Altering the Point of You

Perspectives on Intersubjectivity and Metacognition

Liljenfors, Rikard

2012

Altering the Point of You

Perspectives on Intersubjectivity and Metacognition

LUND UNIVERSITY

Rikard Liljenfors

Department of Psychology

CLIMATE
COMPENSATED
PAPER

Faculty of Social Sciences and the Department of Psychology
ISBN 978-91-628-7765-1

Printed in Sweden by Media-Tryck, Lund University
Lund 2012

Table of Content

Abstract

The aim of the thesis was to examine different aspects of the role of intersubjectivity in metacognitive development and in social understanding. More specifically, it investigates how different theoretical frameworks, such as mentalization theory, the theory of primary intersubjectivity, and interaction theory describe the developmental role of intersubjectivity. The suggestions these theories make in regard to this is also studied. Common to all three papers included in the dissertation is the conviction that intersubjectivity actually is central for, and affects in a basic way, social and cognitive development from the very beginning of life.

The methods employed are theoretical and concern the analysis of empirical studies in developmental psychology, as well as the analysis of, and comparison between, theories concerning different aspects of social understanding.

In the first paper, metacognition is interpreted as a way of managing cognitive resources that does not necessitate algorithmic strategies or metarepresentation. When pragmatic, world-directed actions cannot reduce the distance to a particular goal, the agents involved may engage in epistemic action directed at cognition. Such actions are often physical and involve other people, and thus are open to observation. Taking a dynamic systems approach to development, it is suggested that implicit and perceptual metacognition emerges from dyadic reciprocal interaction. Early intersubjectivity allows infants to internalize and construct rudimentary strategies for monitoring and control of their own and of others' cognitions by means of emotion and attention. The functions of initiating, maintaining and achieving turns make proto-

conversation a productive platform for developing metacognition. It enables the caregiver and the infant to create shared routines for epistemic actions that permit training of metacognitive skills. The adult is of double epistemic use to the infant—as a teacher who comments on and corrects the infant's efforts, and as a cognitive resource for the infant.

The second paper deals with the question of how primary engagement and interaction relate to social understanding, most notably mentalization. The basic hypothesis considered is that primary intersubjectivity and mentalization are complementary and that the latter depends on the former, but the converse to this is not the case. Primary intersubjectivity is the sharing of experiences. It involves emotional engagement in second-person relations that are meaningful to the infant already from the start, whereas the theory of affect mirroring provides an explanation of how mentalization and representational abilities develop from dyadic interaction and contingency detection. A comparison of the theories suggests that, despite of their differences, they can fruitfully be combined. This paves the way for developing an alternative interpretation of affect mirroring, one based on the idea of young infants' understanding the experiential dimension of emotion and using this to understand others. This makes it possible to trace the continuous development of social understanding based on emotion experience and affect sharing, and in addition to elaborate on the role of second-person engagement in attachment.

The third paper concerns the concept of mentalization as it was introduced into psychological science by Fonagy and his associates. The study describes some fundamental aspects of how the development of mentalization is viewed within the framework of this theoretical approach, enabling certain issues that seem difficult to explain in terms of mentalization theory to be more readily understood. A critical discussion of the theory is then undertaken, comparing and contrasting it with the theory of primary intersubjectivity. A suggestion is made concerning the development of mentalization that connects it with the notion of primary intersubjectivity. More specifically, it is argued that mentalization develops originally within the context of primary

intersubjectivity, and that primary intersubjectivity is a basic prerequisite for the development of mentalization and in addition that there is a partial overlap between the concepts of primary intersubjectivity and implicit mentalization.

Key words: Social Understanding, Mentalization, Primary Intersubjectivity, Metacognitive Development, Theoretical Psychology.

List of Papers

Paper I-III:

I. Brinck, I. & Liljenfors, R. (2012). The developmental origin of metacognition. *Infant and Child Development.* DOI: 10.1002/icd.1749. (Will be published in *Infant and Child Development 21(6)* as a Target article together with peer-commentaries and authors' response.)

II. Liljenfors, R. (2012). From primary intersubjectivity to mentalization: On the development of social understanding. *(Manuscript submitted for publication).*

III. Liljenfors, R. & Lundh, L-G. (2012). Mentalization, intersubjectivity and affect mirroring: A critical discussion of some aspects of the development of mentalization. *(Manuscript submitted for publication).*

Acknowledgements

As most of you already know, a piece of work that takes several years to finish tends to involve many aspects of one's life and a number of different persons in one way or another. I would like to express my gratitude to various persons who have been involved in this process.

First, I would like to thank Sven-Birger Hansson and Jean-Christophe Rohner for introducing me during my first period of time at the Department to the intriguing research area of attention, perception, emotion and unconscious processes.

My main supervisor Ingar Brinck has supported and encouraged my work consistently over the years, helping me to better understand various important aspects of research and emphasizing the importance of communicating one's ideas adequately.

Lars-Gunnar Lundh, my assistant supervisor, encouraged me in the development of my own thoughts, already from the beginning of the project, at a time when I needed it the most. His encouragement, at times accompanied together with his critical evaluation of my ideas, have continued ever since.

Without the comments from both of you on the texts I have written and the ideas I have presented, the thesis would never have been finished at all in its present form. I am extremely grateful for the efforts you have made and have learned a great deal from both of you. Thanks a lot!

Thanks also to Henry Montgomery for valuable comments and a fruitful discussion at the final review.

Special thanks go to Robert Goldsmith, who has corrected and improved the language of the thesis, and who has also encouraged me to continue with and finish the work.

I enjoyed and profited from the many valuable discussions of intersubjectivity I had with Jordan Zlatev 10 years and more ago. These discussions fuelled my interest to a degree that made it last ever since. Thanks also to Henrik Levinsson for kind support and for the discussions of metacognition we have had.

Without the social support of friends of mine and of my family, completing a project of this sort would have been unimaginable. I wish accordingly to express my gratitude first of all to my parents, Kent and Ingrid, for being so supportive and encouraging both in ways that are easily formulated – for such concrete things as financing a computer for me as I began my work, for example – and in countless other ways that are impossible to pin down, yet are of far greater significance. Thanks also for your enduring patience between the unfortunately rather rare occasions of my visiting you, which applies to an equal extent to my brother Tomas and his family, namely Elinor, Henning and Sixten. Tomas' continual support along the way has also been very much appreciated, along with everything else you represent. My heartfelt thanks for everything!

My most sincere thanks also go to my in-laws, Lasse and Carina, among other things for your being supportive of and so curious regarding my project, to my brothers-in-law together with their spouses, i.e. to Daniel and to Ellinor, to Olof and to Erica, and to both Jakob and Claes, for your encouragement and support in various ways over the years. It appears appropriate too that I thank my youngest relatives, born between 2005 and 2010 (practically during the same period of time as that of my being preoccupied with the thesis). These include Lovisa, Henning, Oscar, Ludvig, Sixten, and Victor, who embody and remind me of what this thesis is really about, hopefully preventing me from overly intellectualizing things.

At the Department of management at BTH there were a number of colleagues who helped particularly in making my days there enjoyable with their vivacious presence. Similarly, the days I have spent in Karlshamn would have been far less joyous without all of the dear friends of mine there, none of whom I have forgotten. Special thanks to Mia Hemming for encouraging me very much to finish the thesis during a discussion late one evening in January concerning her own project as well. I hope you're next! Many thanks to Marie Å as well for helping me to realize how much discipline means for creativity.

Yet the person who without doubt has meant the most to me in all of this has, of course, been Johanna. Without the blessing of your continuous encouragement and cheerful incentives during what were the more discouraging moments, things would have looked radically different. Your predisposition for finishing things within deadlines (completing them before they're due) was enormously inspiring to me, even if it took me several years to come even at all close to you in this respect. Yet, here it is. As you know, words seldom come easy, but from the bottom of my heart my heartfelt thanks for everything!

1. Introduction

We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time

T. S. Eliot

Some ten years ago, in Marc Hauser's "Evolution of Communication" (1996), amongst passages on avian song, echolocation, and dolphins' imitative capacity, I came across a passage that dealt with 'the human child's discovery of mind.' In that section, a false-belief task was used to support the distinction between implicit and explicit knowledge, conceived as being applicable to the course of children's development of knowledge.

A false-belief task is prototypically arranged in a setting in which a story protagonist places an object in one of two locations. As the protagonist leaves the scene, the object is unexpectedly moved from Location A to Location B. Before the protagonist returns, the children participating are asked to say where the protagonist will look for the object. Typically, for children above the age of 4, the answer seems obvious; as the protagonist has no information of the object's being replaced while he was out, he will search for it in location A where he left it. Children younger than 4 typically answer that the protagonist will search for the object where it in fact is at the moment they are asked, in this case location B. They thus “misrepresent” the kind of knowledge that the protagonist can have given his perception of the situation. In other words, they do not understand that others can have a false belief; one that differs from reality.

In this particular case, however, the experiment took “an interesting twist”, as Hauser expressed it (p. 601). Besides what false-belief studies normally show, it was found for this version of it, developed by Clements and Perner (1994), that 45 % of the children between the ages 2 years, 11 months and 4 years appeared to *look* at location A (correct) but nevertheless to answer *verbally* location B (incorrect). Children younger than 2 years and 11 months both looked and answered incorrectly, whereas children above 4 both looked and answered correctly, yet here – in between – there seemed to be a discrepancy between what some of the children indicated through their gaze and what they indicated by their words. The somewhat mysterious spark of these findings was fuelled by Hauser’s assertion that “[a]lthough Clements and Perner are able to rule out a number of potential explanations for their data, they can only provide a speculative account of the pattern obtained” (p. 602). The explanation suggested did not relieve me nevertheless of my puzzlement and fascination: the younger children being considered to be in a “one fact – one representation stage of development” (Hauser, 1996, p. 602) in which they represent facts about the world, yet fell short of making judgements about the facts. I was eager to find out more about this intriguing discovery that to me seemed to embody a link between humans’ developing mental capacity and other species’ way of dealing with the world, a kind of bodily versus mental understanding, even if such a distinction ought to be taken with some caution. Yet do these later capacities develop from the former, and in that case, how? Perhaps, even more importantly, regarding the kind of understanding that precedes an explicit and verbal understanding – what is it that remains of it once verbalization comes to dominate the manner of responding and of understanding and explaining things?

Passing a false-belief test has often been taken as evidence that the subject "got" a *theory of mind*. However, Astington (2001) begs to differ regarding this conclusion:

> Many authors take great care to distinguish among false belief, theory of mind, and social cognition (e.g. Flavell & Miller, 1988), emphasizing that children's theory of mind develops gradually during the yearly years, and can be characterized differently at different ages; yet one still hears people say that children have 'got' theory of mind when they pass the false-belief task (p. 685)

Astington points out that it would be better to administer a battery of tasks to measure a child's theory of mind instead of relying on a single test, saying too that researchers tend to focus on epistemic states such as beliefs instead of such motivational states as desires, emotions and intentions, which provide the most important reason for preferring multiple tasks. Moreover, this emphasises the fact that a theory of mind, and indeed social cognition too, is exceedingly complex.

A theory of mind can be defined as a mentalistic type of understanding, which means seeing mental states as causes of behaviour. In line with this then, in order to make sense of others in interaction one needs to recognise an underlying mental state as a cause of the other's actions and thus attribute this state to him or her (e.g., that he *believes* that the object is at location X, therefore he looks for it there). Several other terms can also be suggested to refer to the social-cognitive capacity of attributing mental states to someone: metacognition, mentalizing and mind-reading, these being used more or less interchangeably by some (e.g., Cortina & Liotti, 2010; Flavell, 1999; Michael, 2011). Leaving the specific explanations and suggested theories behind, a theoretically non-committed way to address this type of interpersonal understanding of this type is with use of the term *social understanding* (cf. Carpendale & Lewis, 2004, 2006), introduced in Barresi and Moore's presentation of “a theory of social understanding and the forms of representation of intentional relations.” In the article, they describe

> a system of such forms that can be applied uniformly to self and others; it requires the use of an intentional schema that can integrate first person information about an intentional relation with third person information about that relation. (Barresi & Moore, 1996, p. 121).

To be more specific, a mentalistic type of understanding constitutes only one aspect of social understanding. However, mentalization has often come to signify *the* way of conceiving of social understanding. In the study reviewed above, the children who looked at the correct location but answered incorrectly can be suggested to have acquired an *implicit* theory of mind (Garnham & Ruffman, 2001; Ruffman, Garnham, Import & Connolly, 2001). Implicit theory of mind and implicit understanding of belief refer to understanding as indicated by indirect and nonverbal (i.e. implicit) measures, rather than direct and verbal (i.e. explicit) ones.

Since the Clements and Perner study, several other claims of false-belief understanding in younger children have been presented. Onishi and Baillargeon (2005) for example, show that 15 month old infants can detect another person’s false beliefs if this is measured in such a way that the infant’s gaze is able to reveal what expectations he or she has. In a comparable paradigm of Kovacs et al. (2010), 7 month old infants showed sensitivity to false beliefs in others. These findings are still controversial, yet the keen interest in what infants actually understand and what they do not is in itself interesting. Does it matter whether infants can sense others’ mental states at 7 months or at 15 months or already as newborns? Reddy (2008) argues that it “matters profoundly” because the views of science and of philosophy affect people’s actions as well as their ways of thinking.

If one views infants as more or less lacking thoughts, feelings and perceptions this is bound to affect the way one treats and relates to them. Reddy

(2008, p. 6) notes that it was "not so long ago… that medical science asserted (without parents being able to challenge it) that neonates cannot feel pain and thus justified a variety of intrusive practices like surgery without anaesthesia." Thus, it seems that our notions of infants affect the surrounding society and families' attitudes and accordingly how adults *act* towards infants. This is very much a reason for investigating infants and their emotional and mental capacities. Of course, there are other reasons too. Besides expanding our knowledge of infants, which in itself of course needs no further justification, there are reasons that may not be as obvious. In attempts to construct theories concerning social understanding, specifically in infants, we may well discover something new. Investigating how infants understand themselves and the world is an area of research that has come a long way during the past few decades, its being able to tell us something about how adults understand all of this as well. One of the intuitions that has guided this work is that of there being aspects of adult social understanding that are not described or acknowledged by such notions as those of theory of mind, an in line with this that one can compare how adults understand others with how infants do so. The aim of the thesis can be said to be as follows:

> *To examine different aspects of the role of intersubjectivity for metacognitive development and for social understanding.*

There are two main theoretical perspectives, both of them concerned with how infant interaction can create the basis for social understanding, although the perspectives from the two differ and disagree explicitly on several points.

The one theoretical perspective is the *second-person approach* (e.g., Reddy, 2008; Gallagher, 2005; Hutto, 2008), first formulated as an alternative to theories of *social understanding*, most notably to theory of mind, as formulated either in a first- or in a third-person perspective. Barresi and Moore suggested a system reflecting the human capacity to enter into shared intentional relations with others and in so doing to employ an intentional schema integrating first and third person information regarding intentional relations; a capacity they claimed to be “the basis for human social understanding” (Barresi & Moore, 1996, p. 122). Barresi and Moore maintained that a strong case can be made for the integration of the first- and the third-person perspective. Two commentaries at the time regarding this posed the question of what happened to the second-person perspective (Reddy, 1996; Gomez, 1996).

This approach acknowledges the fact that infants' understanding of other persons starts in the second person, which is qualitatively different from the third person perspective, rather similar to Martin Buber's distinguishing between an *I-Thou* relationship from an *I-It* relationship (Buber, 1958). For instance, being the object of others attention and being addressed face-to-face is experienced differently (feels different) than someone talking *about* you with another person (cf. Reddy, 2003). Another relevant concept here is that of *primary intersubjectivity* – the idea that infants are born with a specific motivation to engage in intersubjective communication and the sharing of experiences (e.g., Trevarthen, 1998a). According to this approach, infants are endowed with this ability already from the start, which thus represents their entering into the social world of cooperation.

The second approach, one that deals with the importance of affective relations for social understanding, yet in a different way, is *mentalization theory* (e.g. Fonagy, Gergely, Target & Jurist, 2002), which suggests that mentalization is a developmental achievement that depends on affective mirroring in attachment relations. Primary intersubjectivity is thus rejected here. Yet, at the same time the concept of theory of mind is criticized for obscuring social understanding by being too narrow, particularly for not giving affect and emotion a role in the explanation of how the mind works. Besides making a case for the development of mentalization, Fonagy et al. show how the concept and the theory can contribute to clinical psychology and psychotherapy, and their having launched a specific form of treatment based on their findings (e.g. Fonagy, Bateman & Luyten, 2012). In their recent work, the concept of implicit mentalizing has received more attention, quite likely because current empirical studies concern the understanding of false beliefs in infants. This concept, however, appears to overlap with primary intersubjectivity (see paper III), both in development and in function.

