, 2006). This means, the chemical interaction functions not only as a mediator both for controlling the individuals’ behavior but also the behavior of the whole group, since the remaining chemical traces and the

---

**Figure 4.** Indirect social interaction mediated through chemical substances in the environment.
adaptations made by individuals in the environment have a feedback effect on the behavior at both the individual as well as collective level. In other words, the overall cooperative social behavior of the colony is considered as a collective phenomenon. Despite the coordination between the individual and social levels, stigmergy alone is not sufficient to explain and develop the complexity of social life, social interaction and social cognition.

All in all, the (bodily) social interaction carried out by social insects is rather rigid and inflexible because their chitin morphology leads to a limited expressiveness of embodied social interactions, contrary to species with more degrees of freedom such as mammals. In other words, although social insects are embodied, their means for direct embodied interactions are limited based on their morphology, which we call exo-organismic (social) embodiment, since their chitin morphology actually provide some constraints for the diversity of direct bodily (social) interaction. According to Maturana and Varela, social insects are not distinguished individually by their behavioral variety and have a limited capacity to learn. Similarly, Dautenhahn (1995) notes ‘ants do not have friends’, which means they are not distinguished individually and do not form ‘bonds’ with each other, but rather their connections are based on their particular roles in the colony. Taken together, these aspects explain why the social interactions of social insects are of limited interest from a social embodiment perspective in cognitive science.

5.1.2 Social interaction beyond social insects
Considering social vertebrates and birds, such as horses, wolves, primates (including humans) and parrots, there are fundamental differences in bodily appearance, morphology, and complexity compared to social insects. The main morphological difference is that social vertebrates and birds have a common morphology of ‘inner’ skeleton and ‘outer’ muscles, which results in greater behavioral diversity that subsequently offers more degrees of freedom. Consequently, more degrees of freedom enable a larger variety of bodily states to be formed. Therefore, this kind of morphology is called - endo-organismic (social) embodiment. It should be noted, however, that although reptiles and fish belong to this kind of embodiment, they usually do not use facial expressions and other bodily actions, to the same degree, as mammals, and therefore I not consider them as being the major topic for the discussion.

Apparently, the bodily appearances of mammals’ and birds’ are not locked into the specific roles existing in the current group, as in the case with social insects like ants, in which the various roles of individual ants are rather fixed during ontogeny, as a kind of ‘morphology specialty’. The lesser degrees of difference between individuals’ bodily morphology are compensated by their better capacity for direct behavioral coordination through various bodily postures, visual, tactile and auditory expressions, which in turn give rise to different modes of social interactions (Maturana &

---

28 The material/substance of the outer shell of the physiology of social insects.
Most generally, this kind of embodiment involves a larger number of cells and more advanced physiology from a morphological perspective as well as increasingly extended developmental and learning processes, which result in more complex behavioral systems. Consequently, this kind of embodiment with many degrees of freedom and pro-longed ontogenesis allows individual vertebrates to participate in and form interactions and activities that arise as co-ordinations of behaviors between some organisms (Maturana & Varela, 1987). That is, they are able to form bonds between them and can have different roles depending on learning abilities and the complexity of the group. Accordingly, they are able to establish inter-individual relationships. This kind of more complex social organization is distinguished from the collective kind of organization found in social insects. Maturana and Varela (1987) denote these more flexible phenomena associated with the participation of individuals in constituting and maintaining third-order unities, social phenomena. A pack of wolves, for example, have such great flexibility in their behavioral coordination that together they can kill a large moose, a task of which no single individual alone wolf is capable. That is, hunting in wolves is the result of certain modes of social interactions which allow the generation of a new realm of collaborative phenomenon not achievable by a single individual alone (Maturana & Varela, 1987; Morris, 2005).

At first glance, the behavior couplings in the example of the wolf hunting resemble the stigmergic behavior of social insects, but on closer inspection, there is a fundamental difference. The social insects’ coordination of behavior is the result of responding to certain features in the environment, whereas the wolves’ interactions are directly shared co-ordinations through a greater flexibility depending on their embodied action-perception couplings. As a result of this shared coordination of behavior, hunting as a social phenomenon emerges. However, in order to achieve these shared co-ordinations there is a need for various kinds of social learning mechanisms. A vast set of social learning mechanisms exists, but various kinds of imitation are considered crucial for more advanced forms of social learning (cf. e.g. Dautenhahn, 1995; Maturana & Varela, 1987; Moore, 1992; Tomasello, 1999; Whiten, 2000). Dautenhahn (1995), for example, argues that imitation requires the ability to focus on one particular individual’s behavior and therefore evolved in species that form interactions between individual conspecifics. Furthermore, Whiten (2000) claims that the most cognitive complex learning strategies are imitation, by which he means the ability “an animal may learn from the actions of another” (ibid, p 479). Currently, no commonly accepted definition of imitation exists, however the spectrum generally ranges from more simple mimicry or ‘mirroring’ to ‘real’ imitation (see Tomasello, 1999). Most generally, and without reviewing the large amount of literature on imitation, one could say that what all these variants of imitation definitions have in common is that the differences concern how aware or conscious the learner is of its particular ‘imitating’ act. This can range from no awareness at all to being conscious of others as intentional beings, from who they imitate a particular set of activities to reach a certain goal or end result. The tricky issues of consciousness and
intentionality are not further discussed here, but these issues will be returned to later on. Indeed, I merely emphasize the continuum between responding by instinct to features in the environment and ‘directly’ sharing coordinated behaviors with others that, to various extents, are learned. While the final result in both cases is well-coordinated social behaviors at the overall collective level, with endo-organismic (social) embodiment the individuals’ own behaviors are more flexible and more complex than in exo-organismic (social) embodiment. In other words, more means of expressability as well as more modes of social interactions result in more multifaceted behavioral couplings or co-ordinations appropriate for explaining and developing the complexity of social life, social interaction and social cognition.

Proceeding from the morphological differences of social interaction in vertebrates, compared to social interaction in insects, we now focus on the study of social interaction and cognition in the former. Generally speaking, research on social interaction and cognition concerns how to understand and explain the behavioral processes related to interactions between conspecifics in various species. Some comparative ethologists such as Miklosi, Topali and Csanyi (2004, p. 995), for example, embrace a wide range of aspects in their definition of social cognition, which included:

...recognition and categorization of conspecifics and their emotions, the development and management of social relationships, ‘friendship, the acquisition of novel skills by interacting with conspecifics, (‘social learning’), the manipulation of others by means of communicative signals, the competence to perform joint cooperative actions and the question of ‘mind-reading’ skills’.

However, other social psychologists study how humans interpret and attempt to understand other human beings. Fiske and Taylor (1984, p. 1), for instance, characterize social cognition as “the study of how people make sense of other people and themselves. It focuses on how ordinary people think about people and how they think they think about people”. In accordance with this definition, Singer, Wolpert and Frith (2004, p. xvii), for instance, claim;

...the study of social interaction involves by definition a bi-directional perspective and is concerned with the question of how two minds shape each other mutually through reciprocal interactions. To understand interactive minds we have to understand how thoughts, feelings, intentions, and beliefs can be transmitted from one mind to the other ... how to communicate these thoughts (Emphasis added).

In addition, social psychologists like Quinn, Macrae and Bodenhausen (2003, p. 66) characterize research in social interaction and cognition as follows:

Social-cognitive research, with its adherence to the information-transmission metaphor, is fundamental to the study of process; that is social cognition is the part of social psychology that deals with the psychological mechanisms that mediate the individual’s response to
the social environment. At such, the nature of mental representation and the dynamics of information-processing are central topics of social-cognitive inquiry.

Furthermore, a closely related research area ‘mind-reading’ or theory of mind (ToM), aims to provide explanations for the ability to predict and explain the behaviors and actions of oneself and others. Whiten (2003, p. 376), for instance characterize ToM as “the everyday psychology that we use to understand and explain our own and others’ actions by reference to mental states, such as ‘desiring’, ‘knowing’ and ‘believing’”. The term was coined in 1978, when Premack and Woodruff presented their classical paper that questioned whether a chimpanzee has a theory of mind or not.

An individual has a theory of mind if he imputes mental states to himself and others ... A system of inferences of this kind is properly viewed as a theory because such states are not directly observable, and the system can be used to make predictions about the behavior of others (Premack & Woodruff, 1978, p. 515).

Accordingly, from the above definitions of social interaction and cognition, it is obvious that research in the social domain has been significantly inspired by theories and methods of cognitive psychology, taking an information-processing approach (see subsections 2.6-2.8, and cf. e.g. Fiske & Taylor, 1984; Kunda, 1999; Quinn, Macrae & Bodenhausen, 2003). It should be highlighted that a fundamental aim in studying social interaction and cognition, either in humans or animals, is to describe the “mental representations that emerge in the course of social interactions and how these representations affect and control behavior” (Miklosi et al., 2004, p. 995, but see also Augoustinos & Walker, 1995; Fiske & Taylor, 1984 to mention a few). In contrast, from embodied and/or biological perspectives on cognitive phenomena, Maturana and Varela (1987) portray different means of social interactions which result in various phenomena, and mainly distinguish between social and cultural phenomena. While they designate social phenomena as “those phenomena associated with the participation of organisms in constituting third-order unities” (p. 195), cultural behaviors are “the transgenerational stability of behavioral patterns ontogenetically acquired in the communicative dynamics of a social environment” (p. 201). Generally speaking, despite their emphasis on the importance of different means of social interactions for the emergence of these phenomena, they do not particularly discuss various kinds of social interactions. However, as the views on and definitions of social interaction and cognition described thus far indicate, there are different kinds of social interaction. In the following section, I discuss and portray various levels and kinds of social interaction fundamental for the emergence of social and cultural phenomena in more detail.

5.2 Various Levels and Kinds of Human Social Interaction

There are almost as many proposals for organizing the various levels and kinds of social interactions as there are researchers, but the differences lie principally in terminology and the sizes and numbers of parts, and not in
where the major ‘cuts’ are made. Despite the fact that most animals engage in social interactions, as in the previous wolf-hunting example, the issue of ‘intersubjectivity’ adds a different aspect to social interactions in humans. Generally speaking, intersubjectivity refers to the manifestation of shared meanings constructed by people in their interactions with each other. For that reason, intersubjectivity results in a basic discrimination between the self and others as well as the ability to compare and project one’s own private experience or mental state onto another person (Rochat & Striano, 1999). This means, the shared experience between two interactants is used as a way to interpret the meaning of the social situation, and in order to accomplish this second-hand perspective between “I – you”, it is argued that a sense of ‘self’ is necessary (cf. e.g. Bråten, 2003; Rochat & Striano, 1999; Tomasello, 1999). Briefly stated, they emphasize that intersubjectivity is the cradle of social cognition. Furthermore, it the discussion of concepts such as ‘self’ and ‘intersubjectivity’, the issue of consciousness is raised once again. According to Damasio (2003), consciousness and mind are not synonymous, and claims, “consciousness is the process whereby a mind is imbued with a reference we call self, and is said to know of its own existence and the existence of object around”(p. 184). In addition, there is also a distinction between intentionality and intentions. While intentionality refers to the quality of actions, done on purpose, intention refers to the agent’s internal mental states which represent such actions (Malle et al., 2001).

According to Bråten (2003), the main categorization of human social interactions, which now seems to be established in use, dates back to the work of Trevarthen and Hubley (1978), in which the distinction between ‘primary intersubjectivity’ and ‘secondary intersubjectivity’ was introduced. Since then, their categorization has been elaborated and differentiated further. Table 2 below illustrates some different categorisations of social interaction proposed in the literature.

<table>
<thead>
<tr>
<th>Table 2. Schemes for classifying different kinds of social interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary intersubjectivity</td>
</tr>
<tr>
<td>Secondary intersubjectivity</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

In order to explore these different kinds of social interactions proposed in the literature properly, I primarily proceed from Bråten’s (2003) generalization of social interaction and intersubjectivity, which served as the foundation, and the resulting categories are here termed as follows.
In subsequently portraying these three main levels of social interaction, I also incorporate the above categories of means and kinds of social interactions to provide a more detailed description of them, at each level. In addition, while the ontogenetical and aspects are not explicitly addressed in this section from an embodied perspective, we return to these developmental issues in section 5.5.

5.2.1 The primary level
The primary level is characterized by various kinds of dyadic interaction, which is the mutual and direct sharing of behavior and emotions, usually through facial expression, vocalization, gaze, and turn-taking (cf. ‘Primary level of sensorimotor attunement’, Bråten, 2003; ‘dyadic interaction’, Tomasello, 1999; ‘dyadic engagement’, Tomasello et al., 2005; ‘primary intersubjectivity’, Trevarthen & Hubley, 1978). The age span for this level entails the period of sensorimotor attunement from birth to one month, and the period of primary intersubjectivity between two to seven months (Bråten, 2003; Rochat & Striano, 1999). The work on the social bonding between infant and caretaker has roots to scholars such as Trevarthen and Bruner (Hendriks-Jansen, 1996; Shanker & King, 2002). Meltzoff and Moore’s (1977) classical findings revealed that newborn infants are effortlessly able to mimic tongue extension, mouth opening and lip protrusion behaviors demonstrated by an adult human, implying a close linkage between visual perception and motoric behaviors. Bråten (2003) suggests that this primary level of sensorimotor attunement and interpersonal communion results in an emergent sensation of self and other from the early weeks of life, which in turn enables and develops further in so-called protoconversations between infant and caretaker approximately the second month of life (Trevarthen, 1979; see Bråten, 2003; Tomasello, 1999; and Tomasello et al., 2005). Furthermore, mutual gaze refers to basic social behavior/means, denoting the skills of recognition as well as maintenance of eye contact, for which infants have a strong preference from a very early age. Mutual gazing is frequently used during protoconversations, in which infants’ look into the eyes of the partner.

During these face-to-face interactions, the interactants attend and attune to each others’ movements as well as emotional and facial expressions, creating modes of mutual immediacy. A central feature of protoconversations is its turn-taking structure, in which emotional displays seem to be the glue that holds the interaction together, and not its continuity. Additionally, the same emotion is often expressed through different behaviors by the interacting partners, while one partner might use vocalization, the other partner may instead display the emotion through facial expression (Tomasello et al., 2005). It should be noted, however, that Tomasello (1999) does not use the
term intentionality at all in the dyadic form of social interaction. He argues there is no ‘self’ involved in dyadic interactions, emphasizing that this kind of interaction is merely a form of ‘attentional-sharing’ and not ‘intentional-sharing’.

5.2.2 The secondary level

The secondary level is characterized by various kinds of triadic interaction, which designates the more conventional use of the term intersubjectivity, and entails joint attention to objects or states of shared social referencing, beginning around nine months of age (cf. ‘Secondary level of object-oriented intersubjective attunement’, Bråten, 2003; ‘triadic interaction’, Tomasello, 1999; ‘triadic engagement’, Tomasello et al., 2005). These triadic interactions involve a coordinated interaction between child, objects, and other people. As a result, a referential triangle of ‘shared’ or ‘joint’ attention, between the child, another person and the object or event to which they focus their attention emerges. Gaze-following is an essential prerequisite for triadic interaction, which is the rapid shift between looking at the eyes of another person, following their gaze, and focusing the look at the same distal object or person.

Typical joint activity behaviors are the giving and taking of objects, such as toys or other cultural artifacts, building block towers together, as well as conventional pointing and naming games (Tomasello et al., 2005). These abilities are consequently commonly termed joint (or shared) attention, which involves a whole complex range of social skills and interactions, such as gaze-following, joint engagement, social referencing, and imitative learning (cf. Tomasello, 1999). However, Tomasello et al. (2005) apply the concept, joint perception, instead of the commonly used term, joint attention for triadic interactions. They emphasize that in triadic engagements, such as building block towers, the interactions are mostly individual activities in the form of mutual responsiveness to sharing goals and perceptions of some external object in a triadic fashion. Despite the different terminology, triadic engagement, joint perception, and joint attention abilities, serve as a basic mechanism for sharing experiences with others, in one way or another.

According to Tomasello et al. (2005), however, there is a qualitative difference between triadic engagement and collaborative engagement, and this change emerges around 12-15 months. Generally speaking, the former can be regarded as ‘passive joint engagement’ whereas the latter is instead characterized as ‘collaborative joint engagement’. In the collaborative form of engagement, the issue not only concerns sharing goals and perceptions, but also coordinating and motivating the different roles in the interaction, in which there is an exchange of roles termed role-reversal imitation. Besides sharing perceptions with others, as in the triadic interactions/engagements presented above, collaborative interaction is characterized by the coordination of one’s perceptions with others through joint attention, i.e., choosing what to attend to in the perceptual space. Different forms of gestures, in particular pointing to a target of shared interest, are common ways of establishing ‘active’ forms of joint attention, or as Tomasello et al.
(2005) explain, pointing is an indicator of willingness to share attention with another person. However, there are different kinds of pointing, and a particular distinction is between imperative pointing and declarative pointing. While the former denotes a mere reaching behavior for an object of interest, the latter is characterized by an outstretched arm with a pointing index finger, i.e., pointing at something which has the function of drawing attention to particular objects out of reach for the interactant. As Tomasello et al. (2005) suggest, the basic motive for declarative pointing is to direct the other interactant’s attention only for the sake of sharing attention. Furthermore, the collaborative nature of pointing is also displayed in the ability of assisting others to achieve their goals by focusing their attention to the ‘right’ direction (Tomasello et al., 2005). In summary, there are several commonalities between the conventional use of joint attention and collaborative engagements proposed by Tomasello et al. (2005), such as the triadic nature. Indeed, by collaborative engagements they particularly stress the active, collaborative, and role-reversal aspects of triadic interactions in further detail.

5.2.3 The tertiary level
The tertiary level includes symbolically mediated interaction (beginning at 18–24 months) as well as reflective interaction (beginning at 3-6 years). While in symbolic interaction there are various explanations of the nature of language and symbolic interaction, here I mainly focus on language as a means for social interaction and communication. This means, symbolic interaction is characterized in this thesis as the view of ‘language’ as a conventional symbolic system for discourse; a view similar similarly stated by Myin and Smets (2002, p. 634), “language as a primarily interactive social practice” (cf. e.g. ‘first-order mental understanding of self and others in representational mediacy’ Bråten, 2003; ‘symbolically embodied social institution’ Tomasello, 1999; ‘coordination device for directing the intentions of others’ Tomasello et al., 2005). Tomasello et al. (2005), for instance, stress that language, in the sense of socially collaborative activity, is a certain form of social engagement, since it functions as a ‘bi-directional coordination device’. Indeed, it can be argued that symbolic interaction is a particular kind of triadic and/or collaborative interaction, but in addition, it involves representational and mediating aspects. This means, conversation is an inherently collaborative activity toward a joint goal during which the two different roles of listener and speaker, through reversal imitation, share perspectives on things. Similarly, Shanker and King (2002), citing Savage-Rumbaugh, emphasize that language as symbolic interaction is “inter-individual interactions that come to be coordinated through the use of words” (p. 613).

The representational nature of language is crucial, and Gärdenfors (2004), for instance, distinguishes between two different kinds of representations: cued and detached representations. Cued representations signify something that is at hand in the present situation, while generally speaking, cued representations are referred to as signals. Detached representations thus stand for entities that are neither at hand in the present situation nor elicited by some recent situation, and detached representations are referred
to as *symbols* (Gärdenfors, 2004, but see also the work by Vygotsky on signalization vs. signification in subsection 2.4.4 as well as Mead’s work on non-significant and significant gestures in subsection 2.4.6). It should be emphasized that both symbols and signals are regarded as tools/ a means of communication and while the line between cued and detached representations is indeed not fixed, but rather blurred, it directs our attention to certain features of the issue of representation (Gärdenfors, 2004). For that reason, the ‘language’ of bees (cf. subsection 5.2.1) is not a ‘linguistic’ language since it is based on cued signals, and is not socially shared and detached, nor has representational properties, as in the case with symbolic language. However, the mediating function in collaborative interaction through throphallaxis in social insects has similarities with the mediating function of language in the form of a ‘linguallaxis’, as Maturana and Varela (1987) termed it. However, the fundamental difference between throphallaxis and linguallaxis is the characteristics of the mediating ‘substance’. The *symbolic* representations used in linguallaxis are special since they are i) *intersubjective*, i.e., they are socially shared by others; and ii) *perspectival*, in the sense that each symbol singles out a certain way of considering some situation (Tomasello, 1999). This is not the case with the more pre-given, static and rigid chemical substances used in throphallaxis.

Language serves as an ‘interaction tool’ (see also Vygotsky’s work in subsection, 2.4.4 and Mead’s work in subsection 2.4.6) that liberates us from the immediate experience of ‘here’ and ‘now’ and then offers the possibility of viewing things and situations from different perspectives as well as providing metaphors and analogies (Tomasello, 1999; Zlatev, 2007). The third-party view or perspectival aspect of language, thus, not only functions merely as a way of social interaction, but also provides the ability to view oneself as one interacting partner among others in social discourse. Furthermore, Tomasello (1999) emphasizes that symbolic interaction is a very powerful medium of communication because it enables great specificity and flexibility of reference. Similarly, Gärdenfors (2004) points out although both humans and other animals cooperate to achieve complex societies and common goals (e.g. through stigmergy in social insects and in the earlier wolf-hunting example), the use of detached representations (symbols) enables humans to create *new* goals of cooperation, which subsequently result in new social organizations and institutions. He further argues that there is no need for detached representations in animal cooperation, since the common goal is usually already present in the actual situation, and for that reason, animals need not to form and share joint representations. On the other hand, however, if the goal is distant in time and space, there is a need to create a common representation prior to the ensuing cooperative and/or joint action. As Gärdenfors explains, “cooperation of detached goals requires that the inner worlds of the individuals be coordinated. It seems hard to explain how this can be done without evoking symbolic communication” (p. 4). Hence, symbolic interaction offers means of long-lasting cooperation, also towards future goals, and thus makes it possible for us to share visions and ideas (Gärdenfors, 2004).
However, in order to share visions of the future and use language, a theory of mind may be necessary—a topic which allows us to proceed to reflective interaction. This issue entails the abilities of mind-reading and to have a theory of mind (ToM), as a kind of inner dialogical thinking, that is, social cognition in general (cf. e.g. ‘second-order mental understanding of others’ minds’, Bråten, 2003; ‘inner-dialogic thinking’, Tomasello, 1999). Broadly speaking, without reviewing the extensive literature on the issue of social cognition, ToM, and mind-reading specifically, it seems that reflective interaction is about the understanding of others’ intentions and beliefs, that is, what others are doing, perceiving, intending, knowing and so on. This understanding of others’ mind provides the interpretative medium to determine what and why somebody behaves in a certain way, which depends on the goal and intention of the actor. Through mind-reading, the interactant takes new perspectives on a social situation by reflective meta-discourse, and it is argued by some that this is made possible through the use of language (cf. e.g. Tomasello, 1999; Tomasello et al., 2005).

Furthermore, Tomasello (1999), for instance, suggests that language is a form of cognition - emphasizing it is cognition packaged for purposes of interpersonal communication, and subsequently, the use of linguistic symbols might result in certain kinds of cognitive representations.

Besides interacting socially with others, one can also engage in collaborative social activities, that is, cultural phenomena. Accordingly, Tomasello and colleagues emphasize that both symbolic interaction and having a ToM are necessary for being able to engage in collaborative activities, which involve shared goals and socially coordinated action plans, that is, joint intentions. They stress that interactions of this type require not only an understanding of the goals, intentions, and perceptions of other people, but also, in addition, a motivation to share these things in interaction with others. Thus, according to them, this ‘sharing’ ability is the ultimate understanding of intentionality.

Accordingly, the consequence of this collaborative and socially shared engagement, through symbolically mediated interaction, is a new form of cognitive representations, which is referred to as dialogic cognitive representations, used in linguistic communication and other forms of symbolic interactions involving ‘shared intentionality’ (Tomasello, 1999; Tomasello et al., 2005). As a result, skills of cultural cognition emerge, in which activities are organized and mediated by linguistic symbols and social institutions. Besides the work by Tomasello et al. (2005), Decety and Sommerville (2003), for instance, also portray the issue of ‘shared intentionality’, which is suggested to serve as the foundation for joint actions’ (cf. e.g. Pacherie & Dokic, 2006; Sebanz, Bekkering & Knoblich, 2006). Furthermore, Tummolini and Castelfranci (2006) stress that “joint intentional action is done on the background of such shared intentional states”.

In summary, through subsections 4.2.1 – 4.2.3 much focus is on various kinds of intentionality and internal representations, while the role and
relevance of embodiment in social interaction is disregarded. This can be partly explained by the common view of regarding 'mind' as superior and/or independent of the 'body', which subsequently considers bodily social interactions (such as gesture, posture and so on) merely as the visible result and output of mental intentions. However, if we return to Maturana and Varela’s initial illustration of third-order-structural couplings (see Figure 3), we notice that the emphasis on interaction-couplings with the environment is rather absent in the above levels. Generally speaking, there is a separation between beliefs, intentions, etc., and the actual behavior in the presented levels and kinds of social interactions. Consequently, this leads us to the crucial question: How to access the other person’s intentions? As seen in the this chapter, particularly in the above section (5.2) and its subsection (5.2.1-5.2.3), the most common and dominant view in socio-cognitive research holds that human intentions and so on, are transmitted from one mind to another (cf. e.g. Miklosi, Topali & Csanyi, 2004; Quinn, Macrae & Bodenhausen, 2003; Rochat & Striano, 1999; Sebanz, Bekkering & Knoblich, 2006; Singer, Wolpert and Frith, 2004; Tomasello, 1999; Tomasello et al., 2005; Tummolini & Castelfranci, 2006). However, this implies implicitly that social interaction is a rather passive process between two Cartesian minds, as Gallagher (2005) expresses it (e.g. Tomasello et al.’s (2005) distinction between passive and collaborative engagement in subsection 5.2.2). It should be stressed, however, that there is no room for passivity in social interactions, since interactants are almost always active (e.g. Gallagher, 2005). In order to discuss this issue in more detail, we must consider the common metaphor for how communication and social interaction are consider within socio-cognitive research, and compare it with a more dynamical metaphor.

5.3 Different Metaphors of Social Interaction

In the following two subsections, two different metaphors of social interaction and communication are portrayed and compared. Since the information transmission metaphor is traditionally more used and well-known in socio-cognitive research, the dance metaphor is discussed more extensively since I want to portray its alternative ways of describing social interaction and communication in more detail, given that it is not the most well-known metaphor in the field of socio-cognitive research. Finally, some criticisms of the dance metaphor are presented, as well as some probable ways of developing it further are outlined. It should be pointed out that I here deliberately use the term ‘metaphor’ instead of ‘model’, in order to address the issue of social interaction and communication from a more general perspective.

5.3.1 The information transmission metaphor

Current theories of social interaction are still dominated by traditional mechanical and information-processing models of human communication, and most classical studies of interaction focused rather on the content in the interaction, i.e., information, than on its execution (Shanker & King, 2002). This view of social interaction and communication is usually labeled the tube, conduit or information transmission metaphor (ITM), which views
interaction as a *sequential* process in which agents interact through emitting and sending information to each other via different channels. Therefore, the ITM uses entities such as *signal* and *response, sending* and *receiving, encoding* and *decoding* (Shanker & King, 2002). Accordingly, interaction is treated as a rule-governed system with predictable casual sequences, an objective stance from the researcher’s point of view, mainly focusing on *one* individual at a time. In particular, how that individual processes and responds to the received signal(s), which then are “broken down” into objective and measurable units. Communication, then, is said to occur “when one organism (the transmitter) encodes information into a signal which passes to another organism (the receiver) which decodes the signal and is capable of responding appropriately” (Ellis & Beattie, 1986, in Shanker & King, 2002, p. 607). Therefore, ITM regards interaction as a linear, discrete state of events, meaning that an individual is either sending or receiving, such as two faxes or computers transmitting information back and forth. This view of interaction and communication has its origin in the work on the transmission of electronic signals by Shannon and Weaver in the 1940s (cf. sections 2.5 and 2.6). When Shannon described his method of converting sounds and images into binary strings as *encoding* the *information* in the sent *message* this metaphor was soon imported into the field of human interaction, since it fitted ‘hand in glove’ with the computer metaphor of mind (see chapter 2). Therefore, this metaphor of communication created an overwhelming impact on the way we conceive how communication and interaction actually function (cf. Shanker & King, 2002).

Nevertheless, the ITM has been criticized for several reasons. Gibbs (2006), for instance, points out that this view of communication disregards the fundamental problem of *how* meaning is grounded in everyday experience, similar to in the ‘symbol grounding problem’ (Harnad, 1990, in section 2.7), particularly how meaningful symbols relate to embodiment and referents in the real world. Monk (2003) also criticizes the ITM metaphor of interaction and communication. According to him, it does not provide the whole picture of interaction, since it ignores the social dimension of language use (cf. Clark’s (1996) theory of language use). This means, how a group of people use language in interaction, which is more than putting together the individuals interactants into a whole group. According to Monk (2003), in Clark’s theory, social interaction and communication is considered a collaborative activity, given that the spoken utterances are commonly ill-formed, imprecise, overlapping and contain many repairs and restarts. This is not the same as in the well-defined and correct codes in the ITM of social interaction. Furthermore, King (2004) emphasizes ITM fails to reveal the full story of social interaction, because its central assumption is that the signals containing the information flow back and forth between sender and receiver actually *oversimplifies* and *misrepresents* what actually happens in social interaction and communication. She stresses these processes cannot be reduced to so-called “social information transfer”. The main point here is that information is not an identified and discrete entity which can be sent, through signals, from one agent across time and space to another agent. As Maturana and Varela (1987, p. 196) point out, the tube metaphor of
communication is wrong, since “[b]iologically, there is no transmitted communication in communication. Namely there is not something that is communicated … Saying does not ensure listening and looking does not ensure seeing”. Similarly, Thompson and Valsiner (2002) argue this view of interaction leads us astray in search of ‘mythical substances’ shuttled between ‘mythical storehouses’. Instead, the urgent task is to discover how interacting agents, either animals or humans, accomplish originating another’s interactional pattern through their actions. Likewise, as Vauclair (2002) expresses it, in the conduit metaphor words function as containers for meaning, but whereas signals may be transmitted in a technical fashion, meaning is not. He argues that ITM actually falls short in explaining how interaction works, and its main problem is not its discrete and bi-directional structure of sending and receiving signals, or even encoding and decoding. Instead, the central failure is its assumptions that meaning is transmitted in signals. Hence, Vauclair (2002, p. 643) stresses that meaning is constructed, and “the construction process relies on environmental signals (words and gestures) and conceptual knowledge and processes. Signals, words, gestures and expressions do not mean; they are prompts for the construction of meanings”. A similar argument was put forward by Fogel (1993, p. 76) who stated that

...information is created in the interface between perception and action.
It is not just that I can perform an action to achieve a purpose that is meaningful, but it is what I discover about how I perform the action: what my arms and legs can do, my cardiovascular tolerance, my trust in my ears, and eyes to assist me. It is that last point, the salience of the body... is missing in many theories of meaning.

Therefore, alternative explanations of interaction and communication that address social interaction from a situated and embodied perspective are required, since the traditional ITM of communication can be viewed as a disembodied sender-receiver explanation of pre-given information, missing contextual and bodily aspects. There has been much focus on words or verbal aspects as meaning or information carriers, and the nonverbal and verbal aspects of information coincide with the separation of information and channels. In ITM, it can be argued that the body functions only as an input and output device, i.e., a physical interface between the external world and the internal mental processes. The interacting persons’ postures, movements, gestures and facial expressions are then disregarded as signals or channels for the information to be transmitted. This means, the signal is not a part of the cognitive system other than a wrapping for the sent message through different channels, and it supposes that both the signal and system is static and well-known. But people are not linear in their actions, and it is not possible to separate units that interact in a discrete and sequential manner, or as King (2004) expresses it, the ITM view holds that ‘the sum is equal to its parts’, a stance that is challenged by alternative approaches of interaction.
5.3.2 The dance metaphor

An exciting move within socio-cognitive research has occurred, from using the more well-known ITM to applying the less well-known dance metaphor (DM)(Shanker & King, 2002). The DM views interaction as a dynamical process grounded in dynamical systems approach (DSA) (cf. e.g. Fogel, 1993; Johnson, 2001; King, 2004; Shanker & King, 2002; Thelen & Smith, 1994). As King (2004) explains, social interaction should be considered as a dance, in which each partner must participate, moment by moment, in creating the coordination, which then will create a Gestalt effect, i.e., the whole is greater than the sum of its parts (see subsection 2.4.2). As Ingold (2002b, p. 627) explains “such a shift will ... release biological and psychological studies of communication from the straightjacket of hard-core cognitivism”. Generally speaking, according to the dynamical systems approach, interaction emerges when the situated and interactive partners are constantly interacting with and changing in respect to one another’s, (which emerges in an aggregate pattern of mutual co-action). According to Fogel (1993, p. 29), interaction is regarded as a "continuous unfolding of individual action that is susceptible to being continuously modified by the continuously changing actions of the partner.”

The main focus in the DM is on co-regulated interactions and the emergent and interactive behaviors in the particular social context (Fogel, 1993), particularly how the interacting agents constantly establish and sustain a sense of mutual rhythm and movement. Such an activity of mutual attunements is conducted cross-modally, adjusting to each others’ actions, facial expressions, tone of voice, vocalizations, and so on. Therefore, it has been argued that cross-modal co-ordinations of movements and vocalization in interactions are significant in order to understand how an individual’s particular experiences govern its behavior (Shanker & King, 2002). This means, the DM is blurring the line between verbal and nonverbal aspects of interaction, as well as the assumption that information refers to mental states or representations. The DM puts forward how difficult it can be to actually point out who of the interactive partners functions as the ‘initiator’ of the interaction, and much of that dance is carried out by affective and nonverbal means (Greenspan & Shanker, 2004; Shanker & King, 2002). This means, the participants in the interaction are viewed as continuously active and interactive, functioning as ‘sender’ and ‘receiver’ at the same time, looking at how the whole configuration changes over time (Shanker & King, 2002). Thus, the DM views interaction dynamically, and therefore the interaction pattern is non-linear.