Both of these theoretical approaches (mentalization theory and the second-person approach) have thus been developed partly as alternatives to the more or less omnipresent ideas contained in theory of mind, their being conceived as sufficient to account for social understanding.

The thesis is concerned with these two theoretical perspectives and endeavours to show that they both have significant contributions to make regarding an explanation of social understanding and that bringing them closer to one another may be worthwhile. Despite their differing points of departure and differing standpoints, the phenomena of interest are the same: that of understanding social understanding and how it develops.

Certain conceptual definitions have been purposely avoided. *Understanding* is probably the most central of these as it concerns the aim of the thesis and is involved in most theories and discussions. Over the years I have made various attempts to define it but found most of these to be wanting. I did get some valuable assistance from Georg Henrik von Wright's "Explanation and Understanding" (1971), in which he noted that:

> …'understanding' also has a psychological ring which 'explanation' has not. This psychological feature was emphasized by several of the nineteenth-century antipositivist methodologist, perhaps most forcefully by Simmel who thought that understanding as a method characteristic of the humanities is a form of *empathy* (in German *Einfühlung*) or re-creation in the mind of the scholar of the mental atmosphere, the thoughts and feelings and motivations, of the objects of his study. (von Wright, 1971, p. 6).

This, however, lead into what tended to look like a long odyssey through the history of science in general, and history of psychology in particular, searching for the particular distinction between understanding and explanation, and seeing the two as epitomes for the two types of understanding that were indicated in the study by Clements and Perner (1994). von Wright pointed out that "[o]rdinary usage does not make a sharp distinction between the words 'explain' and 'understand'" (p. 6), which fuelled this adventure further. However, I finally realized that I had ended up in a quagmire of free associations and hidden meanings and so had to leave it behind.

What seemed to me the most rewarding statements regarding understanding appeared instead in a text by Stephen Seligman (2000) who wrote: "Understanding is not exactly *about* experience, it is itself an experience" (p. 1193). Of course, one feels clearly if one understands (cf. a "feeling-of-knowing"; Koriat, 1993). Such an experience may, of course, be wrong – and in that case qualifies as being a misunderstanding, albeit unwittingly. How could we otherwise misunderstand something? Thus, I trust that the reader understands this term and others in accordance with common-sense – whatever that is.

In line with this, the concept "social understanding" could be said to denote the way people in general make sense of each other's behaviour. As interpersonal understanding of this type is typically taken to involve the understanding of self, a metacognitive understanding of oneself is included here in the notion of social understanding.

Some deliberate limitations are also involved. As linguistic development can reasonably be said to interact with cognitive and affective development, in this sense the choice to not consider linguistic development either may set certain relevant issues aside. Yet given the intention of focusing on the association and the link between social understanding and interaction in infancy, and the fact that these intersect within the area of what can be called an implicit sense of social understanding, the choice not to take linguistic aspects into account hopefully appears reasonable. Furthermore, the dissertation concerns for the most part younger children (below 18 months) for whom language acquisition has just begun.

The thesis is organised as follows. Section 2 begins with a review of the theory of mind concept. The theories upon which the thesis is focused are best understood against the backdrop of the theory of mind paradigm. The intention then in sections 3 and 4 is to describe the second-person approach and the mentalization theory as described by Fonagy et al. (2002), and the evidence, grounds for the hypotheses and theories and the motivations for including them. Section 5 deals then with various contrasting theories and their account of infants' social knowledge, in particular with theories of attachment and of primary intersubjectivity. In section 6, a brief review is presented of non-mentalizing functions of social understanding, or, socio-cognitive functions that exist either prior to, or alongside, mentalizing modes. The papers to be presented later in section 8 will thus hopefully appear to be situated in a more familiar context than they might otherwise appear to be. Section 7 turns to the question of methodology and except for discussing methodology in theoretical psychology generally I will describe the methods I have used in the papers. In section 9, finally, after the separate papers have been introduced, a general discussion will be undertaken that ends with certain concluding remarks that summarize the contributions of the thesis in relation to the research area involved, making certain proposals finally for future research.

2. Theory of Mind and Empathy – First Steps to a Conception of Social Understanding

In the first number of *Behavioral and Brain Sciences* in 1978 there appeared a target article by David Premack and Guy Woodruff with the title: "*Does the chimpanzee have a theory of mind?*" They stated their subject matter as follows:

> The chimpanzee's evident comprehension of physical relations makes it of interest to determine at what level these relations are available to the animal and whether there is a sense in which he can be regarded as a lay physicist. But questions of this kind are only indirectly relevant to our present concerns. We are less interested in the ape as a physicist than as a psychologist (every layman, of course, is both); we are interested in what he knows about the physical world only insofar as this affects what he knows about what someone else knows. (Premack & Woodruff, 1978, p. 515).

Premack and Woodruff speculated about a particular possibility, one implying that the chimpanzee may have a "'theory of mind,' one not markedly different from our own." The phrase *Theory of Mind* meant an individual's imputing of mental states to him/herself and to others by inferring from behaviour what mental state may have caused it. Furthermore, they argued that such a system of inferences is properly viewed as a theory because mental states are not directly observable, and also, because the system is capable of making predictions about other organisms' behaviour in particular. As their ambition was to investigate the chimpanzee's comprehension in problem solving, they mentioned three potential a priori interpretations of the results: *associationism*, *theory of mind*, and *empathy*. Since the interpretations of this seminal article have had a huge impact on research on theory of mind, it is worthwhile to take a closer look at these.

Two types of interpretations have commonly been considered in developmental and comparative psychology: either some version of associationism, that is, how a certain problem or task is solved through familiarity with some aspect of it; or, second, a cognitive mechanism or process suggested as being responsible for solving the task. The second interpretation departs from the associationist one in that it claims external independence (i.e. from the environment) to a certain degree. Such an external independence would be highly favourable as regards novel situations and tasks, in which one cannot lean on previous experiences with the task at hand. The former, however, has its strongest merit in involving generalizations from previous experience in situations that resemble the situation at hand, or some aspect of it. Arguably, sometimes it is more economic to lean on external routines or patterns to save cognitive resources and allow for other things to be processed simultaneously, for example. Few would deny that both of these abilities matter to some degree. Storing internal representations is valuable for many reasons. For instance, treasuring memories of light summer nights while freezing in December makes sense to most people as a vivid memory of the good times, that hopefully returns, can keep one going through the hard times. Imagining places and situations that are not here certainly facilitates our going there, in contrast to our needing to follow a path step-by-step.

Premack and Woodruff suggested, as an alternative to associationism, *theory of mind* and *empathy*. Actually, they insisted that empathy and theory of mind are "not radically different views; they are in part identical" (p. 518). Empathy does not grant the subject any inferences about another's knowledge; instead it is "a theory of mind restricted to purpose" contrary to "a more nearly complete theory that takes into account not only the other's motivation, but his cognition as well" (p. 518). In the empathy view, the subject "'puts himself in the place of the actor,' and chooses an alternative in keeping with what he would do were he in the actor's predicament" and does so by imputing a purpose to the actor, and the subject then "takes over the actor's purpose, as it were, and makes a choice in keeping with that assumed purpose" (p. 518). In addition, Premack and Woodruff note that the empathy view can be seen in either of two ways, understanding either resulting from imagining what the subject would do were *he/she* in the other's position, or, from what the subject would do were he/she *the other person* in that situation. The second alternative is not clearly distinguished from a theory of mind as this would assume knowledge of what a toddler is like, which seems equivalent to inferring what the other is like and will do.

The two interpretations – theory of mind and empathy – soon became two contrasting disciplines in the growing research discipline: theory of mind that takes a third-person perspective (a theory referred to by the slightly odd phrase *theory theory*), and *simulation*, which is referred to as a first-person perspective[1]. The distinction between a theory and a simulation approach was part of an intense debate that lasted for some time (see Davies & Stone, 1995a, 1995b), even if hybrid-versions that see inferences and simulation as complementary capacities have gained acceptance (e.g., Perner 1996). Simulation theory claimed support from empirical findings on mirror neurons in monkeys (e.g. Gallese & Goldman 1998, Goldman 2006), a finding that was not predicted from the theory theory. Besides the two dominant approaches, theory theory and simulation theory, there are *modular* approaches to theory of

[1] The simulation theory soon came up with two versions, termed *introspectionism* (e.g Goldman, 1989, 2006) and *radical simulation* (e.g. Gordon, 1986) that differ on how to conceive of the simulation process. This difference is similar to what Premack and Woodruff (1978) anticipated in distinguishing between two views on empathy, see above.

mind that have some similarities with theory theory, most notably the simile of a theory or conceptual framework (e.g. Leslie 1994, Malle 2005, Premack 1990). The modular account stresses innate modules such as a theory of mind-mechanism (e.g. Leslie 1995, Baron-Cohen 1995), whereas theory theory emphasises social development, especially as concerns language and its connections to cognitive development.

The most emblematic mental state as regards mind knowledge is arguably "belief", epitomized in the classic false-belief task set up by Heinz Wimmer and Josef Perner (1983, see the Introduction section above). However, the idea is derived from Bennett (1978), Dennett (1978), and Harman (1978), respectively, who in comments on Premack and Woodruff's article suggested a formal paradigm for studying "children's competence in representing another person's definite belief which differs from what the subject knows to be true" (Wimmer & Perner 1983, p. 106).

The false-belief task became highly influential in developmental and cognitive psychology, so much so that social understanding (conceived as theory of mind) became more or less identified with the ability to understand others' false beliefs despite the fact that it was only identified in children above four years of age. Some years later, however, findings came up that suggested a precursory implicit theory of mind (e.g. Clements & Perner 1994, Ruffman, Garnham, Import & Connolly 2001). Conceiving of theory of mind as implicit has lowered the boundaries for when infants can be granted a theory of mind. As was mentioned above, more recent studies based on implicit measures such as anticipatory looking and violation-of-expectation paradigms suggest a sensibility for others' false beliefs in infants at the age of 15 months (Onishi & Baillargeon 2005) and of 7 months (Kovacs et al. 2010). The issue of what implicit mentalizing might mean is dealt with in papers II and III.

3. A Second-Person Approach – Instead of a Theory of Mind

As we saw, in approaches to theory of mind a clear distinction developed between those claiming a third-person approach and those preferring an explanation from a first-person perspective. The first proposals for a second-person approach to social understanding and the mind were probably two commentaries, Juan Carlos Gomez (1996) and Vasudevi Reddy (1996), to an article by Barresi and Moore (1996). Barresi and Moore argued that organisms have direct first person information about their own intentional relations and direct third person information about other agents' intentional relations. The former tends to be information about "objects in relation to the agent's actions or potential actions", whereas the latter tends to involve "agents and their movements in spatiotemporal relations to objects" (p. 109). The claim was that third and first person information are distinct from one another but can be integrated by means of an "intentional schema" to be applied equally much to activities of both the self and others. The critical question raised by Reddy was why the second-person perspective was omitted in the first place[2].

In the assumptions of a second-person approach there are, according to Reddy (2008), three core features: The first is to reject the gap that theory of

[2] Besides these proposals for the primacy of second-person engagement, there is "Interaction theory" (e.g., Gallagher, 2004) that highlights the significance of second-person interaction as well.

mind assumes, i.e. that other minds are opaque to perception and only accessible through inference or simulation. The second one is to reject the assumption often made that we understand "others" in the same way, in other words that we have one way of understanding others that is not significantly affected by *who* those "others" are. The third is a second-person approach that regards "active emotional engagement between people as constituting – or creating – the minds that each comes to have and develop, not merely providing information about each to the other" (Reddy, 2008, p. 27). The first problem concerns that of other minds, in the sense that what we assume regarding this is what a theory of social understanding must explain[3]. The second point concerns what Reddy calls relational knowing, the third point emphasising then the creation, or construction, of mind and of minded interaction.

I would like to add a fourth feature also inspired by Reddy's work: instead of extrapolating from what we believe about our adult psychological realm to the psychological world of infants, the opposite direction could be preferable.

The way infants learn to understand others is likely to affect the nature of adult social understanding because early functions may not disappear, but are very likely interwoven to some extent in a more mature and developed manner of understanding (certain suggestions concerning this are developed in papers II and III).

[3] In other words, if one assumes that mind is invisible and distinct from behaviour, then it becomes necessary to explain how we reach a conclusion that behaviour has been caused by an internal, mental state. If on the other hand one see behaviour as something that is part of the mental sphere one need not make those inferences, as behaviour is meaningful in itself. An emotional expression, for instance, is part of, or manifests, the emotion in question, rather than its referring to a mental state that is otherwise opaque.

To be sure, the way adults understand their infants is also an important factor in how the infant learns to understand others, but in a radically different sense. As adults, we are biased by all kinds of presuppositions about infants and their understanding, and by the view we have of ourselves and our capacities. What adults, researchers, philosophers, and others think of infants also affects what we grant infants, i.e. how we interpret their capacities, abilities and skills. Thus, we need, as in all research, an as unbiased look as possible at what is being investigated, i.e. at what infants actually do, and we need to think thoroughly about what the empirical findings presuppose.

For instance, just as De Bruin, Strijbos and Slors (2011) assert, there are no a priori reasons why the assumed belief-desire model (BD-model) of adult folk psychology should also be suitable for explaining implicit social understanding as it is evidenced in infants, and thus the psychological world that infants inhabit. It might well be the case that the BD-model is inadequate for our understanding of the infant world.

In the same manner, theory of mind has been accused of making social understanding "overly intellectualized" (Gallagher, 2001; Hutto, 2008; cf. Hobson, 1993). Similarly, Jane Heal (2005, p. 35) declares that "if we misdescribe the nature of the fully-fledged competence, then our theorizing about the earlier stages will be undertaken in the light of a misunderstanding." One might even say that the mature explicit, explanatory model is necessarily, to some inevitably undefined degree, dependent upon social, cultural, and cognitive resources and notions, whereas newborns' rudimentary understanding is likely to be closer to nature, so to speak. Of course, one might never be endowed with such an unbiased and theory-independent gaze, more likely to be engaged again and again in a "continuous interplay between empirical results and theoretical reflection" (Looren De Jong, Bem & Schouten, 2004, p. 276). However, this point may nevertheless be deserving of emphasis every now and then.

The best example of taking the second-person perspective seriously might be the case of attention. Tomasello (1999) argued for a nine-month social-cognitive revolution that concerns what he sees as the start of joint attention:

> As infants begin to follow into and direct attention of others to outside entities at nine to twelve months of age, it happens on occasion that the other person whose attention an infant is monitoring focuses on the infant herself. The infant then monitors that person's attention to her in a way that was not possible previously… From this point on the infant's face-to-face interactions with others… are radically transformed. She now knows she is interacting with another intentional agent who perceives her and intends things toward her. (Tomasello, 1999, p. 89; cited also in Allen, 2003, p. 99).

This is an apt description of how an infant can start to appreciate that there is an intentional agent interacting with him or her. Recognizing that one is the object of somebody else's attention seems to be a crucial step in the infant's development (cf. Reddy, 2003). Tomasello says that this requires that the infant understand others' intentions, which according to him is not possible before so-called joint attention starts. Joint attention, in this view, is triadic attention in which the infant and another person are jointly attending to a third object, as indicated by their directing their gaze at the object and referring to it verbally. In this, the adult's gaze alternates between the infant and the object. Through intending things *about* the object *to* the infant, the adult's intentions about both the object and the infant are "discovered" by the infant according to Tomasello.

However, Reddy (2003) argued for infants making exactly the same appreciation of others' attending to them, only some months before. In her view, realizing that one is the object of others' attention does not require that the infant first recognize the adult's attending to a third object. Interaction during the first months is arguably full of potential moments for experiencing that someone "perceives her and intends things toward her" as Tomasello says. Reddy argues instead that infants' awareness of others as attending beings and as being the object of their attention "must lead to, rather than result from, representations of self and other as psychological entities" (Reddy, 2003, p. 397). Note that Tomasello sees attention and intention as distinct, whereas Reddy sees attending to things as intending those very things (see Brinck, 2001; Brinck & Gärdenfors, 2003).

Infants about 2 months in age react to attention from others in a variety of ways. They show more smiles during eye contact and less when the other looks away, showing distress when they are unable to disengage from a gaze or when a gaze is non-contingent, as in still-face paradigms, showing coy reactions to renewals of attention by a combination of "intense smiling with brief gaze and head aversion sometimes accompanied by raised curving and arm movements, an expressive pattern often considered the archetypal self-conscious displays." (Reddy, 2003, p. 397; cf. Reddy, 2000). At 4 months, infants also make active attempts at directing others' attention to themselves by "calling vocalisations." In the second half of the first year, infants engage in interaction that can involve showing-off, clowning, teasing and repeating acts to re-elicit praise from adults (e.g. Trevarthen & Hubley, 1978; Reddy, 1997, 2001, 2003). The point here is to illustrate the kind of behaviour that the second-person approach considers to constitute the first manifestations of social understanding.

Why disqualify these primary appreciations by claiming that infants *first* need to understand intentions by means of perceiving others attending to a third object *before* they understand intentions in the other's attending to them? It makes no sense. One aspect of this critique is, of course, methodological. Obviously, it is easier to observe when the participants in a study direct their gazes more overtly, as in a triadic attention, at external objects. Besides turning their gaze toward the object, they may turn their heads and arrange their bodies in different ways so as to adjust to an object, move in direction towards it, et cetera. Attention may not seem as readily detected in a meeting of gazes between two persons who mutually attend to each other. On the other hand, if one sees two people looking at each other, the fact that they attend to each other is not far-fetched. Active emotional engagement entails a certain feeling or experience, such as when one suddenly experiences the "breath-catchingness and warmth" of receiving a smile from a person, contrary to what it feels like to simply observe a smile at someone else. It thus "matters powerfully," Reddy maintains, "whether the 'other mind' that you observe is turned towards you in engagement with you or towards someone else" (Reddy, 2008, p. 27). The expressions may be the same but the experience of them can be "phenomenally different", because the expression is both more immediate and more powerful in direct engagement, as it calls forth an obligation to *respond* to the other person's act. Reddy emphasises that social understanding "is not only coloured by my emotion, it is *explained* by the fact that my emotion is engaged with X" (1996, p. 140). Thus, the assumed gap between the first-person experience of one's own feelings and one's experiencing of the other's feelings "needs only emotion, not concepts or interpretive schemas, to bridge it" (Reddy, 1996, p. 140).

Gomez (1996) argued in a similar vein that second-person intentional relations involve specialized systems that may have been selected by evolution and as such are primary to both first- and to third-person intentional relations. The evolutionary argument should be considered fundamentally important, as a system that most likely is a domain-specific system seems much more parsimonious then a domain system of a more general character at a more or less abstract level (such as a theory of mind or an intentional schema). Besides such parsimony, non-human primates are known to be sensitive to eye-like configurations, to which they often react strongly with aggressive or defensive behaviours, and accordingly there appears to be brain circuits in monkeys dedicated for detecting differing types of eye gaze, particularly gazes directed at their own faces (Gomez, 1996). Gomez concludes that being the target of the visual attention of another organism, appears to "have been singled out in evolution to have its specialized information-processing mechanisms" (p. 130). Developmental data also supports the idea that second-person relations are primary to first- and third-person relations, as infants generally engage in second-person interactions well before they engage in third-person interactions (cf. Stern, 1985; Reddy, 2010). This invites one to conceive of second-person intentional relations as being a part of inherent functions, such as the perceptual system for example.

Gallagher (2008) argues that at least a rudimentary social understanding is basically a part of the perceptual system. In this view, we have immediate access to intentional states through direct perception. We immediately perceive others' affective attitudes and the intentionality in their gestures, facial expressions, movement, and actions (Hutto, 2008), which makes inferences or simulation superfluous or at least secondary. Discrete emotional experiences are accompanied by facial expressions, changes in voice, posture, movements and they "emerge in ontogeny well before children acquire language or the conceptual structures that adequately frame the qualia we know as discrete emotion feelings" (Izard, 2009, p. 5). Colwyn Trevarthen does not, to my knowledge, explicitly mention a second-person approach, but his work implies a second-person perspective[4]. The idea of direct and immediate knowledge in interaction is something that Trevarthen expresses clearly on the other hand. The following quotation stands out in sharp contrast to the problem of other minds:

> [T]he effiency of sympathetic engagement between an infant and the adult signals the ability of each to 'model' or 'mirror' the motivations and purposes of the partner, immediately... Perceiving other persons' emotions cannot be divorced from the generation of expressive forms of acting because every person's communicative signals are made by highly specific forms of movement that are adapted to fit human perceivers' sensitivities (Trevarthen, 1993a,b). A smile *is* happy, as is walking with a fast, tripping step; tears *are* sad, as is a slow dragging way of walking and a downcast look, and emotions expressed by one person can lead to instantaneous sympathetic mimicry in an other. Expressions

[4] Trevarthen's theory of primary intersubjectivity is discussed more in section 5 and in the separate papers.

> of the self ‘invade’ the mind of the other, making the moving body of the Self resonant with impulses that can move the Other’s body too. (Trevarthen, Robarts, Papoudi, & Aitken, 1998, pp. 59-60).

Here, we are faced with the notion, expressed by many, for instance by Wittgenstein (1980, see e.g. § 570), that emotions and mental states generally are not primarily mental representations. Rather, they are manifested in overt behaviour and actions. Facial expressions are part of an emotion rather than an effect of it. This more theoretically oriented argument has implications, of course, for how we interpret the empirical evidence that exists. I have more to say about this in section 5.

To conclude this section, arguments for a second-person approach may perhaps be boiled down to the idea that human interaction, in which making sense of minded action is imperative, is not primarily about detached prediction, explanation and control. Rather, minded action is embedded in a pragmatic understanding of situations and concerns affective and intentional attitudes expressed in bodily actions that are “directly” or immediately perceived, without the need for resorting to inferences or simulation (Brinck, 2001; Brinck & Gärdenfors, 2003; Gallagher, 2001, 2004; Gallagher & Hutto, 2006; Hutto, 2008; Reddy, 2008; Reddy & Morris, 2004).

Central to this view is the type of emotional engagement in second-person relations that infants have with their caregivers. However, one can be emotionally engaged in third-person relations as well, for instance as you watch a movie or hear a story of people. In such cases, the engagement is on your behalf, so to speak. Also, you may have occasioned second-person interactions in which you or the other lack even the slightest propensity for engagement, as in trying to talk to someone whose mind is “somewhere else”. These are, despite their apparently violating such a notion, not a threat to the approach; they are not to be understood literally.

Next, we will take a look at a different response to theory of mind, that of mentalization theory. In it, we encounter ideas familiar from the theory of mind framework, yet the notion of theory of mind is nonetheless criticized on some important points here.

4. Mentalization Theory – A Second Alternative to Theory of Mind

In the previous section we encountered some problems evident in the theory of mind perspective. Mentalization theory (Fonagy, Gergely, Jurist & Target, 2002) agrees that the notion of theory of mind disregards emotion, which in their view is a considerable limitation: "The emotional significance of mental states determines the evolution of the capacity [i.e. mentalizing, theory of mind] or structure available for processing, but this is not usually addressed" (p. 30). Moreover, they posit that mentalization is a developmental achievement, and not something that can be assumed. They argue that current models of theory of mind development portray

> a barren picture, which ignores the central role of the child's emotional relationship with the parents in fostering the capacity to understand interactions in psychological terms. The development of children's understanding of mental states is embedded with in the social world of the family, with its network of complex and often intensely emotionally charged relationships, which are, after all, much of what early reflection needs to comprehend (p. 30).

Any theory meant to account for social understanding must thus involve emotion and affect. Fonagy, Bateman and Luyten (2012, p. 3) take the term "mentalizing" to denote "the remarkable and pervasive human tendency to look beyond the visible shell of the body in understanding behavior and seeking descriptions and explanations in terms of states of mind" and is thus the “imaginative mental activity that enables us to perceive and interpret human behavior in terms of intentional mental states (e.g., needs, desires, feelings, beliefs, goals, purposes, and reasons)” (p. 4).

However, in this theoretical framework some of the features are nonetheless included that the second-person approach is confronted with, especially the way in which the distinction between mind and behaviour is invoked. For instance, they argue that the development of social understanding goes from a primary level in which social contingency rather than emotions explains infants’ manner of dealing with the world (without distinguishing between animate and inanimate objects), although including the process of affect mirroring (see below), over what has been termed a teleological mode in which infants understand actions and behaviour in physical reality in terms of reference to future states or goals, their finally entering into a mentalizing mode then in which they begin to mentalize “once representations of future goal states come to be thought of in terms of the agent’s belief about physical reality.” (p. 33). Development to reach a mentalizing mode, it is argued, depends on “the quality of interpersonal interactions between the infant and the parent” (p. 33), this theory too thus baring social understanding on social interaction.

Fonagy et al. (2002) describe their aim at the most general level as that of highlighting the crucial importance of developmental work in psychology both to psychotherapy and to psychopathology, their attempting to point in this way to a new direction for psychoanalysis and psychotherapy. They have developed a clinical approach, and a form of “mentalization-based treatment” as well, that are presented in several pieces of work (e.g. Allen & Fonagy, 2006; Bateman & Fonagy, 2004; Fonagy & Bateman, 2012). The concept of mentalization contains several overlapping constructs besides that of theory of mind, such as empathy, mindfulness, affective consciousness, and psychological mindedness (Choi-Kain & Gunderson, 2008).

The definition of the concept of mentalization has developed further over the years. It was first defined as a “reflective function” that refers to “the operationalization of the psychological processes underlying the capacity to mentalize" (Fonagy et al. 2002, p. 24). That concept is understood in terms of Dennett's "intentional stance" (1987) and is explicitly compared to theory of mind and to folk psychology. The reflective function was initially operationalized as a measure composed of several items from the Adult Attachment Interview (AAI) in which a parent's narrative of his or her own attachment experiences is scored (cf. Slade, 2005).

A second definition of the concept was introduced as the *Interpersonal Interpretative Function* (IIF). The IIF is described by Bateman and Fonagy (2004, pp. 74-75) as "an evolutionary-developmental function of attachment" that consists of a "cluster of mental functions" that serves to process and interpret new interpersonal experiences. The IIF "consists of interpretive functions in the related domains of affect-regulation, attention, and reflective function" (p. 75), its distinguishing two types of interpretive processes: those that concern interpreting cognitions of self and others (IIF-c) and those that are directed at affect states (IIF-a). The IIF-c concerns primarily reasoning about epistemic states, whereas emotional resonance (or empathy) exemplifies the IIF-a. These are associated with distinct but to some extent overlapping neurocognitive systems: "Cognitively oriented mentalization involves several areas in the PFC [prefrontal cortex], whereas affectively oriented mentalizing is particularly related to the ventromedial PFC" (Fonagy, Bateman & Luyten, 2012, p. 29). The ability to predict and experience the emotions of others is mediated, in other words, by a functional connection between the two interpersonal interpretive centres in the brain (Fonagy et al., 2002).

Fonagy, Bateman and Luyten (2012) suggest that mentalization should be broadly construed in terms of four distinct dimensions having two polarities each: Cognitive vs. Affective, Automatic/Implicit vs. Controlled/Explicit, Self vs. Other and External vs. Internal Mentalization. One significant point of this novel conceptualization is that it is easier to assess a person's mentalizing profile, that is "the individual's functioning with respect to each of the polarities underlying mentalizing, particularly because there may be dissociations between these polarities (e.g., impairments within one polarity but not within other polarities)" (Luyten, Fonagy, Lowyck & Vermote, 2012, p. 43). This also facilitates comparison with other, related constructs, such as those of empathy and mindfulness.

Cognitive versus affective mentalization is understood, as was indicated above, in terms of the two sub-systems IIF-a and IIF-c, and is exemplified by the developmental distinction between understanding of desires and beliefs. This dimension is important to the theory because full-fledged mentalization entails the integration of cognition and affect, such as in empathy and mentalized affectivity (Jurist, 2005).

The dimension of Self-oriented versus Other-oriented mentalizing needs no further explanation. This dimension implies that a person can have difficulties in mentalizing in either or both of these polarities. However, there is much evidence that these are often tightly interwoven. For example, disorders characterized by severe impairments in feelings of self-identity, such as psychosis and Borderline Personality Disorders, are also characterized by severe deficits concerning reflections on others' mental states (Fonagy, Bateman & Luyten, 2012). It is a distinctive feature of theories concerning knowledge of mind that these polarities are somehow related (cf. Carruthers, 2009; Gallagher & Meltzoff, 1996). The "sense of self", or subjectivity, is at the very centre in mentalization theory, the development of both subjectivity and mentalization being associated there with one's relations to others, especially attachment relations.

The polarities concerning internal versus external mentalizing are particularly interesting, as these show a clear correspondence in many ways to various ideas that were discussed in relation to the second-person approach. It is also the most novel of the dimensions, even if it may have been implicit in earlier writings. *Internally focused mentalizing* "refers to mental processes that focus on one's own or another's mental interior (e.g. thoughts, feelings, experiences)" (Fonagy, Bateman & Luyten, 2012, p. 22). *Externally focused mentalizing*, on the other hand, concerns mental processes that "rely on physical and visible features and one's own or another's actions" (p. 22). This distinction is illustrated by patients who are impaired in their ability to read the other's mind, while they at the same time show hypersensitivity to facial expressions. Patients with anti-social personality disorder, however, show the opposite pattern: they are often "experts" at reading the inner states of others in order to manipulate them. At the same time, they appear to lack the ability to "read fearful emotions from facial expressions" (p. 23).

Last, but certainly not least, the dimension of automatic/implicit vs. controlled/explicit is considered by Fonagy, Bateman and Luyten (2012, p. 20) to be the "most fundamental polarity underlying mentalizing". Controlled/explicit mentalizing, as they state, is a "serial and relatively slow process, which is typically verbal and requires reflection, attention, awareness, and effort." Automatic/implicit mentalizing, on the other hand, "involves parallel and therefore much faster processing; is typically reflexive; and requires little or no attention, intention, awareness, and effort" (Fonagy, Bateman & Luyten, 2012, p. 20; see also Fonagy & Luyten, 2009). The distinction as one between reflective and reflexive systems suggested as neurocognitive models of social cognition is worth emphasising (Satpute & Lieberman, 2006; Lieberman, 2007). In our daily interactions, Fonagy et al. (2012) say, mentalizing is "predominantly implicit", since in most interpersonal situations we tend to "rely on automatic and unreflective assumptions about ourselves, others, and ourselves in relation to others" (p. 20). Moreover, individuals tend to "relax controlled mentalization and judgements of social intent and social trustworthiness in secure attachment relations" and instead "rely on more automatic, intuitive processes" (p. 20). Thus, while implicit mentalizing is unreflective, it is reflexive and involves ordinary, everyday experience: "When things go smoothly, particularly within secure attachment relationships, relying on automatic mentalization appears to be normal because more reflective processing is unnecessary" (p. 20). Fonagy, Gergely and Target (2007) suggest that the biological basis of implicit mentalization is probably active by the age of 1 year and possibly even earlier. Since implicit mentalizing is particularly interesting for the aim of the thesis, a more detailed description seems justified. Allen (2006) explains that

> [w]e mentalize others implicitly, for example, in conversations: we take turns and consider the other person's point of view, to a large extent – when all goes smoothly – without needing to think explicitly about it. We also mentalize others implicitly when perceiving and responding to their emotional states: we automatically mirror them to some degree, adjusting our posture, facial expression, and vocal tone in the process. Were we to attempt all this explicitly, we would come across as stiff and wooden rather than naturally empathic (p. 10).

Yet if mentalizing others implicitly seems elusive, he says, it is nothing compared to the phenomenon of mentalizing oneself implicitly, or "unreflectively". In a later passage the conclusion is offered that

> [m]entalizing implicitly in relation to oneself, then, entails an emotional state connected to the self – a pre-reflective, felt sense of self that is inextricable from the agentive sense of self, the initiator of purposeful action. Mentalizing implicitly, one has a sense of self as an emotionally engaged agent – 'what it feels like to be me' in the process of thinking, feeling, and acting. (Allen, 2006, p. 11).

In this more recent approach, as you may have noticed, the word "mentalizing" is preferred, as this verb emphasising the concern with *something we do* more than with something we have; in other words it refers to a "mental activity" (cf. Allen, 2006; Fonagy & Bateman, 2012). It also fits in well with the idea of implicit mentalizing as something we do unreflectively, in our every-day encounters with others. Allen and Fonagy et al. appear to agree on this interpretation.

In his outline of the concept of mentalizing, Allen (2003) unwittingly captures the essence of the second-person approach: "We mentalize when we treat others as persons rather than objects" (p. 93). This may seem trivial, but assuredly it is not. At the end of the paper, Allen turns to Strawson (1985) in reminding us that mentalizing can also be made in a detached manner (and I borrow Allen's quote from Strawson):

> To see human beings and human actions in this [detached, objective] light is to see them simply as objects and events in nature, natural objects and natural events, to be described, analyzed, and causally explained in terms in which moral evaluation has no place; in terms, roughly speaking, of an observational and theoretical vocabulary recognized in the natural and social sciences, including psychology. (p. 40) (Allen, 2003, p. 107).