In order to exemplify this dynamical unfolding, Shanker and King (2002) use the study of facial expression. Traditionally, basic emotions and stereotypical facial expressions (interest, surprise, sadness, and so on) have been studied scientifically since Darwin’s days (see subsection 2.2). It has been argued that the construction of these facial expressions, and their evoked responses, are controlled by genetically determined programs that are imitated through stimuli or reflexes that trigger the neural program of facial muscles (e.g. Ekman, 1998). Consequently, the study of facial expressions
has mostly comprised static and fixed expressions, supported by findings that infants are born with a certain ability to perform basic facial expressions, and children both blind and deaf from birth can express facial expressions such as smiling and showing anger (see Eibl-Eibesfeldt, 1989). However, there are problems with this stance, and as Fogel (1993, 28), for instance, expresses, “the apparent discreteness of the smile is actually a perceptual illusion”. Moreover, Shanker and King (2002, p. 607) argue “facial expressions of emotions are constantly changing in response to the changing dynamics of social interactions”. Accordingly, the so-called stereotypical facial expressions associated with ‘basic’ emotions actually develop within the context of socially shared emotional experiences. In other words, from the DM point of view, to study the development of stereotypical facial expressions, and the communication of emotions, looking at the whole picture, in which the complexity of individual’s changing facial expressions of emotion unfolds in co-regulated interactions is required.

According to Fogel (1993), co-regulation occurs when individuals’ joint activities intertwine and achieve a unique and mutually created set of social actions. It should be noted that co-regulation emerges as part of a continuous process of interaction, and it is recognized by its spontaneity and creativity. During interaction with others, we employ culturally accepted conventional means of movement and expression, and “information is created between people in such a way that the information changes as the interaction unfolds” (Fogel, 1993, p. 19). Shanker and King (2002) point out that the distinction between the concepts of co-regulation and interactional synchrony needs to be clarified. Interactional synchrony refers primarily to bodily actions in relation to speech. They further mention that empirical findings indicate that partners in a dyadic interaction are said to be in when they are in a similar affective state and attuned to one another’s communicative behaviors, which happens about 30% of the time. However, as Fogel (1993) explains, interactional synchrony focuses on the combination of elements from two individuals, and therefore overlooks aspects of social relating of crucial importance in co-regulation. This means, the main difference between co-regulation and interactional synchrony is that social partners besides synchronizing or harmonizing when interacting, also make meaning as they interact jointly. However, a common misunderstanding that co-regulation requires symmetry between the interacting partners exists. It has been argued that adults have greater motor-control and social impact, and it is therefore difficult to study interactions between infants and adults. Instead, the DM focuses on joint continuance of action between social agents, and does not insist on equivalence of action (Shanker & King, 2002).

Moreover, there is a strong emphasis on creative as well as rigid and cultural frames in the process of co-regulated interaction (Fogel, 1993). The concepts of frames, according to Fogel et al. (2002, p. 192) refers to “stable patterns of mutually coordinated activity related to the topic, setting and scope of the dialogue”. Generally speaking, they stress that frames are interactions that unfold into regularly recurring routines in order to affect mutual and coordinated actions. A typical cultural ritual, for instance, is the greeting,
which has a past, is culturally accepted, is expressed cross-modally, and requires maintenance by interactive partners in order to still be ‘alive’. However, it should be noted that recurrences of the same greeting or frame over time within social interaction is never exactly the same, despite its regularity (Fogel et al., 2002). There is more or less some variability since the interactive partners co-construct their interaction, moment by moment, rather like improvising than following a script or narrative. (This view is in contrast to the concept of script stated by Shank & Abelson, 1977 see section 2.6). Therefore, frames that do not change much over time are supposed to be rigid, whereas frames that vary are more creative. In order to act jointly with another person in interaction, it is a necessary to continually re-adjust one’s own actions in relation to those of the other partner. This is where creativity transpires, which is required in order to ‘keep up’ and still ‘stay in’ the continuing and unfolding dialogue, which in turn enhances development (Fogel et al., 2002). It should be noted that creativity requires one’s alertness to new interpretations of meaning in the context of interactions. For instance, a stereotypical movement could suddenly take on some whole new dimension of meaning, which demands that the co-participants have ‘open minds’ and are able to re-interpret and treat the significance of that innovative action (Fogel, 1993; Fogel et al., 2002).

Additionally, the shift of considering the nature of interaction as a co-regulated activity has implications with regard to the ‘content’ or meaning of the interaction. According to the DM, information is not pre-given, instead “mutual understanding is something that emerges as both partners converge on same shared feeling, thought, action, intention and so on” (Shanker & King, 2002, p. 608). This means, DM does not follow some predetermined format; instead such a process is intrinsically creative. Similarly, King (2004) emphasizes that information and/or meaning does not literally reside in the energy of a ‘signal’, but represents an emergent property of the combined attributes of the individual producing the ‘signal’, the individual perceiving the ‘signal’, and the circumstances under which the ‘signal’ is emitted, using the vocabulary of the ITM. In other words, the role of social interaction is to create and build up meaning through joint participation. Fogel (1993) s that meaning is then created by reducing degrees of freedom. He stresses that almost unlimited possible and potential patterns of interaction that could emerge when individuals interact exist. The main problem, however, is clarifying why a certain pattern emerges. He suggests complex systems, as time goes by, converge toward quite stable and identifiable patterns of functioning. This convergence occurs in accordance with how the components of the system constrain each other. Initially, the number of degrees of freedom in interaction is very high, which is based on the individual’s physical embodiment, their ability to create communicative actions, personality, past experiences and so on. Despite the many potential ways of co-regulated interactions, a quite stable and small set of interactive patterns emerge, which are both creative and rigid. Following this argument, when the degrees of freedom are reduced, meaning emerges. It should be added, however, that in certain situations communication and social
interaction do function as in the ITM, which is totally acceptable (Fogel, 1993).

The DM originates from interactivist approaches such as Piaget and Vygotsky (see subsections 2.4.3 and 2.4.4, but similar ideas can also be found in Bateson (1972) and Bruner (1990). Stated briefly, Piaget focused on developing through interaction with objects, and Vygotsky in a social surrounding, particularly in the zone of proximal development (ZPD), in which a novice is being guided by a more experienced partner. Most generally, the two metaphors portray two diametrically opposed views on interaction, and their main characteristics and differences are encapsulated in Table 3.

Table 3. The main characteristics of and differences between the ITM and the DM of social interaction (Adapted and altered from Shanker & King, 2002).

<table>
<thead>
<tr>
<th>Information transmission metaphor (ITM)</th>
<th>Dance metaphor (DM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear and discrete</td>
<td>Non-linear and dynamical</td>
</tr>
<tr>
<td>Interaction is rule-governed</td>
<td>The interaction is constrained by its partners’ past, ontogeny, biology and context</td>
</tr>
<tr>
<td>The information is about internal states or representations</td>
<td>The information is not hidden internally or about mental representations, but can be manifested in visible actions that expands the bounds of the skull</td>
</tr>
<tr>
<td>Static</td>
<td>Creative and rigid frames</td>
</tr>
<tr>
<td>Encryption process</td>
<td>Co-regulated and meaning making activity</td>
</tr>
<tr>
<td>Verbal interaction</td>
<td>Cross-modal interaction</td>
</tr>
<tr>
<td>Information is pre-given</td>
<td>Information is created in an emergent manner in meaning-making activity between interacting agents in a particular situation in a certain context</td>
</tr>
<tr>
<td>Disembodied</td>
<td>Embodied</td>
</tr>
<tr>
<td>De-contextualized</td>
<td>Contextualized</td>
</tr>
<tr>
<td>Separate individual’s who exchange information</td>
<td>Individual’s form a single unit creating meaning together</td>
</tr>
<tr>
<td>The sum is equal to its parts</td>
<td>The sum is more than its parts</td>
</tr>
</tbody>
</table>

However, Miklósi (2002), for instance, criticizes the DM excessively focusing on co-regulation, given that all our actions occur in real-time, which does not free us from the passing of time in co-regulation. Instead, Miklósi emphasizes that the main difference between the ITM and the DM is how they respectively treat interaction per se. According to him, the ITM views communication as a one-way process by not focusing so much on the receiver’s feedback to the sender, whereas the DM stresses the feedback effects by modeling communication as a two-way process. Obviously, there is not any prerequisite for regulation in the ITM, but in the DM co-regulation,
such as turn-taking, becomes a critical issue. The problems, as he expresses it, concern at which level of complexity the communicative interaction should be modeled, and he points out that “we need theories and means to describe the mechanism that contributes to dancing behavior” (ibid., p. 633).

Furthermore, Thompson and Valsiner (2002) suggest that the urgent task should be to discover how “communicating” animals go about designing one another’s behavior through their actions, that is, what they mutually produce. According to them, the DM disguises the role of agency, and they explain

...to call a social interaction a dance is to stress the peraction of social agents. When agents peract, they act through or by means of one another. Each has a state of affairs toward which his or her behavior is directed, and that state of affairs requires certain actions on the part of the social partner. The behavior of each actor is therefore directed toward using the other as a tool to produce a particular desirable result. The dialectic between their peractions is the dance (p. 641).

Finally, Waters and Wilcox (2002) believe that a cognitive approach of the DM is needed since such an approach would offer a better explanation of how meaning is communicated. They point out the danger of throwing out the cognitive baby with the information-transmission bathwater. Given that the DM does not address this issue in enough detail, they (p. 645) urge Shanker and King to

...incorporate both the mental and the social into their DS\textsuperscript{29} model. Communicative gestures occur in social settings; guided emancipation is a social behavior. Gestural signals, whether acoustic or optical, do not contain meaning; rather meaning is assigned to gesture by ape or human. The assignment of meaning to perceptual events, though it may be socially constructed, dynamic, and emergent, is also essentially a conceptual task (emphasis added).

Hence, so far, I have argued that the ITM limits the complexity of social interaction as well as how it should be conceptualized (cf. Shanker & King, 2002). Generally speaking, these two fundamentally different paradigms of research represent different theoretical approaches, and therefore afford different kinds of knowledge. Consequently, from a cognitive science perspective, two major points of interest are: i) the role of representations; and subsequently ii) methodological issues. It is appropriate to ask these questions because the DM should not be viewed as a new form of behaviorism (see section 2.3), disguised in ‘new clothes’. Instead of focusing on internal processes, the DM mainly focuses on acting agents in a particular context, and the main goal of studying interaction is to describe what agents do in certain social situations. Accordingly, their behavior should not be viewed as evidence of any underlying mental representations that enable them to behave in a certain way. Indeed, the principal issue at

\[\text{In their reply, DS is the abbreviation of ‘dynamical system’ which here is interchangeable with the DM.}\]
stake is whether the DM with its anti-representational models and explanations can shed any light on cognitive phenomena (Shanker & King, 2002).

5.4 Revisited: Characterizing and Studying Social Interaction and Cognition

In this section, I present and describe alternative views on representations and intentions, in which they are regarded as emergent products of social interaction. These subsequently lead us to view social interaction as relational. Furthermore, I return to the Distributed Cognition approach (Hutchins, 1995, mentioned briefly in subsection 3.2.1) and here portray it as a conceptual framework for considering and studying cognition ‘out in the open’ in social interaction. Taken together, this stance offers potential answers concerning the lack of cognitive representations as well as methodological issues of the DM of social interaction and communication.

Returning to the issue of the nature of social interaction and cognition addressed in the beginning of this chapter (section 5.1), we note that representations, intentionality, and intentions are crucial concepts in mainstream social-cognitive research. It is also argued that these aspects serve as the very foundation of social cognition in several ways. Most generally, individual actions and behaviors are seen in interpersonal terms, as a kind of passing between two Cartesian minds, as Gallagher (2005) explains. However, Gibbs (2001) argues that intentions are emergent products of social interactions, although as discussed throughout section 5.1 and its subsections, it is generally assumed the way we understand the intentions of others, either verbal or nonverbal, is through certain kinds of mental state, traditionally considered to be individualistic and singular (Gibbs, 2001). This means, many scholars tend to view intentions as mental acts that pave the way for the performance of behavioral acts. Given that intentions are viewed as psychological states, the content of the intentions has to be mentally represented. On the contrary, Gibbs (2001) points out that this view of intentions as “determinative, private mental acts that precede human behaviors does not capture the complexity of all intentional actions in social situations” (p. 106). Many examples of intentional meaning are in particular inherently vague and un-determined because they make little sense when claiming that the understanding of others’ intentions does not rest exclusively on recapturing the other person’s encoded intentions. Thus, “[I] argue that this traditional account of intentions must be supplemented by the idea that intentions are, at least, sometimes, emergent products of social interactions” (Gibbs, 2001, p. 106). Accordingly, intentions are, in many cases, viewed as emergent products of interaction between individuals, as well as interactions between the individual and the environment, and for these reasons they are distributed between individuals (see also the similarities to the work of Mead in subsection 2.4.6). Consequently, this stance has major implications for theories of intentionality within social cognitive science.
Morris (2005), for instance, discusses how dogs interact with each other, creating a shared world via an intercorporeal dance of unconscious communication, similarly to Mead’s previous description. Morris, however, suggests that instead of studying each individual dog’s behavior we should study the dogs’ intercorporeal dance as a whole, which demonstrates an interesting phenomenon beyond mere instinctual behavior and unconscious/non-significant communication. As he explains, the crucial point is that in order to show "it is biting that I am not doing" the individual dog must perform biting, at the same time as it does not actually bite. That is, the dog’s performance entails something such as a behavioral ‘as’ act, which demands that the other dog not bite back. This means, the behavioral complex ‘biting’ but not biting in the real sense, is a way of creating friendship. Briefly stated, the dogs perform an interaction pattern which stands for something else, given that the dogs’ dance transcends instinctive interaction patterns. Thus, from this perspective, the dogs’ interaction pattern as a whole actually blurs the distinction between non-significant and significant communication, to use Mead’s vocabulary, which leads us to consider the evolutionary origin of symbolic versus non-symbolic interaction.

The philosophical anthropologist Sheets-Johnstone (1999), for instance, discusses the difference between ‘symbolic’ and ‘survival’ behavior as the result of separating mind and body, where survival behavior stands for the physical functioning (e.g., nest building) and symbolic behavior is the mental functioning (e.g., building homes). The representational power is the ‘essence’ of symbolism, in which the ability of standing in for things that are not present at hand has a referential aspect. Sheets-Johnstone stressed that the emergence of symbolic thinking has to be considered by viewing mind and body as united, and not as two distinct entities, arguing that "analogical thinking is foundationally structured in corporeal representations" (ibid., p. 15). Presumably, these corporeal representations have been influenced by social interaction, but they must be physically possible to perform. Furthermore, she advocates that primatological studies confirm the idea and phylogeny of corporeal representations. She points out that the primatologist Carpenter already in the 1930s, suggested the tongue and mouth are sexual analogues to penis and vagina in a spatio-kinetic way, and calls attention to this behavior being present in other primate species such as langurs and howlers. Furthermore, Sheets-Johnstone stresses that this particular behavior is also present in humans of the tribe !Ko-Bushman. Apparently, this primary mode of symbolization is fundamental for all kinds of communication, not only sexual one (Sheets-Johnstone, 1999). Taken together, she claims there is plenty of evidence illustrating that ‘corporeal representations’ are the biological medium and the ability to represent meaning corporeally in the form of tactile-kinetic gesture is present all over the animal world, as seen in Morris’ dog fighting example. The main advantage of corporeal representations, according to Sheets-Johnstone, is that these behaviors need not be analyzed further since the unambiguous meaning is “embodied in the flesh”. In accordance with this remark, Sheets-Johnstone argues that an in-built semantic dimension in living bodies exists, which enables them to be a ‘biological medium for meaning’. Thus, as
Sheets-Johnstone explains, animate bodies are the primary source of meaning.

In order to understand how behavior which produces symbols, but is not symbolic itself functions, e.g., why pears and mountains can represent female breasts, we need to consider that the behavioral disposition toward iconicity is not conscious, according to Sheets-Johnstone (1999). Instead, she emphasizes “bodily symbols are structured not in reflective acts but in pre-reflective corporeal experience” (ibid., p. 90). It should be noted, however, that although the animate being itself is unaware of the symbolizing behavior as such, it is not unconscious of its actual behavior. For instance, Sheets-Johnstone uses von Frisch’s description of the dancing honeybee, and claims that the bee represents direction by its orientation to gravity, richness of food by the energy of the dance, and distance by spatio-kinetic shapes. Consequently, in communication, it is obvious there are gestures or series of movements that refer to something beyond themselves, played out corporeally, and this is the evolutionary framework for gesture (Sheets-Johnstone, 1999). Hence, her major claim is that human beings are not unique in their ability to perform symbolic behavior. She also maintains symbolic thought did not emerge from a void, since modern thinking and cognition is an extension of this biological matrix suggesting that analogical thinking is the cornerstone of symbolization, which is latent in corporeal representations. This line of argument has connections to Vygotsky’s view on language as the use of psychological tools or signs. Sheets-Johnstone, for instance, stresses that the origin of tool use is only a substitute of tooth use, using the body as the model, in which tools are analogous to teeth. This means, initially tools were used only as substitutive tools and not as a conscious symbolic act, since from the very beginning, the experience of their actual usage, either as teeth or tool use, was quite similar. Only later in phylogeny, would tools stand for something else, and possess representational characteristics.

Furthermore, Morris (2005) discusses the behavior of ‘tooling up others’, when he describes the creativity of interaction in a pack of wild dogs during rabbit hunting. He emphasizes that the dogs function as a group and show remarkably coordinated skills as they move together in anticipating the rabbit’s movements. Although hunting instincts drive the dogs individually, they act on these instincts as a group, in which their bodies are coupled by ‘lending a hand’ during the rabbit hunt, i.e., ‘tooling up’ each other. Thus, they use each other’s bodies as tools in collective behavior, and such spatially and temporally distributed couplings result in successful rabbit hunting. Morris stresses that while the origin of material (non-animate) tool use, such as sticks, can be found in animate helping hands when acting and performing tasks together, the human use of inanimate tools is the human way of equipping ourselves for ‘loner-being’ (cf. Sheets-Johnstone, 1999, above). Morris exemplifies this with the task of digging a big hole, which one can accomplish either cooperatively or individually by using a backhoe or spade. Hence, Morris argues that the biggest difference between animal and human collective behavior is that “a human can do individually what might
otherwise take a *group* of animals. But a human can only do the latter after learning, from other humans, language and ‘how to think’” (p.65). In other words, I suggest that the peraction as Thompson and Valsiner (2002) discuss in subsection 5.3.2 can partly be explained by the phenomenon of ‘tooling up’ others.

According to Gibbs, the traditional view of human intentions as exclusively private mental states in individual minds ignores the dynamic, interactive nature of intentional action, and he states:

> Most notably, the traditional view excludes the important role of others (listeners, readers, and observers) in intentional behavior and language use. The meaning of many communicative exchanges, for example, rests not just in their mind of a person, but emerges from the collaborative process of interaction between participants. Consider … the necessity of looking at intentions as emergent products of social interactions” (p. 109).

Gibbs addresses two major complications for the traditional stance of intentions. Firstly, it has been shown that speakers sometimes change their intentions in relation to the interpretation of the addressee. A speaker, for instance, may make an utterance with one intention in mind, but when the utterance is misinterpreted and the addressee misconstrues it, the speaker then adapts his or her own mind, acknowledging the new construal, and the interaction unfolds. The point is that a speaker’s intentions can partly be derived from the process of negotiating meaning in discourse (Gibbs, 2001). Similarly, Monk (2003) points out that Clark’s theory of language use bridges the cognitive and social aspects, with the central issue of his theory being *common ground*. Common ground means “the things we know about what is known by the person we are talking to” (Monk, 2003, p. 270). Clark and Brennan (1991) declare that all collective and coordinated actions are based on common ground and its accumulations, which is the case in communication. There are different kinds of common grounds, and collaborative activity is characterized as joint action, which is necessary for common ground to develop (Monk, 2003). Grounding interactants try to establish that what is said is also what is understood. Different shapes of the grounding process occur in different situations, but grounding is the fundament of all communicative and collective actions (Clark & Brennan, 1991). Furthermore, although we usually regard ourselves as the agents of our actions, but we rarely act independently, given that many of our seemingly individual actions are instead better characterized in the collective than on merely individual terms. According to Gibbs (2001, 2006), collective behavior does not equal a summation of individual actions, since when humans engage in collective activity, individuals are guided by collective *we-intentions*, which are not based on individual intentions (see subsection 5.2.3). Instead, collective intentionality means that each person believes the other individuals present are possible candidates for collaboration. Thus, there is a fundamental difference between individual and collective intentionality. The collective intention is partly the emergent product of the interactions between individuals and cannot be reduced to each person’s
own intentions per se, since when the issue of intent arises, interactants negotiate and construct a mutually shared social realm with the individuals as well as relational implications (Gibbs, 2001).

Secondly, the work of cultural anthropologists addresses another problem with the traditional view of intentions. As Gibbs summarizes it, the underlying assumption in the traditional view of intentionality is not shared across different cultures, since the main focus on an individual’s intentions by most scholars rather reflects a Western white middle-class bias about the nature of selfhood than a universal phenomenon (similar ideas are also found in Bruner, 1990). Citing Ochs (1984, in Gibbs, 2001, p. 114), he states:

...the emphasis on personal intentions in Anglo society and scholarship is tied to a cultural ideology in which persons are viewed as individuals, i.e., coherent personalities, who have control over and are responsible for their utterance and actions.

Despite the fact many cultures appear to focus on an individual’s subjective mental states such as thoughts, beliefs, desires, and intention etc., some focus instead on the consequences of the individual’s actions, and not primarily on what they intended to communicate. Indeed, it might be argued that individual intentionality is one of the ‘holy cows’ of Western thought, but it actually overemphasizes the psychological state of the individual at the expense of the social context in which the actions unfold (Gibbs, 2001). Samoan cultures, for instance, are societies in which the responsibility of the individual’s actions to interpersonal relationships and situations rather than to single person, is a fundamental attribution. It should be noted, however, that this stance does not imply the Samoan culture does not attribute certain behaviors to individual persons, but only that the causal foundation of these behaviors is not related to some person’s presumed internal mental state (examples from other cultures can be found in Gibbs, 1999; but see also Farnell, 1995; Fogel, 1993; Rogoff, 2003; Rogoff & Lave, 1984 for how cognitive abilities generally are shaped by socio-cultural factors).

Furthermore, Morris (2005) mentions that Waal (2001, 2003, in Morris, 2005) makes a similar remark concerning cultural differences in the history of primatology, between Euro-American and Japanese attitudes with regard to conducting primate research. While the former is inclined to study individuals, the latter is inclined to study groups. In other words, cultural preferences influence how scientific work is conducted, either in primatology or social cognition. Gibbs’s relational view of social interaction and intentionality has, in my opinion, strong similarities to the earlier work of Mead (see subsection 2.4.6). Furthermore, King, who is one of the initiators of the dance metaphor, defines meaning as “constructed through action between social partners rather than through transmission of ideas from one mind to another” (2004, p. 6). This stance is quite similar to Mead’s view which she also acknowledged in her writings.

In order to demonstrate the ‘illusion’ of intentional actions, Gibbs offers the ‘duckling’ example, also portrayed by Hendriks-Jansen (1996). Initially,
Konrad Lorenz (in subsection 2.4.1) described the parenting behavior of mother ducks through various and independent activity patterns, each of which with its own and distinct behavioral triggers in the environment. (The crucial point is that all these behaviors converge on the actual duckling, and not from some internal representation in the mother duck’s head that corresponds to the particular duckling ‘out there’.) This means, the traditional notion of regarding the mother duck’s parenting behavior as internally represented leads us astray, since the parenting behavior is the emergent outcome which exists neither alone in the mother duck nor in the duckling but in their action-perception couplings with the environment. Although the duckling example is illustrative, and it can be argued to have nothing to do with human intentions and intersubjectivity, there are several examples from the human social domain; such as the emergent and intentional behavior of brain, bodies and artifacts in the work of Distributed Cognition (Hutchins, 1995), and studies of mother-infants interactions (cf. e.g. Gibbs, 2006; Johnson, 2001).

The Distributed Cognition approach (DC) proposed by Hutchins (1995), considers how to understand the organization of socio-technical systems, in which the object of study is the way cognition is distributed between people, and the tools they employ. The main focus is on the ways information is propagated and transformed through the different media at a system level. According to Perry (2003), the DC, however, merely extends the traditional notion and theoretical framework of cognition as computationalism, since it still uses the notions of representations and representational transformations for describing human cognitive activity in larger units of study. According to Perry (2003, p. 194) “researchers trained in cognitive science do not have to abandon their theoretical knowledge and conceptual apparatus to understand distributed cognition”. The main difference from computationalism “is in its theoretical stance that cognition is not just in the head, but in the world (Norman, 1993) and in the methods that it applies in order to examine cognition “in the wild” (Perry, 2003, p. 194). However, although Hutchins’ theoretical framework is based on traditional notions of computationalism, it is modified in order to be applicable to the whole socio-technical system as the unit of analysis rather than the single individual’s mind 30. Consequently, taking the whole system as the unit of analysis makes it possible to observe the different kinds of representations, visible or invisible, which are fundamental parts in the socio-technical system. As Johnson (2001) explains,

30 The issue whether DC should be considered as computationalism or not, are not the major focus in this thesis. The interesting point here is the framework’s system level of analysis, and the implications for studying social interaction and cognition from such a perspective.
...in this view, cognition is expanded from an individual enterprise to a distributed activity that involves a variety of socio-cultural elements, including the behavior of multiple individuals, their use of objects, and their shared histories. In such a model, the unit of analysis is typically not mental structures in individual minds, but “real-time” interactions between the various participants and their environments... communication, itself, is a ‘cognitive” process’ (ibid., p. 167).

This means, contrary to viewing cognition as internal processes, the social interactions and materials comprising such systems are considered to be directly observable cognitive events. That is, DC extends the unit of analysis beyond the individual mind. With this crucial change in perspective, much of cognition previously hidden ‘inside’ the skull has now become apparent. According to Johnson (2001), given that much of cognition from a DC perspective is ‘apparent’ in the on-going interactions of the participants, researchers are actually able to observe cognitive events played out in real-time, dynamic social interactions. In other words, “the cognition we need to study has become a group event” (Johnson, 2001, p. 170). Therefore, DC offers tentative suggestions how to methodologically study meaning and cognition, and then incorporate these issues into the dance metaphor of social interaction.

Nevertheless, Johnson (2001) points out that adopting the stance that cognition is best studied as an observable and distributed event rather than an invisible, mental one, does not prevent us from recognizing the role of mental representations. Although the focus in DC is on observable behavior, the DC approach can, to some degree, be used to deepen our understanding of the mental events that come to be connected with the observed actions (Johnson, 2001). In particular, mental representations and events are not assumed to generate or drive the behavior in DC, but to be the product of it. Consequently, by considering mental functions as the outcomes of social interactions, contrary to standard approaches in which mental functions or representations are used as the ‘explanatory tool’ for the actual behavior, DC follows in the footsteps of Vygotsky (see subsection 2.4.4), which Hutchins points out himself.

Returning to the previous levels of social interactions presented throughout section 5.2, the observable social behavior is primarily used as a source of inferences in order to relate the mental representations of intentionality that constitute cognitive processes. According to Johnson, one problem with this inferential mapping is faced in research on imitation, since different types of imitation are primarily distinguished by the types of representational processes supposedly underling them. She emphasizes that researchers will soon lose themselves without any behavioral criteria which is able to distinguish one kind of imitation from another. Instead, she argues that DC, with its focus on observable interaction as being cognitive events in themselves, side-steps these issues to some degree, since its focus is rather on the capacity to do than on the capacity to represent. This means, instead of mapping the actual behavior to its underlying representational format,
research on imitation has to consider the *fidelity* and *novelty* of the reproduced behavior in form of co-regulated interactions.

Johnson points out that empirical data on co-attention, for instance, are rather absent from nearly all accounts of imitation research. She emphasizes that this kind of “information is of special interest since contingencies between co-attention and motor-activity are of particular importance in human cognitive development” (ibid., p. 172). In other words, Johnson stresses the ability to see others as ‘intentional’ as a prerequisite for the types of actions that constitute social interaction and cognition generates unnecessary problems for a parsimonious explanation of these abilities. In particular, the main problem, according to her, is not the idea but rather the *role* of mental representations in these processes. Generally speaking, these intentional models of social interaction assume that when children interact in similar ways to their more skilled interactants, their behavior reflects essentially the same type of representations evoked in their caregivers, e.g., various form of intentionality. This was the case in the examples of different levels of social interaction presented previously in subsections 5.2.1 – 5.2.3, in which the general explanation of these interactions rests on the assumption that the caregivers behave as if the infants are, to various degrees, intentional beings (cf. Tomasello, 1999; Tomasello *et al.*, 2005). Furthermore, Gibbs (2006) identifies another problem with most developmental studies and their conclusions. He argues that despite their ‘brilliance’, they usually consider children as passive observers who understand and experience the social and material world by “visual inspection of real-world events” (ibid., p. 8), and not as actively interactive agents in the ongoing and co-regulated social interaction.

In contrast, taking a DM/DC approach, children would be ‘treated as if they could participate’ (Johnson, 2001). This means, from the very beginning and whatever it does during mother-infant interaction, the caregiver treats the infant by taking turns in a ‘conversation’, filling in the gaps, offering the infant opportunities to respond, and so on, as the ‘dialogue’ unfolds. Initially, the caregiver provides all the meaning, but when the infant is able to fix its gaze and coordinate its action-perceptual couplings, the caregiver usually interprets the infant’s behavior as intentional, such as the grasping behavior which is transformed into act of pointing demonstrated by Vygotsky (see subsection 2.4.4). Thus, as Gibbs (2001) states “[T]he intentionality in the mother-infant interaction does not reside in any individual mind; it emerges as a product of their social interaction. Thus, what is intentional about the mother infant interaction cannot be explained simply in terms of the mother’s and infant’s intentions with respect to each other” (ibid., p. 120). According to Gibbs, by treating the infants as intentional beings, the caregivers ‘bootstrap’ and scaffold them into a socio-cultural environment, which partly rests on the “illusion of intentionality”.

Hence, from a methodological perspective, Johnson emphasizes that the DC model is ‘made-to-order’ for studying this process, because it regards cognition as being created through interaction and manifested in the
observable dynamics of the group. It should be said, however, that although the DC/DM approach is influenced by the work of Vygotsky, it exceeds the unit of analysis of Vygotsky’s so-called ‘zone of proximal development’ (ZPD) (cf. subsection 2.4.4). According to King (2004), much of today’s research following in Vygotsky’s footsteps, still focuses on what happens between agents in the turn-taking and coordination. Instead, both the DC approach and King (2004) exceeds that unit of analysis, i.e., two separate individual’s who exchange words and gestures. Instead, as she explains, they together form a single unit creating meaning together – a unit that in sum is more than its parts, embracing the very creativity and transformation underdeveloped in prior interactionisms’ approaches (but not in Rogoff’s (2003) work on guided participation). Moreover, in the DC approach all observational descriptions are situated in context, and the flexibility and novelty of the behavior offers a straight measure of complexity. However, Johnson points out while the DC approach presents its own methodological challenges, there is one major advantage in making this methodological and conceptual shift. By mapping internal representations to the behavioral outcome in order to study the ongoing participation of co-regulated interactions, allows researchers to side-step some of the difficulties that arise in justifying inferences to unobservable mental events and representations (Johnson, 2001).

In summary social interaction is a relational process in which meaning and intentions are emergent products of social interaction, and in most situations they can be viewed as a kind of distributed phenomenon rather than as individual private mental acts. In other words, we should not consider intentions to be ‘in there’ but instead ‘constructed’ between people and their surroundings. Furthermore, Gibbs’ emergent-interactive view of intentions as well as Hutchins’ distributed cognition framework fit ‘hand in glove’ with the alternative interactivist approach of embodied cognitive science portrayed in chapter 3. Their emphasis is, however, on the socio-relational side rather than on the embodied side of the interactivist coin. For instance, despite the emphasis on interactions between agents and their social surroundings, the DC framework offers little on the embodied nature of human cognition, and is currently peculiarly ‘disembodied’, a fact Hutchins (2006) admits himself. Indeed, Fogel (1993), for instance, emphasizes that embodiment is the basis in co-regulated behavior, its role and relevance is poorly understood and needs further explanation. The following section addresses the ontogenetic perspective from an embodied and relational perspective.

5.5 The Role of the Social Body in Motion in Cognitive Development

The ontogenetic perspective is important since one cannot ignore the complex processes that are involved in the development of a new-born child to a grown up adult (see e.g. Hendriks-Jansen, 1996). However, Rogoff (2003), points out that much research on children’s social learning processes is mostly concerned with ‘socio-cultural’ aspects. However, ‘biological’ aspects of this activity could be the focus of analysis in associated research, and “in this way biological, sociological, and individual aspects of human functioning
can all be seen as contributing to the overall process, rather than as rivals, trying to cut each other out of the picture" (Rogoff, 2003, p. 62). In my opinion, taking an embodied and relational approach, as done in this thesis, is a tentative step in favor of Rogoff's request.