Allen says that patients are naturally highly sensitive to a detached stance such as the one just described. "Overly objectified, they will rightly complain of being analyzed, scrutinized, or put under a microscope" (p. 107). In a detached stance, it is possible to formulate intellectualized interpretations, but they do the patient no good if mentalizing is not simultaneously firmly grounded in an emotional engagement. In case it is short of an engagement of this type, "we have the form but not the spirit of mentalizing" (p. 107). One can doubt, of course, whether having the form but not the spirit qualifies as "true" or "genuine" mentalizing, but such an objection probably misses the point: implicit mentalizing must not be overly intellectualized, but it risks being so if emotional engagement is lacking. One important thing pointed out by Allen (2006, p. 7) is that ”mentalizing is not only something we do; it is also something we can fail to do."

To conclude, mentalization theory presents several types of clinical cases that give us reasons to believe that the very distinction between inner and outer aspects of thinking or behaviour is important for at least some aspects of social understanding. This means that when the second-person approach rejects the distinction between mind and behaviour, stating that the distinction assumes the problem of other minds, the dimension of internal and external are in danger of being dismissed as well. Of course, one can deny that there is a biological difference while still recognizing that there is a cultural one. On the other hand, if it is the case that the ability to mentalize is of crucial importance both in psychotherapy and in adult reflection, this does not mean that it is necessarily crucial for understanding in children. It is still possible that children understand through practices such as those suggested by the second-person approach.

A possible way of solving this could be to suggest two ways of construing mentalization, or in other words of construing the ability to conceive of internal states as opposed to external behaviour in social understanding (i.e. as a "folk psychology", Davies & Stone, 1995). For one thing, mentalization could be construed as a *descriptive* concept. In this view, the questions for this research would be those of where and when and how this ability develops, functions, etc., the concept thus needing a precise definition. A second possibility would be to view mentalization as a *normative* concept, where it could be argued that engaging in mentalizing is desirable because it helps us to understand ourselves and others in ways that are clearly superior as opposed to some other ways of making sense of (and to) others. I imagine this distinction to be similar to how Kahneman and Tversky's descriptive prospect theory (1979) responded to the normative subjective expected utility theory (SEU; Savage, 1954; von Neumann & Morgenstern, 1944). Instead of examining conditions for ideal decisions based on probability theory, Kahneman and Tversky set out to examine how people in fact do make decisions. Their descriptive, empirical account differs largely from the conception of how people ought to make optimal decisions according to SEU. Even for its most determined critics, the normative conception could be seen as warranted. Arguably, the mentalization theory entails both of these considerations.

That mentalization works well in psychotherapy seems beyond doubt, although there is more to find out about this approach, of course. It could be helpful to use it in educational practice concerned with persons who recently became parents, for example, because many studies attest to parents' mentalizing about their own experiences as well as about their infants deeds and doings affecting their children's cognitive development and attachment (e.g. Meins et al., 2002; Peterson & Slaughter, 2003; Grienenberger, Kelly & Slade, 2005). Concerning the descriptive concept there, of course, is much to say; indeed, parts of the thesis are concerned with such questions (papers II and III) and the second-person approach takes issue primarily with the descriptive aspect of things.

Let us first of all acknowledge, therefore, that one can accept the normative sense and yet be sceptical about mentalization as a concept describing in a universal way how people go about understanding others. The descriptive claim is not difficult to come across. Tooby and Cosmides (1995, p. xvii), for example, claim that "humans everywhere interpret the behavior of others in... mentalistic terms because we all come equipped with a 'theory of mind' module (ToMM) that is compelled to interpret others this way, with mentalistic terms as its natural language." Or in Baron-Cohen's words that it is "hard for us to make sense of behavior in any other way than via the mentalistic (or 'intentional') framework" (Baron-Cohen, 1995 p. 3).

Wellman, Cross and Watson (2001) discuss their findings from a meta-analysis of theory of mind development, their concluding that

> [a] mentalistic understanding of persons that includes a sense of their internal representations—their beliefs—is widespread…Even if an understanding of actions in terms of beliefs proves to be not strictly universal, the meta-analysis documents that it is impressively widespread, at least in childhood. This suggests that such a conception is a natural, easily adopted way of understanding persons worldwide; it is cognitively 'contagious,' to use Sperber's (1990) terminology (Wellman, Cross & Watson, 2001, p. 679).

Although their conclusion is more modest and thoughtful, they nonetheless regard a mentalistic understanding as being a "natural" way of understanding. They do admit, however, that cultural factors may be involved, even if they find no support for it in this metaanalysis. In section 6, various notes regarding this will be presented.

5. Interaction in Infancy

In considering dyadic interaction in infancy, two notions are of particular interest: attachment and intersubjectivity. Mentalization theory aims at establishing a link between attachment and the development of mentalizing ability through the processes of social contingency and affect mirroring. Bateman and Fonagy suggest that

> the link of the secure base phenomenon to the development of mentalization will be increasingly understood to be causal rather than correlational in that the group of capacities that underpin adequate social understanding…are evolutionary tied to it…and therefore (3) that deficits in attachment create a vulnerability in the child to later environmental challenges because of deficits of interpretive capacity. (Bateman & Fonagy, 2004, pp. 78-79).

Bateman and Fonagy's model aims at developing attachment theory in the sense that the development of mental representations is one of the primary functions of attachment alongside that of protection (Main, 1991). Thus, mentalization is regarded here as a continuation of attachment. However, mentalization is sometimes referred to as a type of intersubjectivity, a view in which intersubjectivity is defined as the ability to read others' mental states (e.g. Allen, 2006; Cortina & Liotti, 2010; Fonagy, Gergely, Target & Jurist, 2002),

Contrary to this, others suggest that newborn infants are endowed with an ability to partake in, as well as a motive to engage others in, interactions that also are considered intersubjective, yet with a differing definition of intersubjectivity, a phenomenon referred to as primary intersubjectivity (e.g. Trevarthen 1998a). This section deals in some detail with primary intersubjectivity and its commencement in relation to attachment.

5.1 Primary Intersubjectivity

The term *intersubjectivity* was introduced into psychology in 1974 by Joann Ryan (see Beebe, Knoblauch, Rustin & Sorter, 2003) and was picked up by Colwyn Trevarthen (1974; "provocatively because of its Marxist overtones" according to Reddy, 2008, p. 248) for denoting early dyadic interaction between the infant and the caregiver. Evidence from video of "protoconversations" between 2 month olds and their mothers was suggested by Trevarthen to demonstrate "primary intersubjectivity" (Trevarthen, 1974, 1979). Intersubjectivity was also defined as the deliberate sharing of experiences about objects and events (e.g. Trevarthen & Hubley, 1978), their adding the term "primary" meaning that "the infant is born with awareness specifically receptive to subjective states in other persons" (Trevarthen & Aitken, 2001, p. 4).

Besides Trevarthen, whose work is concentrated upon here, there are also other suggestions of infant intersubjectivity that have been made, perhaps most notably by Meltzoff and Moore (1977, 1998) and Daniel Stern (e.g. 1985), the present section dealing primarily with Trevarthen's work[5].

Primary intersubjectivity can be regarded in any case in different ways. At times, it has seemed to me that the focus in at least some of Trevarthen's work is on how a newborn comes to the world "equipped" more than it concerns the intersubjective relationship *per se* (e.g., Trevarthen, 2010). Yet, in looking at matters more closely, this appears to me to clearly not be the case. Trevarthen's concept of primary intersubjectivity cannot stand on its own without a second

[5] See Beebe et al. (2003), who compare their respective theories on intersubjectivity.

person. Yet, explicit phrases about the infant's "motive" and Trevarthen's focus on neonatal neurobiology have at times overshadowed inherent reciprocity in my reading of him. As this perhaps is also the case in others' reading of the concept of primary intersubjectivity, it might be worthwhile to point that out. There are indeed passages to be found in which Trevarthen highlights the importance of the Other. In an article from 1992 with the (apparently) illustrative title "An infant's motives for speaking and thinking in the culture" he writes:

> 'communicating' with psychologically responsive beings is inherently separate and differently organized in the baby's mind from 'doing' with unperceptive, unthinking, unwilling and unfeeling physical objects. Communication must be motivated by some 'representation' of, or readiness to partake in, *the reciprocity of feeling, acting and perceiving that is possible only with another mind.* (Trevarthen, 1992, pp. 104-105, italics added).

As the italicized sentence in this quote makes clear, intersubjectivity requires there to be two subjects who mutually relate to each other *as subjects*, which basically means that thre simply being an infant possessed of intersubjective capacity (the "motive") does not mean the presence of an intersubjective relationship. For this to be the case requires the presence of a mindful, second subject to engage in the infant. In a similar vein, Beebe et al. (2003, p. 818) note that infants: "in the first few months are not conversational unless appropriate receptive invitations are given by the partner. Mother's expressive behavior is adapted to the multimodal perceptual readiness of the infant and conveys animacy, vitality, and energy." Thus, a dyadic intersubjective relationship requires *involved* rather than *disinterested* subjects (cf. Reddy, 2005). The question is how engaged relationships come about.

Studies have shown that interactive strategies such as those of maintaining attention and being maternally sensitive increase referential communication (i.e., reference to a third object) (Carpenter, Nagell, & Tomasello, 1998) in infants that make use of language. Yet this same process can also have importance in non-referential (i.e. face-to-face, in the second-person) communication that takes place much earlier. The concept of affect attunement – a person's being tuned into the other's affect states (Stern, 1985) – concerns, one might say, being non-referentially engaged with another subject. In a study be Legerstee, Varghese and van Beek (2002), affect attunement was found to promote both gaze monitoring at the age of 3 months and coordinated attention at 5, 7, and 10 months. This meant, they claimed, that the development of these specific communicative exchanges results from both innate motivations to engage psychologically with others and successful scaffolding of the attentional abilities in their infants on the part of the parents (Legerstee, 2005). Legerstee concluded that this supports the idea of parental *participation* being a decisive component in the infant's socio-cognitive development.

Thus, a strong claim for primary intersubjectivity also requires that the other party in communication (i.e. the caregiver) relate to the infant *as a subject* (cf. Paper I). In other words, the parents' motivation to engage in their infants is just as important as is the infant's innate capacity for this. Let us turn now to a theory and research discipline that concerns parents' motivation and intersubjective skills, namely attachment theory.

5.2 Attachment

Fonagy, Bateman and Luyten (2012, p. 4) say that acquisition of the capacity to mentalize "depends on the quality of attachment relationships—particularly, but not exclusively, early attachments, as these reflect the extent to which our subjective experience was adequately mirrored by a trusted other." According to this, it is a goal of their theory to explain how the ability to mentalize develops from attachment through affect mirroring, a process in which the caregivers reflect the infant's primary affect states and thereby assist in regulating the affect they have. The mirrored expression is eventually, to varying degrees, internalized by the infant, who thus gains the ability to regulate and inhibit affect and arousal. The ability to regulate affect and arousal is also seen as being pivotal to the functioning of attachment. Mentalization theory argues, in a very relevant way, that impairments in mentalizing abilities that sometimes appear later on in life are grounded in attachment. However, although Fonagy et al. claim that these abilities are grounded in attachment relations, they reject the possibility of primary intersubjectivity.

According to Bretherton (1992) attachment is the special tie between mother (or some other caregiver) and child, the disruption of which is involved in separations, deprivation and bereavement (cf. Sroufe, Egeland, Carlson & Collins, 2005). Whereas attachment was first seen as a "stress-reducing and safety-promoting" behavioural motivational system (Bretherton, 1992, p. 767), the concept of an *internal working model*, or a mental representation, was then introduced (cf. Main, 1991). Bowlby, the founder of the theory (1969, 1973, 1982), later suggested that the individual's representations of self and of the attachment figure, acquired through interpersonal interaction patterns, are complementary. From an attachment figure that acknowledges the infant's needs for comfort and protection, at the same time as he or she respects the infant's need for autonomous exploration of the environment, the likely representation (working model) of the self is both valued and self-reliant. A caregiver who frequently rejects the infant's requests for comfort and/or exploration most probably affects the working model of the infant's self in the direction of the infant's feeling unworthy or incompetent (Bretherton, 1992). From such representations, that result from interactional patterns, children anticipate the others' likely behaviour and plan their own responses in accordance with these. Importantly, as Bretherton notes, Bowlby argued that "dyadic patterns of relating are more resistant to change than individual patterns because of reciprocal expectancies" (Bretherton, 1992, p. 768).

5.3 Attachment and Primary Intersubjectivity – Two Concepts, Two Functions?

Fonagy et al. (2002, p. 210-220) disagree with the thesis of primary intersubjectivity. In their discussion of it they identify three different positions: strong, weak and 'no-starting-state-intersubjectivism', all three of which they dismiss. Along with Trevarthen, the authors mention Tomasello (e.g. 1999), Stern (1985) and Meltzoff and Moore (e.g. 1998) as proponents of intersubjectivity. Besides a strong and a weak position, they see also a third position that centres on the notion of a "sharing of experiences", a position that at best, they say, can be called "objective intersubjectivity" because

> although the subjective states of parent and infant may become aligned due to emphatic parental mirroring or infantile imitation of parental affect expressions, this does not imply that the young infant is aware of sharing the subjective state with the other, or, for that matter, that the other experiences a subjective state at all. (Fonagy et al., 2002, pp. 219-220).

Without the ambition of sorting this out here, I could note that in order to scrutinize their arguments the terms "aware" and "subjective" need to be defined, and that once this is done the disagreement may have disappeared. However, the view that the sharing of experiences is evidence of primary intersubjectivity means instead, according Fonagy et al., that an "infant's readiness to engage in affective interactions with caregivers function to establish and maintain proximity to the attachment figure" (2002, p. 218).

The critique of the first two positions is partly misguided because they seem to attribute certain Cartesian assumptions to intersubjectivists and it is clear that the intersubjectivists do not share Fonagy et al.'s view of that matter (in papers II and III this issue is discussed more specifically). Fonagy, Gergely and Target (2007, p. 291) state that a "Cartesian view of the mind" has "influenced" the theories of primary intersubjectivity. According to them, this view “presupposes direct introspective access to subjective intentional and emotional mind states”, presumably conceived as internal, opaque states, and it “implies the existence of prewired, universal and subjectively equally accessible intentional and emotional self states in all human individuals”. This means, according to the authors, that primary intersubjectivity “involves a ‘rich’ mentalistic interpretation of the nature of the young baby’s subjective experience of her own as well as of the caregiver’s mind states during the organised patterns of mother-infant interactions from birth” (2007, p. 292). The claim that primary intersubjectivity involves a “rich mentalistic interpretation” is, however, clearly in dispute with how intersubjectivists see things (see paper III).

Concerning their reasons to reject primary intersubjectivity more generally, Fonagy et al. say that “[a]t the heart of this argument is the view that the presence of subjectivity in the infant cannot be assumed but, rather, must be considered as acquired in the process of interaction” (Fonagy et al., 2002, p. 218). If they argue that subjectivity already from the start cannot be “assumed” but must be “acquired in the process”, then it follows that at least a certain amount of time must pass for it to develop, which means that infant intersubjectivity is already impossible for that reason. Thus, primary intersubjectivity would seem by definition to be ruled out. In any case, after making a thorough review of the issue, Fonagy et al. conclude that there is no compelling evidence to support the intersubjectivist notion. The point here is not to review this entire matter again (see Fonagy et al., 2002, pp. 209-222). Instead, the point is to examine the two ideas – primary intersubjectivity and attachment – as two aspects of early infancy and examine what they would “look like” if one were to subject them to a theoretical analysis.

To conclude what has been said thus far, the notion of the “sharing of experiences” and the central place it is given by intersubjectivity theorists is questioned by Fonagy et al.. They ask in what more specific way there is a dyadic sharing of emotions, as “engaging in protoconversational turn-taking is neither a typical nor an effective response when the baby is in need of soothing”

(Fonagy, Gergely, & Target, 2007, p. 293). Here, a typical aspect of interaction is emphasised – that of soothing – an aspect that is likely to be associated with an "attachment perspective". However, although no one denies that infants need soothing, they are not in need, of course, of soothing all the time. They also engage in chatting, teasing and playing, behaviours that tend to be emphasised in research on intersubjectivity (e.g. Reddy, 2000, 2008). Trevarthen and Aitken express their perspective as follows:

> The expressive behaviours in affectionate chat and play have no immediate role in the regulation of the neonate's physiological state, comfort, or survival. They are distinct from maternal breast-feeding, stroking, holding, rocking, vocal comforting, and the like. The caregiver responds to neonatal signals that are very different from appetitive movements, distress cries, or gestural signs of fear anger, anger, or fatigue. (Trevarthen & Aitken, 2001, p. 6).