As previously seen in this chapter, the interest in mainstream research of social cognition usually focuses on the outcome, that is, when children reach certain stages or pass different tests, but not on how these capacities develop. Furthermore, the research is rather disembodied. Consequently, in order to provide an embodied perspective on the various ‘mentalistic’ explanations for the levels of social interaction and cognition portrayed in section 5.2, I here offer a developmental perspective. This means, socially embodied actions develop through an unfolding process of action-perception couplings with the physical, material, social and cultural environment. As these socially embodied actions progress, socially interactive cognition builds on earlier advancements and adaptations to the possibilities and constraints of various kinds of action-perception couplings. The relational nature of social interaction is, as Johnson (2001) claims, both the source and cause of what ultimately must end up ‘inside the head’, following in the footsteps of Vygotsky and Mead (see subsections 2.4.4 and 2.4.6). Most generally, socially interactive cognition is always relational, first external and visible, and only later internalized and invisible, and studying these transformations might best capture social cognitive development. Furthermore, I suggest that the four functions of embodiment in social interaction presented in section 4.4 provide the mechanisms and/or functions to explain how socially interactive cognition develops, and therefore I do not explicitly refer to these embodied aspects in every subsection that follows. Thus, I integrate the embodied nature presented in chapters 3 and 4, with the relational nature of social interaction presented in chapter 5 to child development. It should be worth mentioning, that I do not intend to offer a ‘complete’ picture of the ontogeny of socially interactive cognition, but to emphasize and highlight some crucial aspects of embodiment in the course of development.

Prior to that, however, the use of the term ‘level’ could be questioned. Fogel (1993), for instance, emphasizes that the idea of levels in development should not be regarded as discrete phases or stages, but as metaphors for development rather than as accurate stages of development (as seen in the work of Piaget in subsection 2.4.3). This means, a continuous model instead of a discrete model of socially interactive cognition is adopted (cf. Fogel, 1993). The distinct levels could instead be regarded as more stabilized crossmodal interaction patterns (cf. Fogel’s consensual frames in subsection 5.3.2) which have emerged through continuously changing and co-regulated processes, and not as the result of the controlling influence of some internal representations (Fogel, 1993). Furthermore, affective and emotional aspects, for instance, play important roles in regulating social interaction (e.g., Greenspan, 1997, Hendriks-Jansen, 1996). As Greenspan (1997) explains, “emotions, not cognitive stimulation, serve as the mind’s primary architect” (p. 1), and the most crucial role for emotions might be to create, organize and coordinate the majority of our cognitive functions. He emphasizes that
every sensation that children notice also gives rise to affect and emotion, and as infants’ experiences increase, sensory impressions become increasingly tied to feelings (cf. Damasio, 1999; 2003). Briefly stated, paying attention to the subjective state of our body will almost perceive an emotional tone within one’s own body. This inner emotional mood continuously adapts to in the innumerable variations that are used to categorize, and, most important of all, make sense of our experience. This process involves the whole body, since the emotions are created and brought to existence through the expressions and gestures we make within the muscle system of our faces, arms, and legs – smiles, frowns, slumps, waves and so forth. Thus, this dual coding to understanding explains how emotions organize intellectual capacities, and subsequently create a sense of self (Greenspan, 1997). Instead of using the term dual coding, as Greenspan does, I take an embodied point of view which unites rather than bridges these aspects.

5.5.1 Initial cross modal interaction and the primitive ‘self’

Human infants are ‘ultra’ social already from birth, and it is suggested that the role of these early social interaction patterns is 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). Thus, children are treated as if they could participate (cf. section 4.4) Furthermore, these social interaction patterns are cross modal and relational from the very beginning, given that Meltzoff and Moore’s classical finding of neonatal mimicking portrays a close link between action-perception already at birth. This shares a basic neural format, probably through the mirror neuron system, although researchers do not yet know whether it is present at birth or not. From an embodied perspective, this cross modal connection seems to be a basic capacity that children are born with or develop very early. This crossmodal linkage is crucial for the development of social interactive skills, given that it establishes interpersonal bonds between ‘you and me’ as a kind of primitive self-other ‘space’. The capacity to automatically establish a meaningful interpersonal link emerges around two months of age, providing the basis for empathy (Gallese, 2003), in which embodiment functions as a social resonance mechanism.

Furthermore, children make sense of their sensations, through a course of discovery of how to organize their sensations and their bodies’ response to them, such as guiding their movements of the head, arms, eyes, and so on (e.g., Fogel, 1993; Greenspan, 1997, and see also section 3.3 ‘Body in Motion). Through guiding these movements, children are able to establish social interaction patterns with others. This means, caregivers interpret these movements as initial signs of intentionality, reading desires, wishes motives, and so on. At this point in time, bodily movement is one of the major buds of desire or motivation. With a growing ability to accomplish more advanced action-perception loops, through the increasing ability to coordinate its muscles due to the environmental constraints, ‘communicative’ movements result. Children are able to participate in affective but non-symbolic socially embodied interaction through gestures, gaze, facial expressions and pre-verbal vocalization. These relational
interactions develops from two to three sequences of interaction to nearly fifty to sixty during the first year of life, resulting in a more developed sense of pre-self (Greenspan, 1997). Thus, embodiment serves as both means and end in relational social interaction.

5.5.2 The shift from dyadic to triadic interaction and the emergence of ‘self’ 31

As stated previously by Tomasello (1999), the so-called ‘nine month’ revolution marks the shift from dyadic to triadic interactions (see subsection 4.2.2). However, while that describes the shift in understanding that occurs at about nine months, it does not explain how and why it occurs. Tomasello admits that the personal experiences necessary for this understanding remains unclear, and therefore the question can be raised - How does this link between self and others emerge? Tomasello notes that in coming to understand others as intentional agents, another crucial factor arises; the ability to simulate the other person’s intentional actions by analogy to one’s own actions, which as a result causes the self to become intentional. Tomasello stresses 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. However, the idea of understanding intentions by simulating the other person’s viewpoint leads to the question: How do children distinguish between the experience of their own actions and the actions of others? The ‘explanation’ Tomasello (1999) offers is that the time when the child starts to understand that other people have intentions and goals such as itself, is a result of our species’ ‘ultra’ social ability. On the contrary, I suggest that neither this nor simulation theories alone can explain how this intentional understanding emerges in the child. Instead, I argue that self-initiated and self-experienced locomotion is required for the emergence of a social understanding of the self.

As described previously in section 3.3, the body in motion is a rather neglected factor, and here the significance of self-induced locomotion in the child’s social and emotional development is portrayed. It should be stressed that locomotion is not necessarily a direct causal factor. Instead, the child’s cognitive and emotional changes emerge from the experiences that result from the child’s own locomotion capacity. When the human child starts 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 child must adapt to the new situation, paying close attention to the environment and to self-induced movement with respect to that environment. The consequences of this affect the interaction between the child and its surroundings. Campos et al. (2000) use an analogy based on a French saying, which states that ‘when the finger points at the moon – the idiot looks at the finger’. The empirical data they present suggests that children without self-induced locomotion experience still perform such as the ‘idiot’ in the saying, whereas children with locomotion experience are able to follow, to various degrees, referential gestures towards a distal target (Campos et al. 2000). Hence, their proposal is that crawling is the cradle of

---

31 This section has appeared in similar form in Lindblom & Ziemke (2005a, 2005b, and 2006).
the ‘social referencing phenomenon,’ given that it enables children to receive social signals that have an obvious distal referent. As a result, there is a sudden increase in the behavioral pattern of checking back and forth to the caregiver. This behavior is a crucial feature of the meaning making activity of social referencing, which makes it possible for children to understand how distal communication affects the regulation of social interaction.

Hence, it is via these regulations of interaction that children develop shared meanings with the caregiver. At around nine months children are able to respond to gestures when the target is not visible. This means children are able to differentiate between their own visual field and that of the gesturer and, in so doing, display a nascent ability for perspective taking. This social ability then develops further and encompasses communicational signs from others, which enable children to grasp that other people also have intentions. For example, while performing Piaget’s ‘A-not-B-task,’ crawling infants appear to be more attentive and more actively searching for communicative signals from the experimenter (Campos et al. 2000). Children with locomotion experience perform better on tasks assessing the tendency to follow referential gestural communication than pre-locomotor children (e.g. exhibiting gaze-following, head turning, and pointing). Similar results were found in studies with Chinese children, who begin to locomote at a later time than European/American children owing to cultural factors. Similar results were also reported in children having motor disabilities which delay their self-induced locomotion (Campos et al. 2000). Taken together, empirical evidence shows that there is a significant developmental change in referential gestures around the age of nine months and that self-induced locomotion experience is involved in that shift.

Thus, the answers to - ‘why does this revolution take place at about nine months?’ and – ‘how do children create and distinguish between the experience of their own movements and those of others?’ are as follows. The guiding issue here is the role of self-induced locomotion for the emergence of the ‘nine-month revolution,’ the point in time when the child begins to understand that others are also intentional beings. Hence, that point in development when children begin to understand themselves and others as intentional, ‘coincides’ with the onset of self-induced locomotion. I propose that this is not a coincidence. Instead, it is primarily through the experience of self-induced locomotion and perceiving the (physical, emotional and social) world and acting upon it that infants develop the capacity of understanding others as having different perspectives and intentions. This means when children begin to locomote, they acquire an individual experience of the surrounding world through their own actions, emotions and perceptions. As a result, a primitive self emerges as they begin to distinguish more clearly between self and world.

Consequently, when children can put themselves in someone else’s place, they can relate to both their own perspective and the other person’s. This perspective-taking is grounded in the experiences of self-induced locomotion, which might be a fundamental aspect of distinguishing between the
movements of self and other. This emerging understanding is bootstrapped through socially-scaffolded bodily experience, which gives the child access to the meaning of the situation. Subsequently, that understanding of perspective-taking could be used during embodied simulations, making it possible for the child to simulate offline what it would be such as to be in the other person’s situation, based on its own self-induced locomotor experience. In that sense, the sensorimotor and social dynamics of bodily experience function as a crucial driving force in cognitive development. But self-locomotion is not solely responsible for this emerging social understanding, given that most living creatures are able to move around by themselves. The ability to perform and distinguish embodied simulations and/or embodied practices of these experiences are also required.

Concerning intentionality and the shift from movement to action, the emergence of these abilities is based on the view that the social environment functions as a scaffold for the development of pointing, transforming a simple bodily movement into an intentional action (see the work of Vygotsky in subsection 2.4.4). Thus, cultural customs affect, but do not determine, the organization of social interactions under the constraints of embodiment (cf. Farnell 1999). Accordingly, Greenspan (1997) argues that when children are able to connect sensation and emotion to intentional action, they proceed to a more advanced level of intentionality, in which increasingly more complex pre-symbolic communication guides the way of social interaction. As he explains, while movement provides a means of defining and expressing wishes or desires, the combination of affect and purposeful movements is the crucial aspect that characterizes intent.

During development, children are able to recognize and create larger and more advanced patterns of interaction, though creative and rigid frames as ways to create meaning (Fogel, 1993). Furthermore, children are also capable of reproducing whole patterns than mere bits and pieces of different frames. In greetings, for example, Greenspan suggests that children learn to say ‘hallo’ not by a set of rules but by associating saying ‘hallo’ with the particular emotion of seeing someone his family knows and the accompanying friendly feeling. Thus, the social practices and pragmatic situations of social interaction and cognition, as well as their defined roles (cf. Mead in subsection 2.4.6), play a significant role in the process of shaping children’s understands of others. That is, the social and cultural environments function as scaffolds to characterize meaningful actions (e.g. Fogel, 1993; Gallagher, 2007b). According to Greenspan, one key to the puzzle of categorization lies in the fact that emotion organizes experience and higher cognitive functions, thus offering a solution to how we are able to generalize. Most generally, the ability to recognize and experience these emotions through embodied resonance functions as the mechanism that makes it possible for us to navigate through the social sphere, before children have acquired language. Greenspan points out that when the ability to recognize larger interaction patterns appears, at around 18 months, children are more skilled in ‘reading off’ the feelings of others and learn to deal with them in a more effective way, through the course of social
interaction. Younger children know how others will react to them, but not because they can think by mental concepts, but because their emotions are linked into extensive patterns based on experience, which provide them with clues for how to understand and interpret the actual situation. Thus, children’s experienced emotions, which are now linked to frames of interactions offer continuous ‘commentaries’ on their acting in social interactions in relation to the experiences of their own social bodies in motion.

5.5.3 Symbolic and reflective interaction enter the scene
Another fundamental milestone that has close links to the social body in motion is language, given that language production itself is based in quite complicated muscular movements (cf. Iverson & Thelen, 1999). From an embodied point of view, Özcaliskan and Goldin-Meadow (2005), for instance, stress that gesture provides stepping-stones to more complex linguistic constructions, as in moving from one word constructions to multi-words constructions. Children produce one-word utterances at a certain period in language development, but the ability to combine them into sentence-like strings takes a considerably long time. During this time span, however, children use gesture as a means of expanding their communicative repertoire, and they use them to convey increasingly complex ideas (Özcaliskan & Goldin-Meadow, 2005). Thus, these gestures function as psychological tools à la Vygotsky (see subsection 2.4.4). Furthermore, gesture is the first indication of development in a variety of tasks, including early language-learning. For instance, children are able to produce a ‘predicate + argument’ construction through gesture-speech combination. This indicates that their inability to produce the construction entirely in speech does not stem from a lack of understanding how to do it, since children can communicate this relation cross modally (Özcaliskan & Goldin-Meadow, 2005). Özcaliskan and Goldin-Meadow (2005) offer several possibilities for manual preference. Firstly, it might be easier to convey meaning in the manual modality than in the verbal modality, given that children use gesture in word-like ways several months before they make word-like sounds for the same purpose. Secondly, it might require less fine motor control of the hand to produce recognizable manual gestures than using the mouth and tongue to produce recognizable sounds. Thirdly, gesturing may put less effort on memory than words and signs, because their ‘conventionalized’ and abstract forms must be memorized and recalled in situ. The form of a pointing gesture, in contrast, does not vary with its referent, and it is easier to embody physically and to remember. Even iconic gestures can be generated at once with the resources available at the moment (Özcaliskan & Goldin-Meadow, 2005). This means, children might find it easier to embody their language constructions through a spontaneous gesture than through a conventional speech utterance. Özcaliskan and Goldin-Meadow’s (2005) findings put gesture as an integral part of early language development, given that gesture both precedes and signals oncoming changes in speech. Thus, as they explain, gestures are way stations on the road to abstract rules, and this process culminates in a verbal format which puts the actual meaning into explicit awareness.
Greenspan and Shanker (2004), for instance, emphasize that the grasping of symbolic meaning is rooted in children’s experienced emotions and feelings that now find a new and additional way of expressions. According to them, when children learn to separate perception from action via affective interaction in the second year of life, the results are freestanding ‘images’, or re-activations from an embodied perspective, that the children explore according to their emotional experience. In particular, wishes and so on, are represented internally as multi-sensory images, which presumably are based on embodied simulations/re-activations. Initially, these images emerge in a free floating manner such as un-connected balloons, but they become emotionally meaningful ideas or internal symbols through emotional interaction (Greenspan, 1997). During these interactions, action words begin to be used instead of the actual action and the words have content, and feelings are also conveyed via words. When these freestanding images are bridged with each other more advanced thinking and language skills develop, since symbolic expressions become a more economical way of creating and organizing the experiences, objects, categories and so on. Through the manipulation of these symbolic expressions, children allow themselves to expand their understanding of the world and themselves, and the ability to represent feelings emerges. For example, Greenspan wrote;

...the ability to abstract a feeling and give it a name... to know that the tightness in the chest is ‘fear’ – allows her to bring emotions to a new level of awareness and express them symbolically rather than by acting on them physically. She can tell dad she feels scared rather than shrieking in fear. The child who does not attain at this level can experience her feelings only at the level of behavior or visceral reaction”(ibid., p. 76).

Proceeding to reflective interaction, such as ToM and mind-reading, language could be a crucial aspect in order to achieve these abilities, since the use of symbols makes it easier to put forward different perspectives of a situation than only by using sensorimotor simulations or re-activations. Thus, the use of symbols could change our way of thinking, offering us a new thinking tool (see Vygotsky’s work on psychological tools in subsection 2.4.4), although these symbolic tools are deeply rooted in various kinds of bodily experience, which is illustrated by the vast amount of physical and emotional metaphors in language. Chesnokova (2004), for instance, points out that the work of Carpendale and Lewis (2004), quite paradoxically, is the first serious attempt to discuss the development of social understanding in a social context, and they conclude in their review that “social understanding is the emergent product of social interaction”, which is a remark in line with the relational nature of the proposed framework (see section 6.2). They present empirical findings which show that children who have closer positive relationships with others demonstrate a higher level of social understanding than in children that have few close relationships. With these findings as their foundation, Carpendale and Lewis state that it is now time to move beyond the deadlock of the “out of date” theories of ToM. They suggest that one direction of study might be to investigate ‘multiple nonverbal’
interactions which occur in the natural social setting and play a major role in promoting social understanding, especially at early stages of language development. Their central claim is that the focus of research should be shifted from investigating mental states and mentalist concepts as the consequences of children’s understanding of others’ minds to carefully investigating and analyzing social and cultural contexts that shape the content of children’s social cognitive skills. According to Gallagher (2007b), children’s ability for intersubjectivity are accomplished through embodied practices in which they “do not try to get into the other person’s mind” but instead “try to get into their world, or more precisely, into a world we already share with them” (ibid., p. 2). Children are active constructive agents within social interaction, and do not passively adopt transmitted pre-given information. Thus, to have the inner experience children need the outer experience.

Furthermore, Fernyhough (2004), for instance, points out that although Vygotsky did not explicitly addressed ToM, numbers of aspects concerning ToM are present in his writings (see Vygotsky’s work in subsection 2.4.4). In the zone of proximal development (ZPD), for instance, children are involved in activities that are scaffolded by peers in order to acquire mental terms and concepts, which only later they are able to ‘fully’ understand. Of major importance here is language as a psychological tool because private and inner speech may have important roles in mediating children’s own reflections on the characteristics of mental states and processes through embodied practices and/or embodied simulation (Gallagher, 2005). Moreover, Vygotsky’s General law of cultural development (GLCD) provides a viable approach which considers social cognition, such as ToM and mind-reading, as genuinely rooted in social experience. That is, through internalizing dialogue with others, children internally reconstruct the dynamics of that dialogue with all of its characteristics, which involves different perspectives of a certain situation. In line with this remark, Fogel (1993, p. 4) argues

...that the working of the mind and the ways in which we perceive and understand ourselves is remarkably such as the form of our personal relationships. The life of the mind is a dialogue between imagined points of view. The points of view of the mental discourse bear a close resemblance to positions taken by two different persons in a discussion, or to the physically embodied positions of individuals engaged in non-verbal communication

Thus, ToM reasoning represents the pre-eminent example of dialogic thinking. In the process of internalizing the ‘outer’ social interaction, children have to handle others’ ‘mental states’ and then coordinate these different perspectives with their own intentions. Furthermore, Gallagher (2005) points out that there is a lot of empirical evidence of the full-blown ability of ToM or ‘second-person embodied interaction’ as Gallagher calls it. In addition, this ability is grounded in the embodied practices that generate our primary access for understanding others, and continue to do so to a large extent even when children are able to attain ToM abilities with other
persons perceived as significant others. Thus, this internalization process which results in ToM and social cognition is explained as offline cognition that indeed is body-based. The child is able to do the same thing offline as in the real on-line situation.

Furthermore, according to Gallagher (2007b), children’s ability for intersubjectivity and pragmatic interaction is scaffolded, and then enhanced, via the development of narrative practices around the age of three to four years. These narrative practices enable children to put persons and actions in certain situations and contexts together, and through the construction of less to more advanced narratives, the children develop more subtle and sophisticated understandings of others (Gallagher, 2007b). Gallagher points out that this kind of narrative practice, combined with our embodied practices and the actual social and cultural contexts, mostly provide sufficient scaffolding for our understanding of others. However, this does not mean that he neglects the role of language for a more fully account of intersubjectivity, but his major point is that embodied practices in situ offer a relatively reliable and elemental understanding of others, rejecting the idea that mental concepts (such as 'beliefs', 'reasons', 'desires', etc.) serve as the foundation for ToM. Instead, he suggests that narrative competency may in fact serve as the foundation for the use of folk psychological concepts, in those situations when we are unable to directly grasp puzzling and odd behaviors (Gallagher, 2007b).

Finally, concerning self and intentionality, Gallagher (2005), for instance, states that in the same way that the human body structures human experience, it also shapes the human experience of self, and perhaps the very possibility of developing a sense of self, arguing that “if the self is anything more than this, it is nonetheless and first of all this, an embodied self” (p. 3). Furthermore, he points out “it may even be possible to say that bodily movement, transformed onto the level of action, is the very thing that constitutes the self...embodied movement contributes to the shaping of perception, emotional experiences, memory, judgment, and the understanding of self and of others” (Gallagher, 2005, p. 8). Moreover, recent work on the activation of the mirror neuron system situated in a certain context (see section 4.2) reveals that a certain kind of mirror neurons, so called logically related mirror neurons, may lay the foundation for intentionality. Conventionally, the description of an action and the interpretation of the reason why that particular action is performed are suggested to rely on two different mechanisms. The mirror neuron system, however, provides an alternative solution, given that the logically related mirror neurons automatically code the motor acts that are most expected to follow the observed action in the particular context (Iacoboni, et al., 2005). This means, the ability to infer the forthcoming new goal, i.e., intentionality, is already inbuilt in the mirror neuron system and explaining intentionality by two different mechanisms is both unnecessary and biologically implausible. During the course of ontogeny and through socially scaffolded interaction, the mirror neuron system develops further to more advanced forms of intentionality, both online and offline. The mirror neuron systems is
also considered to cooperate with other brain areas, which suggest how humans with a damaged prefrontal cortex, which regulates emotions, motor planning, and the sequencing of behavior, have problems making judgments (cf. Greenspan, 1997 concerning the relationship between motor planning and making judgments).

Metaphorically speaking, many ways lead to Rome and there is no royal road to cognition. Socially interactive cognition develops, and is not pre-given. Instead, it becomes enacted, and concerning the self, it is also an invention rather than some pre-given internal entity. Thus, the *relational* and *embodied nature* of the ontogeny of socially interactive cognition and self portrayal in this section is in sharp contrast to individualistic and dis-embodied approaches presented earlier in sections 5.1-5.2.

**5.6 In Summary**

In this chapter I portray different characteristics and levels of mainstream social interaction and cognition, and also identify two major forms of organismoid embodiment in social interaction, namely exo-organismoid embodiment and endo-organismoid embodiment (I return to the question of what kind of embodiment is necessary for cognition in section 8.1). Furthermore, I present and contrast two different metaphors concerning the nature of social interaction and communication that are used within socio-cognitive research (section 5.3). Taking a dynamical stance, that is, adopting the view of the dance metaphor of social interaction, I propose an alternative way of considering social interaction as well as intentions, and some other approaches for characterizing and studying social interaction are presented (section 5.4). In section 5.5 I argue that infants’ development of cognition and self is firmly grounded in the dynamics of the experience of socially scaffolded experience of their feeling, moving and interacting bodies. Crucial to the embodiment of cognition is the elementary and intertwined relation between the experiences of one’s own body and its interplay with the physical and social environment, which describes the relational and embodied nature of the different kinds of social interaction and cognition. Human cognition and self are not pre-given, since they develop and are through this process scaffolded by the social and material environments in the child’s particular culture. Thus, the self is an invention, its neither mind nor body but the emergence of becoming human by interacting, through one’s embodiment, with one’s social and material surrounding.
6. Situating Embodied Action within the Social and Material Sphere

The majority of the ideas presented in the previous chapters is now combined and appear in a new guise, thus expanding our theoretical and conceptual horizons. This means, the relational and distributed nature of socially interactive cognition that is previously portrayed is here justified by the embodied nature of these aspects. Generally speaking, the embodied nature of social interaction discussed in chapter 4 rather confirms than contradicts the dance metaphor of social interaction and communication. Thus, by integrating these perspectives, a deeper understanding of socially embodied actions in human cognition can be obtained. The result of this approach, however, is both different from, and greater than its individual parts, given that it encompasses not merely social embodied phenomena lodged in the individual, but also how other participants’ embodied actions are situated within human interaction and context, as a way of acting and creating meaning together. In other terms, we now proceed from ‘embodiment and social interaction’, which is the title of chapter 4, beyond embodiment in social interaction, to view it as ‘embodiment situated-within-social interaction’. This shift, from embodiment in social interaction to ‘situated embodiment-in-social interaction,’ is a fundamental one, I believe, and therefore the title of this chapter is ‘Situating embodied actions within the social and material sphere’. While section 6.1 motivates this standpoint both from a historical perspective and ‘state-of-the-art’ perspective, then section 6.2 presents the specific framework. Finally, section 6.3 discusses its contribution and uniqueness.

---

32 Parts of this chapter will be published in Lindblom and Ziemke (forthcoming).
6.1 Putting it all Together

From a historical perspective, the present standpoint can be regarded as the re-integration of previously presented ideas, such as viewing cognition as a dynamical and on-going process rather than a theory of content (see chapter 2). More specifically, I intertwine the insights of Mearly-Ponty (see 2.4.2) concerning the embodied nature of human cognition, with the relational nature of human cognition as described by Vygotsky (see 2.4.4) and Mead (see 2.4.6). Although Merleau-Ponty is considered as the 'knight of the Body' (Sheets-Johnstone, 2003), he overlooked, in my opinion, the first and foremost relational nature of the human mind. Indeed, he did not dismiss social aspects, like language and culture as such, given that he stretched the individual’s embodiment to the edges of the cultural domain (Varela, 1994). Varela, for instance, pointed out that although Merleau-Ponty identified factors such as language and gesture, it does not mean that he conceptually grasped the body as a socially and cultural entity, i.e., the moving and embodied body as a socially enacted body. Hence, Varela criticized Merleau-Ponty’s notion of the lived body, arguing that this notion of agency “is as ghostly as the force of the Cartesian mind!” (p. 219). This means, although ‘the lived body’ was a rejection of the Cartesian ‘trick’ of privileging the agent mind by disregarding the body, the lack of a person is present in both Descartes and Merleau-Ponty’s views. The point is that intentionality belongs to a person, given that “people intend, not bodies; minds do not intend; people do” (ibid., p. 220). In other terms, humans are socially enacted agents given that as people they are social, and use their intentionality through mental and bodily actions in the course of lingual and gestural interaction. Thus, the apprenticeship of “being-in-the-world” is not an individual enterprise, and as Varela explains, Cartesianism implicitly presupposes, paradoxically, the person as an agent, and thus the fundamental social and relational nature of being human.

Nevertheless, Varela did not entirely reject Merleau-Ponty’s work, admiring how advanced he was as a philosopher in thinking about the relational nature of human cognition, taking a vast leap in the right direction, although not reaching all the way. In other words, the major problem with Merleau-Ponty’s embodied theory of cognition is that it lacks a clear idea of the relational nature of the person as a socially enacted agent (Varela, 1994). Furthermore, Morris (2005) argues that the comparative approach of Merleau-Ponty is flawed, since he investigated the behavior of the ‘lone animal’ in his work on agent-environment interaction, despite the fact that animal and human behavior is rather a group phenomenon. The same criticism can be applied to the work of von Uexküll (subsection 2.4.1). By way of contrast, both Mead and Vygotsky, independently of each other (Valsiner & van der Veer, 1988), realized that matter much earlier. Mead, for instance, regarded the human individual an embodied agent within the social sphere, and, according to Varela (1994), had already in 1913 finished his conception of the social nature of mind and self. But as presented in subsection 2.4.6, Mead focused mainly on vocal gesture (speech) and little on manual gesture. In sum, the linking of manual and vocal gestures constitutes the being of personhood in the social act of conducting
communicative interaction (cf. Farnell, 1995, 1999). Similarly, Vygotsky stressed that human cognition is the consequence of the intertwining of sociocultural and biological factors, as portrayed in subsection 2.4.4. Thus, the integration of Merleau-Ponty’s insights with Mead and Vygotsky’s ideas, results in the shift from the individualist perspective to the relational perspective of the nature of socially interactive cognition.

From a ‘state-of-the art’ perspective, contemporary theories of embodiment, as described in chapter 3, have largely disregarded the mind as first and foremost social and relational. Hence, there is a need to move beyond the current emphasis on the interactions between the organism and the physical environment, to interactions between agents and their social environment. Others recognize this as well, which implies that the theoretical knowledge of embodiment needs to be extended beyond current notions and/or levels of ‘individual’ embodiment to also encompass the social dimension (see section 3.4).

Concerning social interaction and cognition, many of the dominant theories in these areas tend to overlook the role of the body in social interaction by treating embodied social interactions such as body posture, gaze and gesture as nothing but the visible consequences of mental intentions (see sections 5.1-4.2; subsection 5.3.1). Indeed, the stimulating shift in social interaction studies from the Information transmission metaphor to the dance metaphor (section 5.5) illuminates the dynamically emerging, creative and interactive behaviors of a particular social situation, instead of discrete and linear processes. In other words, the dance metaphor focuses on the emergence of information in the dyad between agents rather than pre-given information transmitted back and forth between them (subsection 5.5.2). Furthermore, the distributed cognition approach treats these social interactions and the external materials involved, as directly observable cognitive events in the situated interaction (section 5.4). With this crucial change in perspective, much of cognition previously hidden ‘inside’ the skull has now become apparent. That is, cognitive functions are manifested in artifacts, and dynamic and social processes can be directly observed and studied. However, the main focus in DC is on the flow and transformation of information through different media at a more general level, rather than the particular role of the body in these processes. This means, despite the emphasis on interactions between agents and their material as well as social surroundings, the DC approach offers little on the embodied nature of human cognition, as usually conceived, and is currently peculiarly ‘disembodied’ (see Hutchins, 2006).

Although the role of embodiment in social interactions is only peripherally addressed in mainstream embodied cognitive science (see chapter 3), a great deal of work from various disciplines addresses the significance of embodiment in social interaction and cognition (see chapter 4). However, much of the material concerning embodiment and social interaction presented and discussed so far (chapter 4), is mostly from an ‘objective’ third-hand perspective and somewhat de-contextualized (cf. ‘4.1 Embodied
social psychology’ and ‘4.3.2 Gesture in language’). Crucial missing aspects are, for instance, the relational nature of human social interaction given that much research on gesture was carried out and described from information-processing models (see 5.3.1) and ‘objective’ third-hand perspectives. For example, in their respective works on gesture, McNeill (1992, 2005) and Goldin-Meadow (2003), mainly describe and study shapes of the hands, trajectory of movements, location and orientation of hands in relation to the body and so on, but neglect, to various extents, the situation at hand in which the gestures occur in an unfolding interaction. Although Goodwin’s (2003) work on symbiotic gestures focuses on bodily aspects in an unfolding contextual situation (see subsection 4.3.2), understanding the situated embodied social interaction could be amplified and deepened, by more fully integrating how socially embodied actions, such as gesture, speech, facial expression, gaze and posture matter when they are within their spontaneous context of use; that is, socio-culturally situated in an unfolding dialogue. It is exactly this unified view that I present in the framework for the embodied nature of social interaction and cognition.

In summary, the above perspectives aim to motivate the framework for the various kinds of internal and visible embodied actions within situated social interaction as a dynamical process of organizing meaning and action. What distinguishes my approach from the historical writings of Mearley-Ponty as well as Mead and Vygotsky, is that the present framework I propose utilizes recent theoretical work and empirical evidence from cognitive neuroscience and embodied cognitive science. Hence, this approach enables more detailed explanations of the embodied nature of socially interactive cognition. Thus, the historical insights of Dewey, Darwin, Piaget, Vygotsky, Mead, and Merleau-Ponty, to mention a few, are brought together with the profound contribution of recent embodied cognitive science.


This section presents the interdisciplinary framework for the embodied and relational nature of situated interaction and cognition. Generally speaking, the framework integrates so-called ‘inner’ physiological processes such as mirror neuron activity and embodied simulation with so-called ‘outer’ aspects such as posture, speech gesture, facial expression, and gaze with the individual, mutual and joint actions of interacting agents in the unfolding of interaction in situated socio-cultural contexts. It should be pointed out that this framework does not intend to bridge the correlation gap between the ‘inner’ and ‘outer’ perspectives, since it attempts instead to portray a thorough and integrated understanding which supports and explains the relationships actually existing among these aspects of embodied cognition. This means, instead of building a bridge, I try to re-characterize these issues from an embodied cognition viewpoint. To clarify the framework, I am inspired by Rogoff (2003) to develop a series of images that aims to illustrate the framework, which depicts some of the different aspects of the embodied and relational nature of social interaction and cognition. The images are
from my own empirical study at a horse ranch, which is presented in detail in chapter 7.

Figure 5a illustrates that much prior research has mostly been focusing on one or some specific aspect(s) of embodiment such as posture (cf. e.g. Barsalou et al., 2003; Niedenthal et al., 2005a, 2005b; see section 4.1), gesture and its relation to speech (cf. e.g. Gallagher, 2005; Goldin-Meadow, 1999, 2003; McNeill, 1992, 2005; see section 4.3), facial expression (cf. e.g. Barsalou et al., 2003; Niedenthal et al., 2005a, 2005b; see section 4.1), kinesthetic experience/body in motion (cf. e.g. Adolph & Berger, 2006; Farnell, 1995, 1999; Gallagher, 2005; Hurley, 1998; Sheets-Johnstone, 1999; in section 3.3) or specific neurological processes (cf. e.g. Arbib, 2006; Gallagher, 2005; Gallese, 2004; Rizzolatti, 2005; Rizzolatti & Arbib, 1998; see section 4.2). However, in this thesis I try to integrate them more fully instead, thereby providing a more comprehensive picture and synthesis of socially embodied actions in social interaction and cognition (see Figure 5b). This means, I regard them as a unified and holistic cross-modal interaction system. Furthermore, by using the term action instead of behavior (cf. e.g. Farnell, 1995; Varela, 1994), the framework highlights that socially embodied actions are a set of movements with meaning for the actual person or agent. The concept action, contrary to movement, implies intentionality or a goal. This means, embodied actions should not be regarded as appendages of the thinking ‘mind’, since they have a crucial impact on the agent’s cognitive processes. That is, the framework proceeds from the experienced body to the enacted body, suggesting that the agent or person must be primary in our conceptual understanding. However, these perspectives alone do not provide us with many clues about what is happening in Figure 5b.