Instead, the interactions, or the aspects of interaction that Trevarthen calls special attention to, are calmness, enjoyableness, and the dependence on sustained mutual attention and rhythmic synchrony. Thus, the two aspects of behaviour referred to can differ in many ways its being likely, for example, that seeking protection and soothing on the one hand, and affectionate chat, on the other, differ in terms of both positive and negative affects.

Bretherton (2005) highlighted the idea that attachment denotes a social bond, but not just any kind of social bond. Rather, according to attachment theory, a parent (or other caregiver) comes to function as an *attachment figure*, which means that this role (i.e. the attachment) has a special status and is distinct from the role of a *playmate*. According to Bretherton (1985, p. 4), Bowlby actually meant that these two roles were conceptually distinct: "A child is said to seek the attachment figure when under stress but to seek a playmate when in good spirits". As the roles are not incompatible, however, they could be both be filled by one and the same person. Thus, there are reasons for considering these roles as being conceptually distinct.

Cortina and Liotti (2010) argue that attachment and intersubjectivity differ in their evolutionary history and in the motivational functions they serve. Whereas the goal of attachment is protection, intersubjectivity aims at cooperation. These are two aspects of the interaction and communication, of course, that take place in infancy and, even if they are conceptually distinct, they in effect are intertwined. Regarding attachment behaviour, for example, distress cries that are properly responded to by the caregiver, establish the secure haven from which the infant can depart to explore the world. In this exploration, the social world has it its own interests; this being where intersubjectivity comes in.

Thus, when Trevarthen argues that taking part in "affectionate chat" is something quite distinct from attachment security and nutritive needs, these two roles are most likely what he means. Trevarthen and Aitken (2001, p. 7) maintain that "infant survival and development depends on communication with a caregiver to service the baby's needs for an emotional *attachment,* but also to maintain and develop an intimate emotionally expressed *companionship* in changing purposes and conscious experiences" (cf. Trevarthen, 1998b). Thus, the interaction that the infant partakes in can be viewed differently, depending on the perspective. The well-known phrase "in the eye of the beholder" can come to mind here. On the other hand, in paper II it is suggested that intersubjectivity and mentalization are two separate functions, which counters the view of Cortina and Liotti (2010, p. 411), who use the terms intersubjectivity and mentalization interchangeably:

> We think concepts of theory of mind (ToM), metacognitive monitoring, mentalization or mindreading abilities, and advanced intersubjective abilities—that have different research and intellectual origins—are nonetheless all pointing to the same basic phenomena, namely the ability to read the intentions, emotions, and goals of others and the ability to observe and reflect on one's own internal experience.

One could say that intersubjectivity is distinct from explicit mentalization, but this leaves the question of how to understand implicit mentalization unanswered. Mentalization as a concept ought best to have a definition that encompasses, both in implicit and explicit terms, the most important aspects of it.

Primary intersubjectivity and attachment may thus reflect different aspects of interaction in infancy. If this is the case, focusing only one of these would inevitably bias the view of the abilities and capacities that infants possess. If not the existing evidence would be sufficient to make room for both of these approaches, it would be possible to test these hypotheses by examining infant-caregiver interaction under conditions of positive affect (which should predict intersubjectivity measured, for example, as taking turns in what Trevarthen termed 'affectionate chat') and negative affect (which should predict attachment behaviour, such as seeking proximity, comfort, protection)[6]. Accepting the idea that mentalization is an advanced form of intersubjectivity, there are some studies exploring the association between attachment and intersubjectivity (measured by means of theory of mind tasks), but with mixed results (see Meins et al. 2002, 2003; Laranjo, Bernier, Meins & Carlson, 2010). One plausible alternative hypothesis would be that this association could involve *primary* intersubjectivity and thus the opposite trajectory, in which intersubjectivity predicts attachment possibly testable by use of a longitudinal design involving measurements first at, say, the age of four months and then at the age of about a year. Jaffe, Beebe, Feldstein, Crown & Jasnow (2001) did that and found that vocal rhythm coordination at four months predicted attachment at 12 months. Importantly, the contingent interaction with a stranger resulted in a more certain prediction than the infant-mother interaction did. This seems consistent with the distinction between two different roles played by others in interaction.

[6] There would, of course, be many variables to consider. The suggestion here is only meant to provide a sketch of the general idea, not to specify the hypotheses.

To conclude what has been said thus far, we have seen that the two opposing approaches – mentalization and the second-person approach – are both worthy of consideration but for differing reasons. Thus, early interactions can be interpreted from the point of departure of either of these two theories. However, just as there seems to be room for both attachment and intersubjectivity in early infancy, there may be room for accepting both mentalization and non-mentalizing aspects of social understanding. The last suggestion may appear at first glance to provide a somewhat speculative account, but it can be seen as worthwhile and promising when viewed within the larger framework of social understanding. The penultimate section before the papers are introduced that follows now is devoted, accordingly, to providing a brief glance at non-mentalistic ways of understanding.

6. Some Notes on Non-Mentalizing Modes of Social Understanding

This section deals with non-mentalizing modes of social understanding. In section 4 above, mentalizing was defined as an imaginative mental activity that enables us to perceive and interpret human behaviour in terms of intentional mental states (Fonagy, Bateman & Luyten, 2012). As we have seen, there are some who argue that mentalistic understanding is the only way that people can genuinely understand each other. The attempt will be made here to demonstrate that this is certainly too strong a claim.

There are several non-mentalistic explanations suggested besides those that are involved in the second-person approach (primary intersubjectivity, direct social perception, interaction theory). These are sometimes termed "pre-mentalistic", which indicates their existing prior to the development of mentalization. However, this contrasts the view that non-mentalizing approaches such as primary intersubjectivity can exist alongside one's mentalizing ability once it has developed and that expression will therefore be avoided.

Most notably we find here *teleological explanations*, in other words explanations that do not consider mental states but rather more concrete goals and purposes[7]. There is a simple example that illustrates the contrast between a mentalistic explanation and a teleological explanation, one from Csibra and Gergely (1998, p. 255):

Q: *Why did the chicken cross the road?*

A1: *To get to the other side.*

A2: *It wanted to be on the other side.*

It is in A2 alone that a mentalistic causal explanation is provided. A mental state ("wanted") is attributed to the actor there as being a cause of the behaviour involved (crossing the road). Note that the teleological explanation is not less right than the mentalistic one. Everyone understands A1 as a perfectly sensible answer, different though it is from A2. Importantly, the crucial aspect of things is not that these explanations need to be verbalized by a person, but rather that the explanation needs to also be distinguishable at the implicit level. Therefore, the mentalistic claim is that mentalistic explanations are the way most individuals, at least above four years of age, understand each other; a claim that is thus universal (Baron-Cohen, 1995; Tooby & Cosmides, 1995; Wellman, Cross & Watson, 2001).

Mentalization is supposed to involve metarepresentations, or secondary representations according to Fonagy et al. (2002). One important feature of secondary representations is that they can be decoupled from reality. The basic idea is that a primary representation is stimulus-bound and represents objects in the real world; it is experienced as being "for real", whereas a secondary representation is decoupled from reality, meaning that it can be used for

[7] There of course are also other candidates such as behavioural rules (or behavioural abstraction) (see e.g., Gallagher & Povinelli, 2012; Perner, 2011; Ruffman & Perner, 2005) but we leave these aside to focus only on teleology.

exploring alternative means of representation. The object it refers to[8] can then be experienced as being both "not for real" and "nonconsequential" (Fonagy et al. 2002, p. 299). In the case of affect mirroring, the caregiver's mirrored expression permits the infant, for example, in pretend play, to engage in a "corrective emotional 'rewriting' of the negative affect memory by reexperiencing it in the marked 'as-if' mode with a modified emotional content" (Fonagy et al., 2002, p. 299).

Such a model needs to explain how the subject can be in touch with the primary state at the same time as the second-order state is decoupled.

What is needed, therefore, is a theoretical model designed to encompass both mentalizing and non-mentalizing explanations in a way that does not construe the non-mentalizing approaches as being undeveloped and only precursory to achieving a "real" understanding, but rather that can account for the development of, and interaction between, mentalizing and non-mentalizing understanding (i.e. understanding through interaction or attention to external features, be it through external and implicit mentalizing, or primary intersubjectivity). Papers I – III are concerned with this project in ways that differ from one another. Certain ideas appear to be particularly potent in the construction of such a model, although what is provided here will be only a rough sketch. These are ideas concerned with the role of representations in the development of mentalization, issue being taken here with various features of mentalization theory.

[8] There are differing views of how a secondary representation refers to an object in the external world, or to a primary representation. Perner (1991) argues that one should say that it refers to the object, doing so via the primary representation of it, whereas Leslie (1987) stipulates the presence of an "anchoring" relationship, which means that the secondary representation is anchored in "parts of primary representations" (p. 418; cf. Perner, 1991 p. 293). Fonagy et al. refer to Leslie's theory of representation.

Joëlle Proust (2003a; 2007) explicated two senses in which metacognition can be understood and asserted that they are functionally distinct, the one involving metarepresentation and the other being of a more basic kind, one that involves *procedural reflexivity* instead of metarepresentative reflection. According to this view, a metacognitive control system can be precursory to (an explicit) "mentalizing ability, not only because it offers procedural knowledge to a potential redescription mechanism, but also because the resulting enhancement of executive capacities offers the control structure that decoupling requires" (Proust, 2003a, p. 352). In other words, to run a model decoupled from reality means that some other model must also be present that "hooks up" to reality. This move tackles the problem identified above concerning how decoupling can be explained as being critical for the engaged reflection that mentalizing entails.

Proust contrasts 'having a capacity' and the need to 'exercise it'. This leads us to a second important issue. She argues that even if one can make use of mentalizing abilities, the possibility of approaching an issue in non-mental terms by simplifying the problem is always an alternative:

> There are deep forms of engagement available to an agent who has access to metarepresentational thinking, because normally such an agent independently possesses metacognitive capacities. But she does not need to exercise them, in particular when under time pressure, or in routine situations… A subject who already has these mental concepts in her repertoire may replace them by their shallow, non-mental counterparts. (Proust, 2007, p. 287).

Proust compares this with Perner's (1991) situation theory in light of the fact that many of our daily encounters can be understood sufficiently on the basis of situation theory alone.

According to Perner (1991, p. 255) a situation-theory of behaviour is a "'mentalistic theory of behaviour' in which mental states are construed as theoretical constructs…[and] as relations to situations". A situation theory provides a kind of teleological explanation that can be distinguished from a mentalistic one. However, Perner and Roessler (2010) go a step further and argue that the custom in psychology of linking practical reasons with mental states is actually unnecessary. Not that there is anything wrong with using beliefs and desires as reasons, but reasons *need not* be conceived as mental states. As far as I can see, this amounts to a new, and better, situation-theory than the 'old' in which 'mental states are construed as theoretical constructs'. Reasons can, in Perner and Roessler's view, be external, practical and objective – *facts* in other words – and as such get involved in the explanations that also adults, well endowed with mentalizing abilities, in many cases prefer and see as more apt. Moreover, the teleological origin of social understanding (or common-sense psychology) is, according to Perner and Roessler, the best way to explain how mentalization comes about:

> It is not enough to think of certain mental states as the causes of bodily movements. What matters is the ability to see how some of the agent's psychological properties provide her with considerations that from her point of view can be seen to amount to a practical reason. Understanding the subjective reason informing someone's intentional action requires delineating what, from her perspective, presents itself as an objective reason. (2010, pp. 32-33).

According to Perner, non-mentalizing understanding is even more frequent than mentalization stating, "we stay situation theorists at heart. We resort to a representational theory of mind only when we need to" (Perner, 1991, p. 251). Perner and Roessler's view puts more emphasis on this statement. Gallagher (2001) regards Perner's notion of situation theory as being similar to "an embedded cognitive practice that relies on those pre-theoretical embodied capabilities that three-year-olds have already developed to understand intersubjective situations" (2001, p. 95), in other words the capacities involved in primary and secondary intersubjectivity (cf. Gallagher, 2004).

However, whereas teleology is a non-mentalizing mode, it is far from being primary intersubjectivity. Thus, if development starts with primary intersubjectivity, the model needs to explain how primary intersubjectivity connects to teleology (that is, situation theory). Between these capacities there needs to be a bridge of sorts and Proust's model of procedural reflexivity, which links primary cognition with metarepresentation through the intermediate level of procedural reflexivity, may possibly assist in the enterprise of linking primary interaction with teleology.

Recursive metarepresentation, Proust says, is an operation that "relies on the syntactical phenomena of natural languages. It is indeed a universal formal property of human languages that they admit embedded clauses" (Proust, 2007, p. 290). This means that metarepresentation "is *not as such* cognitively demanding. It is implicitly mastered through language use" (Proust, 2007, p. 290). The implication of this is important, namely that metarepresentation can be either deep or shallow.

Metacognition is often defined as a dynamic model of monitoring and control processes, one that passes information on from an object level to a meta level, as well as from meta level to object level (e.g. Nelson & Narens, 1990). There is no truth-functional independence from one another of control and monitoring, which means that "a metacognitive control loop aims at establishing a coherent and reliable picture of the presently available cognitive capacities" (p. 287). Thus, metacognition can never be shallow, but needs instead to always engage both levels in order to operate properly, where "it cannot predict or evaluate without simulating, which means spending significant time and resources on running a dynamic model for the task" (p. 286). Metarepresentations, on the other hand, are generally truth-functionally independent from their embedded representations. The attribution of a belief to a person may be true even if what the person believes is false. Deep metarepresentation thus involves engaging metacognitive simulation, whereas shallow metarepresentation takes on a simpler mode, one that does not involve the full implications of the proposition in question – its not running the simulation so to speak. However, when metarepresentation "redescribes" metacognition, in line with Karmiloff-Smith's suggestion to this effect (1992), metarepresentation "automatically receives a deep reading" (Proust, 2007, p. 287). Proust maintains that "as long as you don't need to evaluate the truth of an embedded content to correctly apply a mental concept to it, you don't need, a fortiori, to test and evaluate your own judging or learning capacity in order to form that metarepresentation." (p. 287).

Compared to what mentalization requires according to the mentalization theory, Proust's distinction certainly points in a different direction. Given that children do learn language and that the semantic function of language in one sense constitutes the ability of "doing" recursive metarepresentation, metarepresentation cannot represent a requirement for (implicit) mentalization. If metarepresentation can be shallow, which mentalization cannot be, then it seems wise to try to define mentalization in some other way than by metarepresentation. What could be more suitable for this than procedural reflexivity? Proust's theory suggests that procedural reflexivity rather than metarepresentation is ample for mentalization. "The difference between an implicit, nonmentalistic form of metacognition and its 'redescribed' or explicit form, is that reflexivity occurs not only at the process level, but also at the semantic-intentional level" (Proust, 2003a, p. 352; 2003b).

To conclude what has been said thus far, non-mentalizing understanding is probably more important than mentalization theory seems to recognize. A study that can illustrate this idea concerns how mentalization is cultivated in different social contexts.

In a study of associations between culture and somatic symptoms, Ryder et al. (2008) used spontaneous problem reports, structured clinical interviews, and symptom questionnaires to investigate how symptoms of depression were presented by nearly 300 outpatients in China and Canada. In line with the hypotheses made, with the results of previous studies and with existing theories, the Chinese outpatients "reported more somatic symptoms on spontaneous problem report and structured clinical interview compared with Euro-Canadians, who in turn reported more psychological symptoms on all 3 methods" (p. 300). Importantly, and also in line with the predictions, both somatic and psychological symptoms were frequently expressed in both places. A relevant pathological construct – Alexithymia– is often assessed by use of the

Toronto Alexithymia Scale (TAS). This questionnaire is used to “measure the tendency to not clearly experience or articulate emotional states, with the negative pole often being associated with psychological mindedness” (Ryder et al., 2008, p. 305). Higher scores on the TAS were observed in the Chinese participants. Yet this effect, they added, was “carried exclusively by *externally oriented thinking*, which does not measure a difficulty but instead measures a tendency to not value inner emotional experience as particularly important” (Ryder et al., 2008, pp. 309-310, italics added). These patients were thus able to both experience and to express their emotions, but they did not focus on them and did not make them central to their life, according to the authors. In contrast, ”Western culture stands out for its unusual emphasis on the personal experience and interpersonal communication of emotion” and the independent self-construal and values of self-expression that is “common to Western cultures emphasize an internal focus in contrast to the external and interpersonal focus found in many other parts of the world” (Ryder et al., 2008, p. 310).