Figure 5. Image 5a illustrates that much work on socially embodied action has yet focused on isolated parts, and image 5b illustrates the integrated and holistic approach in this thesis that focuses on cross-modal interaction.
An agents’ embodied actions are always in relation to others, which means that it is literally useless to distinguish between individual and social actions, since they are two sides of the same process rather than two separate processes (cf. e.g. Fogel, 1993 in subsection 5.3.2; Mead, 1934, in subsection 2.4.6; Vygotsky, 1978, in subsection 2.4.4). Thus ‘being-in-the-world’ is primarily based on being an enacted body in a socio-cultural surrounding with others. This socio-cultural world is not any extract ‘added on’, but an integral part of social interaction. As previously mentioned, embodiment constrains (e.g. how we can move our hands), and cultural norms affect (how to interact in a certain cultural setting), but do not determine, the structure of embodied social interactions. However, as figure 6 illustrates, including other people, still makes it difficult for us as readers/observers to fully grasp what is actually happening in the image, which is not the case for the people involved.

Figure 6. The including of other persons is not enough for us as observers to grasp what is going on.

The nature of social interaction is relational and meaning and intentions are emergent phenomena of these social interactions (cf. e.g. Fogel, 1993; Gibbs, 2001; Hutchins, 1995; Johnson, 2001; see section 5.4) which in most situations can be regarded as some distributed process rather than individual private mental acts. To put it another way, we should not consider intentions to be ‘in there’ but instead ‘constructed’ between people and their surroundings. This means, the cognition we have to study is ‘out in the open’ and become a ‘group event’. Thus, the unit of analysis is situated embodied actions in co-regulated social interaction, which goes far beyond the bounds of internal mental processes (see section 5.3).
What actions should be counted as intentional ones depends on whether they are interpreted as intentional actions, regardless if they are intentional or not from the beginning.

**Embodied actions are situated in cultural and material contexts,** meaning that we do not examine embodied actions apart from their context of use, cultural and material, because the actual situation at hand adds meaning and intentions to the embodied actions (cf. e.g. Adolph & Berger, 2006; Clark, 1997; Farnell, 1995; Fogel, 1993; Hutchins, 1995; Mead, 1934; Vygotsky, 1978). By *cultural context*, I refer to the different culturally accepted ways of doing things, like how we interact socially (see Rogoff, 2003). For instance, pointing with an extended index finger is not a universal action, as some might believe, given that in some cultures people instead point with their lips or with their feet (cf. e.g. Kendon, 2004; Kita, 2006; Rogoff, 2003). This means, an embodied action is culturally dependent given that it is recognized and interpreted ‘communicatively’ by other people in that particular culture. Thus, culture is not some ‘add-on’, but a fundamental aspect incorporated in ongoing activity. By *material context* then, I mean the role of external structures such as the different tools, artifacts and etc that we use as scaffolds, offering important resources to social and cognitive action (cf. e.g. Clark, 1997, 1999b; Hendriks-Jansen, 1996; Susi, 2006). This means, the ‘gap’ in interpreting what is happening in Figure 6 above, for us as readers, is bridged when we have the whole ‘picture’ presented. The material environment, in this case a filly in a paddock, is of major importance for us to understand what the people in the picture actually are doing. In this case, as illustrated in Figure 7, the individuals are attempting to lift a filly (i.e. young horse) that is lying down in the paddock at a ranch.

![Embodied actions are situated in cultural and material contexts.](image)

**Figure 7.** Embodied actions are situated in cultural and material contexts.

*Interacting socially through embodied actions is a dynamical process that unfolds at a temporal horizon,* which means that cognition should
be considered a process rather than a static object of content. However, the pictures and texts provided are ‘stiff’ media that do not make justice to the representations of these dynamical properties. However, when the reader instead views the images as snapshots in a sequence of movements, the key to the dynamical approach of socially embodied actions is illuminated. Furthermore, it is important to regard how these embodied actions are integral parts of the relational interaction always occurring (cf. e.g. Fogel, 1993; Shanker & King, 2002; in subsection 5.3.2). That is, actions are always in relation to some prior information, and socially embodied actions do also play a fundamental role in this process. Figure 8 illustrates what happens a few moments later. The temporal horizon also explains why the filly needs lifting support. She is unable to rise by herself because she has been weakened by medical treatment and it is not healthy for a horse to lie down for long periods of time.

Figure 8. The dynamical and temporal aspects of socially embodied actions are parts of the ongoing activity.

**Socially embodied action is not pre-given, since the social mind develops.** That is, the functional significance of embodiment and its possibilities for action depend on the fit between the actor’s body and the properties of the surrounding environment (cf. e.g. Adolph & Berger, 2006; Fogel, 1993; Vygotsky, 1978; and section 5.5). Briefly stated, the central and inescapable feature of embodied action is that the opportunities and constraints of the body as well as the social and material environment are continually changing during the course of ontogeny.

**The main focus of analysis is embodied actions in social interaction and cognition,** which means we have to consider what we particularly want to analyze and investigate. It is necessary to make a trade-off between different aspects of a phenomenon, given the impossibility of focusing on everything at the same time. This does not imply an either-or choice, but rather a decision about what to place in the foreground and what to place in the background. As a way to highlight this stance, I have decided to show a ‘whole’ image, as shown in Figure 9, but in which the aspects that I consider
as the background is blurred. The aspects that represent the main focus of analysis are indeed foregrounded. This way of illustrating my stance does not imply a clear distinction between the foreground and background, as one sees in Figure 6, which makes that either-or cut.

![Image](image.jpg)

**Figure 9.** The main focus of analysis is embodied actions in social interactions and cognition is not blurred, as a way to represent that they are placed in the foreground. The aspects that represent the background are blurred.

In my opinion, the background has a crucial impact on what happens in the foreground, but the main focus of attention, both in this framework and throughout this thesis, is on the **role and relevance of the body in social interaction.** The following four fundamental functions of embodiment serve as the underpinnings for these social “agent-environment interactions”. However, I would like to point out that although these four fundamental functions are of major and necessary importance, they are not of exclusive and sufficient importance.

**Embodiment functions as a social resonance mechanism,** which means there is no need to encode or represent embodied social stimuli to more ‘advanced’ or high-level cognitive states since the bodily states in themselves are actually cognitive or affective states, as related work portrays. Furthermore, the social resonance mechanism does not only function in direct or online social interaction, but also in offline interaction when the social stimuli are more abstract, such as in words and concepts. Hence, this first function portrays how cognitive and bodily states of the interacting partners are reflected both in themselves and in-between them (see subsection 4.4.1).

**Embodiment functions as a means and an end in communication and social interaction.** Mirror neurons and the mirror neuron system are suggested to provide the linkage between ‘action’ and ’action-perception’, and apparently, this implies that the body and its sensorimotor processes are ‘cognitive’ in themselves, and not only bound by the brain. The great benefit of this action-understanding linkage, besides the parsimony of the account,
is the inbuilt dual ability of grasping both the ‘what’ and ‘why’ aspects of the present action, i.e., what the action is about as well as catching the intention behind the movement. Different degrees of activation of and couplings with the mirror neuron system form a continuum between basic levels of action-understanding to more advanced ones, such as gestures and language. Although the current knowledge of the underlying mirror neuron mechanism is unable to explain in detail the whole complexity of human social interaction, it does shed light on how the interacting partners are able to share the communicated meaning in the dialogue. Hence, this second function offers a tentative explanation of that particular linkage, thereby unifying the ‘inside and ‘outside’ perspectives of socially embodied interaction (see subsection 4.4.2).

Embodied actions and gestures function as a helping hand in shaping, expressing and sharing thoughts. Besides speech, gesture is a significant (embodied) aspect of social interaction, which may provide important information to the listener, since gesture offers speakers the means of expressing thoughts difficult to articulate in speech. Through gesturing, we are able to generate and embody dynamical associations between different matters, which can offer new insights into the present situation or problem at hand. In addition, gesture sometimes serves as an explicit instance of the action-meaning embodied in speech, suggesting that hand movements are physical externalizations of the speaker’s ideas. Hence, this third function stresses how bodily actions operate both outwardly (inter-subjectively) and inwardly (intra-subjectively) in social interaction and cognition, in which our embodiment (paradoxically) both constrains and enables the possibilities of expressive actions (see subsection 4.4.3).

The body functions as a representational device. In addition to speech, there is the more controversial claim of the body also having representational properties, where certain kinds of gesture, portraying representational aspects, are the least provocative and most obvious examples of the body as an (external) representational device. The neurological roots of this ability could be the activity of the mirror neurons, since their linkage between ‘action’ and ‘action-perception’ might propose a kind of ‘action representations’ directly enacted in social interaction. Furthermore, since mirror neurons are able to compensate for missing ‘information’, and yet seem to ‘understand’ the goal of the action, it can be argued that the grasping of the action does not require a declarative understanding, since it is meaningful in itself. Hence, the cognitive processes of social (cognitive) interaction are, to varying extents, suggested to be grounded in embodied ‘representations’ that have representational content. It should be noted, however, that we are not denying mental concepts as such, but questioning their organization, since they may be the result of, and grounded in, embodied interactions and not the underlying requirement for cognitive processes (see subsection 4.4.4).
6.3 General Remarks

Taken together, the above issues constitute the framework, which unifies the so-called ‘outside’ and ‘inside’ perspectives of the body in socially interactive cognition from an embodiment viewpoint. In addition, the framework situates socially embodied actions in the cultural and material context, as well as considering the dynamical and temporal aspects. Furthermore, while many approaches to embodied cognition and social cognition focus primarily on phenomena residing within the individual agent, this framework is conceptually different and considers the relational nature of social interaction, meaning, and intentions. Accordingly, many cognitive processes are distributed and situated ‘out in the open’. Focusing on the functions of socially embodied actions, the framework emphasizes that they do not simply serve to express our internal cognitive processes, but they are part and parcel of social interaction and cognition.

It should be noted that almost all of the issues in the framework have been presented previously, either by myself or by others, and I have given some references to most of them. The framework is unique, however, because it weaves together theoretical ideas and empirical findings from a number of major disciplines. I believe that the framework I propose knits all these ideas together, but not as separate or alternative entities. On the contrary, the framework reveals how properly and flawlessly they all fit together, occurring as hand in glove. Thus, my framework offers a coherent and compelling whole, in which the sum is more than the individual parts.

As far as I know, no one else has provided any credible, holistic and detailed explanation of how social interaction and cognition arises from the actions of the body in its cultural and social sphere, from a ‘radically’ embodied perspective. In my opinion, it appears fruitless to approach this problem from a neurological/sensorimotor perspective alone, or exclusively from the social and contextual side. Indeed, I am attempting to explain the role of the body in social interaction and cognition from an embodiment perspective in situ.

What ties all these issues together is the idea of the social mind as being relational (e.g. Mead, 1934) and ‘radically’ embodied (Clark, 1999a), in the social and material sphere. Accordingly, the social dimension meets the physiological dimension, and this framework reaps the best of both worlds without neglecting the effects of embodiment for social interaction. Thus, I now provide a coherent picture for the role of the body in social interaction, in which so-called contextual, cognitive and physiological terms are deeply intertwined. What unites all these issues is how profoundly embodiment shapes social interaction and cognition through unfolding socially embodied actions in material and cultural contexts. Thus, the socially embodied actions go ‘beyond the flesh’.

As stated above, the key to this coherent union is the way our social mind is embodied, a fact that we cannot neglect or trivialize. Metaphorically speaking, the different ideas presented here are like putting together pieces
in a puzzle, were the motive finally emerges in a way that you could not imagine in beforehand as well as could not be predicted from any of the individual pieces alone. However, I do not claim that my framework shows the detailed or complete picture of the role and relevance of the body in social interaction and cognition, but illustrates a rough sketch of the overall ‘picture’. Thus, it provides a solid foundation from which to proceed from.

It should be noted, however, that providing a concrete and single image of the entire framework here is not possible, given that it would provide a rather ‘false illusion’. I have tried my best to depict my main ideas in the previously images, but are not able to make the ‘one and only’ image (section 7.3 presents a shorter summary of the framework).

To conclude this chapter, the framework is now presented, and in the next step, it is further evaluated in a case study of spontaneous social interaction ‘in the wild’. Accordingly, the framework acts as a guide for the following empirical work in chapter 7, which aims to clarify and illustrate the understanding of socially embodied actions *in situ*. 
Chapter 7

Sometimes the most obvious things are hardest to see because they are right in front of your eyes.
Johann Wolfgang von Goethe

Not everything that can be counted counts, and not everything that counts can be counted.
Albert Einstein

7. Empirical work

This chapter presents the empirical part of this thesis, with the aim of illustrating and providing some detailed observational fieldwork on socially embodied actions captured in spontaneous interaction in situ. Taken together, questions that need to be addressed in current theories of embodied cognition are, for instance, how does meaning emerge in social interaction, and what role and relevance do embodied actions have? Moreover, how does the body affect social interactions, which social processes are affected, and what functional roles do the body serve in social and cognitive processes?

The remainder of this chapter is structured as follows. Section 7.1 presents the chosen approach for the empirical design, data recording and the data analysis, and introduces the participants of and the setting in which the empirical work was conducted. Three different episodes of social interaction and cognition that somewhat differ in focus are described in section 7.2 (subsections 7.2.1-7.2.3). Section 7.3 offers some discussion and conclusions. Subsection 7.3.1 discusses the relationships between the empirical work and the theoretical framework. Subsection 7.3.2 presents some additional aspects that arose during the analysis of the empirical data. Subsection 7.3.3 concludes with a synthesis between the theoretical framework and the empirical work that is subsequently generalized into different dimensions of embodied action in situ.

7.1 Case Study: Empirical Design

In order to illustrate some aspects of the previously presented theoretical framework (section 6.2), to further the understanding of embodied action in spontaneous socially interactive cognition, naturalistic inquiry (cf. e.g.
Lincoln & Guba, 1985) was chosen. According to Lincoln and Guba (1985), naturalistic inquiry is characterized by observations conducted in a natural setting. Naturalistic inquiry focuses on “in depth” and detailed “fieldwork descriptions of activities, behaviors, actions, conversations, interpersonal interactions, organizational or community processes, or any other aspect of observable human experience” (Patton, 2002, p. 4). Furthermore, in naturalistic inquiry, one also incorporates the context in the analysis, given that it provides much of the meaning in a situation (Lincoln & Guba, 1985; Patton, 2002). Generally speaking, in order to achieve scientific rigor it is of utmost importance that the researcher ‘collects good data’ and ‘draws sound conclusions’ (Patton, 2002). With regard to sample size, Patton (2002) points out that “[t]he validity, meaningfulness, and insights generated from qualitative inquiry have more to do with the information richness of the cases selected and the observational/analytical capabilities of the researcher than with sample size” (Patton, p. 243). This means, the quantitative number of case studies is not a measurement of research quality, instead the measure of quality lies in the study itself.

For the naturalistic inquiry a case study approach was chosen, but the process differs from the typical scientific approach, since the collected data is from my private life, initially intended for private use and not as research material. However, I subsequently discovered the recorded material’s potential for scientific purposes. That is, I compared it with other empirical data-collections I had carried out, which primarily had the aim to provide the data for my empirical work, and I then realized that my private material could be a better source of empirical data. Before I decided to use my private material, I carefully investigated and analyzed whether it fitted the demands for serving as scientific data. The following reasons motivate my conclusion that my own material could be used as a ‘scientific’ case study.

A case study is considered to be “information rich” and illuminative, offering useful manifestations of the phenomenon of interest (Patton, 2002). More specifically, an instrumental case study is particularly feasible in conducting basic research, given that “a particular case is examined mainly to provide insight into an issue...the case is of secondary interest, it plays a supportive role, and it facilitates our understanding of something else” (Stake, 2000, p. 437). Accordingly, the main enterprise of scientific work is to gain the best possible explanations of the phenomena of the study, and given that the phenomena are given ‘out there’, the actual case is a tremendous opportunity to study the phenomena of interest (Stake, 2000). Furthermore, inspired by Hutchins’ cognitive ethnography (1995), and due to the topic’s situated nature, the case study should be conducted in everyday social activities ‘in the wild’ (cf. Hutchins, 1995), with the focus on how humans embody their thoughts in visible socially interactive cognition through crossmodal interactions such as gesture, speech, facial expression, posture, and gaze. Given that my own material well matched the above criteria, I decided to use it in this thesis.
My empirical data were gathered with video recordings (e.g., Xiao & Mackenzie, 2004) which could be considered as being carried out through participant observation. Other strong motivations for choosing my private video-recordings were that my data significantly fitted with the following sampling techniques, namely purposeful sampling, intensity sampling, and typical case sampling. As phrased by Patton (2002, p. 230), “[t]he logic and power of purposeful sampling lies in selecting information-rich cases for study in depth. Information-rich cases are those from which one can learn a great deal about issues of central importance to the purpose of the inquiry”. Intensity sampling, then, is a sampling technique that “seeks excellent or rich examples of the phenomenon of interest, but not highly unusual cases” (Patton, 2002, p. 234). Finally, in typical case sampling, the focus of attention is on illustrating and highlighting aspects that are typical for certain phenomena (Patton, 2002).

In addition, despite the fact that the video-recordings initially were not intended for scientific purpose, the chosen approach provides some significant positive aspects. Firstly, it can really be considered being spontaneous and in situ, given that it was collected for private use. Secondly, because of that, there is a reduced risk of slipping into any biases of observational effects (cf. e.g. Patton, 2002; Shaughnessy & Zechmeister, 1997) that negatively might alter the participants’ performance and actions. Thirdly, I, as the person holding the video-camera, did not were my ‘scientific glasses’ during the video-recording, and as a result not implicitly influencing the data collection in service to scientific perspectives. Accordingly, a negative consequence of the chosen approach is that the video clips are sometimes not as optimally recorded as desired, for which I do apologize and hope that the readers are tolerant of this circumstance.

The data gathering, here used as the case study, was carried out during a guided tour, in which I participated, at a ranch that maintains and preserves the Blackfeet’s herds of Spanish Mustang horses. The visit at the ranch was one item on the agenda during a two-week vacation trip in Montana and Idaho, in which the central themes were horse culture and horse back riding. The ranch, named Blackfeet Buffalo Horse Coalition (BBHC), hereafter referred to as the ranch, is situated outside Browning, Montana, USA. Part of the mission of the ranch is to give the horse back to the Native Americans across North America, as a way to remind them of their past as well as to pass on the important relationship between the horse and the native American culture to the younger generation. The horse is a central source of power, religious and cultural meaning for Native Americans in general, and the Blackfeet in particular. The different tribes of Blackfeet are located on both sides of the US-Canadian border, but the Blackfeet view each other as one whole community not separated by the frontier.

33 More information about the Blackfeet Buffalo Horse Coalition (BBHC) can be obtained on their web site, available at http://www.buffalohorse.org.
34 According to Gibson (2003), the Blackfeet confederacy consists of four separate nations, namely the Blackfeet Nation in Montana, USA, as well as the Siksika, the Piegan and the Kanai nations that are located in Alberta, Canada (see the Appendix).
established in more recent times (see the Appendix for a map over the territory).

Furthermore, all the participants have subsequently allowed me to use the material for scientific work and publishing. The participants in the study were first and foremost the head of ranch, Bob BlackBull (Bob), and a group of Swedish visitors, all of whom are experienced in handling horses. Besides Bob, the participants are Dag, BoChief, Peter, Katrin, Åsa, Mats, and myself. Katrin is the leader of the visiting group, and a friend of Bob previously. It should be noted that the group visits as Katrin’s friends and not as tourists. Furthermore, although Bob does not speak Swedish, and all the visitors understand and speak English well, within the group they mostly communicated in Swedish. The duration of the entire visit was a couple of hours, in which I video recorded for a period of approximately 58 minutes. It should be pointed out that the selection process carried out consisted of finding social interaction episodes that were possible to analyze properly.

The video-recordings were analyzed at the micro-level, through frame-by-frame analysis. The aim was to investigate, from an embodied point of view, cross-modal embodied interactions in social interaction and joint actions. Accordingly, by situating these socially embodied actions in a social and material context, the contextual situation at hand provides meaning and intentions to the visible embodied actions. Thus, the unit of analysis was situated embodied actions in co-regulated social interaction, which extends far beyond the bounds of internal mental processes. However, the dynamic and situated nature of socially embodied actions result in some problems when representing and illustrating them in more “stiffer” mediums such as verbal descriptions and two-dimensional photographs, which are the media used here. I decided to use the present tense in the verbal descriptions of the episodes, although obviously they have occurred in the past. In doing so, readers are offered more ‘dynamics’ and ‘flow’ in the writings than had I used the past tense. Indeed, the dynamics of embodied actions are better displayed in the actual video recording, and ideally, it would be preferable if the readers of this thesis were able to actually view the video clips being analyzed so they could observe my analysis of the sequences themselves. In some limited situations, however, I was forced to speculate. On the one hand, the actual video-recordings did not always capture the desired aspects or recorded from the wrong angles. On the other hand, while I could not carry out any experimental investigations at the neurological and physiological levels, in those cases I kept as close as possible to the empirical data presented in the literature in order to describe the phenomena of study. It should be pointed out that throughout the following analysis, I use the earlier identified four functions of the body in social interaction (described in section 4.4 and in the framework (section 6.2), as the underpinnings for describing and explaining how embodiment is part and parcel of social interaction and cognition. They are, however, not always mentioned explicitly throughout the analysis, since such an approach would require in lots of reiterations which would withhold adequate descriptions and analyses of the social interaction at hand.
7.2 Case Study: Analysis and Results

In what follows, I present descriptions and analyses of some selected short episodes that illustrate the role and relevance of embodied actions in socially interactive cognition. Firstly, Bob informs us where in the USA there are Spanish Mustang horses which originate from the herd at the ranch (subsection 7.2.1). Secondly, Bob explains the relationship between himself and the horses, and here a new kind of speech-gesture mismatch is identified (subsection 7.2.2). Thirdly, in the episode with the filly that is unable to rise by itself, the focus is how embodied actions lend a hand; both in a practical and a literal sense, in joint action (subsection 7.2.3).

7.2.1 Different herds of Spanish Mustang horses

All the visitors are sitting on Bob’s veranda, having their lunches that consist of some take-away food, while Bob begins to present the ranch, its program, and mission. The overall scene is illustrated in Figure 10. The whole sequence analyzed below unfolds during approximately 16 seconds, and was video recorded from my position, sitting next to the railing (the person wearing a dark green vest in Figure 10).

![Figure 10. The visiting group is sitting at Bob’s veranda while Bob is standing beside the vehicle (photo taken by and used with permission from Katrin Gunnarsson).](image)

During the conversation Bob usually looked at Katrin, who took the picture in Figure 10, and the analysis begins when Bob changes topic, from answering the previously posed question by BoChief (the man sitting on the left of me in the above Figure 10) to telling us the different places in North America that nowadays keep herds of Spanish Mustang horses originating from the ranch’s herds. Next is an excerpt from the spoken interaction between Bob and Katrin as well as the rest of the audience.
1 Bob “...and now we have the herds”
2 Bob “...I have...let’s see”
3 Bob “...we have two herds up in Canada”
4 Bob “...we have two herds...on the Salish-Kootenai” [Katrin agrees, saying ‘uhm’]
5 Bob “...we have a herd for the Northern Cheyenne”
6 Bob “...we have a herd for the Crow”
7 Bob “...and we have a herd being put together for Fort Belknap”,
8 Bob “for the Assiniboine and the Gros Ventres” [while Katrin hums in agreement]
9 Bob “...and we certainly will have more on the Northern plains again”
10 Bob “....all that’s part of that dream”
11 Katrin “oh, that’s wonderful”
13 Bob “... that make them su...survive”
14 Bob “but it is something else”

Initially, the actual change of topic is manifested in his bodily actions, given that he alters his bodily position and gaze, from leaning against the vehicle and looking at the audience on the porch to directing himself to the prairie and gazing out over the landscape (as seen in Figure 11).

Figure 11. Just before Bob changes the topic, and the beginning of the utterance ”...and now we have”.

During the shift in bodily posture and gaze, Bob begins to tell us about the different herds of horses by saying, quite slowly in a thoughtful tone of voice, which is also manifested in his facial expression, ”...and know we have the herds” (as seen in image (b) in Figure 11).
Figure 12. When Bob utters "...I have", which continues with "...let's see...".

While he continues his utterance, after a short pause, saying "...I have...let's see...", his hands begin to narrow each other (see image (a) in Figure 12) and when he utters "let's see..." his hands are in touch with each other. His left index finger rests on the index finger his right hand, although he does not look at them while they make contact (see image (b) in Figure 12). This forming and contact of his fingers serve to mark that he has finished his re-enactment of the different places that have Mustang horses from the ranch, and that he now is able to spell out their locations. That is, the actions of his hands reflect and manifest the internal re-enactment of his memory as well when the process is sufficiently completed. Thus, these images illustrate that the act of remembering is facilitated by putting the body in motion through gesturing, as a profound way of shaping and expressing thoughts as well as showing the status of the internal mental search, presumably through embodied re-activation (see subsection 4.2.2).

Figure 13. When Bob starts saying “we have two herds up in Canada,” and his tapping with the index finger.
When he continues, after a very short ‘thinking’ pause as to make sure his re-enactment of the different places is correct, he says in a more rapid and definite tone of voice “...we have two herds up in Canada”. During this utterance, he lifts his left index finger and taps it on the right index finger (see the images in Figure 13). This action is a way of highlighting the fact there are two herds in Canada. While the initial position of the index finger functions as a reminder of the Canadian location, the tapping action signifies the ‘two-ness’ of the actual number of herds in Canada. This tapping action signifies the representational aspect of gesturing and therefore conveys meaning in its own way, given that the actual gesture is a sign of another aspect other than the actual tapping movement. Following the arguments of McNeill (2005), this gesture is a material carrier of the ‘two-ness’ concept which is also accompanied in Bob’s speech. That is, the information conveyed in speech and gesture is the same, a so-called speech-gesture match (cf. Goldin-Meadow, 2003, in subsection 4.3.2). The tapping action highlights the central information in the utterance, namely that there are actually two herds up in Canada. It should be noted that when Bob utters “two” his left index finger is in motion, disconnected, but connects when he says ‘herds’, thus serving as an indicator of the importance of the correct number of herds in Canada. Furthermore, when he pronounces “Canada” he actually grips the right index finger with the fingers of his left hand. This is once again a way of focusing the other significant aspect of the utterance, namely the importance of the correct location of the herd. Thus, the two most important aspects of the whole utterance, the number and location, are manifested in speech-gesture matches.

Bob then pauses for a short moment, and continues by saying, in the same tone of voice, “...we have two herds...on the Salish-Kootenai”. During the first part of the utterance, i.e., “we have” he moves his left hand from the index finger and instead touches the right hand’s middle finger with his left index finger. During “two herds” he moves his middle finger downwards just slightly (see the images in Figure 14).

**Figure 14.** The touching of the middle finger and the moving of the middle finger during the pronunciation of “two herds”
The actual shift, from index to middle finger signifies the change of location, from Canada to the next place, and the number of herds is expressed by moving the middle finger downwards just slightly, as a way of signifying the ‘two-ness’ of the herds, as in the above example. This means, these two additional matches also illustrate the most central aspects of his utterance so far, i.e., the *shift of location* and the *correct number* of herds.

During the expression of the second part of the utterance, i.e. “on...the Salish-Kootenai”, Bob makes several changes in his bodily actions. He moves the direction of his gaze, from looking out over the landscape to presumably looking at Katrin, and he also makes a pointing gesture. The gaze toward Katrin may serve as confirmation, given that she may know something about Salish-Kootenai or is familiar with the herds located there. The *deictic* pointing gesture with his thumb, serving as indicator of direction, goes upward and backward, while he utters “Salish-Kootenai” (as seen in Figure 15).

![Figure 15. The shift of gaze and the pointing gesture, and the continuance of the pointing gesture upward and backward while pronouncing “Salish-Kootenai”](image)

However, from the holistic perspective of the framework, this embodied action is actually a *semiotic gesture* (Goodwin, 2003, in subsection 4.3.2), since the context in which the gesture is carried out organizes and scaffolds the gestural expression. In this particular case, the direction of the pointing gesture actually provides additional information, namely where Salish-Kootenai is situated in relation to Bob’s own position. That is, the semiotic gesture adds additional information, namely the location of Salish-Kootenai in relation to Bob’s actual position. Salish-Kootenai is, according to the semiotic gesture, situated behind Bob’s back. Furthermore, the quite wider scope of the gesture could signify the greater distance to Salish-Kootenai, which in reality is approximately 100 miles. It should be noted, however, that although Bob did not say where in Canada the herds were situated, in the previous sentence (line 3), the distance to the US-Canadian border is...
approximately 30 miles, i.e., much closer than to the Salish-Kootenai. Maybe this is why he did not make any deictic gesture when he talked about Canada (see Appendix for a map over the territory). It should be stressed that Bob’s distinct upward movement highlights another, but significant aspect, namely the fact that the Salish-Kootenai are situated on the other side of the Rocky Mountains, which means one must ‘go upward’ and travel through the mountain passes of the Rocky Mountains, in order to reach the other side (The Rocky Mountains are visible in the background of the images with Bob in the front). This means, the additional information conveyed in the semiotic gesture, but not in the speech, is also a mismatch (Goldin-Meadow, 1999, 2003, in subsection 4.3.2), containing the additional information of direction, the need to pass through the Rocky Mountains as well as the distance to Salish-Kootenai. Although the distance in miles is not particularly great, the crossing of the Rocky Mountains adds considerable effort to the trip. Thus, the entire gaze-gesture-speech-context system provides different, but complementary, information to the construction of shared meaning.

However, before he continues his account of the different herds, he changes his gaze again, and looks down at his hands, while also moving his left hand downward (as seen in the first three images of Figure 16).

![Figure 16](image)

**Figure 16.** The beginning of the utterance concerning the Northern Cheyenne.
Meanwhile, his facial expression appears more concentrated, maybe serving to indicate the verification of the correct name of the next location. However, he really seems to ‘know’ it, given that he begins the utterance almost immediately, looking satisfied and relaxed, which is also manifested in his tone of voice, saying confidently “We have a herd for the Northern Cheyenne”. It should be pointed out, however, that when he starts to utters “a”, he holds his right ring finger with his left hand, and rapidly shifts his gaze upward and looks at Katrin (as demonstrated by image (d) in Figure 16). During the articulation of “Northern Cheyenne” he moves his ring finger downward, as a way of manifesting and expressing this third location of Mustang horses, similar to the above examples. As also illustrated in the above images, he uses his fingers as a way to put the places in order. Canada is represented by his index finger, Salish-Kootenai by his middle finger, and the Northern Cheyenne is represented by his ring finger.

In the following moment, without any hesitation, he utters in a confirmed/satisfied tone of voice “we have a herd for the Crow...”, during which he gazes at his hands/fingers once again (see the first images of Figure 17). However, as soon as he has grabbed his middle finger, he stops looking at it, but then moves it more downward than before (as illustrated in Figure 14). He also lifts his gaze upwards, when he utters “a”, then looks out over the landscape during the rest of his utterance (as seen in the last image in Figure 17). While this action is very similar to that indicating the Northern Cheyenne, he uses a different finger to signify the Crow. Given that he has already used his middle finger, he differentiates between the places by the extent of the movement of the middle finger, that is, it is much wider in scope than in the previous Salish-Kootenai example. That is, he makes another speech-gesture match.

**Figure 17.** The shift in gaze and the wider movement of the middle finger in the Crow example.
He continues with his listing, saying “...and we have a herd being put together for Fort Belknap”. Once again he shifts his gaze from looking out at the landscape down to his hands. But when his left index finger touches the right little finger (it is actually his little finger, as his ring finger is hided behind his index finger in image (a) of Figure 18), the actual shift of his gaze occurs to coincide with the articulation of “a”. The change of gaze, however, is really two shifts. First he changes direction from looking at his hands to gazing out over the landscape, but then very quickly alters his gaze and starts looking at Katrin instead. During the verbalization of “being put” he pulls his little finger downwards slightly, and during “together” he begins the pointing gesture. He points ahead and upward with his index finger (as indicated by images (b-e) of Figure 18).
While Bob pronounces “for”, in a quite prolonged way by stressing the vowel sound in ‘for’, the very peak of the upward movement of his semiotic gesture occurs (as seen in image (d) of Figure 18). The articulation of “Fort Belknap” is thus delayed until his left hand has completed the ahead/forward movement and moved down to touch his right hand’s little finger again, moving it slightly, as a way of confirming the correct location. In addition, during the ahead/forward and downward movement of his left hand, he also changes his gaze, from Katrin to me (as illustrated in image (f) of Figure 18), which is also a way of manifesting and sharing his meaning of the utterance with people other than Katrin. Thus, the change of gaze marks my inclusion in the discursive dialogue. As with the Salish-Kootenai example, the overall gesture-speech-context system provides the ‘whole’ meaning. Fort Belknap is, according only to his gesture, situated in the opposite direction of Salish-Kootenai. This direction is correct, given that the Rocky Mountains are behind him (to the West), and Fort Belknap is located approximately 200 miles to the east, i.e., in front of him. This means, both matches and mismatches are present in this utterance.