The phrase “externally oriented thinking” refers to a subscale of the TAS. This subscale carried the full effect of Alexithymia as neither the Difficulty Describing Feelings subscale nor the Difficulty Identifying Feelings subscale showed a significant effect, which according to the authors suggests a less pathological interpretation concerning the TAS and Alexithymia. According to Ryder et al., research literature on the subject has tended to construe somatisation in China as being a culture-bound phenomenon rather than considering that psychologization is unusually common in the West.

In light of what has been said about mentalization above and the distinction to be drawn between mentalistic and teleological explanations, these are interesting results, externally vs. internally oriented thinking being involved. It is tempting to view these findings in the light of the distinction between teleological and mentalistic explanation. Apparently, a culture characterized by externally oriented thinking would emphasise the stating of one’s purpose and one’s motives in teleological terms, as these focus on concrete matters better than terms for mental, private states do. One should recall that both Proust and

Perner specifically asserted that a non-mentalistic understanding of many of our encounters is often sufficient to act in an appropriate way in the world generally. The situation-theory of behaviour, in Perner's view, represents a teleological understanding that amounts, in Proust's words, to "using the world as the model from which to predict behavior, instead of using another subject's representation of the world" (Proust, 2007, p. 287). This is tremendously interesting given that externally oriented thinking does precisely that – uses the world instead of others' mental ideas about it, as the guiding model.

What does it mean to say that a "culture is characterized by externally oriented thinking" or for that matter, characterized by mentalization? One way to conceive of this is to say that teleology has a firm grip on the common-sense psychology in these contexts. One question that is crucial to the perspective here concerns the relationship between early intersubjectivity and teleology. One possibility is that teleology concerns the reasons that people provide for their actions. Early intersubjectivity is not concerned with reasons in that sense, but it may constitute the ground on which teleology can develop. Before a child can understand objective facts in the world in relation to other people, they need to understand more basic things such as movement, attention, communication, etc. – capacities that are involved in primary intersubjectivity. When these matters become comprehensible, practical reasons can become involved.

Another way of conceiving of it is suggested in Brinck (2008). Brinck claims that two types of intersubjectivity develop in parallel in the first year, interattentionality and interaffectivity (cf. Stern, 1985). The one concerns the sharing of attention, the other the sharing of emotion (affect). These are independent, but in typically developing subjects interact and enable a gradual development of social understanding in the first few years in life. In case one of them malfunctions, this will cause specific impairments. Problems with interaffectivity for example, are likely to cause difficulties in social orienting and attention to distress.

A likely scenario is that narratives are involved in the development of reason-type explanations and understanding. Daniel Hutto (e.g. 2008) argues that our folk-psychological understanding of others, that is, the social understanding that is shared by members of a specific culture, society, or group – is more or less forced upon us by the narratives that pervade in that group, culture or society. A narrative in this sense is not only verbally introduced to us in stories and tales, but also exists in notions, figures of thought and the kind of embodied interaction that members of a social environment engage in. It is likely indeed that not only how people talk about how they understand things, but also how they think they understand them, is of central importance here. Is it the way they in fact do understand things? If we turn the question the other way around, what reasons have we to believe that people's actual understanding is in perfect harmony with how they think or say they understand things? Also, is it conceivable that we understand things in some other way than we ourselves believe we are understanding them?

7. Methodological Considerations

A doctoral thesis in psychology usually has at its core one or more empirical investigations that are reported on in detail. In one sense this dissertation is no exception to this. Yet the label theoretical psychology is also quite warranted in the present case, in view of the fact that the studies included here will only refer to other empirical studies, carried out by others and reported on by them earlier.

The section begins with a review of different perspectives in theoretical psychology, historical as well as current ones, that are applicable to the studies that are presented in the next section. However, before moving on to the next section, the methods employed in these studies are presented in light of the theoretical psychology that has been reviewed.

7.1 Theoretical Psychology

Theoretical psychology can concern fundamental issues as well as conceptual and theoretical ones. It analyses, discusses and compares different explanatory models and paradigms and their ontological and metaphysical assumptions, their scope and their value. It also analyses fundamental issues pertaining to large areas of research, such as social cognition, for example, and analyses and criticizes the way questions are asked within the area or areas in question. Some of the questions that have occupied me while writing this dissertation are the following: What is social understanding? How do we ask questions about the development of social understanding, and how else could we do so? What is there to gain in reformulating these questions and changing the ways in which we conceive of young children's understanding of the mind?

The methods used in theoretical psychology are of course conceptual and theoretical. They can consist in part of conceptual analyses, for revealing and eliminating misconceptions, for fine-tuning existing concepts, for developing new concepts and conceptions that can highlight new factors and new aspects of known phenomena. Conceptual analysis is not merely a matter of analysing what is already known, but can also lead to new knowledge through re-interpretation of old facts and the generation of new hypotheses. Theoretical psychology also involves analysing existing data, often published by other researchers, to suggest new interpretations, or to put the data into new contexts and relate them to other data, the relationship which has not thus far been investigated, and possibly also to place them in a new theoretical context. This is a way of generating new and testable hypotheses. What theoretical psychology achieves is supposed to be fed back into empirical psychology.

Over the years, there have been many suggestions of how best to conceive of the place for theory within psychology. At times, one can even get the impression that not much has happened as regards theory within psychology. For instance, in 1940 Kurt Lewin wrote a paper entitled "Formalization and progress in psychology" in which he stated that in

> recent years there has been a very marked change in the attitude of American Psychology. During the 1920's and early 1930's psychologists were, on the whole, rather adverse to theory… Today, a definite interest in psychological theory has emerged… The need for a closer fusion of the various branches of psychology demands tools which permit better integration (Lewin, 1951, p. 1).

The concerns about integration in an otherwise fragmented discipline were shared by Slife and Williams about half a century later (1997, p. 117) in their maintaining that "[i]ncreased signs of disciplinary fragmentation as well as threats to mainstream psychology's philosophy of science have presented challenges that call for thoughtful disciplinary discussion" which is (partly) why they suggest theoretical psychology as a subdiscipline that can facilitate such a discussion. Slife and Williams also refer to Koch's (1959) history of the development of psychology in which it is argued that psychology settled on its way of answering questions, that is, its methods, *before* the theoretical questions were developed. In the natural sciences, problems and question tended to come first, they say, and methods to come as a response to theoretical problems. In scientific psychology, thanks to a determination to apply positivist methods to humans, those questions that could generate empirically testable hypotheses tended to be valued. However, Slife and Williams assert that "there is certainly no logical reason why theories should generate empirically testable hypotheses. The only reason is a privileging of method in general and positivistic method in particular." (1997, p. 119). The recognition that empirical psychology emphasises method rather than theory tallies with my own experience. Yet, whether it also means that the custom of giving priority to method results from 20th century positivism remains an open question as far as I can see.

What other reasons are there for pursuing theoretical psychology? Let us review some differing answers to this question. We can start with André Kukla (1989, 1995, 2001), whose suggestion has much in common in many respects with the "five major tasks" for theoretical psychologists that Koch (1951) suggested in an influential paper on the subject.

Kukla proposed that theoretical psychology be a specialized branch of psychology, "a theoretical wing comparable to the well-established theoretical disciplines that exist in other scientific disciplines… an active sub-discipline with a well-articulated research program and a growing corpus of special methods and results." (1989, p. 785). He views psychology as "the most aggressively empiricist of all academic disciplines" (p. 785), which he finds ironic as psychology constitutes a "particularly promising arena for the exercise of nonempirical methods of research – far more so than, say, geology or bacteriology" (p. 785).

Kukla suggests several tasks for the theoretical psychologist, *theory construction* being the first important task, as he sees it, in light of his view that data "do not yield up theories of themselves, nor will theories emerge by adding more data to the lot. There is no alternative but to invent a theory" (p. 785). A second task, as he sees it, is to *derive new empirical predictions* from existing theories, whereas confirming the predictions he considers to be a job belonging to empirical psychology. Third, *coherence analysis* tests whether a theory is internally consistent, that is, whether it contains internal inconsistencies such as a proposition "P" and its negation "not-P" both being presented as correct, or any form of inconsistent consequence that is false already on conceptual grounds. A theory can involve "circularity, infinite regress, ambiguity, non sequitur, or nonindependence among its fundamental assumptions" (p. 787), in that case the theory being in need of correction. The fourth task, *conceptual analysis,* can thus test a theory's internal coherence to ensure that the observations and hypotheses are not formulated in a way that makes them necessarily true, and that the concepts involved are distinct. Studies may even set out to test hypotheses that logically are necessarily true, in that case what is tested being anything *but* the hypothesis, but rather simply the procedures employed in the study, for example (cf. Smedslund, 2002). Fifth, *conceptual innovations* can lead to new models that can be tested. When Kukla discusses the place of *a priori* propositions in psychology, his disappointment with empiricism turns into a rationalistic credo in which he understands rationalism as the view that "there is some a priori knowledge", whereas empiricism is identified as the view that "only necessary propositions can be known a priori" (p. 791).

To explain further what he means by this he refers to Kuhn (e.g. 1962) and Lakatos (e.g. 1970), who both suggest that "scientific theories inevitably contain some propositions that are too basic to be submitted to empirical test, even though these propositions are not logically necessary truths" (p. 791).

Examples of such presuppositions are contingent beliefs on a priori grounds that one needs to adopt in order to be able to engage in scientific work at all and thus that underlie all scientific theories. Kukla takes the example of a disagreement between behaviourists and phenomenologists on how to interpret introspective reports. Where a behaviourist would maintain that 'a subject S reports *about* an experience E', the phenomenologist would say that 'subject S *experiences* E'. Such a disagreement cannot be settled by yet another experiment, Kukla says, as both parties "would systematically interpret the results of any experiment in accordance with their own methodological percepts" (p. 792). These presuppositions are rarely declared explicitly in "the beginning of a scientific enterprise" (p. 792); one has rather to work backwards from what is actually said and done to "the system of presuppositions that seems to warrant these practices" (p. 792).

While this example concerns a strong conception of a priori propositions, Kukla also postulates a "pragmatic a priori", which means that a priori constraints on scientific theories "are largely cognitive and social in nature" (p. 792). The difference as compared with more strict a priori convictions is that, regarding pragmatic ones, "we are free to reject the presuppositions of any particular theory" (p. 792) and we can of course search for alternatives, but "the pragmatic a priori imposes constraints that must be adhered to by all conceivable theories in psychology. The presuppositions of particular theories can neither be confirmed nor disconfirmed by experiment" (p. 792), thus pointing out the necessity of theoretical work.

In a second article, Kukla (1995) advances his suggestions by adding two further tasks for the theorist: amplification and simplification. According to Kukla, these have the following properties: "(1) between them, they account for most of the research time of theoreticians; and (2) neither type of work requires the theoretician to be acquainted with the data relating to the theories." (p. 202). The first point, although difficult to understand, probably denotes the fact, as described earlier in the paper, that scientists in disciplines such as theoretical physics, for example, spend most of their time working with formalized theories.

Amplification is a subcategory of what Kukla terms "coherence analysis". It involves assigning a probability value to a scientific theory by analyzing both its relation to other, comparable theories and its empirical consequences, "i.e. all the empirically testable propositions that follow from the theory, whether or not they have already been tested" (p. 202). This probability value is taken to describe the epistemic standing of the theory at a specific time. Amplification consists of "performing operations that increase or decrease the probabilities assigned to extant theories" (p. 203), its being "a logical, as opposed to empirical, discovery about a theory *T* as a result of which *p(T)*, the probability assigned to T prior to the discovery, has to be changed" (p. 202-203).

Simplification, in turn, concerns basically the constructing of new theories, but by means of simplifying earlier ones. Demands for simplicity and parsimony can be seen in such notions as those of *Occam's razor* and *Morgan's canon*.[9]

In this second article, Kukla's approach appears to concern formalized work more than his previous paper did (1989), his acknowledging too that, whereas a theory is most likely to be constructed in close contact with data, both amplification and simplification are theoretical activities autonomous of and distinct from empirical work.

There is of course much to say regarding Kukla's suggestions. First of all, his intention to present theoretical psychology as a highly worthwhile enterprise is certainly welcome. The very fact that theoretical work is emphasised to this

[9] "In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale" (Morgan, 1894, cited in Kukla, 1995, p. 210). See Brinck (2009) for a discussion of Morgan's Canon.

degree can undoubtedly be helpful for psychological researchers generally. His idea that the theoretical psychologist has specific tools and methods to work with is clearly valuable as well. However, criticism has been directed against his view, particularly that of its construing theorizing as being an autonomous activity, distinct from empirical work (Looren De Jong, 2010; Looren De Jong, Bem & Schouten, 2004)[10]. I will review some of these criticisms here with the aim of presenting a broader view than one might otherwise take of theoretical psychology. The section ends with certain conclusions before going on to the more specific methodological concerns of the thesis.

Looren de Jong, Bem and Schouten (2004), think the main problem regarding Kukla's view is his assuming there to be a neat distinction between observation and theory, one they feel is untenable. According to them, theorizing in psychology does not depend upon there being a strict dichotomy between theory and observation. Instead, they argue that the armchair analysis Kukla considers to be the core of theoretical psychology is "of limited use" (p. 276). For example, Kukla's suggestion is that the theoretical psychologist propose new conceptual innovations without use of any new data, yet Looren De Jong (2010) argues that such innovations "more probably… result from a mix of new concepts and new observations" (p. 749). My interpretation of Kukla's suggestion is that the very intention of proposing new concepts is most likely *preceded by* one's *taking part of* observations or data, not that one needs to present new data simultaneous to presenting one's conceptual innovations. It is possible, however, that contrary to my reading of him, Looren De Jong means that one should also present new data to support one's innovations. As far as I can see, Kukla does not oppose the interpretation I make, since he also

[10] Both Looren De Jong (2010) and Looren De Jong, Bem and Schouten (2004) are primarily concerned with Kukla's book "Methods of Theoretical Psychology" (2001), but in (2004) they note that the main text of Kukla's book is basically identical to his early articles (Kukla 1989, 1995) a fact they consider "unfortunate" (p. 277).

acknowledges that, when we are engaged in the innovative development of new concepts while sitting in our armchair "we rely… on *prior* observation" (1989, p. 790).

Looren De Jong, Bem and Schouten (2004) maintain that Kukla's suggestions are valuable as general recommendations to theorists that they be aware of the importance of concepts, presuppositions and perhaps also of a priori truths. Yet as sole guidelines for doing theoretical psychological work their suggestions risk making theoretical psychology too distinct a discipline, that is, a discipline that empirical psychologists feel they need not take account of.

I consider this critique to be more relevant to Kukla (1995) than to Kukla (1989), where in the former he outlines ways of assigning probability values to existing theories without dealing with data.

The naturalist suggestion of Looren De Jong, Bem and Schouten (2004) can be seen as fostering a "continuous interplay between empirical results and theoretical reflection", their feeling that the theoretical psychologist should "reflect on empirical data, integrate results, and uncover presuppositions and historical roots of empirical and theoretical work in psychology" (p. 276). Since both reflecting on empirical data and integrating results are probably involved in most researchers' activities, their suggestion puts theoretical and empirical psychologists on an equal footing regarding this. Highlighting historical roots, however, is valuable. It pertains in particular to theoretical psychology and is perhaps something that Kukla does not particularly stress.

Looren De Jong's (2010, p. 750) main worry seems to be that theorizing "does not have to be an autonomous activity, in the sense of being separated from empirical work, and concerned with formal techniques" in order to be of interest. He offers a naturalistic position that considers the distinction between theory and observation to be less distinct.

According to the various authors just referred to, naturalism involves a different view of knowledge and science, their being considered to be "about the world and have their effects in the world" (Looren De Jong, Bem &

Schouten, 2004, p. 284). In other words, since they take naturalism, pragmatism and functionalism to denote more or less the same ideas regarding this, they adopt a pragmatist view in which science is seen to not search for knowledge “for its own sake” but for pragmatic reasons, i.e. for being able to change things in the world.

An important thing about naturalism in this context is that methodological statements there are considered to be equivalent to empirical statements in the sense that they are an object for scrutiny and thus are accepted, pragmatically speaking, if they work out well, that is, if they turn out to be effective methods of providing us with new knowledge. This poses a problem for the naturalistic view, a problem we will return to below. Now, we will attempt to broaden the view taken beyond what we have considered thus far, examining various other tasks besides those Kukla took up.

Looren De Jong (2010) argues that the various perspectives theoretical psychology considers can be regarded as a *continuum* ranging from *constructive to deconstructive* approaches, each position on that continuum being at least potentially valuable. The naturalistic approach adopts a position in the middle, emphasising theoretical psychology’s proximity to empirical psychology. Specifically, he argues that this contrasts with Kukla’s view, which is at the construction-end of the scale, social constructionism instead being situated at the deconstructive end of the continuum.