Bob then continues his listing, saying “for the Assiniboine and the Gros Ventres”. During the articulation of “Assiniboine”, he shifts his gaze and looks out over the landscape, at the same time pulling his right little finger, verifying the correct location (as seen in image (a) of Figure 19). During the verbalization of “and”, he changes his gaze again, looking at Katrin. While he utters “the Gros Ventres”, he makes another, but very tiny, tapping action on his right little finger. Then Katrin makes a humming sound, and her positive tone of voice is confirmed by Bob with a slight nod, that manifests their agreement through co-regulation interaction.
Then Bob pauses for a moment, and says “and we certainly will have more on the Northern plains again”, while he looks away, out over the landscape, almost from the very beginning of the utterance (as seen in image (a) of Figure 20). This shift in his gaze might indicate the major change of topic. Although still talking about the horses, he has proceeded from enumerating the places that keep horses originating from the herd of the ranch to the general mission of the ranch. While making his utterance, he also moves his hands in a concerned way, moving some piece of material in his hands, which earlier was kept inside his left hand. His concern is also reflected in his tone of voice and facial expression. He continues, after a couple of seconds, saying “all that’s part of that dream”, and when he has uttered “that’s” he makes a very determined grimace with his lips, stressing the significance of his mission for the ranch. Katrin then hums in agreement, immediately saying in a quiet but empathic tone of voice “oh, that’s wonderful”. Bob continues to say “that make them su..survive”, during which his facial expression alters with an even more serious and thoughtful look and he gazes downwards (as depicted in image (b) of Figure 20). His tone of voice now alters to very serious, quite sad and he hesitates slightly just before he begins to pronounce “su..survive” which he utters in an almost cracked voice. During the whole utterance, he continues to manipulate intensively the piece of material in his hands.

Figure 19. The shift in gaze during the utterance of “for the Assiniboine and the Gros Ventres”.
Figure 20. The first two images illustrate Bob’s bodily actions when talking about the future of the horses, and the last one shows his altered bodily actions when he has changed the topic.

Bob pauses for a second or two, still in the same position, manipulating the piece of material and with his very serious facial expression. Then he begins to say “but it is something else”, and when he verbalizes “is”, he adjusts his bodily position to again lean on the side of the vehicle, along with shifting his gaze, to start looking at the audience, and tells us about something else. The new topic concerns some pieces of advice about how to take pictures of the horses when we are out on the prairie, and the entire utterance concludes with a joke. This means, the shift of topic is manifested in gaze, tone of voice, bodily position, and facial expression (i.e., the body as a social resonance mechanism). When he has told his joke, in a rising and teasing tone of voice, then laughing a little, the rest of us are also laughing. As in the beginning of this episode, the change of topic is once again manifested in the cross-modal changes of the cross-modal embodied actions.

Generally speaking, all his bodily actions, i.e., facial expression, tone of voice, the downward gazing as well as the bodily posture reflect the compatibility between his mission with the ranch and his embodied actions. As pointed out earlier, the horse has a very significant role in the Blackfeet culture. Blackfeet were once a very large and successful nomadic tribe, and the horse was, for instance, used in buffalo hunting, warfare, as a measure of wealth, and considered to have great spiritual powers (cf. Ewers, 2001). Moreover, the horse was considered equal to man, being a creature with a soul. In view of these aspects, it is easier to understand why Bob put so much effort and emphasis in his expression about the survival and future of the Spanish Mustang horses.

Taken together, at a general level, this example demonstrates the act of scaffolding (cf. e.g., Clark, 1997; Hendriks-Jansen, 1996; Wilson, 2002; Wood et al., 1976) or situational decoupled strategy (Wilson, 2002 in subsection 3.2.1) or surrogate situatedness (Clark, 2005, in subsection 3.2.1) by using one’s fingers as artifacts/representational devices during the enumeration of the different places that have Spanish Mustang horses originating from the herd at the ranch. In total, there are nine different herds situated at seven locations. His cognitive strategy of off-loading the act of remembering into a visual and external representational format through
embodied actions is very obvious and observable in the above examples. However, the very action and experience of moving/touching his fingers actually brings forth the names, by facilitating the shaping of the numbers and locations, instead of functioning as a way of externalizing the already existing names of the places.

More specifically, this strategy has far wide reaching consequences than stated above. The explanation thus far, focuses on what one could term individual scaffolding, but it should be noted that Bob is actually both in relation with himself and with the others. By using his fingers he creates and experiences action-perception loops through his body in motion, given that the actual movements provide him with the kinesthetic experience of movement as well as the felt sense of touching when he grips and senses his fingers during the enumeration of places, in which every embodied action he makes creates a spectrum of embodied experiences to him. Thus, he constructs and experiences sensory-motor information at the same time through the crossmodal integration of his embodied being. This type of relational embodied scaffolding functions as a way of being in dialogue with oneself as well as with others, and is accomplished by the activation of his own mirror neuron system through Mead’s loop (see subsection 4.3.2). This means, most of the enumeration and the act of remembering is an intra-personal interaction but there is also an inter-personal theme present. This is evident during the enumeration of herds, in which the interaction changes between being intrapersonal and intrapersonal. When Bob is within co-regulated interaction, he looks and gazes at the others, mostly at Katrin. But when he lowers his gaze or breaks the mutual gaze, he is instead in relation to himself, constructing the correct name and location, which is shown through his facial expression and gaze that indicate he is not ‘being here’ in dialogue. When he has enacted the correct location in the process of remembering, he enters into the intercommunicative dialogue with the others again by re-establishing eye contact and making his utterance. During most of the co-regulated interaction, Bob looks at Katrin, which is not surprising since they know each other from earlier on, which manifests the temporal and personal relationship of the unfolding dialogue. Furthermore, Bob ‘knows’ that Katrin ‘knows’ about the situation of the Spanish Mustang horses as well as the significant role of the horse in Native American culture.

The information is conveyed in gesture-speech utterances, sometimes as matches, and sometimes as mismatches which provide additional information. It is interesting to note, that the number, location, and shift of location are manifested in speech-gesture matches, as a way of highlighting the most significant issues in the expressions (cf. subsection 4.3.2). However, the most distant and/or demanding locations are also manifested in mismatches. This is illustrated in the semiotic pointing gestures for Salish-Kootenai, which is located on the other side of the Rocky Mountains, and in the case of Fort Belknap that is situated far away. Furthermore, the symbolic aspects of gesturing are shown, particularly in the ‘Up-ness’ and “Two-ness” gestures (cf. subsection 4.3.2).
Moreover, despite this episode occurring in spontaneous social interaction ‘here and now’, it also demonstrates several forms of offline cognition (cf. Subsections 3.2.1; 4.1.1-4.1.2). On the one hand, the different herds are not present in the situation and neither are the names of the actual locations. On the other hand, at the conclusion of the above episode, Bob considers the future, declaring one of the major visions of the ranch, i.e., that the Spanish Mustang horses should be spread across the Northern plains again. As manifested cross-modally, through tone of voice, prosody, facial expression, bodily posture, gaze and verbal utterance, it is very apparent that this is of major importance to him personally and as a Native American. He identifies strongly with the horses and their fate. This means, his attitude (cf. Subsection 4.1.2) is highly influenced and reflected by embodied states; whether the performance is carried out online (embodied practice – the body as a social resonance mechanism) or offline (by some kind of embodied simulation/re-activation of embodied practice (i.e., the body as a means and end). In addition, the above example demonstrates how so-called integrated online and offline cognition occur, and the effortless changes between the different processes is explained by the view that human cognition is fundamentally grounded in embodiment. In summary, the act of relational embodied scaffolding in the above episode significantly contributes to the cognitive processes of memory and categorization, whereas the final phase demonstrates the ability to imagine the future.

7.2.2 They own me
Despite the detailed account of embodied actions in the above presented episode (subsection 7.2.1), it is, however, not optimally illustrated given that it can be criticized for not properly representing the dynamics of co-regulation in socially interactive cognition. It mainly focuses on Bob and little on the interacting partners of the unfolding dialogue. In the following episode, the video-recording is captured from a different angle, thus providing a better possibility for a greater picture of social interaction among the interactive partners. Before this episode occurs, we have all traveled by car into the prairie and stopped at some of the different herds ranging at the ranch. During these stops, we were able to observe, photograph and encounter firsthand the Spanish Mustang horses. During each stop, Bob told us, in a very enthusiastic manner, about the horses and their pedigrees, various incidents that have occurred, and the herds’ social organization. In the following analyzed sequence, lasting approximately 6 seconds, Bob tells us about his relationship with the horses, and the dialogue proceeds as follows:

1. Bob "everyone says to me - how many horses do you own?"
2. Bob "I don’t - they own me" (then Katrin says ‘yeah’)
3. Katrin "...How many horses own you? (followed by laughter)
4. Bob "yeah...really"

Initially, the change of topic is marked by an altered posture, functioning as an indicator of the change of subject. Before that, Bob was looking out over the herd, but now he turns slightly toward Katrin and Peter while also focusing his attention on them and on some of the other members of the
group (as seen in the images of Figure 21). Even Katrin moves her head, now facing towards the horses. Taken together, they create a referential triangle of shared attention, and during the entire episode, Bob actively alters his focus of attention between the horses and the audience, as a way to maintain that the interactive dialogue unfolds without interruption and/or hesitation. That is, the actual dialogue is created and manifested through bodily actions \textit{in situ}, and not established beforehand.

\begin{figure}[h]
\centering
\begin{tabular}{cc}
\includegraphics[width=0.4\linewidth]{image1} & \includegraphics[width=0.4\linewidth]{image2} \\
a) & b)
\end{tabular}
\caption{The change of topic is manifested in both bodily posture and focus of attention.}
\end{figure}

However, just before Bob begins to make his utterance, he alters his bodily position, now to face the horses, as a way to indicate that he offers them his focus of attention back. This might serve as an indicator that the horses, which surround them in a half-circle to the left, have a central role in the subject to be discussed (as seen in the shift from Figure 21 to Figure 22).

\begin{figure}[h]
\centering
\begin{tabular}{cc}
\includegraphics[width=0.4\linewidth]{image3} & \includegraphics[width=0.4\linewidth]{image4} \\
a) & b)
\end{tabular}
\caption{The topic to be discussed, the relationship between Bob and his horses, is expressed through his bodily posture, gesture and focus of attention.}
\end{figure}

Before Bob begins his utterance as well as during the verbalization of the first part, he leans his body slightly forward toward the horses (as seen in image (b) of Figure 22). Meanwhile Bob verbally utters “everyone says to me”, in a fairly alert and easy-going tone of voice that is also reflected in a delighted facial expression, while also making a little and quick gesture with
his hands. The very slight movement of the hands can be viewed in the images of Figure 22, but the gesture is quite loose and unspecific in its form. It may only function as an accompanying action to the slightly forward motion of his posture.

When he continues the expression, saying "...how many horses do you own?", still in the same tone of voice and facial expression, he once again moves his hands. This time, however, the gesture is bigger in scope and more manifested, given that it is molded by both his hands, having its 'peak' during the pronunciation of "horses". The entire gesture unfolds in less than a second, but seems to be a significant part of the cross-modal and co-regulated interaction pattern. This means, the gesture-speech match highlights horses, which can be interpreted as the central issue of the utterance, but the gesture also serves as an indicator of the numbers of horses, providing an 'answer' to how much 'many' actually is. That is, the gesture-speech co-regulation both serves as a mismatch and a match in this case, because the wider gesture with both hands indicates that the amount of horses is not very small or tiny, which is reflected in the wider scope and the distinction of the gesture’s movement. The gesture’s peak can be seen in image (a) of Figure 23.

![Figure 23](image)

**Figure 23.** The gesture performed during the verbalization of horses in the utterance "...how many horses do you own?"

Bob is now back in his initial position (illustrated in image (b) of Figure 23), but before he continues with the next utterance, he turns his head and gazes toward the group, offering them his focus of attention, but not explicitly looking at Katrin and Peter (as seen in image (a) of Figure 24 as well as in Figure 21 which also shows Peter to the left of Katrin). This indicates how the interaction is co-regulated, given that Bob does not take anything for granted and the shift in focus of attention and slightly altered bodily posture are ways of establishing and maintaining the unfolding interaction. Although the illustrated figures are unable to provide rich information about Katrin’s and the other members of the group’s attention, Katrin’s slightly altered bodily posture toward Bob indicates that she is still an active partner in the interactive dance. That is, it takes two to tango! Bob
subsequently utters, in a teasing and ironical tone of voice "I don't - they own me". During the first part of the sentence "I don't" Bob once again makes a similar quick and outward gesture with his hands as in Figure 23 above, but this time it is more loose and ill-defined. This gesture can be interpreted as highlighting that he does not own the horses, but the interesting part unfolds when he says “-they own me” as illustrated in Figure 24. The entire pronunciation takes less than three seconds.

**Figure 24.** The gesture in form of bodily actions toward and from the thorax performed during the verbalization of “-they own me”
This is what occurs. After the minor gesture performed during the verbalization of "I don’t", he makes a short pause before saying "- they own me". Meanwhile, he moves his hands upward and toward his thorax, but without touching it (as seen in images (a) and (b) of Figure 24). However, when he utters "they", the toward motion ends and takes the opposite direction instead, moving away from his thorax (see images c and d in Figure 24). That is, the change of direction of the movement occurs during the verbalization of "they", and during the rest of the verbalization, namely “...own me” the outwardly action continues (as seen in images (d-f) of Figure 24). In addition, when Bob has finished his utterance, he holds his hands motionless in that particular position for a while (as seen in image (f)) and Katrin then says “yeah”. Her agreement is also manifested in her tone of voice, which is positive and rising, but rather quiet. Moreover, she shifts her focus of attention to Bob.

Next, Bob starts to move his hands again into a more widespread gesture which he maintains for a time (as seen in Figure 25).

![Figure 25. The co-regulated bodily actions performed by Bob and Katrin while Katrin utters "...How many horses own you?"

At the same time, Katrin paraphrases Bob’s utterance by saying, in a quiet, but rising and joyful tone of voice, ” ...How many horses own you?” during
which she very rapidly changes her focus of attention from Bob, toward the horses, and then looks back at Bob (as illustrated in images (b-d) of Figure 25). In more detail, Bob’s participation in the co-regulated interaction is indicated by the changes in gesture; bodily posture and facial expression (see images in Figure 25). Bob’s gesture ends when Katrin utters “own” and he holds his hands along side his upper body when she finishes her question. During her utterance, he also shrugs his shoulders, holding them in the upward position for a moment. Additionally, his facial expression changes, displaying the emerging meaning they create together. His face changes from a stricter, resolute expression into a grin with closed lips. Meanwhile, he lowers his shoulders (as seen in image (d) of Figure 25). When Katrin has finished her verbal utterance, Bob nods in agreement. Then Bob says “yeah...really” in a more ironic and falling tone of voice. As indicated in image (d), both Katrin and Bob lean toward each other and then they all laugh, and the laughter occurs just after the end of the verbal expression. This is interpreted as a way of showing their agreement of understanding and meaning-making activity through co-regulated embodied interaction.

Bob then immediately looks away from Katrin, paying attention to the herd, and his bodily position is slightly altered toward the horses as well (as seen in image (a) of Figure 26). Katrin also looks away, but just a moment later she makes a distinct action toward Bob, and really bends over in his direction as well, as she turns her attention to him (as seen in image (b) of Figure 26).

![Figure 26. The bodily actions performed after the spoken dialogue.](image-url)
However, Bob seems to take no notice of her significant action, he is still focusing on the horses, and as a consequence Katrin leans back again and starts looking at the horses too. Bob’s action of not paying any attention to Katrin’s movement toward him actually ends the whole conversation, which Katrin obviously understands. However, she comes back into the ‘bigger’ picture or theme, namely the horses, by also focusing her attention on them. This means, she accepts the conclusion of the conversation, but still holds close the central theme of the dialogue, namely the horses. A couple of seconds later, Bob begins to talk about something else, although the horses are the central theme, and the communication starts when Bob turns his attention back to the visitors.

Taken together, at a general level, this episode illustrates how ubiquitously present speech-gesture matches and mismatches are in human social interaction, and that they sometimes serve as deliberate attention devices for certain kinds of information. This means that the actual gesture has a significant and conscious communicative intention, not only serves as a way of shaping and expressing Bob’s own thoughts.

Moreover, further analysis of the mismatch performed by Bob during and after the utterance “they own me” reveals earlier unforeseen but significant aspects of socially interactive embodied actions (as seen in Figures 24 and 25). In more detail, Bob’s upward gesture towards his thorax (as seen in images (b–c) of Figure 24), is considered an example of the fourth social embodiment effect, i.e. compatibility between cognitive and bodily states enhances performance, although Barsalou et al. (2003, see subsection 4.1.2) did not consider gesture and language. This means, the motion towards his chest is easier to make when the issue at hand is something positive than something negative. In this case, the horses are very positive for Bob which is also manifested in the semiotic gesture, given that the horses graze in front of him. However, he then changes the direction of the gesture very quickly, just before he actually would have touched his chest, and the ‘toward gesture’ takes the opposite direction, that is, outwardly. The actual change of direction happens when he utters ‘they’ which represents the horses. When the sentence is completed, he holds his hands still for a while, and then makes a bigger and more widespread gesture which he ‘freezes’ during a pro-longed moment while he shrugs his shoulders. These embodied actions are ways of representing the bizarreness of the utterance, namely that the horses should own him. In order to highlight the conflicting meaning expressed, he makes a certain kind of mismatch. In other words, Bob makes a communicative and intentional mismatch, as opposed to the more unconscious/unintended mismatches that Goldin-Meadow reported (see subsection 4.3.2). It is important to point out that Katrin actually grasps the underlying meaning of the odd utterance, manifesting this in several ways. She paraphrases his verbal expression by saying “How many horses own you?” in a joyful and teasing tone of voice. In addition, the following laughter also serves as a way of manifesting the emerging meaning through co-regulated interaction as well as Katrin’s leaning towards Bob. These actions undoubtedly demonstrate the relational aspect of social interaction and
communication which is based on the four fundamental functions of the body in social interaction and cognition (subsection 4.4.4).

Furthermore, the whole issue concerns abstract and high-level cognitive entities, i.e., offline cognition, namely ‘ownership’ but also ‘irony’. At no single moment during the interaction is the issue of irony explicitly discussed, but instead emerges during the unfolding dialogue. Regarding the reference to ‘ownership’, an alternative explanation is that Bob assumes the perspective of the horses during the ‘toward’ phase of the gesture, and then switches the perspective to his own during the ‘outwardly’ phase. However, this alternative explanation is at odds, given that the irony can be interpreted both in a literal and practical sense. From a realistic perspective, the horses cannot own him, however, in an emotional sense, his devotion to the ranch and its horses implies they ‘own’ his full attention.

7.2.3 Lending a hand in joint action

While above two presented episodes mainly concern different aspects of social interaction and communication of interactive partners, the following episode concerns about joint actions (cf. subsection 5.2.3). Generally speaking, joint action is when several interactants should jointly perform something together. In this particular case, they should together lift a filly that is unable to rise by itself. The above episodes have been more straightforward in the verbal dialogue, as demonstrated by the excerpts presented at the beginning of each subsection. In this episode, however, the excerpt of the verbal expressions alone offers little information about the episode; that actually consists of several sub-episodes and that the verbal interaction is rather random. The excerpt proceeds as follows:

1. Bob            “Ohoa… get up!”
2. Bob            “…Come on my girl”
3. Bob            “…Get up!”
4. Bob            “….they’ve just been dewormed [Katrin says ‘yeah’], …and that is one of the reasons she’s in that bad shape…but she’s pretty alert” [pause]
5. Katrin         “you’ll have to take each other by the hands and together lift …from underneath”
6. Bob            “Let’s go, let’s go”
7. Bob            “Come on”
8. Bob            “Up! …Up!”
9. Bob            “she don’t want to”
10. Peter         "take… take each others' hands... from opposite sides"
11. Myself        “from underneath”
12. Bob           "Katrin, come over here, we need you so we can pick her up"
13. Bob           “Up!”

Despite the lack of verbal information in the unfolding dialogue, this episode offers significant aspects given that it actually is a spontaneous situated

35 Parts of this section are presented in Lindblom (2006).
activity of joint action. In order to outline the entire picture of the episode, I chose to illustrate all the different sub-episodes since they provide much of the meaning of and to the successful lift. The episode is analyzed from both an embodied, and from a social interactive perspective. It should be added, however, that the descriptions of the relationship between facial expressions, tone of voice, posture and so on, are less detailed here from an embodied perspective than those presented previously, since the major focus is on the joint action.

Normally, the several herds of Mustang horses at the ranch range freely on the prairie, but sometimes they need to be looked after. On one occasion we entered one of the paddocks at the farm, where some fillies were kept near the dwelling house for surveillance (as seen in image (a) of Figure 27).

![Image](image_url)

**Figure 27.** The fillies kept in the paddock for surveillance, and Bob patting the filly which is unable to rise by herself.

One of the fillies is lying on the ground, and makes some unsuccessful attempts to rise on its own. Meanwhile, Bob accompanies the filly’s attempts by saying in a tender but encouraging tone of voice “Ohoa… get up!” Nevertheless, she continues to lie down and while Bob approaches her, he utters emphatically “Come on my girl”. Once he is close to her, he sits on his heels and starts patting the filly on its shoulder and back in order to establish a bond with her (as seen in image (b) of Figure 27). He then rises taking a couple of steps backwards while encouraging the filly to rise once again, by clapping his hands. The filly starts immediately to rise and almost succeeds in her efforts, but then lies down even when Bob claps his hands once more, encouraging her by gently saying “Get up!”. Then he approaches her, squats and begins to pat her on the shoulder once again looking quite concerned, but the filly continues to lie down. After a while, he stands up and tells us, in a neutral informing tone of voice, that “they've just been dewormed [Katrin says 'yeah']....and that is one of the reasons she’s in that bad shape...but she’s pretty alert”. At this point, the joint intention to assist the filly to rise emerges, although nobody explicitly expresses the notion. The situation at hand is obvious; the fact that the filly is unable to rise by her self is obvious, and everybody knowing it is not healthy for a horse to lay down for too long are some significant factors for the joint intention to emerge. Moreover, nobody is able to lift the filly him/herself given here
weight and size. However, there are several persons available to actually lend the filly a hand. The accomplishment of the helping filly to stand on her own legs unfolds during the following episode, which is structured into two sub-episodes.

The visitors are discussing the filly’s condition, when Katrin says (in Swedish) “you'll have to take each other by the hands and together lift ...from underneath”. At this moment, Bob is standing beside the filly and begins to bend down at the same time as the filly starts yet another rising attempt (as seen in the images of Figure 28). It should be noted that it is difficult to say who is influencing who here, the filly or Bob. The filly might notice Bob’s downward action as an intention to rise, or Bob might be grasping her intention to rise, and for that reason he bends down to assist her. Together they create a co-regulated activity with the intention to 'get up'. Bob’s intention is also manifested in his speech, and during the lifting attempt, he utters rapidly “Let's go, let's go” in a low but energetic tone of voice.

![Figure 28. Bob's and the filly’s lifting/rising attempt.](image)

Bob places his right hand under her belly (see image (d) in Figure 28), but just when it seems their joint effort could succeed, Bob needs to adjust his grip and in the moment he does that, the filly stops trying. As a consequence, the lifting attempt fails. During Bob’s and the filly’s rising effort, BoChief and Dag approach in order to assist them (BoChief is seen in image (d) of Figure 28).

However, Bob initiates a new lifting attempt at once (as seen in Figure 29), and now he adjusts his grip to be more from above than from the side, probably an adjustment in accordance of the prior try. It should be noted
that BoChief and Dag enter with big strides as well as almost in step (as seen in image (a) of Figure 29). Without paying any particular attention to each other, for instance, they do not establish eye contact since everybody's attention is directed toward the filly. However, Bob notices that their entrance, and utters “Come on” while together they take their positions for the actual lift (as seen in images (b-d) of Figure 29). Thus, they create the joint intention of lifting the filly, and in order to do so they take a hold on the filly from different directions. Bob from the right side, BoChief from behind and Dag from the left side (as seen in image (d) of Figure 29).

Without any hesitation, they begin jointly to lift the filly up, while Bob in an alert and positive tone of voice says “Up! ...Up!” presumably to the assisting men, but he also klicks his lips and tongue, as a way of encouraging the filly to exert herself in the lifting attempt. These kinds of sounds are often used as encouraging commands when handling horses. However, the filly makes no attempt to exert herself and the men’s effort fails as well, seemingly because they did not really cooperate with each other or did not know exactly how to perform the joint action of lifting, and consequently the unregulated interaction, lasting about 5 seconds, comes to nothing (as seen in Figure 30).
When the men realize that their lifting attempt has failed, they all back away from the filly at once (as seen in Figure 31). This way of acting serves as an indicator that they intuitively understand that fact, and want to step away from the unsuccessful situation. Furthermore, the men’s reverse action from the filly might demonstrate the first social embodiment effect, in which perceived social stimuli results in both cognitive and bodily states. In this case, the still lying down filly could elicit the social stimulus of failure, which results in two of the men’s backing away. While they back away, Bob states “she don’t want to”, but not in a judging or disappointed tone of voice, but rather verifying the actual situation. It is obvious they need to make another attempt, and in order to explain how they should make it this time, Peter says (in Swedish), watching the men from the side, ”take... take each others’ hands... from opposite sides” (image (c) in Figure 31), and I add (in Swedish) “from underneath”.

At that point in time, Dag aligns his bodily position towards Bob’s, just for a short moment, which can be seen as a way of establishing a joint focus of attention and agreement of the task ahead. Then Bob turns around, and while pointing at Katrin and looking at her, he says ”Katrin, come over here, we need you so we can pick her up” (as seen in Figure 32). More specifically, while Bob utters the above sentence, he moves his right hand, which initially points at Katrin, toward the filly, and this action is now accompanied by his left hand which unites in a movement that can be interpreted as a kind of lifting gesture (as seen in images (a-d) of Figure 32).
Figure 32. The first two frames (a-b) show Bob’s shift in pointing from Katrin to the filly, and the following ones (c-f) display Bob’s lifting gesture and then Dag’s accompanying gesture.

The lifting gesture begins while he lowers his upper body and hands slightly, and then he continues the whole gesture upwards. Thus, this gesture illustrates how he expresses and shares his thoughts in action as well as in speech. Moreover, by actually performing the lifting gesture he uses his body as a tool or representational device for the sake of communication.
It is important to notice that the timing of the upward action of his hands and torso coincides with the last part of the utterance ("pick her up") and the particular forming of his hands (Figure 32b-32e) into a general gesture of the imaginary grip. That is, he makes several matches. Furthermore, it can be viewed as a way to assure Dag they should lift her jointly, which is represented by the gesture signifying the closed grip of each others’ hands. Then (Figure 32e) before Bob lowers his hands, Dag who has been looking at Bob, starts to make a lifting gesture, which is slightly different, as a way of sharing Bob’s idea of how to lift, as well as indicating that he actually has understood what to do in the next try (Figure 32e). However, Dag does not make the same representational gesture for lifting. Instead, he expresses how he should act from his perspective – how he should use his hands in the coming lift, whereas Bob’s gesture instead reflects how to grasp each others hands jointly from a third-hand perspective. In short, different perspectives of the lifting are expressed in gesture, but they are both ways of shaping and externalizing thoughts in socially embodied actions.

It should be noted that apart from creating, showing and expressing thoughts, Bob’s lifting gesture has kinesthetic and phenomenological aspects (see section 3.3). This means, besides the gesture’s end result, the dynamics of the action itself is a fundamental part of the cognitive function since the phenomenological experience of the bowing action are part of Bob’s action understanding as well as Dag’s action recognition, via presumably through a embodied simulation/re-enactment mechanism or embodied practice (see section 4.2). The entire sequence in Figure 32 takes less than three seconds, in which Dag’s gesture lasts less than one second, which is obviously enough, since before attempting the next lift (as seen in Figure 33) Dag repeats his lifting gesture, but now also towards BoChief while Bob is looking at them. This illustrates how the body functions as a social resonance mechanism facilitating the understanding of another person’s actions as well as means and end in social interaction. Furthermore, Dag’s repeated lifting gesture, lasting less than a second, serves as both intra-communicative as well as inter-communicative representational device. By repeating it, he wants to assure both himself and the others how they should act in the new lifting attempt in order to succeed. Thus, the body serves as a representational device.
Figure 33. Dag’s repeated lifting gesture, which now both represents his coming grip as well as their first-hand as well as third-hand perspective.

It should be noted that there is another significant difference in his repeated iconic gesture, since both first-hand and third-hand perspectives are represented this time. As seen in images (c-f) of Figure 28, Dag first makes the lifting gesture from his first-hand perspective, but this gesture is then transformed into a more overlapping gesture of his hands, as a way to
represent the third-hand perspective. As partly seen in images (33f-g), the men focus their attention on Dag’s gesture and then Dag looks down, but it is difficult to see whether he looks at his hands or at the filly. That is, his gesture functions as a way of assuring how he and they should jointly act in order to succeed.

In the successful lifting of the filly, Katrin joins the men taking hold of the front of the filly, whereas Bob and Dag grasp each others’ hands under the filly’s belly, and BoChief takes a hold on her hind quarter (as seen in images (a-b) of Figure 34). It should be noted that the men have already taken their positions when Katrin joins them (as seen in images 34a-34b). When they have all taken their positions, Bob utters “Up” and their joint action of lifting the filly begins at that very moment (as demonstrated in images (c-d) of Figure 34). This time, they are all co-regulated and as a result, the filly finally stands up and is able to join the other horses (as seen in images (d-i) of Figure 34). More specifically, Katrin in front of the filly, pulls it upward, while Bob’s and Dag’s hands together form a joint grip under the filly’s belly (as seen in image (f) of Figure 34). Meanwhile, BoChief has a steady hold of her hind quarter. Although the positions of the successful lift and those of the failed attempts do not differ very much, the differences are significant enough for the joint action to succeed. This means, their jointly co-regulated activity is the difference that makes the difference.
When the filly is up on her ‘hooves’, she is immediately very eager to trot away, and when she does that Bob utters “Yeah...“, which is followed by a huge, deep and loud sigh of relief (in image (i) of Figure 34). When the filly joins her companions and whinnies, everybody laughs and exhales relieved. Thus, the undertaking is completed effectively. It should be added, that the men alone were able to provide the helping hand for the filly, which means that Katrin’s physical involvement in the lift was not of crucial importance for its success. However, her involvement rather served as a social scaffold for the men, supporting the idea that they would succeed this time.

Figure 34. The successful lift and Bob’s huge sigh of relief in the final image.
Taken together, the previously uttered vocal suggestions how to ‘act’, without paying any attention to each others positions and actions in the first attempt to lift the filly were not enough for a co-regulated lifting action to emerge. Instead, when the participants were able to express and share their thoughts through cross-modal interaction, these socially embodied actions (including speech) were realized in the successful lift. All in all, embodied actions provide a ‘helping hand’ in social interaction, and in a practical sense, gesturing actually enabled the men to embody their thoughts in actions and metaphorically speaking, their hands also lent the filly a hand. Beside the role of the four presented functions of the body in social interaction and cognition, the successful lift also reveals how crucial embodied actions are in the emergence and engagement of the shared social situation, actually ‘lending a hand’ for the shared situation to occur and be manifested. Thus, the successful lift illustrates the relational nature of social interaction and cognition.

After the successful lift, the rescue team looks at the filly and her herd mates (as illustrated in image (a) of Figure 35). Then BoChief and Dag leave the paddock, but Katrin and Bob remain. It should be noted that they both place themselves into similar bodily position, each standing with their hands in their jacket pockets holding them into their sides. This means, they unconsciously mirror each others bodily actions (the second social embodiment effect, see subsection 4.1.1) using their bodies as social resonance mechanism (as seen in Figure 35).
The fillies then trot to the other side of the paddock, and as a result both Bob and Katrin turn around in order to follow the horses’ actions. However, the actual turning motions between Bob and Katrin as well as their bodily postures are very co-regulated. Thus, their actions almost look like some choreographed dance movements. These co-regulations are realize by using of the body as a resonance mechanism as well as a means and end in interaction.

All in all, the entire lifting episode is an example of the human capacity to engage in joint cooperative activity, which is based on some kind of we intentionality. That is, the ability to share intention plans, and so on with others, and this ‘sharing’ ability is the ultimate understanding of human intentionality (see subsection 5.3.2). In order to accomplish the joint action, i.e. the successful lift of the filly, they do not have to grasp what to do and how to do it beforehand, but to also co-coordinate their embodied actions in relation to each other (including the filly’s position and action) on a temporal horizon. Taking the entire group as the level of analysis, however, offers an explanation in which their shared understanding is the emergent outcome of embodied actions in social interaction in situ, in which there is no inferential gap separating visible embodied actions from the descriptions of its internal representational form. This means, the ‘shared intention’ of lifting the filly and its accomplishment is the emergent result of the co-regulated interactions between the interactants ‘out there’ and properties in the surrounding rather than the consequence of some advanced mental intentions residing within each individual’s mind. Thus, the joint action is a distributed and collective phenomenon. A crucial characteristic here is that the embodiment of and the embodied actions carried out by the members of the rescue team as well as the properties of the actual situation, the filly lying down, provide the general structure of possibilities and constraints for the particular joint action to be carried out. Thus, the we intentionality emerges out of the entire situation which is greater than and different from its involved parts.

Broadly speaking, given that the visitors are aware of the unhealthy situation for a horse to be lying down for too long, which is basic knowledge for people familiar with handling horses, the intention of jointly providing a
helping hand emerges. It should be noted, however, that none of the participants explicitly encourages anyone else to assist Bob, but as a consequence of the emerging intention of joint action, some straightforward advice about how to actually perform the lift are verbally expressed. Katrin, for instance, says “you’ll have to take each other by the hands and together lift ...from underneath”. Meanwhile, Bob interacts with the filly with encouraging calls as well as by patting and hand clapping, as a means and end for social interaction. It should be noted that in this particular case of human-horse interaction, touch serves as a fundamental means and end for the interaction, given that the filly does not understand human language, so to speak.\textsuperscript{36} In the first lifting attempt, in which Bob is alone with the filly, it is quite difficult to decide whether the ‘up’-movement is initiated by Bob or the filly. We might suppose the filly begins to move, and her rising movement is recognized by Bob’s visual and motor regions beside his mirror neuron system which in turn could be activated. As a result of this spread activations of his mirror neuron system and/or related areas, Bob interprets her slight movement as an intention to rise, and immediately decides to assist her. (Whether the filly grasps his attention to assist him is beyond the scope of this thesis.) However, the reason for the failed lifting attempt lies in the non co-regulated interaction between Bob and the filly, because they were not able to timely coordinate the forces of their combined upward movement. Their individual force intensities vaxe and vane in a way that fails to result in a successful lift from a system level of analysis. However, Bob’s initial grip was not optimal for the entire lifting action, given the filly’s bodily shape and weight, which he actually ‘knows’ and experiences during the lifting attempt. This information is created from his continuous bodily actions and embodied experiences during the initial phase of the lift. For that reason he tries to adjust his grip, but when he tries to alter his hold of the filly’s belly, she stops trying and instead leans back, which definitely ends the attempt. The interaction between Bob and the filly creates the information that Bob is not able to succeed on his own, since the filly cannot ‘understand’ and assist him in the effort to help. Taken together, these factors increase the ‘we intention’ of carrying out a joint lifting attempt.

Meanwhile, BoChief and Dag approach the filly and Bob. Generally speaking, their different action-perception couplings and/or activations correspond to and converge on the filly lying down ‘out there’, and do not reside in the interactants’ individual heads alone as some kind of internal mental intention. The forthcoming joint lifting attempt is the emergent outcome of the situation at hand and the previous episode. Moreover, it should be noted that the act of joint lifting is a cultural activity. This means, the action of lifting something jointly requires that it can be recognized and understood by other members of the particular culture (cf. Fogel, 1993). Furthermore, the underlying factors for the decision to really lend the filly a hand are also influenced by cultural norms. Normally, the Spanish Mustang horses range freely on the prairie such as wild animals and are not taken care of like

\textsuperscript{36} This observation is consistent with other observations of human –horse interaction done during the tour at the ranch.
domestic animals. Bob and the visitors, however, are influenced by their cultural beliefs and opinions that impel them to feel obliged to help the filly.

However, Bob’s attention is still on the filly when BoChief and Dag arrive to assist, since he is busy finding another, more suitable grip on the filly’s body in order to help her up. The men assisting take their holds from proper positions, but their individual actions are not jointly coordinated at all, despite Bob’s accompanying verbal expression ‘Up’. In fact, this lifting attempt is even more unsuccessful than Bob’s previous one despite the assistance of the two other men. Although that they actually know what the act of joint lifting means, the filly probably does not understand. Bob’s and Dag’s individual actions are by no means coordinated, although they stand on opposite sides of the filly, which is a very logical initial lifting position. However, the actual joint grip, under the filly’s belly which is of major importance for the accomplishment of a successful lift, never occurs. They do not achieve a proper hold, and when all the men experience the non-rising filly’s body through their action-perception couplings with her, the phenomenological ‘embodied knowledge’ or intercorporeality (cf. Gallagher, 2007, subsection 2.4.2 and section 4.2) creates the information that the actual lifting will not succeed, and they stop simultaneously. As a result, the men back away from their failed lifting positions. The bystanders’ recognition of the failed attempt, just through witnessing the unsuccessful try, activates their embodied action-perception couplings. As a consequence, the men receive some more verbal advice about how to accomplish the lift, “take each others’ hands...from opposite sides” and “from underneath”. It should be stressed that these verbal instructions are directed explicitly to Bob and Dag, although their names are not expressed in the utterance. Their positions one on each side of the filly and the failed result of the previous lifting attempt actually provide the meaning that fills the gap of the missing ‘information’ of the verbal utterances. The ‘information’, that they (Bob and Dag) should grasp each others hands under the filly’s belly is so obvious regarding the current situation; there is no need to include the subjects in the verbal utterance. Furthermore, nobody explicitly explains from where they should take each others’ hands, since this information is already present ‘out there’. However, Bob feels they need some support, and asks Katrin to join them. There are several motives for asking Katrin. She has previously visited and spent time at Bob’s place, and she is the person, who just a before offers them some advice about how to perform the lift.

It should be noted that the first time gestures are used in this episode occur when. Bob points at Katrin and when he turns to Dag and performs the lifting gesture from the third-hand perspective. This means, the verbal expressions alone have not been sufficient to provide the information about how to accomplish the lift. The men need to readjust their previous joint lifting actions, and as a way of entering into a creative frame (cf. subsection 5.2.2) of lifting, gesturing emerges. By using different kinds of lifting gestures, Bob and Dag provide visible, distributed, and shared representations from different perspectives of the forthcoming grip, from Dag’s own perspective while also one interactive partner with the other
participants. For this reason, Bob’s gestures, and Dag’s initial and repeated gesture serve as a way of creating the model/representation of how to do it this time. That is, gesturing shapes the ‘bigger picture’ of the situation ‘here and now’; what they should do and how they should do it by their active participation in the construction of the fulfillment of the task ahead as a way of overcoming the previous failures. Generally speaking, the ‘solution’ of how they could succeed this time is created, expressed, investigated and confirmed right in front of their eyes as well as through their embodied experiences. With regard to socially embodied action, both Bob and Dag use their own bodies as the accurate models of the grip, and thus there is no need to represent or simulate the perception of the others’ movement into a higher ‘cognitive’ level, given that their own bodies function as social resonance mechanisms (cf. subsection 4.4.1). For instance, Dag’s perception of Bob’s action already provides a first-person model of the joint lifting grip under the filly’s belly which Dag experiences as he would have perceived his own action. This means, instead of decoding the incoming visual stimulus, and then manipulating it via some kind of internal symbols in order to make a model of what Bob is thinking and knowing, and then make a decision of what he thinks he knows about Bob’s thoughts and intention. Consequently, Dag makes a decision, the result of which is Dag’s intention, which is decoded back to a motor command that expresses the actual lifting gesture. A more parsimonious explanation is that Dag’s perception of Bob’s action is the understanding of the lifting gesture, which does not require the extra step of internal simulations or mental representations (see subsection 4.2.1) In addition, Dag and Bob are members of a similar human cultural sphere, and through their ontogeny they have been bootstrapped into the meaning and accomplishment of the particular concept of ‘lifting’, which is represented in the men’s’ gesturing (see Vygotsky, in subsection 2.4.4).

The crucial difference that makes the difference here is Mead’s loop (in subsection 4.3.2), which explains why the gesture itself both functions inter-communicatively and intra-communicatively. In other words, the relational aspect of social interaction and cognition, which is grounded in embodied experience, has the dual characteristics of so-called conversation of significant symbols as Mead (subsection 2.4.6) termed it. This means, Dag’s understanding of the meaningfulness of Bob’s gesture emerges from his ability to perceive Bob’s gesture in himself. It might be possible that Dag’s own mirror neurons system functions as the mechanism for Mead’s loop, and the activation of his mirror neuron system provides Dag with the ability to take the perspective of the other simultaneously. As a consequence of the activation as well as the relational aspect of social interaction and cognition, Dag immediately constructs his gesture, illustrating his perspective of the actual grip, which is then repeated and transformed into Bob’s initial third-hand perspective. Therefore, there is no need to verbalize anything concerning how they should take each others hands and lift. The present joint action is already ‘out there’, together with the interaction experiences of the previously unsuccessful lift, making it possible for everybody involved to get a grasp, metaphorically speaking, of what to do next. Thus, the gesturing functions inwardly and outwardly at the same time. This observation and its
interpretation strengthen the embodied nature of language and communication, proposing an embodied linguistics. Thus, Mead’s loop creates a connection of gesture to social interaction and communication.

The glue that binds together the verbal expression, brain functions and kinesthetic experiences in communication is the movement of the body (section 3.3, and Gibbs, 2006, in subsection 4.3.1). In this episode, concepts of bodily movements include, for example “up”, and the overall activity of the entire episode is indicated by the verb ‘lift’, which means to move something upwardly. Furthermore, Bob’s and Dag’s motor schemas (cf. section 3.3) of bodily movement structure these concepts of “up” and “lift” in the filly episode, since both of these concepts make use of significant parts of the sensorimotor system. That is, language, including speech and gesture, is fundamentally embodied, since it is realized by the same neural organization used to plan and guide action (subsection 4.3.1). The link between action-understanding and the use of predicates in language are also evident here. In speculation, the understanding of the upwardly lifting action does not require a declarative understanding, because its meaning is constituted in a fundamental level of action-understanding. This means, the capacity to act and the capacity to represent is unified through our social body in motion.

Before and during the successful lift, Dag and Bob’s hands actually take hold of each other, creating a joint grip under the filly’s belly. Furthermore, the we intention of the joint action is already present, and the logically related mirror neurons possibly activate the coming next actions, which are co-regulated with the men’s embodied experience of really lifting the filly. Since all of this is happening here and now, they must readjust their own actions from the continually changing kinesthetic experiences of their own movements as well as the changing actions of their interactive partner, in conjunction with their own action-perception loops of the filly’s body. During the entire successful lifting episode, Bob and Dag must maintain the dynamics of creativity of their bodies in motion, to continually ‘keep up’ and ‘be there’ in the unfolding lift. Given their embodied experiences, they actually ‘know’ and ‘understand’ that the lift is proceeding in the right direction, which they experience kinesthetically through their embodiment. This means, the filly is lifted upwards while the joint force intensity of their joint grip under her belly is constructed and functions accurately. Thus, the lifting of the filly is accomplished through co-regulated interaction at the system level of analysis. It should be noted, that the timing of Bob’s utterance ‘Up’, illustrates how it emerges jointly with Bob’s ‘understanding’ that the actions and experiences to be taken by the members of the rescue team should be properly coordinated to accomplish the task. Thus, meaning and intentionality are constructed on-line in situ, in the visible public sphere of embodied actions through interactions between several social and cultural beings as well as their physical and cultural surroundings. The entire lifting episode also reveals how crucial embodied actions are in the emergence and engagement of the shared social situation, given that they are actually ‘lending a hand’ for the shared situation to occur and be manifested.
The use of the verbal expression ‘lend a hand’ to describe the participants co-regulated action exemplifies another common human practice which makes use of body movements for metaphoric purposes, which in turn is based on embodied experience. Literally speaking, by lending the filly a hand, it can be argued that the participants use each others’ bodies as scaffolds, i.e., as external hands belonging to their own bodies, which is the behavior of tooling up others (section 5.4). Accordingly, in the creativity of interaction of the rescue team, their bodies couple by ‘lending a hand’ to each other during the lift, i.e., ‘tooling up’ each other. Thus, they use each other’s bodies as tools in the joint action, and such spatially and temporally distributed couplings result in the successful lift of the filly. From an embodied perspective, the ability of tooling up others has similarities with the act of embodying (section 3.2), which might be based on the ability to extend one’s own body schema and body image (see section 3.3) to incorporate another person.

In the unfolding joint action episode of lifting the filly, several cognitive processes occur such as reasoning, planning, decision-making, categorization, attention and problem-solving, as well as the significant involvements of the four fundamental functions of embodiment. Bob’s declarative pointing gesture at Katrin, for instance, not only functions as a way of focusing the group’s attention on her, but also as an action which selects from whom he wants help. Thus, the pointing action draws attention to certain objects or locations in the environment, present or not, which then become more significant than others. Thus, the body focuses as a director of attention, as well as selection and categorization device. Furthermore, the pointing gesture in itself has meaning for the sake of communication, and it momentarily functions as a ‘helping hand’ or scaffold for the emergence of meaning in the particular situation. Furthermore, Dag’s lifting gestures also play a crucial role in planning the forthcoming lifting attempt. His gestures not only imitate Bob’s gesture but also construct the idea of how he should hold his hands, which in turn is a kind of categorization process. The idea enacted by his gesturing is broadly “I will make my part of the joint grip in this manner”. This is also a way of reasoning, since the other men are able to see what he is thinking and planning to do next, which means they are being offered some suggestions about how the lift could be accomplished. Nobody objects to Dag’s idea, and this mutual agreement demonstrates that the other men accept Dag’s thought. This means, these interactions provide significant elements for the shared meaning to emerge, which thus facilitates and makes the reasoning process visible and takes place ‘out there’. Moreover, Bob’s and Dag’s lifting gestures, for instance, function as bearers of meaning. Although their hands’ lifting gestures by their hands are not the same as in a real lifting act, despite their similarity, they do have representational content. Thus, these gestures play significant roles in the emergence of the overall meaning of the forthcoming lift. Dag’s final gesture before the actual lift, for example, enacts two different Perspectives simultaneously. Firstly, he makes the representational gesture from his perspective, but it is then transformed into a third-hand view. That is, his gesture offers representational content as well as perspective-taking, which
are two crucial characteristics of human language. The evidence that Dag expresses these characteristics through gestures only strengthens the relationship between gesture and speech from an embodied linguistic perspective. This way of gesturing supports the relational aspect of human interaction and cognition, which combines intra- and intercommunicative processes in the construction of meaning. Thus, humans are able to seamlessly swap between external and internal embodied actions, and also rely on surrogate situatedness.

Furthermore, one might ask why Dag and Bob gesture instead of expressing verbally how they plan to act in the lift. One tentative reason is that the embodied action functions as a facilitator of the current task. The understanding is put forward through both the content and the movement of the gestures, which means that the gestures in a more straightforward way unify and connect their different perspectives, two separate representations, by binding them together. The expressed linguistic representations are mostly concerned with what they should do, e.g. “up” and “get up”, which offer some overall theme/structure for the forthcoming joint action, while the gestures show how the men should coordinate their hands in order to accomplish the lift of the filly lying in that particular position. If the filly had been lying in a different position, they might have performed a different lifting gesture. This means, the men’s gestures illustrate and enact in more detail the general idea addressed in speech, and the perspectival dialogue through gesture between the men is fundamental to ensure the success of the joint action. The body transforms the more general representational theme into more concrete and detailed representational forms. As a consequence of this facilitating process, from more abstract ideas expressed in speech to more concrete gestures, the solution to the problem is offered. This means, bodily actions facilitate the reasoning process by altering, or simplifying the general idea into more concrete and visible embodied actions, which are realized through intertwining the four fundamental functions of the body.

This way of acting explains how profoundly high-level cognition, such as reasoning and problem-solving, takes place in the visible and public sphere of social embodied interaction. Thus, the pattern of embodied actions enacts and performs cognitive processes executed in the external and shared world.

### 7.3 Case Study: Discussion and Conclusions

This section begins with a discussion of the relationships between the proposed theoretical framework and the conducted empirical work, as well as summarizing the results of the case study at a more general level (subsection 7.3.1). Secondly, some interesting issues that appeared during the analysis of the empirical data are discussed (subsection 7.3.2). Finally, some general dimensions of the nature of embodied action in social interaction and cognition are outlined and described from a meta-perspective, which can be viewed as the synthesis of the theoretical and empirical aspects (subsection 7.3.3).
7.3.1 The relation between the theoretical framework and the empirical work

The main purpose of the empirical study, as stated previously, is to illustrate, and thereby partially evaluate, the proposed theoretical framework (see section 6.2), by providing some detailed observational fieldwork on socially embodied actions captured in spontaneous interaction in situ. However, since I was not able to measure brain-activity and the firing of neurons, I relied on the four fundamental functions of the body in social interaction and cognition (cf. section 4.4) that are based on empirical data from other studies. Nevertheless, the role and relevance of gesture, speech, bodily posture, gaze and so on, which are visible ‘out there’, are examined in more detail. This approach does not imply, however, that the ‘inner’ effects of embodiment are not considered here. On the contrary, they serve as the very underpinning for these visible actions and the relational nature of social interaction and cognition could emerge. The chosen approach of naturalistic inquiry, i.e. studying gesture speech combinations and other embodied actions in their present material and cultural sphere, results in filling the content gap which is dimmed in work that studies these issues in isolation from their context of use. The framework for the embodied nature of social interaction and cognition can be summarized as incorporating the following aspects:

- Embodied actions are viewed as a general cross-modal interaction system rather than being considered separate parts of facial expressions, gesture, bodily positions and so on.
- Embodied actions are a set of movements that have meaning for the actual person or agent, suggesting a move from the experienced body to the enacted body.
- Embodied actions are always in relation to others.
- The nature of social interaction is relational and meaning and intentions are emergent phenomena of these social interactions.
- Embodied actions are situated in cultural and material contexts.
- Interacting socially through embodied actions is a dynamical process that unfolds with a temporal horizon.
- Socially embodied action is not pre-given, since the social mind develops through the social body in motion.
- The unit of analysis is situated embodied actions in co-regulated social interaction.
- The main focus of analysis is embodied actions in social interaction and cognition.
- Four fundamental functions of embodiment:
  - Embodiment functions as a social resonance mechanism.
  - Embodiment functions as a means and end in communication and social interaction.
  - Embodied actions and gestures function as a direct helping hand in shaping, expressing and sharing thoughts.
  - The body functions as a representational device.

The definition of a framework used in this thesis (cf. section 1.2) states that the framework acts as a guide, in explaining or clarifying the issue, rather
than being correct in detail, serving as a point of view for dealing with a scientific phenomenon (see Crick & Koch, 2003). However, there are no reasons to reiterate what is analyzed and described previously (subsections 7.2.1-7.2.3), but I discuss some general issues here.

At a general level, the results of the case study show how humans use their embodied actions in social interaction as ways of facilitating and coordinating different social and cognitive processes. The different episodes from the conducted case study also offer a further understanding of the nature of embodied action in social interaction and cognition. They also provide more detailed descriptions of how meaning emerges through dynamical and coordinated processes that incorporate the social and material sphere during an unfolding temporal horizon. This shows that visible embodied actions do not merely ‘lend a hand’, but are instead an integral and crucial part of how meaning is constructed during an unfolding temporal horizon that constructs new meaning and coordination on the previous ones. In particular, the result emphasizes the relational aspects of embodied social interaction and cognition, because whenever humans act/think/communicate they are always in interaction, either with themselves or with other interactants. During this meaning creating activity, embodied actions play a crucial role, in which embodiment is the part and parcel of social interaction and cognition in the most general and specific ways. Accordingly, the empirical work thoroughly investigates and shows us how meaning emerges as a joint product of the embodied actions taken by all the interactants together in the lifting of the filly (see subsection 7.2.3). Thus, the analysis and results of the different episodes offer careful and in-depth descriptions of the embodied nature of socially embodied actions that are situated within social and material contexts.

Thus, the analysis and results from the empirical work of the previous sections, in my opinion, illustrate the framework’s potential and truthfulness, given that most aspects of the framework are displayed at a general level. However, the aspect “Socially embodied action is not pre-given, since the social mind develops”, however, is one of points least addressed in the empirical part quite obviously because it does not deal with developmental aspects from an ontogenetical perspective (a topic that is discussed in more detail in section 5.5). On the other hand, it could be argued that the developmental aspect is indeed present to some extent, given that the rescue team viewed as a group develops an understanding of how to accomplish the lift of the filly.

### 7.3.2 Additional empirical findings and conclusions

Furthermore, the empirical study also addresses additional findings that were revealed in the different episodes, which add further aspects that are highly compatible with the actual framework. For example, it is shown that embodied actions contribute to the emergence of understanding in social interaction and cognition beyond the close relationship between speech and gesture in language, i.e. matches and mismatches. Furthermore, the analysis illustrates that Goldin-Meadow’s (2003, subsection 4.3.2) general
use of the term mismatch is too broad and unclear, since there are different ways to describe how the relationship between speech and gesture convey meaning. In subsection 7.2.2, for instance, Bob constructs a certain kind of mismatch, which is performed ‘consciously’, since he ‘freezes’ the ongoing action for a moment. Such a focused aspect is significant for how the mismatch provides meaning in the interaction, which in this case reveals the irony of the utterance. Hence, there is a need to further investigate and analyze different kinds of mismatches, as a way of deepening the understanding about the ways social interactants construct meaning in different situations. In more detail, matches mutually display the same content in both speech and gesture, but they are not superfluous. Instead, the different modalities, in which the meaning is displayed, have qualitatively different characteristics, resulting in a parsimonious and sophisticated way of emphasizing the core meaning of the communicative activity. As seen in the case study, gesture seems to underscore the most significant meaning in the more general linguistic expression, usually in conjunction with other embodied actions. In subsection 7.2.1, for instance, the crucial meaning, i.e. the location and number of herds are also manifested in gesture.

One might ask why humans construct matches at such high frequency. My tentative suggestion is that the crucial elements of the utterance constructed cross-modally simultaneously provide different means of action-perception loops, thus providing a greater and deeper ‘impact’ in the emergence of meaning. In order words, by providing different dimensions of meaning, cross-modal interaction patterns are profoundly grounded in embodiment from a communicative perspective that is suggested to range from the underlying mirror neuron activation, via gesture and other embodied actions, to the use of abstract concepts and metaphors.

Furthermore, humans use embodied actions for different purposes, and some ways they were used in the previous case study are:

- Embodied action is used as a way of establishing focus of attention
- Embodied action marks changes of topics and ideas
- Embodied action functions as a coordination device and organizer of actions
- Embodied action simplifies cognitive complexity
- Embodied action enacts ideas
- Embodied action functions as a perspective taker
- Embodied action links and integrates various thoughts and ideas
- Embodied action functions as an engager to participate in social interaction and joint actions

Given that embodied action seems to have these above functions and characteristics in social interaction and cognition, (which in turn are suggested to be rooted in the four fundamental functions of embodiment identified in section 4.4), I draw the following conclusions:

- Cognition and social interaction are sometimes the very same process
• Humans are able to swap seamlessly between internal and external representations of embodied actions
• Joint action is accomplished by the behavior of ‘tooling up’ others, in which body schemas and body images are extended to incorporate (an)other person(s)
• Embodiment is the underlying foundation for the individual and social mind rather than merely linking or bridging them from these two different perspectives

It should be noted, however, that the particular ways human interactants attribute meaning to certain embodied actions is context-dependent, in which one has to consider; i) the task to be accomplished, ii) the persons involved and their actual embodiment, iii) prior actions, iv) material resources, and v) cultural setting, which altogether are of crucial importance for the meaning creating activity to emerge.

7.3.3 Issues concerning the analysis of the empirical data
Besides the above discussion concerning the empirical result, some other issues arose during the analysis of the empirical data. For instance, as a result of the ‘static’ medium used in the analysis and description of the empirical study in this thesis, i.e. snap-shot images combined with written verbal descriptions, might incorrectly give the impression that we were always aware of what happened, and consciously ‘saw’ and ‘noticed’ all of the embodied actions carried out. However, these actions unfolded very rapidly and dynamically in real time, and we did not always ‘consciously’ grasp what we experienced ‘unconsciously’ as a result of being embodied, which has some significant consequences for the analysis of the empirical material. For instance, there is a vast difference between the ordinary ‘observations’ and experiences of a participating partner who is ‘being out there’ in co-regulated social interaction, and the observations of a researcher in the micro-analysis of the video-recording in the computer program. While the, in scope bigger and wider embodied actions, such as obvious pointing gestures, significantly altered bodily positions, and obvious shifts of gaze are visible and ‘there’ in both conditions, what of the tiny and rapid ones that are so frequent and ubiquitous in social interaction and cognition? For, instance, the tiny movement and touching of Bob’s fingers when he uses them as scaffolds the naming the herds in subsection 7.2.1, are almost ‘invisible’ and perceived unconsciously in real time. Due to the ways our cognition is embodied, the body ‘knows’ and ‘grasps’ directly what is going on. Thus, the effects of embodiment do the job for us. Indeed, one might ask – why do humans make them so frequently when they are almost ‘invisible’ and therefore not necessarily communicative. What role and relevance do they actually have if they are not used for social interaction and communication in the first place?

My tentative answers to these questions are that these actions are not first and foremost inter-communicative, but instead function inwardly. Thus, they are both intra- and inter-communicative, stressing the relational aspect of social interaction and cognition that is profoundly manifested by our
embodiment. Thus, the embodied nature of social interaction and cognition unifies the individual and social perspectives. For instance, in Mead’s loop, gesture and language are displayed but their relational characteristics are the same – they are both external actions that we can act upon in the public sphere and internal embodied actions used to organize and structure our internal and sometimes abstract and decoupled thinking, though still grounded in embodied experience. This means, the issue concerning the location of cognition, inside the head or ‘out there’, is not the pivotal task to address. Instead it is what kinds of characteristics the embodied actions have. That is, whether the representation is internal or not is not the crucial aspect but whether they function equivalently in the construction of meaning in human cognition. Sometimes the embodied action, by gesture or shift of gaze, initiates the spoken word, but in other cases the actual spoken word enacts related gesture and related embodied actions. Literally speaking, the dispute concerning where cognition actually takes place would be similar to the argument about whether the capacity to walk depends on either the right or left leg.

Hence, it is through these crossmodal interaction patterns that human understanding and cognition emerge and unfold moment-by moment in situ. However, the vast amount of writings needed in order to merely describe what happens simultaneously during a few seconds provides tentative evidence for such a suggestion. This result stresses the disadvantage of only relying on linear and sequential characteristics, as usually done in verbal interaction, either oral or written. Instead, there are crucial beneficial and parsimonious aspects of acting and using cross-modal interaction patterns, given that they are more rapid, more accurate, and understood ‘at once’, without any ‘higher’ intervening interpretation. The verbal interaction in subsection 7.2.3, for instance, shows how much of the meaning is constructed in the interface between action and perception through embodied experience is not described from a speaker-centered view. That is, the different embodied actions carried out in the material and social spheres altogether construct meaning and understanding that is bigger than the actions’ individual parts, but which disappear in the ‘stiffer’ representational formats provided in the thesis, namely snap-shots and written text. However, the dynamical and emergent aspects are much more apparent and obvious in the video-captured sequences. Thus, the sum is more than the parts, as emphasized in the Dance metaphor of interaction (subsection 5.3.2). However, by taking an embodied perspective, the DM is strengthened and elaborated further.

Another interesting observation occurred when the video-recordings were shown at higher speed. The mirroring of each others’ embodied actions was then made more salient, and sometimes I had the impression that the interacting participants performed some kind of synchronized work out program or dance together (as illustrated at the end of subsection 7.2.3). Moreover, it was not necessary for the interactants to directly be able to see each other from ideal positions as in face to face interaction, given that the
mirroring frequently occurred when they were standing or sitting next to each other.

### 7.3.4 Synthesis of the theoretical and empirical work

This final subsection concerns the synthesis of the theoretical framework presented in section 6.2 and the empirical work described in this chapter, which offers a new theory of the embodied nature of social interaction and cognition. Broadly speaking, the body’s enactment with the social, cultural and material surrounding as well as bodies kinesthetic experiences of this enactment, are the underlying basis of human social interaction and cognition. Embodied actions, however, can be described by some generalizations of social interaction and cognition; situated in social, material, physical and cultural contexts of use, in which they vary along the following dimensions (see Figure 36).

**Scope** – the actual ‘movement’ of embodied action can vary in scope, from really large ones as in wide gestures or bodily movements, via the tiny activations of muscles in the body to the activations and/or simulation processes of neurons in the brain.

**Intensity** – the execution of the very action vary from very slow almost being totally still, to very fast, such as the rapid eye shifts in gaze from a third-hand perspective.

**Location** – embodied actions can be carried out internally. i.e., through embodied simulation or activation ‘inside’ the body which is not perceived from a third-hand perspective. On the other hand, they can be enacted externally, i.e., through ‘visible’ embodied actions such as gesture or altered bodily posture. Humans are able to vary along this dimension ‘seamlessly’ or enact actions that combine internal and external embodied actions at the same time.

**Duration** – the actual period of the embodied action can vary from very short intervals to longer time spans. Some actions are conducted so rapidly they are ‘invisible’ to conscious experience, although they can have a strong impact on the cognitive understanding and meaning-making activity at a more non-conscious, but still significant, level of experience.

**Performance** – embodied actions differ in how they are enacted, i.e., whether they can be enacted throughout the spectrum between online via surrogate situatedness to off-line cognition. Online cognition concerns embodied actions that are enacted by issues present at hand through action-perception couplings with the surrounding social and material environment. In offline cognition, the vary same mechanism as in online cognition enacts some aspects not present ‘here’ and ‘now’ such as abstract concepts, higher-level cognition. Through scaffolding, e.g., counting on one’s fingers, surrogate situations are created, in the ‘middle’ of the continuum.
**Kind of activity pattern** – the different kinds of embodied actions emerge into crossmodal interaction patterns, which are based on different kinds of action-perception couplings with the surrounding social and material world. These interaction patterns are relatively short basic in the beginning of ontogeny and then develop into more long-lasting, complex and intertwined patterns during ontogeny in which also other partners’ action are considered.

**Kind of relation** – embodied actions are always relational or communicative, but can vary whether they are intra-communicative or inter-communicative at first hand, but usually they serve both sides. Bigger and wider actions such as pointing gestures, for instance, are usually experienced as inter-communicative, but they are also intra-communicative, also serving as an intra-communicative action for the person performing the action. Thus, embodied actions provide a means for affecting others’ actions, but they do also serve as a way of mastering ones own actions.

**Level of abstraction** – the actual action can portray different levels of abstraction, from merely showing a concrete mood in combination with posture and facial expression, ranging to abstract and more advanced crossmodal interaction patterns such as pointing gestures and linguistic expressions and other embodied actions that emerge in joint actions.

Figure 36. The integration of the different general dimensions of embodied actions in social interaction and cognition in situ.
By integrating these different dynamic dimensions on top of each other, the emerging result is a wheel of embodied actions, with no beginning or end, but in which embodied actions themselves serve as the wheel’s centre. Thus, metaphorically speaking, socially embodied actions actually move the mind.
Chapter 8

There is no burden of proof.
There is only the world to experience and understand.
Shed the burden of proof to lighten the load for the journey of experience.
Halcom’s Law of Inquiry.

We need new ways of interacting with computers,
ways that are better tuned to our needs and abilities.

8. Discussion and Conclusions
This chapter first contains some reflections on the overall results of the present work as well as the chosen empirical approach. It then poses the question what kind of body is required for cognition and discusses the uniqueness of human cognition (section 8.1). Section 8.2 concerns implications for socially interactive technology, and discusses what it is required for an artificial system to be socially embodied, and the topic is approached from the perspectives of cognitive modeling and engineering. Section 8.3 summarizes the contributions of this thesis, and discusses some vital implications for cognitive science and related domains. Section 8.4 identifies some issues for future work, and section 8.5 concludes the thesis with some final remarks.

8.1 Reflections
Most generally, this thesis has provided theoretical ideas and empirical evidence that present a fuller and deeper understanding of the role and relevance of embodied action in social interaction and cognition than is the case in previous writings of cognitive science. Obviously, the complete picture of the embodied nature of social interaction and cognition is yet to be completed, but this thesis shows that the embodied social mind is now seriously back on track in cognitive science. Furthermore, I hope that I have avoided falling into the dualist trap set by the philosophical underpinnings of my own cultural heritage, although one cannot be sure.
Broadly speaking, I suggest that my framework is an attempt to develop a view into something that is understood to synthesize the present understanding of the body’s interactions with the social and cultural environments, ranging from a variety of different domains into a single coherent framework, that unities the so-called ‘inside’ and ‘outside’ perspectives. In addition, I have offered this framework to those cognitive scientists and philosophers that are interested in a useful and plausible alternative to the traditional and still dominant view of the social mind as an individual and dis-embodied entity, as usually conceived in mainstream socio-cognitive research.

In order to verify these claims, I discuss and compare my own work with related work conducted by other researchers which at a first glance seems to be very similar to the ideas proposed in this thesis, particularly the empirical part. There are some prominent scholars proposing ideas that overlap considerably with mine. The first is Hutchins’ (1995) framework for distributed cognition that previously has been briefly presented in different places in the thesis. The main issues that I have drawn from the DC approach are the stances that cognition is a socially distributed phenomenon which can be studied ‘out in the open’ (see subsections 3.2.1, as well as sections 5.4, 6.2 and 7.1). Although embodiment is considered a central issue in the DC framework, it does not, in my opinion, provide any detailed description concerning the role of the body in cognition. In a recent paper by Hutchins (2006), he poses a similar opinion and writes

In the last chapter of Cognition in the Wild (Hutchins, 1995), I argue that cognitive science made a fundamental category error when it mistook the properties of a person in interaction with a social and material world for the cognitive properties of whatever is inside the person. One enduring problem with this claim is that it demands a description of how cognitive properties arise from the interaction of person with social and material world. Cognition in the Wild provides a profoundly incomplete answer to this question. In the years since its publication, Cognition in the Wild has been criticized for saying so little about the people in the navigation setting. It describes the tools of the trade, and the historical development of the tools. It describes social processes and the cognitive properties of those social processes, but it says almost nothing about the embodied practices of the navigators as flesh-and-blood people. For the most part, the cognitive processes described in Cognition in the Wild, and in other treatments of distributed cognition, are presented without reference to the role of the body in thinking. That is, in spite of the fact that distributed cognition claims that the interaction of people with things is a central phenomenon of cognition, the approach has remained oddly disembodied (ibid., p. 1, bold emphases added)

Thus, the above quote illustrates that the distributed approach, as currently formulated, is not similar or equal to my proposed framework. This means, it does not provide any detailed explanations and descriptions of the embodied, sensorimotor mechanisms that
might underlie the social dynamics of distributed systems or the nature of individual cognition as non-computational. However, this does not mean that DC framework and my work should be considered as alternative approaches, but rather as complementary views.