Although Kukla’s views surely contribute to a *constructive* approach to theorizing and he criticizes empiricism, Looren De Jong believes that Kukla is not that keen on criticizing theories. This is, a problematic statement, however, because Kukla’s methods of amplification concern the evaluation of theories, and if one finds a theory wanting, this is arguably a way of criticizing it. Yet there can be differences in terms of how one criticises a theory, or what point of departure for doing so one selects. Whereas Kukla’s manner of evaluating a theory can be regarded as an inside-perspective, one that helps science to move forward, there is also an outside-perspective that can be taken, one that “deconstructs” theories instead.

Criticising theories is emphasised in the *social constructionist paradigm* which, like Kukla's view, constitutes a response to the demise of logical positivism. Whereas positivism was concerned with determining objective and context-free criteria for scientific rationality and with the justification of epistemological claims, social constructionists reject objectivism and suggest that knowledge, as well as truth, justification and reality, be deconstructed as products of social negotiation.

As an example of the deconstructive approach, Looren De Jong refers to social constructionists[11] who argue that psychological knowledge does not mirror mental or behavioural facts, but rather that psychological concepts are produced in social interactions, so that psychological reality is "in essence negotiable" (p. 754). Theories are seen as performative in the sense of their constituting realities, or forms of life, the way Wittgensteinian tradition has it, their not, as the positivists believed, being linguistic structures that mirror reality. All seemingly objective statements of facts can (and should) be deconstructed and unmasked as products of negotiation, power, and manipulation, just as Kuhn unmasked the inherent sociality of scientific paradigms.

Although Looren De Jong argues that naturalism is a middle-of-the-road position, there are extreme positions as well that he suspects can turn into "naturalistic fundamentalism", views that consider "whatever scientists do" to represent "the plain truth" (p. 753), this illustrating a major problem in naturalism. The problem is this: if no independent criteria for sufficiency and justification are to be found, but only those practiced in empirical work, it is necessary to ask what will happen to "critical reflection upon presuppositions, scope, generality, and adequacy of knowledge claims" (p. 753), in other words

[11] He exemplifies with Gergen as a social constructionists, but mentions also Harré and Searle, whose theories differ from Gergen's on several significant points.

those qualities that both Kukla and social constructionists highlight. This basically means that there is a need for theorizing as an autonomous activity as well, to put it bluntly. Thus, not even this side of Kukla's view can be dismissed.

Looren De Jong concludes his view on social constructionism as being more or less the mirror image of exaggerated naturalism, whereas the latter seems "to accept anything from mainstream Nobel Prize neuroscience, social constructionism seems to accept just anything new and enriching" because of their lacking criteria for legitimacy and justification. Although Looren De Jong appears to appreciate the need of criticizing theories, he seems much less ready to appreciate the way in which social constructionism can contribute. He believes that deconstruction as such can indeed be destructive, meaning that it is not clear what it leads to, despite its being claimed to lead to "emancipation". He argues that without independent criteria "there is no principled way of telling where the emancipatory direction is. Unmasking traditional taken-for-granted psychology as a construction may leave nothing behind, and yield nothing to replace it" (p. 757).

A more reasonable view then is as Looren de Jong (2010, p. 758) states, a naturalism that "[consider itself a continuation of science, and in principle accepts entrenched theories—but only provisionally, for the time being, like 'piles driven into a swamp'". This statement also summarizes the view of Looren De Jong and his colleagues, what could be considered to represent a mild version of naturalism, one that seeks to avoid the dangers of a too naïve naturalism.

Summing up this section, one can note that theoretical psychology can be many things, and can involve quite differing perspectives on both psychology and the philosophy of science and can sometimes involve ideological issues. The approach taken in the thesis can be seen as being situated rather close to a naturalistic approach, above all in the sense of regarding there to be no absolute distinction between theoretical and empirical psychology. However, Kukla has pointed to various methods the theoretical psychologist can employ that can be very useful in research, regardless of whether one conceives of these as being autonomous activities or not. We will return to these in describing the methods employed in the thesis.

More generally speaking, whether one works entirely with theoretical or with empirical studies is seen as representing no basic difference in terms of the philosophy of science. This is not meant to deny vast differences being possible in the perspective one can take, but is simply to assert that no basic differences are *necessary*. Of course, a theoretical study differs from an empirical study in many practical ways far too obvious to be in need of any specific comment.

7.2 The Theoretical Methods Employed

The aim of the thesis has been to examine different aspects of the role of intersubjectivity for metacognitive development and social understanding. More specifically, the thesis investigates how, within different theoretical frameworks, specifically mentalization theory, the theory of primary intersubjectivity, and interaction theory, the developmental role of intersubjectivity is described, the suggestions these theories make being evaluated. Common to all three of these theories that the dissertation takes up is the conviction that intersubjectivity is central to, and strongly affects, social and cognitive development from the very beginning of life.

In the first paper, the hypothesis that metacognitive abilities and skills start to develop between 2 and 4 months of age in episodes of dyadic interaction is investigated. To examine the basis for this hypothesis we analyse the concept of metacognition, there being several alternative constructions for this, developed more or less independently of one another represented in the literature. By analogy with an extended concept of cognition, resulting from tests of implicit cognitive capacities associated with emotion, motivation and attention rather than with verbal competence, inferential reasoning and symbol processing, we find there to be support for an extended concept of metacognition. This could be said to involve, at least to a certain extent, a conceptual innovation in Kukla's terms in its conceiving of metacognition in new terms, ones consistent with, and opening it up for, our hypothesis. The analysis carried out concerns to a considerable extent examining what the specific conceptual construction presupposes, what it implies and how this compares with alternative conceptions that are possible. We then define three aspects of metacognition, one of which is the traditional one, resting on metarepresentation, the other two depending upon an extended conception of metacognition. Metarepresentational metacognition is frequently described in the literature by explaining how this aspect of metacognition relates to the other two, but we could also show how metacognition can be grounded in infancy. Dynamic systems theory permit us to conceive of metacognition in a way making it plausible to employ in a social, interpersonal context. In this way, through analysing the concept and highlighting its social functions, we also find – in examining the history of the concept of metacognition – that the social functions involved appear already in the original definition of it by Flavell, its thus being firmly grounded within the framework of existing research. We construe in this way a theoretical framework in which the hypothesis can be seen as credible, reasonable and justifiable.

Finally, in examining relevant empirical evidence, we argue that the hypothesis is clearly supported by existing evidence. In addition, we provide a description of the concept making it appear promising to operationalize and to put to test.

The second paper examines the hypothesis that two aspects of social understanding – those of primary intersubjectivity and of mentalization – can be considered as complementary, the latter depending upon the former. The methods employed involve first of all a critical scrutinization of existing theories regarding this matter. Two such theories are opposed to one another in terms of which aspects of social understanding are regarded as being primary, whereas a third theory agrees with the proposition that the other two theories should be considered as being complementary to one another, although it construes the developmental trajectory as being the opposite of that conceived of here. The respective theories involve differing presuppositions in regard to certain central issues, such as what the understanding of emotional expressions implies, and what functions an adequate theoretical model of social understanding should be expected to explain. Critically analysing these presuppositions shows that the matter to be of theoretical rather than empirical character, an interpretation being provided of what the evidence obtained regarding infants' understanding of emotional expressions implies.

By further analysing the central claims and the concepts involved in the respective theories, showing them to agree on certain central points, despite their presuppositions differing, it is argued that the two theories closely related to one another can be considered as complementary in the sense of their describing two equally important functions of social understanding which both need to be accounted for.

A review of the empirical evidence supported the hypothesis suggested that they be considered as complementary. In Kukla's terms, the main part of the study can be said to consist of a conceptual and a coherence analysis of the theories and of their respective explanations of social understanding, as well as of the presuppositions involved. Analysed in these terms, the result of the study suggest that social understanding and the separate functions it performs should be included in one and the same model, not only showing there to be two separate functions, each in their own right, but also suggesting how they interact.

The third study involved primarily a coherence analysis of the mentalization theory, subjecting it to a critical analysis of how the conception of mentalization has changed from the early formulations of the theory to the most recent form it has been given. In addition, the theory is contrasted with the theory of primary intersubjectivity, the latter making certain inconsistencies in the former obvious. On the basis of a conceptual analysis, it is argued that one of the mentalization theory's most central concepts, that of implicit mentalization, lacks a proper explanation in terms of that theory, a lack which is particularly salient in view of the centrality of this concept. Since current empirical studies of infants' sensitivity to false beliefs, as indicated by implicit measures of it, speak in favour of an implicit sense of mentalization, it can be seen as highly important to define and explain this concept more adequately. Interestingly enough, the concept of implicit mentalization overlaps with the concept of primary intersubjectivity, making it important indeed to decide between the two theories. The mentalization theory denies the existence of primary intersubjectivity, yet this denial, can be shown on both conceptual and theoretical grounds, to reflect a misunderstanding of what primary intersubjectivity presupposes. We argue, accordingly, in contradiction of both theories, that primary intersubjectivity is compatible with the theory of mentalization.

8. Overview of the Papers

I. The Developmental Origin of Metacognition

The point of departure for the first paper was the fact that although a number of theoretical approaches argue that intersubjectivity constitutes the developmental foundation for awareness of other minds, for socio-cognitive development, and for social understanding, the view that intersubjectivity also plays a role regarding infants' gaining an understanding of their own minds has not been explored to the same extent, despite the fact that many persons working within this basic areas appear to recognize the fact of the conceptions of the self and of the other going hand-in-hand during development (cf. the dimension of self/other mentalizing taken up above).

Our approach permits conceiving of metacognitive development as parallel to cognitive development from early on, and for re-interpreting some of the achievements of young infants as genuinely metacognitive. It allows the formulation of new explanatory hypotheses about cognitive development in the first two years, and is intended to stimulate the development of new experimental paradigms for investigating metacognition in the form of epistemic (inter)action between infant and adult within shared contexts.

We claim that metacognition has its developmental origin in certain features of early intersubjectivity that permit infants to internalize and construct rudimentary strategies for manipulating their own and others' cognitions. More specifically, we argue that metacognitive skills start to develop between 2 and 4months of age in episodes of dyadic interaction. The argument takes its starting-point in the conception of metacognition as the management of cognition. We adopt Kirsh's (2005) framework for operationalizing metacognition that pictures metacognition as the management of cognitive resources internally in the mind and externally in the task environment. Following Kirsh and Maglio (1994), we distinguish between two kinds of action: Pragmatic actions are needed to perform a task or solve a problem and move the agent closer to the goal. They are physical actions that change the task environment. Epistemic actions are used to search for a solution to the problem or select a strategy or procedure to perform the task.

We maintain that when pragmatic, world-directed actions cannot reduce the agent's distance to the goal, there is a need for epistemic actions, directed at cognition. A monitoring mechanism will alert the system that its overall cognitive state is inadequate for reaching the goal. Once the deficiency is identified, the control mechanism will implement a strategy for improving performance, for instance, to re-organize available information, search for new information, or activate memory. Thus, learning how to deal with threats such as breakdown and inefficiency of communication fosters learning of metacognitive strategies. Crucially, intersubjectivity provides the infant with the necessary motivation for this kind of active learning.

We distinguish between three types of metacognition (cf. Brinck, 2006): *implicit* metacognition, which concerns the monitoring and control of hierarchical cognitive processes in activities that require purely causal strategies for reaching goals or completing tasks; *perceptual* metacognition, which requires emotional and attention-based strategies, and *metarepresentational* metacognition, which involves higher-order propositional or symbolic strategies.

Three features make intersubjectivity apt for initiating early metacognitive development: First, shared monitoring and control of cognition are integral to it; second, it enables learning and training of actions that realize monitoring and control functions; and third, feedback is immediate.

We maintain that intersubjectivity allows infants to internalize and construct rudimentary strategies for the dynamic monitoring and controlling of their own and others' cognitions in real time—procedurally, on an implicit level, as well as by means of emotions and attention. The functions of initiating, maintaining, and achieving turns make proto-conversation a productive platform for developing metacognition. The caregiver and the infant can jointly create shared routines for epistemic actions that facilitate learning of metacognitive skills. The adult thus comments on and corrects the infant's efforts, as well as representing a cognitive resource in its own right for the infant.

We cite empirical evidence from a variety of sources that visual attention and facial expression of emotion are the principal means for sharing experiences around the interaction as well as for regulating it. Attention and emotion constitute important behavioural indicators of metacognition in both infants and adults, and observations of gaze-related behaviour and facial, bodily, and vocal emotion expression reveal the manner in which the infant engages with the adult—pragmatic or epistemic.

That metacognition begins externally and later can be internalized by the individual does not mean that metacognition will end up internalized. Rather, metacognition continues to be inherently social but with the possibility to be exercised as the more traditional way of conceiving metacognition has it; as an individual operation in problem-solving and monitoring of internal cognition.

II. From Primary Intersubjectivity to Mentalization: On the Development of Social Understanding

The second paper poses the question of how primary engagement and interaction relate to the development of social understanding, and more specifically of mentalization. Many discussions of this question end up defending either 'interactionism' (primary intersubjectivity, direct social perception, interaction theory) or mentalization. One interesting example is that of Michael (2011) suggesting that interactionism and mentalization should be seen as complementary. However, his suggestion puts mentalization first, and his arguing that interaction depends on some form of (implicit) mentalizing ability. Instead, this second paper examines the hypothesis that intersubjectivity and mentalization are complementary, in a manner such that the latter depends on the former, but not the other way around.

In connection with a detailed comparison of primary intersubjectivity and mentalization theory, the suggestion is made that, despite their obvious difference in perspective and in point of departure, they have a great deal in common, both of them calling attention to the role of affect and emotion in social understanding, and their both treating theory of mind as being too narrow a concept, although they differ markedly in what are regarded as the reasons for this.

The point of this article is very simply to maintain that if one accepts both infants being potentially capable of intersubjectivity and their possessing capacities for achieving it, and their developing the ability to mentalize in pretty much the way that Fonagy et al. suggest, then one should explore the possibility, at least, of the two of them being combined, and more specifically endeavor to explain how mentalization can develop from primary intersubjectivity.

III. Mentalization, Intersubjectivity and Affect Mirroring: A Critical Discussion of Some Aspects of the Development of Mentalization

The third study takes as its point of departure the concept of mentalization as it is construed by Fonagy and his associates. Mentalization theory provides a novel perspective for obtaining an understanding of psychopathology, psychotherapy and child development. The study investigates the suggestion that mentalization develops from affect mirroring and contingency detection in attachment relations and considers specifically four separate dimensions of mentalization (cognitive versus affective, implicit versus explicit, self-oriented versus other-oriented, and externally versus internally focused). We claim that the construal of mentalization in terms of these four dimensions prepare for a comparison between mentalization and primary intersubjectivity. We suggest that external and implicit mentalization lie closer to primary intersubjectivity than one might think, and argue that there is a partial overlap between the concepts of primary intersubjectivity and implicit (and external) mentalization.

Consequently we conclude that mentalization theory can be combined with the hypothesis about primary intersubjectivity despite Fonagy et al.'s explicit dismissal of this hypothesis—and that bringing the two theories together contributes to explain the developmental roots of social understanding. More specifically, we submit that mentalization is seen as originally developing within the context of primary intersubjectivity, and suggest that primary intersubjectivity is a prerequisite for the (normal) development of mentalization.

9. General Discussion

The way in which we allow ourselves
to engage with others circumscribes
the way in which we can know them

Vasudevi Reddy (2008, p. 6)

We mentalize when we treat others
as persons rather than objects

Jon G. Allen (2003, p. 93)

Human beings understand one another
intimately and at many levels

Colwyn Trevarthen (1979, p. 321)

At the outset of the thesis, the aim was stated as that of examining the relation between interactions in infancy and metacognitive development, specifically as regards social understanding both of the self and of others. This aim evolved on the basis of a variety of considerations, a study by Clements and Perner (1994) suggesting the first and most perceptible, if only approximate seeming measure of the vague notion of this that had entered my mind. This vague notion involved the distinction, in types of understanding involved, between understanding things “with the body” and understanding them with the mind or through thinking. There was, of course, a long way yet to go in transforming such notions into researchable questions. Yet the distinction between primary intersubjectivity and full-blown mentalization (or theory of mind) finally became the focus of the studies undertaken. Examining these different theoretical perspectives, with their differing assumptions and presuppositions, led to two quite differing views on the kind of understanding that exists alongside more conscious reflections and explanations, as well as earlier, in infancy, before they develop. It seemed difficult to explain from these two theoretical positions, how these differing capacities – mentalizing and “embodied pragmatics” – can interact in one and the same person. The challenge was to show how these quite differing aspects of social understanding interact.