The second well-known researcher is Goodwin (e.g. Goodwin 2000, 2003), whom for many years has used the terms “embodied actions” and “situated human interaction”. In my opinion, Goodwin (2000) considers carefully the relevant public visiblility of the body, such as a dynamically unfolding locus for the production and display of meaning and action within human social interaction. However, he offers almost no detailed ideas about the underlying embodiment effects in social interaction. His main focus is, in my opinion, rather on how the co-participating body systematically is performing certain actions within a temporal horizon, but from a semiotic perspective. Furthermore, in Goodwin (2003), he points out that the role of the body in social interaction and cognition can be studied from several perspectives. Concerning the embodied perspective, he argues that most research within embodied cognition has mostly concerned the individual perspective (cf. chapter 3), criticizing that the social dimension is missing within current embodied cognitive science research. Thus, both Goodwin and I have studied the body as an important entity for meaning-making activity in situated human interaction, but the does it from a semiotic perspective, and I from a more radically embodied perspective. As in the case with Hutchins’ DC approach, Goodwin’s framework is complementary to my work, but differs in the interpretation and description of the body. Furthermore, the work by Alač (e.g. 2005a, 2005b; Alač & Hutchins, 2004) concerns the significant roles of the body in more advanced forms of socially distributed interaction and cognition, such as scientific reasoning. Her work can be regarded as a complement to the DC approach by incorporating the role of the body in thinking. Although she nicely incorporates embodied actions such as gesture and gaze into a holistic and distributed framework, she views the body as a semiotic entity, following Goodwin’s view. Therefore, there are significant differences between her work and mine, given that she is not providing any detailed descriptions of the underlying effects of embodiment.

Roth (2007) proposes a dialectical account of human interaction and cognition, in which he avoids the dualism of verbal and nonverbal communication, exemplifying the embodied and situated nature of human communication by describing a framework that exemplifies the central role of the body in communication beyond words. However, although Roth refers to the work of Merleau-Ponty (cf. subsection 2.4.2), his own descriptions of the body and its embodied mechanisms are not, in my opinion, genuinely or radically embodied. Firstly, they are rather brief and not provide any details of the underlying
sensorimotor activities. Secondly, they depict a view of the body as a mere output channel rather than as a cognitive entity.

The final scholar that I discuss here is Kendon (e.g. 1990, 2004). Kendon has done lots of research in human interaction during the years, and he has focused much on gesture in relation to speech in utterance production (cf. section 4.3). In a recent review of one of his books (Kendon, 2004), Wilkins (2006) admires how properly Kendon has described and analyzed the detailed fieldwork on gesture-use captured in situ in spontaneous interaction, and thus expanding the theoretical, conceptual, descriptive and methodological horizons of utterance construction. However, Wilkins put forward two minor objections. Firstly, Wilkins criticizes Kendon’s omission of Goodwin’s work (presented above), which he refers to as the ‘semiotic’ turn in cognitive science. The second objection concerns “the book’s speaker-centered descriptive view” as Wilkins (2006, p. 140) puts it, i.e., not seeing social interaction as a social and dynamical process.

More significant, however, may be the book’s ‘speaker-centered’ descriptive view. Despite his recognition of the importance of gesture in interaction....Kendon does not show us how utterances emerge as a joint product of the actions taken by all the communication participants (speakers and interlocutors) together. Similarly, we gain no perspective on how conversational partners monitor, reflect, and acknowledge and interpret the gesture behaviors that are so carefully described. ... Does Kendon want us to see utterance production (and so gesture construction) as the output of an individual’s cognitively instantiated competence in a simple sender-receiver model of communication?"

In order to discuss the relationship between my work in this thesis and the work conducted by Kendon, I agree with Wilkins’ objection, and uses it as a way of highlighting the differences between my work (cf. subsection 7.2.3) and his work. Further arguments in favor of the uniqueness of my work, although I do not focus exclusively on gesture, are also provided by Wilkins. He writes

The semantic potential of gesture has rarely been so thoroughly investigated, and Kendon’s work bears out de Jorio’s contention that it is only by examining gestures in their contexts of use that one can discover their true meanings. To be sure, de Jorio’s intricate program has yet to be completed, but I am very glad it is back on track. I look forward to future works in which Kendon and other researchers inspired by his vision address de Jorio’s full program of integrating manual gesture, body posture, facial expression, gaze, the unfolding of conversation in context, and the actions and reactions of conversational partners (Wilkins, 2006, p. 143).

Taken together, neither Kendon, nor any of the other discussed researchers, do describe and investigate gestures or other embodied actions from a unified relational, contextual, and radically embodied
perspective, in the ways as I do. Thus, the major contribution of my work is that it provides a solid and coherent explanation and description of the underlying embodiment effects that make it possible for humans to behave and act as they do in social interaction and cognition.

Most of the work in this thesis is theoretical, and As Kukla (2000) points out, theoretical work is quite often not viewed as equally ‘scientific’ or ‘advanced’ as empirical work. On the contrary, Patton (2002) emphasizes that the most important contribution to knowledge is theory construction, namely explaining the phenomenon under study. This means, the overall purpose of basic research is “knowledge for the sake of knowledge” and its aim in turn is “to understand and explain” (Patton, 2002, p. 215). Broadly speaking, the main task of any scientific work, either empirical or theoretical, is largely to increase or decrease the plausibility of existing theories, either theoretically or empirically. Consequently, another purpose of the framework described in this thesis is to serve as a foundation to make more detailed hypotheses that can perhaps be tested against already existing experimental findings by other researchers, and, especially, to suggest new empirical work.

Considering the empirical work, readers not familiar with naturalistic inquiry might raise some methodological objections. Various criticisms have been raised against the qualitative approach, both from academic and disciplinary perspectives, and it has been called ‘soft science’, ‘unscientific’ or just ‘exploratory’ and ‘subjective’ (Denzin & Lincoln, 2000). Qualitative research can be regarded as a reaction against the positivist (experimental) sciences such as physics, chemistry and psychology, which quite often are viewed as the major achievements of Western science and civilization. The general assumption of positivism is that there is an “objective truth” without personal bias and opinion. Positivists presume that ‘reality’ is stable and unchangeable, and can be studied by objective empirical methods (Denzin & Lincoln, 2000). Consequently, qualitative research is seen as an attack on the positivist stance, and proponents of positivism argue that qualitative studies are fiction rather than science, since qualitative researcher do not strive to either verify or falsify their findings. Accordingly, the opposition between qualitative research and positivism can be viewed as a struggle concerning what ‘truth’ actually is, which has resulted in qualitative and quantitative research methods. However, Bekoff (2002) argues that the so-called “soft” (qualitative) and “hard” (quantitative) sciences should be unified in order to produce a deep science instead, in which subjective experience is intertwined with more conventional scientific practices. Janesick (2000) concludes that we need to go beyond this methodological discussion between quantitative vs. qualitative approaches, and instead focus on the results from well conducted qualitative studies (Janesick, 2000). Hence, the classical struggle
between qualitative and quantitative research methods has to a great extent been solved, since it has been recognized that different methodological approaches are needed and credible. The challenge is to properly match the right method(s) to the question at hand, instead of holding on to a “narrow methodological orthodoxy” (Patton, 2002, p. xxii).

Denzin and Lincoln (2000), for instance, define qualitative research as a “situated activity that locates the observer in the world”. It has been argued, however, that qualitative research does not have any research questions; instead the researchers just hang around until something just pop-ups. This is not the fact, since a competent qualitative researcher has guiding questions, which are simultaneously revised and offer new directions. According to Janesick (2000), “the essence of good qualitative research design turns on the use of a set of procedures that are simultaneously open-ended and rigorous and that they do justice to the complexity of the social setting under study” (Janesick, 2000, p. 379). Hence, the researcher becomes the actual research instrument - having the ability to really see, and not just watch. Broadly speaking, “[q]ualitative researchers have open minds, but not empty minds” (Janesick, 2000, p. 384).

Furthermore, the quality criteria of quantitative research are not appropriate for qualitative research, because it is like comparing apples and oranges (cf. e.g. Fischer, 2006; Gummesson, 2001; Janesick, 2000; Lincoln & Guba, 1985; Patton, 2002). According to Janesick (2000), the main cornerstone of qualitative research is the description of persons, places and events, and this is the real reason for conducting qualitative research. However, scientific rigor or trustworthiness is of major importance, which is based on the following criteria for qualitative research (Guba & Lincoln, 1989; Janesick, 2000; Lincoln & Guba, 1985), namely credibility, transferability, dependability, and conformability.

Credibility or validity (parallel to internal validity) in the field of qualitative research concerns the fitness or matching between the description and the explanation, namely the search for a credible explanation. There are several techniques available in order to obtain plausible fitness, such as different kinds of triangulation and crystallization approaches (Janesick, 2002). Guba & Lincoln (1989) suggested ‘persistent observation’ as a way to “identify those characteristics and elements in the situation that are most relevant to the problem or issue being pursued and to focus on them in more detail” using “rigorous methods… that yield high-quality data that are systematically analyzed with attention to issues of credibility” (p. 237). In my own empirical work presented here, credibility was increased by triangulation (Patton, 2002), in which one explores whether the inferences from the empirical data are valid, but it is not just a matter of checking the validity of the data (cf. Long & Johnson, 2000). I
triangulated different theoretical approaches, i.e., *theory triangulation* (embodied, relational, and distributed approaches described in chapters 3-5), and *data collection techniques* (participant observation and video-recordings). Some parts of the case study have been published (Lindblom, 2006, Lindblom & Ziemke, forthcoming) and the proposal and content of this thesis have also been discussed with colleagues and other researchers. More generally, many parts of this thesis have been published (see the Publication list for further details).

**Transferability** (parallel to external validity or generalizability) is concerned with the richness and thickness of the description for the topic of interest. This means, to provide plausible guiding questions for the study, to write and report extensive and careful descriptions on the focus of analysis, the setting, the context the culture and so on. That is, the burden of transferability lies in the description of the study, and only the person that tries to conduct a similar study is able to judge whether or not the study is transferable. Concerning transferability in my empirical work, my role as the researcher is to provide rich descriptions of how the study was conducted, the data collection techniques, and how I analyzed the corpus data (see sections 7.1-7.2). It should be mentioned that the theoretical framework (section 6.2) is illustrated in the empirical work acts as a guide or suggests points of view at a higher level of abstraction, and should be independent of the actual setting of the study. That is, people might use gesture in order to reason ‘out in the open’, but exactly in what ways they gesture and which objects and people they rely on in the material and social environment is context dependent.

**Dependability** (parallel to reliability and replicability) refers to the ‘stability of data over time’, and that the process of inquiry should be tracked and be traceable. A study conducted *in situ*, however, is impossible to control and determine as it unfolds, and therefore it cannot be replicated. That is, the real value of naturalistic inquiry is its uniqueness, but the researcher has to offer a detailed and rich description of how the study proceeded, and account for any changes in the object of study or the guiding question. Furthermore, the researcher’s own role should be discussed. In my work, I must be alert and open to what actually unfolds in front of my eyes and to situate the embodied actions in their social, cultural and historical contexts.

**Conformability** (parallel to objectivity) means that the results and the interpretations of the data should be rooted in the collected material, and not some fabrication of the researcher’s own mind. This means, the findings must be rooted in the data and not in the method. To increase conformability, the findings should be traceable to their initial data sources, and the analysis process from raw data to results and conclusions should be as available as possible for inspection and confirmation by persons other than the actual researcher. My own
findings and conclusions can be directly evaluated, to some degree, in the actual descriptions, given that they contain many images from the actual video-recording. Ideally, the readers should be able to view the analyzed episodes. However, the data collection, interpretation and analysis were at times problematic. It was difficult to provide the ‘whole’ picture of the social situation, and in order to capture a better view I should have video-recorded from different angles at the same time. I sometimes also failed to video-record the ‘right’ actions, and therefore missed some valuable corpus of data, because I did not focus on the right things in the actual moment. However, it is quite often very difficult to know beforehand what should happen next, and if one as a researcher uses several cameras from different angles, the spontaneity might decrease, and it would also be difficult to know where exactly to place the other video-cameras. The disadvantages addressed above are compensated by the benefits of capturing spontaneous interaction in the wild without any hesitation from the participants given that the video-recording initially was for private use only.

Another crucial issue to consider is - what kind of body is required for social interaction and cognition? As discussed previously (cf. section 4.1), Maturana and Varela (1987) emphasize that the variety of social interactions rests on behavioral couplings afforded by the organism's nervous system, and similarly Ziemke (2003) discusses what kind of body is necessary for cognition (see subsection 3.2.3). This issue can be considered from two perspectives, namely what are the properties of it, and what is the characteristics and uniqueness of human social interaction and cognition. However, these issues are intertwined. In order to investigate the characteristics and uniqueness of human social interaction and cognition, comparative studies between humans and other species, mostly non-human primates, have been conducted. The common suggestion is that humans have the most advanced and specific mental properties such as intentions, intentionality and language which (perhaps) other animals lack. Consequently, a tentative approach is to compare different kinds of bodily morphology between humans and other non-human primates and animals, in order to trace the phylogenetic development of embodied social interaction and cognition. Falk (2004) compares early mother-infant interactions in the visual, gestural, tactile and vocal domains between chimpanzees and humans, and she appreciates their parallel roles. For example, she described that infant-directed communications from human mothers of 3- to 4-month-old infants are frequently accompanied by exaggerated facial expressions that have precursors in other primates and which signal affiliation and invitation for contact such as raise eyebrows, eyebrow flash, smile, and nodding (cf. Dissanayake, 2000, in Falk, 2004). However, human infants display a large number of activity patterns that appear to be species-unique during ontogeny (cf. Johnson, 2001). 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 and more complex emotional, gestural, prosodic, and tactile face-to-face interaction patterns, which are absent or rare in nonhuman primates (cf. Falk 2004; Hendriks-Jansen 1996; Johnson 2001). As Johnson (2001) points out, a crucial difference between human infants and their nonhuman counterparts is the capacity to ‘tune in’ to others and to get others to ‘tune in’ to them, which require some kind of ‘attentional cue’. Johnson claims that humans use eyes and gaze as attentional cues in the form of direct eye-contact, which in other primates functions as threat. She explains

"This may represent an important distinction since humans, at a very early age, can separate head orientation from gaze... This may be related to the loss, in our species, of the brownish pigment that darkens the visible parts of other primates' sclera. By making gaze direction more salient, this anatomical change may have been adaptive in facilitating the ontogeny of joint visual attention... in my research on social gaze in captive bonobos, who, when mature, show gaze aversion, gaze following, and can use the eyes as a cue at close range... the most marked difference I have observed between the gaze repertoires of our two species is the lack, in bonobo mothers and infants, of such gaze coordination "games" (Johnson, 2001, p. 176).

Other morphological differences between humans and their nonhuman relatives are that humans have more facial muscles, no outstanding forehead that blocks eyes and gaze to some extent, as well as no hair covering the whole face except eye brows which increase human facial expression (Johnson, 2001). This kind of co-attention and the gradual elaboration and coordination of these activities are suggested to form the foundation of joint attention and perspective-taking, supposedly necessary for more advanced socio-cognitive skills. Furthermore, as pointed out by Tanner and Byrne (1999), only the great apes (humans included) can produce movement in all directions from shoulder, elbow, and wrist joints. Accordingly, great apes have the capacity to make movements of embodied actions similar to humans. They also point out that spontaneous gestural communication is present in a group of zoo-living gorillas in San Francisco. These gorillas have developed “species typical” gestures that are shared by all the members in the group whereas other gestures seem to be individually unique. The individual gorillas’ use of the gestures varies in time and in different social environments. An amazing discovery made by Tanner, who had previously been working with signing gorillas, was that the gestural communication in these zoo-gorillas resembled untaught or even taught signing gestures in human-instructed signing gorillas. These studies, among others, might suggest that there is some kind of innate potential for using

---

37 Bonobo is another denotation for pygmy chimpanzee, one of the species of the great apes. The great apes include orang-utan, gorilla, chimpanzee and bonobo.
communicative gestures in great apes (see King, 2004 for more details).

However, to follow in the well-worn path of comparative studies research, as described above, striving to find the differences between humans and other species that might result in the unique human social mind, one might face the same dilemma as Darwin did in the 19th century, and run the risk of repeating his mistake. However to avoid his dilemma is easier said than done (cf. section 2.2). As previously written, Darwin developed two parallel, but not compatible, lines of explanation for the development of the human mind that still influence today’s research. On the one hand, he provided a phylogenetic explanation that stressed descent, which became the dominant view. This stresses a single universal line of development with roots from lower animals to fulfillment in the supposedly ‘highest’, the human being. That is, he tried to explain the human mind from a phylogenetic perspective, in which the human being is the unique species on top of the ‘ladder of development’. On the other hand, he also put forward an adaptationist explanation that emphasizes selection, in which he had stressed the role of environmental conditions and adaptations, resulting in several lines of evolutionary development. However, while this perspective has merely been factored out within mainstream psychological and comparative research, recent studies show interesting findings in favor of the adaptationist explanation.

It was earlier a common belief that only primates and great apes were able to follow gaze and similar skills to some extent, but resent studies indicate that also goats and dogs are able to show human-like social skills (cf. e.g. Miklós & Soproni, 2006 for a recent review). Dogs, in particular, outperform great apes in following humans’ gaze and pointing gestures which are findings that run counter with the phylogenetic explanation (cf. Hare & Tomasello, 2005). Call (2004), for instance, attacks the so-called ‘chimpocentrism’ in comparative studies, arguing that although valuable information can be gained from our closest relatives they should not be the only source of finding diverse solutions. Similar environmental conditions may result in similar ‘solutions’ in non-closely related species. As he puts it, closely related species are not necessarily similar in cognitive capabilities. Similarly, Bekoff (2002, p. 13), for instance, criticizes the strong emphasis in favor of comparative and phylogenetic studies, claiming “[i]t is also important to emphasize that little is to be gained from comparing, for example, the cognitive or emotional “level” of a chimpanzee to that of a human child, for each individual does what he or she needs to do to adapt to the demands on his or her own world,… but surely a young child could never survive in the chimpanzee world.” According to Call (2004), if the ability to use sonar was the pinnacle of evolution, as in dolphins, the human ladder of development is threatened to be turned upside down. The tricky
question is – what should be considered complex and advanced cognitive skills? Obviously, there seems to be a human bias towards the capacities that humans possess, in which there is a course of progressive development where the boundaries are free will and language (Call, 2004). This way of reasoning, however, falsely implies a major increase of capabilities to more complex ones and so on, but evolution also shows a decrease in capabilities since the course of evolution is not determined beforehand (Call, 2004). For instance, parasites have lost certain body organs and instead live and use their host’s organs. Furthermore, dogs\(^{38}\) show a 20% reduction in brain size compared to wolves with equal head size, and they have lost some problem-solving abilities in comparison to wolves, although they have acquired some communicative abilities with humans (Call, 2004). Concerning social cognitive skills and competences, the mammals’ neocortex has been associated with social functions and capabilities, and the frontal cortex is considered the center of higher and complex functions. However, that is only the case for mammals; social birds such as parrots for instance, display complex social and cognitive abilities both in the wild and in captivity (cf. Pepperberg, 2001, 2002), although they lack a neocortex. However, there are striking similarities between primates and parrots, namely that they both live in and interact in complex social and physical surroundings. These examples from different species illustrate that there are many ways that lead to Rome and the human (primate) route is not the one and only. Broadly speaking, there is no ladder of intelligence decoupled from its particular environment, and ontogeny does not re-capitalize phylogeny. Therefore, different animals have developed the kinds of intelligent actions that are necessary in their actual physical and social environment. As Bekoff (2002) describes, dogs are dog-smarts, monkeys are monkey-smarts and thus humans are human-smarts, and all of them are able to enact species-specific action-perception loops in their respective environments (cf. the work by von Uexküll in subsection 2.4.1).

Considering the question of what kind of body is necessary for social interaction and cognition, I argue that the question is ill-posed, given that it implicitly suggests the body alone is the key that provides the answer. Instead, I emphasize the crucial importance of the interactions and experiences of the body in its social and material environment. That is, stressing the adaptationist line of explanation, instead of the phylogenetic line. However, it is interesting to consider why dogs are so successful in human-like social skills, in comparison to great apes, given that they obviously do not have ‘similar’ bodies. A tentative suggestion is domestication, in which dogs became situated within the human socio-cultural context that resulted in biological and social selection processes in accordance with human needs (cf. e.g. Hare & Tomasello, 2005; Miklósi, Topali & Csanyi, 2004).

\(^{38}\) Dogs and wolves shared a common ancestor app. 100 000 years ago, and dogs become domesticated app. 5 0000 years ago (see Call, 2004).
Furthermore, there are some great apes that have been reared by humans, so-called “enculturated apes,” which became to some extent situated in the human socio-cultural environment. As a consequence, they have acquired some human-like social behaviors and mechanisms that they do not actually develop in the wild. This means, the presence of a human cultural environment makes it possible for the apes to go beyond their current level of ape performance and become more “human-like” through social interactions and scaffolding (see Lindblom & Ziemke, 2003). Some scientific work has been conducted on the bonobo or pygmy chimpanzee Kanzi, which is one of the most famous enculturated apes today. Kanzi has learned to communicate via symbols representing words and is able to use about 240 signs (Savage-Rumbaugh et al., 1998). Initially, the major goal was to teach Kanzi’s mother how to use the symbols (in the form of lexigrams) to communicate her desires and needs. At the time Kanzi’s young age meant he did not want to be separated from his mother during her training sessions so he was present too. After a while Kanzi showed that he had acquired communication ability, without explicitly having been trained and actually performed much better than his mother. Kanzi’s language comprehension has been argued to be equivalent to that of a two-and-a-half-year old human child and he is also able to interpret spoken sentences, even when hearing them for the first time (Savage-Rumbaugh et al., 1998). Broadly speaking, recent findings, especially in great apes, indicate that humans are not the only ones performing acts of internalization as in the case of Kanzi, which are contrary to Vygotsky’s view. However, as pointed out by Hearne (2000), although these great apes are enculturated, they are not fully participating members of our society, but which dogs are to a greater extent because they have been domesticated and not only enculturated.

My intention is not to review what different animal species can do or cannot concerning their performance and interpretations of human social skills, such as the pointing gesture, but to highlight the fact that we are still trapped by our heritage and ontology of our Western society. The Clever Hans debacle (cf. section 2.2) resulted in a scientifically verified suspicion against the mental capabilities of animals, which is also manifested in our cultural and religious heritage. Humans are supposed to be superior to animals, a task that Darwin seriously attempted to explain in his theory of evolution, which resulted in his phylogenetic line of explanation. To regard humans as the crowning work of creation is the common stance in Western societies, but this stance is, however, not universal given that other cultures have different worldviews. In Native American ontology, for instance, animals are considered at least equal to humans, having souls, and mostly viewed as superior to humans (cf. e.g. Farnell, 1995; Freke & Wa’Na’Nee’Che, 1996; Neihardt, 1988). According to Native Americans, having a language is not the special capability that
distinguishes so-called ‘superior’ humans from ‘lower’ animals. Instead, animals do not have anything similar to (human) language because they can communicate flawlessly without it, and therefore have no obvious need for one. Only imperfect humans have the need for a language, in order to deal with inconsistencies, misunderstandings and ambiguity. Animals, on the contrary, manage to live and communicate perfectly well without language, clothing, fire and other attributes of human societies. Animals therefore serve as role models for humans, to be learned from, to be used as guides for assistance and to respect in general (Farnell, 1995).

This means, that currently, I cannot provide the answer to the raised question what kind of body is required for social interaction and cognition, given that it depends on what kinds of skills one might consider, and I believe nobody else can either. Only future research can provide additional pieces to solve the puzzle of social interaction and cognition, but in this thesis I attempt to describe it from an embodied perspective in humans. However, this still leaves us to consider the question what characterizes the uniqueness of embodied social interaction in humans?

Based on the contributions of this thesis, I suggest that humans possess what I denote plasticity embodiment, as a way of describing how humans are able to incorporate all these different abilities; such as taking the role of others, engage in joint actions, being able to tool up others, to use both external and internal representations and so on. Humans are not only able to tool up other agents but also to tool up agents of other species, such as dogs, in order to enact well-functioning action-perception couplings at a system level, as in collaborative hunting with dogs. Thus the human act of embodying is of crucial importance, and it is by incorporating another agents or objects into our own body-image and/or schema (cf. section 3.3) we are able to enact action-perception loops that go beyond the bounds of our ‘individual’ embodiment. Thus the human species is unique, but other species are also unique, and the better and more interesting question to address is - How they do use their bodies in social interaction and cognition. The most advanced form of social interaction and cognition, in my opinion, is when living creatures are able to develop the most intertwined and complex action-perception loops in accordance to their own embodiment’s possibilities and constraints. A related issue is the concept chumfo, which is a term originally found in Long (1919), but more recently pointed out by Bekoff (2005), in his preface to the republishing of Long’s book. Accordingly, chumfo is a term originated from an African tribe, and Long (1919/2005, p. 54-55, 60-61) put it as follows:
According to these natives, every natural animal, man included, has the physical gift of touch, sight, hearing, taste, smell, and *chumfo*. I must still use the native term because it cannot be translated, because it implies all that we mean by instinct, intuitive or absolute knowledge and (a thing which no other psychology has even hinted at) the process by which such knowledge is acquired. This *chumfo* is not a sixth or extra sense, as we assume, but rather the unity or perfect co-ordination of the five senses at their highest point ... the perfect co-ordination of his senses working together as one.

It is, therefore, still within the range of biological possibility that a man should here with his fingers or smell with his toes, since every cell of both finger and toe once did a work corresponding to the present functions of the five animal senses.

Although Long wrote this nearly 90 years ago, his thoughts are closely intertwined and in line with what nowadays, and in this thesis, are referred to as embodied cognition science, in which there is no separation from biological/physiological process and cognitive processes. Thus, *chumfo* is another philosophy or word for describing and explaining embodied cognition. The successful lift of the filly in the end of subsection 7.2.3, can be considered as the realization of social *chumfo*, in which the men’s senses both were perfectly coordinated within themselves, but also between themselves.

### 8.2 Implications for Socially Interactive Technology

As portrayed earlier in this thesis, one of the insights (re-) gained by recent research in embodied cognition is that the mind is not, in fact, largely independent of the body, but instead strongly determined by it. For AI research striving to model human intelligence this had radical consequences. However, although embodiment was regarded as one of the cornerstones in *new AI*, embodiment was viewed as being physically instantiated in a robot, which in turn is situated in a physical environment. It can be argued that the mobile robot *Herbert* (mentioned in section 3.1 and described in subsection 3.2.1) was physically embodied, having an ability to move, alter and perceive the environment, but the robot did not interact socially with either humans or other robots, and hence, the ability of social interaction was absent. Obviously, if cognition is dependent on body and sensorimotor capacities, then the only way to achieve or model truly human-level or human-like intelligence in artifacts is to equip them with humanlike bodies and sensorimotor capacities, that is, to build humanoid or android robots. Humanoid and android robots are robots equipped with a human-like body and sensorimotor capacities and which interact socially with human care-takers. There are currently a number of projects that have taken this approach, such as Brooks’ well-known *Cog* project (Brooks *et al.*, 1998), Kozima’s *Infanoid* project.

---

Both Cog and Infanoid are upper-torso humanoid robots, i.e. approximately human-size robotic torsos equipped with stereo-vision heads, arms and hands with degrees of freedom roughly similar to those of human bodies. The Cog project began by implementing the following basic social behaviors: pointing to a visual target, recognizing a beginning to joint attention through face and eye finding, imitation of head nods and regulating interaction of expressive feedback (Brooks et al., 1998). Furthermore, the vision and emotive response platform Kismet can engage in various forms of basic interaction behaviors, grounded in a ‘drive system’ (fatigue, social and stimulation). The platform’s ‘mood’ becomes replicated as emotional and facial expressions (anger, calm, disgust, happiness, interest, sadness and surprise). As a consequence of not being stimulated the system ‘expresses’ boredom, when over stimulated it ‘expresses’ fear, otherwise Kismet ‘is’ interested (Breazeal & Scassellati, 2000, 2002). Next, in Infanoid the initiation of a shared attention ability has been implemented for instance, namely the capability of detecting human faces, finding their eyes, and then catching the gaze direction in order to find the object of interest (Kozima, 2000). Finally, the Android project has a slightly different approach. The major difference between androids and humanoid robots is in appearance and shape. The appearance of androids is much closer to a human-like form and the scale is of a full-body model. A major aim in android science is to replicate human likeness, but the capacity for humanlike social interaction is as yet very limited. They are currently used to test how long they are indistinguishable from real humans. However, the interaction lasts only a couple of seconds before a human subject recognizes that the android is not a real human (e.g. Ishiguro, 2006; MacDorman & Ishiguro, 2004).

Being human-like or human-resembling and equipped with stereo-vision heads, arms and hands with degrees of freedom roughly similar to those of human bodies, obviously only solves part of the problem of modeling human intelligence. Even if a human-like body nowadays is considered by many a necessary condition for a human-like mind, it could hardly be a sufficient one. The remaining question is, broadly speaking, how to get a mind “into” the robot’s body.

A viable approach is to let the robot undergo some kind of ontogeny, in order to be ‘intelligent’. As explained in section 2.5, AI pioneer Alan Turing dedicated a section of his seminal 1950 paper *Computing machinery and intelligence* to the issue of “learning machines.” Turing himself carried out some experiments with simple “child machines,” but after his death in 1954, despite some successes in early neural network research, most AI researchers focused on other approaches. However, both the Cog and Infanoid projects are to some degree inspired by Turing’s child-machine idea, combining it with Vygotskyan
theories (cf. subsection 2.4.4). The aim is to let the robots undergo some kind of artificial ontogenesis in physical and social interaction with their environment. Both approaches also particularly emphasize the role of social interaction with human caregivers, based on theories of social learning in infants.

This means, Cog and Infanoid, for instance are supposed to acquire or develop sensorimotor and cognitive capacities, and ultimately a mind, in some kind of long-term interaction similar to the ontogenesis of human children (note, however, that it is only the software, not the hardware/body, that develops). The hypothesis is - if a humanoid/android robot “grows up” in close social contact with human caregivers then it might develop similar cognitive abilities to human beings, that is, in some sense becomes an “enculturated” robot. Turing (1950) himself seems to have neither considered human-like embodiment nor the full range of human social interactions as particularly crucial for the child machine’s mental development. With reference to the famous case of Helen Keller (cf. e.g. Leiber, 1996), Turing argued that “we need not be too concerned about the legs, eyes, and so on,” as long as “communication in both directions between teacher and pupil can take place by some means or other”. Furthermore, he envisioned the machine as being tutored by humans, but also argued that “one could not send the creature to school without the other children making excessive fun of it”. Taking this approach to the extreme, one might argue such as Zlatev (2001, p. 155) that such “robotogenesis could possibly recapitulate [human] ontogenesis, leading to the emergence of intentionality, consciousness and meaning” in robots. Zlatev further argues that there is “no good reason to assume that intentionality is an exclusively biological property ... and thus a robot with bodily structures, interaction patterns and development similar to those of human beings would constitute a system possibly capable of meaning” (ibid).

However, the ‘static’ kind of embodiment, in form of appearance and bodily shape of the robot (cf. 3.2.3), has received much attention in discussions of social robotics, but the crucial role of the body-in-motion has received very little, although research in various domains has shown the relevance of locomotion experience to human cognition (cf. section 3.3, and subsection 8.1.2). Hence, AI researchers should consider in more detail the effects of developing motor actions on perception, cognition, language and social interaction, given that motor actions are now viewed as both a facilitator and beneficiary to psychological functions. Similarly, Smith and Grasser (2005) ask – ‘why bother about motor development when discussing the development of artificial intelligence?’ They point out, however, that there are several lessons to consider for developing embodied intelligent agents from research in developmental psychology, proposing connections between theories of embodiment, developmental psychology, and artificial intelligence. Briefly stated,
they suggest that embodied intelligence develops. Human children’s cognition emerges as they explore their material and social world, which results in different regularities between the infants and the environment through couplings between themselves and the world around them (see section 5.5, as well as the work of Piaget, Mead, and Vygotsky described in section 2.4). As children develop, their intelligence builds on earlier progresses and adaptations to the constraints and possibilities of different kinds of embodiment and bodily actions, such as self-induced locomotion and language learning. As a result, human children achieve cognitive capacities far beyond any current robot, and the lesson from babies is that “intelligence isn’t just embodied; it becomes embodied” (Smith & Grasser, p. 27). In accordance with this remark, Adolph and Berger (2006) stress that ‘the perfect robot simulation of infant motor development would require a machine whose body and environment develop’ (p. 164). Thus, it can be argued that current approaches of robotics have to pay more attention to the role and relevance of the ‘social body in motion’. Crucial to the embodiment of cognition, according to this account, is the interplay between the experience of one’s own moving body and the physical and social environment. If current theories of embodied cognition are to move forward, they need to address the crucial impact of the body in socially-scaffolded motion. I therefore suggest that the experience of self-induced locomotion needs to be addressed more in current research on embodied cognition and artificial intelligence.

It should be noted, as several researchers have also argued, that robotics research, or AI research in general, can be viewed from at least two different, though intertwined, perspectives. Firstly, that of science, mostly concerned with the understanding of natural systems, and secondly, that of engineering, mostly concerned with the design of interactive systems. This is, obviously, also the case for socially interactive robotics, which on the one hand can be considered an approach for developing powerful tools for cognitive-scientific modeling (Brooks et al., 1988; Ishiguro 2006; Kozima, 2000; Kozima & Yano, 2001; MacDorman & Ishiguro, 2004, 2006). On the other hand, it can be viewed as an approach for human-robot interaction (HRI) or, more generally, building better socially interactive technology (e.g. Breazeal, 2002; Fong & Nourbakhsh, 2003; Fong, Nourbakhsh, & Dautenhahn, 2003). In other words, the need for a better understanding of embodiment in social interaction is relevant for both perspectives.

### 8.2.1 Scientific modeling perspective

Cognitive modeling and AI research have a tradition of considering two distinct, but related, approaches to building artificial minds (or models of mind): one that puts together relatively complex systems with certain cognitive capacities more or less manually (e.g. constructing robots, writing computer programs), and, another that
builds somewhat simpler, but adaptive systems which to some degree self-organize their own cognitive capacities (e.g. using computational learning techniques). This distinction between approaches can be traced back at least to Turing’s (1950) seminal paper, in which he realized the difficulties of attempting to program an adult-like artificial mind and envisioned as a possible alternative, ‘child machines’. Today’s robotic research may combine elements of both approaches, since, with current technology, physical robot bodies need to be pre-built to some degree, before computational development and learning can start. One can then further ask whether humanoid and/or android science is intended to be, in Searle’s (1980) terms, a ‘strong AI’ that builds actual humanlike minds with mental properties, that is, original intentionality, or is it intended to be a ‘weak AI’ that builds better models of human minds than what might be possible with other tools, such as non-android humanoids or non-robotic computational models?

Many elements of Vygotsky’s theory are surprisingly up-to-date, although some aspects of Vygotsky’s work have been criticized and some positions have turned out to be incorrect, but considering its age, it is in line with contemporary research (see subsection 5.5.3). In particular the central points of his theory, the view of social scaffolding as a necessary requirement for the development of individual intelligence, and more specifically the observation that “every function in the child’s development appears twice: first, on the social level, and later, on the individual level,” are still cornerstones of current theories, and not least also of current work on humanoids and androids. Given that human children become enculturated through social interactions with humans and develop the hallmarks of social cognition and intelligence (cf. section 5.5), one might ask to what degree this might also apply to robots. Obviously, the experimental work on Cog and Infanoid is still in its early stages, that is, they simply have not yet gone through any prolonged epigenetic development. Nevertheless, one might already want to address the question of exactly what could be expected to be the outcome of such a process. Will social situatedness and interaction with human caregivers lead to internalization in Vygotsky’s sense? Consequently, will it lead to the “emergence of intentionality, consciousness and meaning” in humanoid robots, as Zlatev (2001) envisioned? However, it has been argued in detail by other researchers that this would not be the case (cf. e.g. Sharkey & Ziemke, 2001; Ziemke, 2001, 2002).

This means, even if robots such as Cog, Infanoid and androids or their successors did develop human-like behavior, it would still only be human observers interpreting this behavior as meaningful. One of the reasons is that the behaviors currently exhibited by these artificial systems and the mechanisms underlying them, have not emerged ontogenetically as in human infants, but instead have been “built-in” to the robots. For example, the implemented ability to point to a visual
target in Cog, is just a built-in behavior, derived from a computational mapping between hand and eye coordination, and is not actually a result of shared attention as in human beings. Instead, the robot actually simply points toward the object at the center of its visual field, without actually sharing attention toward a target of mutual interest with a human collaborator. Similarly, Infanoid can seemingly accomplish joint attention to some extent with a person, focusing on an object of shared interest, but actually the creators have been forced to build in some tricks to implement this behavior. In this case, there is a “color preference” for red so that the robot can distinguish and locate the object of shared interest, a red or pink toy (Kozima, personal communication). No doubt, it is certainly not impossible to implement such behaviors without any “tricks,” but would that make the behavior intrinsically meaningful to the robot itself?

Despite obvious remaining technical difficulties, work in humanoid robotics has reached a stage where it seems possible to envision human-like robots developing in relatively “natural” social interaction with humans, that is, a much richer type of interaction than what Turing envisioned for his child machines. When it comes to epigenetic development, on the other hand, robotics research has not yet fully realized its own visions and ambitions. Much initial effort in humanoid robotics projects has been invested into equipping robots such as Cog and Infanoid with the basic behavioral capacities, such as joint attention, that could bootstrap social interaction with human caretakers. As a consequence, there are very few examples, if any, of robots that have actually gone through several stages of epigenetic development, although there seems to be no reason to believe that this would be impossible in principle. When it comes to naturalistic embodiment, it seems obvious to us that the bodies of today’s robots are “naturalistic” only in the eyes of human observers. This means, a crucial difference between living bodies and their robotic counterparts might be the autopoietic, that is, self-creating and self-maintaining, organization of living systems (cf. subsection, 3.2.3).

It could be argued that many robotics researchers are to some degree repeating the “mistake” of Turing who believed that “there is so little mechanism in the child-brain that something such as it can easily be programmed.” Consequently, he conceived of his child machines as going through two separate stages: an initial design of the machine’s relatively simple hardware and adaptive software followed by a possibly complex, prolonged course of education. In terms of Vygotsky’s theory, this separation might be justified with reference to the distinction between biological factors, which determine the first months of life, and socio-historical factors, which dominate later stages of development. However, Vygotsky also pointed out the inseparability of these two processes: “Cultural development is superimposed on the process of growth, maturation, and the organic development of the child: It forms a single whole with these processes.
It is only through abstraction that we can separate one set of processes from another” (cf. subsections 2.4.2 and 5.5.3). Although some robotic work takes into account computational abstractions of bodily maturation (e.g. Lungarella & Berthouze, 2002), I would such as to argue that developing robots such as Cog and Infanoid might very well be exposed to the right socio-historical factors, but they simply lack the necessary biological factors, which in my opinion cannot be reduced to computational mechanisms. It might be worth pointing out that I am well aware that this line of argument is not unproblematic. On the one hand, admittedly, researchers do not know yet exactly what these biological factors are, but the theoretical framework of the role and relevance of embodiment in social interaction addressed in Section 5.4 might provide a deeper understanding of some crucial factors in the ontogeny of socially interactive cognition portrayed in Section 8.1. The major claim, however, is that it is exactly the biological mechanisms underlying cognition in living systems that make a crucial difference between animal cognizers and robots. In sum, somewhat simplified, we believe that social situatedness and scaffolding alone will not suffice to facilitate the development of true intelligence and intentionality in robots, as long as they are built from “dead” material and computer programs (cf. e.g. Sharkey & Ziemke, 2001; Ziemke, 2001).

This means, android science and humanoid epigenetic robotics are not making any major steps towards actual artificial minds, because they still make a distinction between the physical body that is pre-built and the computational mind that is supposed to develop during interaction with people. The example discussed in subsection 5.5.2, of the role of crawling 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.’ If cognitive development (e.g. of self, intentionality) depends on or emerges from bodily development and the development of skills for controlling the body, the division into physical body and computational mind, or into a physical construction phase and a computational development phase, is not viable for developing artificial minds with original intentionality, and so on.

I suggest, however, that the problem when constructing humanoid or android systems is that the goal typically is to construct the end result, for example, walking androids with certain cognitive and communicative capacities. For that reason, the importance of bodily and cognitive development occurring in parallel is often overlooked (cf. section 5.5). This means, it would be more plausible from an epigenetic point of view to build ‘infant androids’ that were able to develop physically and cognitively such as a human child. That is obviously impossible with current technology because it depends on the development of adaptive, growing materials. This means, even a project such as RobotCub (Sandini et al., 2004), which builds a
crawling humanoid the size of a two-year old human as a platform for studying the emergence of embodied cognition from sensorimotor interaction and exploration of the environment, is limited because although it can provide a better robotic model of a particular phase of embodied cognitive development, the development is limited to the short-term computational development of sensorimotor and cognitive capacities. The point I want to make is that to produce an artificial humanlike mind it is not enough to construct an artificial humanlike body and let it develop cognitive capacities, because having a humanlike form is not the same as humanlike embodiment. Thus, 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.

When it comes to android science as a ‘weak AI,’ that is, the modeling of human minds rather than the building of artificial ones, a crucial question is whether androids are more powerful tools than other robots, such as non-android humanoids. This is, of course, an empirical question, but one may still speculate. So what is it that androids provide better models of? A humanlike appearance makes a crucial difference from the engineering perspective for the android’s potential users. But does it make a difference to the scientist modeling human cognition and behavior? I have argued above that, when it comes to the development of, for example, self and intentionality, androids have shortcomings just such as non-android robots, although people (including cognitive modelers) might be more likely to attribute intentionality to them. But they are unlikely to develop (and thus serve as a model of) humanlike intentionality.

Hence, the major contribution of android robotics to cognitive modeling as ‘weak AI’ is that androids can be used in more realistic experiments of social interaction with people, just because they elicit more natural responses, that is, responses more such as those of human-human social interaction. In that sense, the contribution of androids to cognitive modeling might not so much lie in the androids being better models (e.g. of human self, intentionality), but in their contribution to developing better models of human-human social interaction.

8.2.2 Engineering and human-robot interaction
Let us then consider the ‘engineering’ perspective, that is, humanoid/android robotics as an approach to more natural human-robot social interaction (cf. e.g. Cowley & MacDorman, 2006, Sugiyama et al., 2006, Walters et al., 2006), or more generally, improved socially interactive technology (cf. e.g. Benyon et al., 2005; Picard, 1998; Preece, Rogers & Sharp, 2002).
Generally speaking, robotics is a field in rapid progress and new advances in computer technology, AI, and other areas have resulted in advances in robotic technology that have major implications for the wider field of human-computer interaction. Personal service robots which are supposed to assist and entertain people in domestic settings or in recreational activities have the highest expected growth rate (e.g. Thrun, 2004). Examples of such robots range from robotic vacuum cleaners and toys to robotic assistants for elderly or disabled people (Turkle et al., 2006). All in all, these robots are supposed to carry out social and intellectual tasks. It has been noted, however, that many of these robotic systems have to interact with users who have not been trained to operate them. Therefore, as Thrun (2004) points out, the need to find an effective means of interaction is more crucial in this new area of robotic technology than in professional service and industrial robotics. Ideally, personal service robots are supposed to behave in a humanlike way, offering and creating ‘easy’ and ‘natural’ interaction with the users. But Kiesler and Hinds (2004) argue that “although roboticists are gaining practical experience with mobile and autonomous robots in settings such as museums..., we lack a principled understanding that will accomplish those more ambitious goals” (p. 2).

The physical realization of androids offers a unique user interface, they may appeal to people differently from mechanical-looking artifacts and to some degree humanoids. Central to the usability of these social robots is effective social interactions with humans. This interaction includes a wide range of social features such as expressing and perceiving emotions, communicating verbally and non-verbally, displaying personality, recognizing interaction partners and learning socially (cf. Fong & Nourbakhsh, 2003; Thrun 2004). I believe the main contribution of android robotics lies in developing more advanced social robotics, since there are some major benefits in designing technology that supports human sensitivity to 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 et al., 2003; Kozima & Yano, 2001), which humans are able to grasp rather easily. Hence, the more natural the interaction with androids is or any kind of technology, the more useful it will be for people in their daily lives. Androids might then be a more suitable design solution than mechanical-looking robots. In my opinion, their main benefits and disadvantages are as follows:

It is generally considered that similarities in bodily shape, appearance and expressive behavior of humanoids, particularly androids, offer a number of advantages for human interaction. For that reason, the user’s acceptance of human-robotic systems might be accomplished more easily with humanoids/androids. Moreover, a humanlike morphology is well-suited to function in human environments, and consequently, it might be more enjoyable and rewarding to interact
with these robots. In addition, human-android interaction could heighten subjective feelings about the quality of the interaction, increasing user satisfaction (e.g. Benyon et al., 2005). Given that humans are ‘ultra’ social animals, and consequently experts in social interaction, the necessity for costly training programs for using interactive systems might decrease.

It should be noted, however, that our everyday life involves a highly complex web of tasks and social skills. To assist us as ‘truly’ humanlike partners, these robots have to ‘understand’ our intentions. This in turn requires them to be able to interpret our intentions by recognizing bodily movements and accomplishing joint attention. A possible approach would be to engage robots in various processes of social learning, such as imitation and co-operative learning, and to teach the robot in the same way as when instructing another person or a child (as envisioned by Turing 1950, and described in section 2.5 as well as in subsection 8.2.1). Hence, there seems to be a need for a kind of ‘enculturation’ process similar to the epigenesis of human children. However, it may be difficult to achieve the human kind of social learning in humanoids/androids, because their quest suffers from the same shortcomings discussed above for cognitive modeling in the sense of ‘strong AI,’ that is, the lack of ‘intrinsic’ intentionality. Although humanoids/androids appear to express emotions and perform basic joint attention behaviors, they do not experience these abilities themselves. The experience lies in the eyes of the beholder, namely the human user. This means, what ‘looks’ such as an intentional understanding to an observer has no correspondence in the robotic system itself; it is just the human observer who interprets the behavior as ‘intelligent’ and ‘meaningful.’ This implies that androids do not interact socially through their living bodies in the sense that humans do, but rather through their perceived bodily appearance. Their sociality is in the eye of the beholder.

What consequences will that ‘intrinsic’ lack have for human-robot interaction? It would make it easier for us to interpret the android’s behavior than for the android to interpret our’s because it would have difficulty interpreting our intentions. Also, a very humanlike appearance may offer a too promising impression of the android’s abilities, which may be disappointing to the user (cf. the ‘uncanny valley,’ MacDorman & Ishiguro 2004, 2006; Ishiguro 2006). However, if one weighs the pros and cons of androids from a human-robot interaction perspective, I believe that regardless of whether android robots will be truly intelligent in the ‘strong AI’ sense, this type of technology will allow people to better situate themselves socially in the world of technical artifacts. This means, the real strength of android robotics is not its role as a robotic ‘strong 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.
In summary, section 8.2 discusses some of the implications for robotic research and argues that it is unclear how humanlike intelligence in humanlike robots should be approached. On the one hand, building and programming humanoids/androids to behave such as humans is far from straightforward, because we do not sufficiently understand the relevant underlying mechanisms in humans, although the proposed framework in this thesis offers an deeper understanding (cf. section 6.2). On the other hand, letting humanoids/androids develop humanlike cognitive and behavioral capacities is hindered by the need for an extended period of co-development of morphology, bodily skills, behavioral and cognitive capacities, as follows from the argument presented in section 5.5. From an engineering perspective (subsection 8.2.2), truly humanlike androids/humanoids would allow for modes of human-machine interaction that other interface technologies cannot facilitate. Human-robot interaction is an emerging field, and there is much work to be done. Rapid technological development provides new implementations that were not possible only a few years ago. The uniqueness of human-robot interaction, in comparison to more traditional interactive technologies, offers both opportunities and pitfalls. To conclude, android/humanoid robotics certainly has the ambition and the potential to contribute to an integrated science of cognition and human-technology interaction. Nevertheless, apart from the technical and conceptual difficulties involved, more research is required regarding how androids may influence us as human users, our environment, our way of thinking and being, and to what degree the benefits can be expected to outweigh the costs and risks.

8.3 The Road behind Us - Contributions and Implications

Throughout the chapters in this thesis, I have tried to fulfill the task I set out to accomplish in the introduction. Accordingly, I have presented and integrated several threads that have been woven through the tapestry of embodied action in social interaction and cognition. In search of the sources and functions behind the embodied social mind the journey sometimes embarked into unfamiliar terrain through the winding and rocky roads of a complex historical background. Thus, there is no royal road to cognition given that many ways lead to the social mind, but embodiment is the road to follow. Hopefully, the following contributions and implications broaden some of the conceptual, methodological, theoretical, and practical issues for embodied cognition in general, and the social dimension in particular.

The relationship between body, embodiment and cognition has been disentangled and discussed (cf. subsection 1.3), which contributes at a more general level to a better understanding of the role of the body in social interaction and cognition. I have also characterized what I mean with the terms, body, embodiment, and embodied cognition.
The thesis has offered a rich, though not complete picture, of the ebb and flow of theories of embodiment from a historical perspective. Most generally, it considers the relationship between body and mind from the ancient Greeks until today’s embodied/situated/distributed theories of cognition (cf. chapter 2). In particular, it describes the historical and philosophical roots of agent-environment interaction in work by scholars such as von Uexküll (subsection 2.4.1), Merleau-Ponty (subsection 2.4.2), Piaget (subsection 2.4.3), Vygotsky (subsection 2.4.4), Mead (subsection 2.4.6), and Dewey and Bartlett (2.4.5). It also provides a historical overview of the study of behavior and has described why the theories of agent-environment interaction were abandoned by behaviorism and in the early days of cognitive science (cf. Lindblom & Ziemke, 2007). Furthermore, some arguments for the re-discovery of theories concerning agent-environment interaction are described, and the characteristics of the body in computationalism and embodied cognitive science respectively are compared and contrasted (section 3.1).

The current theories of embodied cognitive science have been developed further by moving on from the more ‘static’ body to consider the body in motion (section 3.3), and its implications for embodied experience (cf. Lindblom & Ziemke, 2005a, 2005b, 2006).

Social interaction is characterized and compared to physical agent-environment interaction. Two different kinds of social interaction have also been described, which results in endo-embodiment and exo-embodiment (section 5.1).

Regarding the nature of social interaction and communication, I have compared two different metaphors of social interaction, and offered arguments for the dynamical Dance metaphor of social interaction (section 5.3). Furthermore, the concepts of intentions and intentionality are analyzed and re-characterized from relational and distributed perspectives (section 5.4).

Concerning the role of the body in social interaction and cognition, I have identified four fundamental functions of embodiment (Lindblom, 2005b, 2006, Lindblom & Ziemke, forthcoming), which are used to describe and explain how and in what ways humans use their body in social situations (section 4.4). The functions suggest that the body is used; as a social resonance mechanism (subsection 4.4.1), as a means and end (subsection 4.4.2), as a helping hand in shaping, expressing and sharing thoughts (subsection 4.4.3), and as a representational device (subsection 4.4.4).

The major contribution of this thesis is the interdisciplinary framework described in section 6.2, which addresses the embodied nature of social interaction and cognition. The framework is based on an integration of research literature, which ranges from areas such as
cognitive science, artificial intelligence, primatology, neuroscience, developmental psychology, cultural-historical work, phenomenology, social psychology, gesture studies, communication, and linguistics to ethology. The framework takes into account interactions between agents and environment, in particular the social dimension, and describes them from a cross-modal perspective. It also includes four fundamental functions of the body in social interaction (Lindblom, 2006, Lindblom & Ziemke, forthcoming) that explain how and in what ways humans rely on embodied action in social interaction and cognition. The framework also considers the co-regulated nature of social interaction and communication. Summing up, the framework addresses the following issues:

- The *unit of analysis* is situated embodied actions in co-regulated social interaction. That is, cognitive processes are visible and can be studied ‘out in the open’.

- Embodied actions such as posture, gesture, and so on, are integrated into a *cross-modal interaction system*, in which they are viewed as a whole rather than as isolated parts.

- The term *action* is used instead of behavior, given that socially embodied actions are a set of movements that have agency for the actual agent. This implies a shift from an experienced body to an *enacted* body, suggesting that the agent or person as such has to be fundamental in the conceptual understanding.

- Agents’ embodied actions are always in *relationship to others*, and it is useless to distinguishing between individual and/or social actions, given that they are two sides of the same coin.

- The nature of *social interaction and cognition is relational*, given that we are always in dialogue - either with our selves or with other individuals.

- Meaning and intentions are emergent phenomena of embodied social interactions. One should not consider intentions to reside inside an individual’s mind, but as constructed ‘out there’.

- Embodied actions are *situated in social and material contexts*, which means that one cannot investigate and analyze embodied actions apart from their context of use, because it provides much of the meaning.

- The co-regulated interaction is a *dynamical process* that unfolds during a *temporal horizon*. This means, cognition should be considered as a process rather than manipulation of static content. The embodied actions are always unfolding, and every action is considered on behalf of the prior actions.
Socially embodied action is not pre-given, since the social mind develops. That is, the central and inescapable feature of embodied action is that the opportunities and constraints of the body as well as the social and material environment are continually changing during the course of ontogeny.

The focus of analysis is on the embodied actions constructed in social interaction and cognition. It is difficult to consider every aspect in the framework, and therefore a trade-off is done to decide what to place in the foreground and what to place in the background. The focus is on the four embodied functions of the body in social interaction.

The empirical work that has been reported in this thesis contributes to the theoretical framework by illustrating some of its major characteristics (section 7.2; and Lindblom, 2006; Lindblom & Ziemke, forthcoming), but it also uncovers some unforeseen issues regarding public and visible embodied actions in social interaction and for the emergence of meaning. Most generally, the three episodes that are described in the case study also consider embodiment from a visible and relational social interaction perspective, instead of focusing on internal individual processes. The analyzed episodes in different social situations are also important in themselves given that they offer some insights into and deepen the understanding of the crucial role and relevance socially embodied actions play in spontaneous social situations and cognitive processes. The latter episode (subsection 7.2.3) describes joint action from an ‘embodiment-in-social interaction’ perspective, thus explaining how meaning emerges and in what ways higher level cognitive processes are constructed ‘out in the open’ and scaffolded by social and material factors.

An explanation of the behavior of tooling up (section 5.3) each other by means of extending one’s body schema and body image contributes to the embodied perspective of social interaction that this thesis offers (section 3.3)

Furthermore, the empirical work has identified different kinds of mismatches (subsection 7.2.2). According to Goldin-Meadow’s work on gestures (section 4.3), gesture in relation to speech is generally categorized into matches and mismatches. The second episode reveals different kinds of mismatches, namely unconscious/intentional respectively conscious/intentional mismatches. An intentional mismatch is constructed on purpose and serves mainly as an inter-communicative device in order to provide additional meaning not conveyed in speech.

The relationship between speech and gesture from an embodied linguistics perspective has been strengthened. The case study
reveals that gesture has perspectival properties as in verbal expressions of language. Furthermore, the representational characteristics of gestures are described. Empirical evidence from the case study shows that matches are very frequent in human interaction, and thus provide more evidence for the relational aspect of human interaction and cognition. These findings contribute to the view of hand and mouth as a unified system in language and communication, which functions both inwardly and outwardly.

In the matter of developmental aspects, the work on the social body in motion and its influence for the emergence of self and joint attention contributes to a better understanding of how crucial, complex and intertwined bodily and social aspects are in the ontogeny of the social mind (section 5.5.2; Lindblom, 2005a; Lindblom & Ziemke, 2005a, 2006). Furthermore, the thesis provides alternative explanations of how the different kinds/levels of social interaction develop from an embodied and relational perspective (subsections 5.5.1-5.5.3), which are in stark contrast to the traditional individual and internal explanations (subsection 5.2.1-5.2.3).

The thesis has also presented some implications to artificial intelligence, in the form of socially interactive technology. It discusses New AI from the perspectives of cognitive modeling and engineering, particularly humanoid and android robotics (section 8.2; Lindblom & Ziemke, 2003, 2005b, 2006).

I have discussed and reframed the issues concerning the characteristics and uniqueness of human social interaction and cognition, as well as what kind of body is necessary for cognition. I have also coined the concept of plasticity embodiment, as well as compared it with chumfo (section 8.1).

Besides these contributions, the current thesis has some implications. From an embodied cognitive science perspective, this work is in line with the request of Clark (1999a), and provides some steps forward in developing a “mature science of the mind” (p. 349). This means, the present thesis contributes to extending the understanding of the social dimension of embodiment in current theories of embodied cognitive science, by providing a more integrated picture of the embodied social mind. In other words, how the bodily basis of social interaction and cognition are used in meaning-creating activity and joint actions. It also complements the current research in theories of situated and distributed cognition and offers a foundation for how to expand the study of social interaction and cognition, in which socially embodied action plays a crucial role.

The field of child development also gains insights from the work on the social body in motion. The emergence of joint attention and a primitive self is based on the view that the social environment
functions as a scaffold for the development the body in motion. Thus, cultural customs affect, but do not determine, the organization of social interactions under the constraints of embodiment (subsection 5.5.2; Lindblom, 2005a; Lindblom & Ziemke, 2005a, 2006).

Studies of communication and social interaction are provided with a coherent and compelling explanation of how human social interaction and communication can be manifested as a relational and dynamical process, by addressing the effects of embodiment.

There are also implications for gesture studies. Firstly, this thesis describes gesture from an embodied perspective, which links gesture and language together as an integrated hand-mouth system. Empirical evidence from the case study also reveals that there are different kinds of mismatches. Secondly, it analyzes gesture in the context of use, and illustrates how gesture contributes in meaning-creating activity that unfolds under a temporal horizon. Thirdly, it integrates gesture with other embodied actions such as posture, gaze and so on, and thus highlights the role and relevance of gesture among other embodied actions in spontaneous social interaction.

Although the thesis has discussed some implications for robotic research and other forms of socially interactive technology (section 8.2, Lindblom & Ziemke, 2003, 2005b, 2006; Ziemke & Lindblom, 2006): several implications remain to be considered. As pointed out in the beginning of this thesis (section 1.1), the concepts of embodiment and epigenesis are sometimes used as a buzzword in the field of socially interactive robotics, but their roles are still poorly understood. The work in this thesis provides a unified and deeper understanding of how embodiment matters in human social interactive cognition required from a cognitive science perspective, which subsequently, can be applied and/or modified to the domain of socially interactive robotics (cf. e.g. Breazeal, 2002; Brooks et al., 1988; Fong & Nourbakhsh, 2003; Fong, Nourbakhsh, & Dautenhahn 2003; Ishiguro, 2006; Kozima, 2000; Kozima & Yano, 2001; MacDorman & Ishiguro, 2004).

Additionally, in the field of HCI, embodied interaction is argued to provide a new model of interaction, which is suggested fundamental for the future development of that domain. Some researchers have advocated that interactive modes which are physically and socially more similar to how tasks are performed in the ‘real’ world should be plausible (e.g. Dourish, 2001). In the subfield of CSCW, there is also a need to develop technology that both supports individual and cooperative work, and bridges the social-technical gap. The embodied nature of social interaction and cognition described in this thesis provides valuable insights on how human interaction unfolds spontaneously. Furthermore, this conceptual understanding could contribute to new metaphors, models of explanations and so on,
which at a higher level of abstraction can provide a further understanding of different usability issues for the future development of socially interactive technology. As pointed out by Rogers (2004), there is not much to gain from developing guidelines based on theoretical/conceptual knowledge, which would be used by practitioners who usually do not have the deeper theoretical knowledge from which the guidelines were developed. As a result, they will not be used in practice at all or applied incorrectly. However, the reverse is indeed a viable approach, namely that practical experience can provide valuable insights to the theoretical construction of knowledge.

8.4 The Road Ahead - Future Work

There are several issues for future research based on the results that have been described in this thesis. The interdisciplinary framework can be elaborated in several aspects. In its current form it includes a number of crucial issues to consider in more detail regarding the role and relevance of the body in social interaction and cognition situated in its material and cultural sphere. For instance, the four fundamental roles of the body in social interaction need further elaboration. The use of the body as a representational device, for instance, can be accomplished in several ways, and needs further elaborations in regard to what different ways embodied action is used as a representation as well as what kinds of representational characteristics it contains.

In addition, the framework could be used as a methodological approach for studying social interaction and meaning-making activity. One tentative domain for such an application is developmental issues. On the one hand, from an ontogenetical perspective, Johnson (2001) points out that “data on co-attention are conspicuously absent from nearly all accounts of imitation research...this sort of information is of special interest since contingencies between co-attention and motor-activity are of particular importance in human cognitive development” (p. 172). The framework provides a suitable approach to investigate these kinds of issues given that it has the system level as the unit of analysis, in which the material and social environment are crucial elements of the system. The framework could be used to study the gradual elaboration of co-regulated interaction, and the coordination of these activities that emerges between the child and its care-taker. It should be mentioned, that co-regulated interaction has been studied more in great apes than in human children (Want & Harris, 2001). Such an approach should also provide an opportunity to reveal further coincidences between motoric and cognitive development.

When discussing developmental aspects, the framework could also be used for comparative study. One possible comparison could be whether and to what extent the four fundamental functions of embodiment are used in non-human primate social interaction and cognition. In the field of ethology, much study of behavior and social
interaction is conducted from a quantitative perspective, but it would be interesting to study animal behavior to more extensively from a qualitative perspective as proposed in the framework.

Furthermore, another purpose of the actual framework is to serve as a foundation for more detailed hypotheses that can perhaps be tested against already existing experimental findings by other researchers, and especially, to suggest new empirical work. Several suggestions for future work arose during the analysis of the case study. Possible empirical issues to study include the division between gesture and speech in language production, the relationship and scope between matches and mismatches, and the frequency of respective category in social interaction. Such studies might reveal more finely-grained categories of matches and mismatches, than evidence found in my empirical work. An additional aspect to consider when analyzing socially embodied actions is what the appropriate level of granularity is. On the one hand, one can scale up the level and compare different cultures in order to investigate whether there are cultural differences in the humans interact socially through embodied actions. Greenfield (2002), for instance, describes cultural differences for the means for embodied social interaction.

The communicative function of touch is one of the least researched areas in infant development (see Hertenstein, 2002). Although there are indications for its importance in the child’s cognitive and social development, it plays much more salient roles in non-Western cultures (Greenfield, 2002). On the other hand, in order to make more detailed analysis, additional means for studying and collecting data are needed. For instance, it would be useful to see exactly where and for how long people look in social interaction and the use of some kind of eye-tracking technique could be considered. Do we really notice the tiny and fast gestures conducted by one’s interactive partner or do people mostly pay attention to the eyes and mouth? Furthermore, how are the shifts of gaze accomplished when one’s partner gestures and speaks at the same time? The relationship between gesture and speech could also be elaborated further, measuring the pitch, intonation and so on, in relation to gesture, as well as investigating cultural differences of speech-gesture combinations.

In addition to the above methodological and empirical suggestions, tool use in social interaction and cooperation needs to be considered into the present framework. While much research on tool use has focused on one or the other aspect, often in relation to the individuals’ cognition, there is still limited understanding of these aspects and their role in cognition, as well as their integration, especially in situations of social interaction (but see e.g. Hutchins, 1995 and Susi, 2006, as examples of work that do consider tool use in social, cooperative and distributed contexts). However, some initial steps to incorporate tool use into the present framework have been conducted
but it is still a work in progress (see Susi & Lindblom, 2005). Concerning tool use, a tentative way to investigate it from an embodied viewpoint is to conduct more detailed analysis of "the act of embodying" (section 3.2).

Another future research task is to use the present work as a tentative approach for the scientific challenge of creating an integrated cognitive science of HCI. Furthermore, in order to develop this integrated discipline, an embodied approach is been proposed (cf. e.g. Benyon, Turner, & Turner, 2005; Dourish, 2001; Hollan, Hutchins & Kirsh, 2002, Mingers, 2001). Dourish (2001, p.1), for example, argues that “we need new ways of interacting with computers, ways that are better tuned to our needs and abilities”. Hence, some guidance for how to proceed in order to reduce the socio-technical gap can partly be found in this work as well as how to build artificial agents such as humanlike robots that induce humans to feel more comfortable in technical surroundings, but also how to actually create better models of human embodied cognition.

Another crucial issue to address in future research is the level of granularity, given that in order to discover how, e.g., social embodiment effects occur, one might have to consider the biochemical level of embodiment. Much focus has been on the neural level, but physiological research reveals that there is a need to consider the biochemical level of the neural and physiological activity (Pert, 1997). For instance, emotions trigger the release of different hormones in the body, and to study emotions properly, there is a need to incorporate the interaction between the nervous system, the immune system and the neuroendocrine system (cf. Karren et al., 2006). This means that more and more research disciplines might become involved in cognitive science, which in turn raises higher demands on the researchers that are active within the field. Only the future will tell us whether the field of cognitive science will expand to further directions.

8.5 Closing Remarks

I hope this thesis provides a solid foundation from which to launch further steps in that direction, and I look forward to seeing future work on my path. The conclusions from this thesis draw their strengths from the very nature of embodiment in social interaction, which deeply stresses that cognition is moved by the sensing experiences and actions of our social bodies in motion. However, there are still several blind spots on the embodied map, and some suggested paths might turn out to lead us astray. I do not know exactly what waits around the corner, but let us never forget the embodied roots of our human mind and communication. If we cut the body from the mind and our social, cultural and material sphere, we will no longer be human cognizers. The body is always communicating to us, so to speak, through its embodied actions. Most generally, our embodiment moves us to the experience of the social mind, and the gap between
body and mind is fictitious. Instead, mind the body in social interaction and cognition.

It might be obvious to the reader through the thesis that I use a metaphor of action, in the form of a journey, as a way of exploring and explaining social interaction and cognition from an embodied point of view. I hope that you have enjoyed the trip, and although my part ends, this is not the final destination on the embodied road to cognition, and I invite you to join the journey.

Lastly, the task I set out to accomplish in the introduction of this thesis, namely to clarify the role and relevance of the body in social interaction and cognition, has provided intellectual challenges and some steps forward, but we still appear to be on the edge of an unknown territory. However, I'm still struggling to fill the gaps in my own understanding of the issue, and as I believe, all scholars are, and the work that has been presented here is part of an ongoing learning process. In order to express one of the lessons learned from writing this thesis, I would like to end with a quote by the Swedish poet Thomas Tranströmer whom writes “Du blir aldrig färdig, och det är som det skall”\textsuperscript{40}.

\textsuperscript{40} In English “You’ll never be complete, and that’s as it should be” (translated by Robert Bly (2001) The half-finished heaven –the best poems of Tomas Tranströmer. Saint Paul: Graywolf Press).
References


Agar M. H. (2005) Personal communication at the advanced ethnography workshop at the Skaraborg Institute, Skövde and University of Skövde April 22, 2005.


Watson, J. B. (1913) Psychology as the behaviorists views it. *Psychological Review*, 20, 158-177.


Appendix

Figure 37. Map over parts of USA and Canada
Figure 38. Map over Montana and Blackfeet country (Adapted from Gibson, 2003, p. 7).
Department of Computer and Information Science
Linköpings universitet
Dissertations
Linköping Studies in Science and Technology


No 281 **Christer Bäckström:** Computational Complexity


No 582 Vanja Josifovski: Design, Implementation and
| No 808 | Klas Gäre: | Tre perspektiv på förändringar och förändringar i samband med införande av informa-
No 894

No 895

No 896

No 897

No 898

No 899

No 900

No 901

No 902

No 903

No 904

No 905

No 906

No 907

No 908

No 909

No 910

No 911

No 912

No 913

No 914

No 915

No 916

No 917

No 918


Linköping Studies in Information Science