In paper I, we argue specifically that metacognition develops from intersubjective interactions starting at an age of between 2 and 4 months. The common assumption that the development of understanding of the self and of others come in tandem is further elaborated on is suggesting that implicit and perceptual metacognition is intrinsically social.

Paper II develops the view that primary intersubjectivity and interaction can contribute to the development of mentalization, their providing a more complete model of social understanding. Such a perspective contributes in particular to theories emphasising social interaction (as broadly construed) through its being argued that mentalization theory has important contributions to make, even if some of the ideas and presuppositions involved are clearly in conflict with how things are conceived in terms of interactionism.

Paper III makes a similar suggestion, only this time from the perspective of mentalization theory and what primary intersubjectivity can add to this theory. The two studies carried last thus approach the combining of mentalizing theories and interactionism from two directions, arguing that since the phenomena of primary intersubjectivity and mentalization that are suggested to be important here appear to exist, a model of social understanding ought not only to include the phenomena involved but also to explain how they interact.

The view defended in the three papers is that social understanding entails various abilities and capacities and that these can be combined, and ought to be, within a single theoretical model. These range from primary intersubjective interactions in which affective and intentional attitudes are communicated and regulated through mutual vocal, facial and bodily expressions, to the more mature conscious reflection of oneself and others' reasons, motives, feelings, ambitions, actions and what these express symbolically, unconsciously, and objectively, the perhaps most qualified ability being that of "mentalized affectivity" (Jurist, 2005). Thus, a theoretical model ought ideally to include the full range of these abilities and explain how they interact. The most promising theory in this respect was seen to be mentalization theory (Fonagy et al., 2012). The varying set of implications for development and for clinical work that are accounted for by it is impressive. The aim of uncovering the associations between attachment relations in infancy and metacognitive development appeared to be a promising one of very considerable potential. However, as has been argued in the papers, the theory also has important shortcomings as regards how metacognitive development is imagined to start. Furthermore, it tends to downplay the importance of non-mentalizing modes of understanding. As the study by Ryder et al. (2008) indicates (section 6), mentalization – or the tendency to value inner, mental experiences rather than externally oriented ones – seems not to be the predominant and universal way that people obtain an understanding of the reasons for things (cf. Hutto, 2008; Perner, 1991; Perner & Roessler, 2010), despite claims to the contrary by theory of mind proponents.

In paper III, it is argued, however, that the most significant point is perhaps that the mentalization theory falls short of explaining its most critical notion, namely that of implicit mentalizing (and to a certain degree external mentalizing as well). This is understandable, given that research on children's and infants' knowledge and awareness of others has developed from explicit measures such as obtained on the basis of answering questions, or of elicited response tasks, to implicit measures, or spontaneous response tasks, that tend to focus on where subjects direct their gaze, as in connection with an anticipatory-looking or a violation-of-expectations paradigm (see e.g., Baillargeon, Scott & He, 2010; Low & Perner, 2012; Thoermer et al., 2012). However, even if the implicit sense of mentalizing is considered the most important aspect of mentalization, it is not sufficiently explained in the theory. The time is thus ripe to concentrate on an explication of implicit mentalizing, or better still to complement a dimensional view with clear definitions of each polarity that can be submitted to an empirical test as regards their respective developmental history. There are already some suggestions that have been made on how to conceptualize early forms of mentalizing, notably implicit forms, reported on in the literature (e.g., Fogel, 2011; Shai & Belsky, 2011), though more are indeed needed.

In papers II and III alike, Fonagy et al.'s rejection of primary intersubjectivity was found to not rest on solid ground. It appears that Fonagy et al. (e.g., 2002) misunderstand what primary intersubjectivity entails. At the very least, the proponents of primary intersubjectivity do not consider their proposal mentalistic. As they explicitly dismiss the very idea of mentalization and mentalism, it would appear strange if they simultaneously argued that primary intersubjectivity implies such a reading.

On the other hand, proponents of non-mentalizing alternatives to social understanding tend to downplay mentalization and to sometimes reject its being vital for social understanding (e.g. Reddy, 2008; Gallagher, 2005).

In the second paper, a similar suggestion is presented, that of combining the second-person approach generally, and primary intersubjectivity specifically, with mentalization. Here, the focus is on emotional engagement. Emotional aspects of social understanding tend to be neglected in the theory of mind literature and, in response to this, both the second-person approach and mentalization theory have highlighted its significance for social understanding and the development of it. One critical topic as regards emotion concerns the question of how to interpret emotional expressions. What does it take to "understand" an emotion? A closer analysis revealed that the two theories arrive at very different interpretations regarding this issue. Fonagy et al. assume a cognitivist reading in which understanding an emotion implies the need of a second-order awareness of the emotional state, present in infants when they cannot yet understand their own emotions but instead have to learn through the process of affect mirroring. Trevarthen and other primary intersubjectivists, on the other hand, claim that emotional expressions are intrinsically meaningful. Trevarthen (1998b) argues that the distinction between affect, which the infant is granted as having, and emotion, which involves cognition/appraisal or other interpretations that come slightly later in development, is a remnant from behaviourism, in terms of which the infant starts with only reacting to stimuli and is granted no activity on its own. Trevarthen argues that infants instead are biologically prepared to take part in emotional communication: "Like colour vision, emotions in communication are differentiated by experience, but the fundamental values and contrasts were there from start" (Trevarthen, 1998b, p. 271). Fonagy et al. (2002) deny that emotion in infants entail conscious emotional experience and claim there to be no evidence to support such an assumption and claim further, with Gergely and Watson (1999), that the question of what is felt and is not felt by infants is probably not empirically resolvable. This is a good example of an important issue being in need of theoretical arguments. The most obvious objection to it is the lack of evidence to support the conclusion that infants who display emotional expressions such as those of adults do not at the same time experience "emotion-specific conscious feeling states" (Fonagy et al., 2002, p. 150).

Another suggestion in paper II is that intersubjectivity and mentalization should be considered as being separate functions. In section 5 above, the suggestion made by Cortina and Liotti (2010), that of attachment and intersubjectivity serving separate functions was discussed. Cortina and Liotti use the terms intersubjectivity and mentalizing to denote one and the same function, mentalization thus amounting to intersubjectivity. Fonagy et al. (2002) also think that mentalization is intersubjectivity, this being why they emphasise its being a developmental achievement, and as such its being in opposition to innate intersubjectivity. At first appearance, this seems to involve a contradiction. Yet the clue to its being otherwise lies in the concept of implicit mentalizing. Note that the broader notion of mentalization entails a notion of implicit mentalizing which in paper III is claimed to overlap with primary intersubjectivity. In accordance with this, intersubjectivity starts in its primary form and then gradually develops into what is known as mentalization. The suggestion that they are separate functions would mean that explicit mentalizing—defined initially as involving conceptual and metarepresentative elements, reflections upon inner, covert mental states in the self and in others, and the subject's conceiving mental states as separate from behaviour—is distinct from intersubjectivity; intersubjectivity in the sense of an embodied, ongoing understanding in interaction. Or, couched in Allen's description, we "mentalize others implicitly when perceiving and responding to their emotional states: we automatically mirror them to some degree, adjusting our posture, facial expression, and vocal tone in the process" (p. 10). When we mentalize implicitly, he says, "we do so intuitively, procedurally, automatically, and non-consciously" (p. 10). Presumably, what Proust (2007) calls procedural reflexivity is involved in this intersubjective, intuitive process of making sense in and of social interaction. The remark by Fonagy et al. that mentalizing is "predominantly implicit", as well as the dimensional view in general, harmonises with conceiving of intersubjectivity as a gradual process, (explicit) mentalization constituting a "special case" of intersubjectivity. It qualifies as a special case because of its abstracting mental states from behaviour in a way that is distinct from how intersubjectivity works otherwise. Whether we describe intersubjectivity as gradual, and mentalization as being a special case of it, or say that intersubjectivity and mentalization are separate, is worthy of a moment's contemplation. The general idea nonetheless seems clear in both cases.

A second alternative as to how to construe mentalization (and intersubjectivity) is the suggestion made in section 4 above that the concept of mentalization can be conceived of as being either a descriptive or a normative concept. The descriptive approach, according to the suggestion made here, permits the claim that mentalization is not developmentally primary but is a developmental achievement, as the mentalization theory implies. As a normative concept, mentalization appears strongly recommendable because of the assumed benefits of reflecting on the mental states, beliefs, desires, and emotions that may have served as reasons for others' actions, also in terms of self-mentalizing. The point is that the descriptive and the normative claims can be kept apart and be dealt with in turn. One can, of course, embrace the one and reject the other.

Paper I explores metacognition as being involved in social interaction from the age of 2 months and onwards, not ultimately aimed at internalisation. In accordance with the dimensional view of mentalization, metacognition (ranging from implicit and external to explicit and internal self-mentalizing in Fonagy et al's terminology, and from implicit and perceptual, to metarepresentative metacognition in ours) is not to be conceived as a phenomenon in which internal, metarepresentative functioning is the end-state of development, or the gem in the great chain of becoming. Rather, metacognition is inherently social, most often implicit, procedural and automatic, as the dimensional view on self-mentalizing would allow for.

As we have seen from the idea of theory of mind as an answer to the question of how people's common-sense social understanding functions, an idea that is still generally accepted, and taking mentalization theory as well as interactionist theories seriously, strongly suggests that the notion of social understanding should be broadened. Although the suggestions of the two theories vary and they emphasise different matters and even at times explicitly dismiss each others' theories, there nevertheless are ways to overcome these conflicts. What can be gained by such an enterprise is to arrive at a model that comes closer to the phenomena they aim at explaining. Human social understanding is hardly an easy notion to deal with; it has also to employ the very object of study to begin with. Kelly's (1963) notion of man-the-scientist is also true turned backwards – when scientists engage their whole arsenal of social understanding functions, yet in accordance with the profession risk "overly intellectualizing" social understanding.

If the phonological merge between a perspective (a point of view) and the meaning and significance of the second person can be excused, we can conclude that by altering the point of you, a variety of different aspects of social understanding will be highlighted. In other words, as the quote at the beginning of this section has it, the manner in which the relation between the subject and object of understanding proceeds circumscribes the way in which we come to know the other person. How the relationship proceeds depends upon inner, mental structures in both persons, as well as the context in which they are situated.

References

Allen, J. G. (2003). Mentalizing. *Bulletin of Menninger Clinic,* 67, 87-108.

Allen, J. G. (2006). Mentalizing in practice. In J. G. Allen & P. Fonagy (Eds.), *Handbook of Mentalization-Based Treatment* (pp. 3-30). West Sussex: Jon Wiley & Sons Ltd.

Allen, J. G. & Fonagy, P. (Eds.), (2006). *Handbook of Mentalization-Based Treatment.* West Sussex: Jon Wiley & Sons Ltd.

Astington, J. W. (2001). The future of Theory-of-Mind research: understanding motivational states, the role of language, and real-world consequences. *Child Development, 72(3)*,685-687.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants, *Trends in Cognitive Science, 14(3),* 110-118.

Baron-Cohen, S. (1995) *Mindblindness: An Essay on Autism and Theory of Mind.* Cambridge, MA: Bradford, MIT-Press.

Barresi, J. & Moore, C. (1996). Intentional relations and social understanding. *Behavioral and Brain Sciences, 19*, 107-154.

Bateman, A. & Fonagy, P. (2004). *Psychotherapy for Borderline Personality Disorder: Mentalization-based Treatment.* Oxford: Oxford University Press.

Bayley, N. (1993). *Bayley Scales of Infant Development* (2nd ed.). San Antonio, TX: Psychological Corporation.

Beebe, B., Knoblauch, S., Rustin, J. & Sorter, D. (2003). A comparison of Meltzoff, Trevarthen, and Stern. *Psychoanalytic Dialogues, 13(6),* 809-836.

Benett, J. (1978). Some remarks about concepts. *Behavioral and Brain Sciences, 1, 557-560.*

Bowlby, J. (1969). *Attachment and loss (Vol. 1): Attachment.* New York: Basic books.

Bowlby, J. (1973). *Attachment and loss (Vol. 2): Separation, Anxiety, and Anger*. New York: Basic Books.

Bowlby, J. (1982). *Attachment and Loss (Vol. 3): Loss, Sadness, and Depression*. London: Hogarth Press and Institute of Psycho-Analysis.

Bretherton, I. (1985). Attachment Theory: Retrospect and Prospect. *Monographs of the Society for Research in Child Development, 50*(1/2), 3-35.

Bretherton, I. (1992). The origins of attachment theory: John Bowlby and Mary Ainsworth. *Developmental Psychology, 28(5)*, 759-775.

Bretherton, I. (2005). In pursuit of the internal working model construct and its relevance to attachment relationships. In K. E. Grossmann, K. Grossmann & E. Waters (Eds.), *Attachment from infancy to adulthood. The major longitudinal studies.* New York: Guilford Press.

Brinck, I. (2001). Attention and the evolution of intentional communication. *Pragmatics and Cognition, 9(2)*, 255-272.

Brinck, I. (2006). Attention-based metacognition. Presentation. *ESF Exploratory Workshop*: *Metacognition and Mental State Monitoring*. Paris, France.

Brinck, I. (2008). The role of intersubjectivity for the development of intentional communication. In J. Zlatev, T. Racine, C. Sinha, & E. Itkonen (Eds.), *The Shared Mind: Perspectives on Intersubjectivity* (pp. 115-140). Amsterdam: John Benjamins.

Brinck, I. (2009). From similarity to uniqueness: Method and theory in comparative psychology. In L. S. Röska-Hardy & E. M. Neumann-Held (Eds.), *Learning from animals? Examining the Nature of Human Uniqueness* (pp. 155-170). London: Psychology Press,

Brinck, I. & Gärdenfors, P. (2003). Co-operation and communication in apes and humans. *Mind and Language, 18(5),* 484-501.

Buber, M. (1958). *I and Thou* (R. G. Smith, Trans. 2nd ed.). Edinburgh: T. & T. Clark.

Carpendale, J. I. M. & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences, 27*, 79-151.

Carpendale, J. I. M. & Lewis, C. (2006). *How Children Develop Social Understanding.* Oxford, UK: Blackwell.

Carpenter, M., Nagell, K. & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63(4),* i-vi, 1-143.

Carruthers, P. (2009). How we know our own minds: the relationship between mindreading and metacognition. *Behavioral and Brain Sciences, 32*, 121-182.

Choi-Kain, L. W. & Gunderson, J. G. (2008) Mentalization: ontogeny, assessment, and application in the treatment of Borderline personality disorder. *American Journal of Psychiatry, 165(9),* 1127-1135.

Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis,* 58(1), 7-19.

Clements, W. & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9,* 377-395.

Cortina, M. & Liotti, G. (2010). Attachment is about safety and protection, intersubjectivity is about sharing and social understanding. *Psychoanalytic Psychology*, 27(4), 410-441.

Csibra, G. & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science, 1*(2), 255-259.

Davies, M. & Stone, T. (Eds.). (1995a). *Folk Psychology. The Theory of Mind debate.* Oxford, UK: Blackwell Publishers Ltd.

Davies, M. & Stone, T. (Eds.). (1995b). *Mental Simulation. Evaluations and Applications.* Oxford, UK: Blackwell Publishers Ltd.

De Bruin, L., Strijbos, D. & Slors, M. (2011). Early social cognition: alternatives to implicit mindreading. *Review of Philosophy and Psychology, 2*, 499-517.

Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences, 1,* 568-570.

Dennett, D. C. (1987). *The Intentional Stance.* MIT-Press/Bradford Books.

Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology, 50*, 21-45.

Fogel, A. (2011).Embodied awareness: neither implicit nor explicit, and not necessarily nonverbal. *Child Development Perspectives, 5(3),* 183-186. DOI: 10.1111/j.1750-8606.2011.00177.x

Fonagy, P. & Bateman, A. (Eds.) (2012). *Handbook of Mentalizing in Mental Health Practice.* Arlington, VA: American Psychiatric Publishing, Inc..

Fonagy, P., Bateman, A. & Luyten, P. (2012). Introduction and overview. In: A. Bateman & P. Fonagy (Eds.), *Handbook of Mentalizing in Mental Health Practice* (pp. 3-41). Arlington, VA: American Psychiatric Publishing, Inc..

Fonagy, P., Gergely, G., Jurist, E. L. & Target, M. (2002). *Affect regulation, mentalization, and the development of the self.* London: Karnac.

Fonagy, P., Gergely, G. & Target, M. (2007). The parent-infant dyad and the construction of the subjective self. *Journal of Child Psychology and Psychiatry, 48*(3/4), 288-328.

Fonagy, P. & Luyten, P. (2009). A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Development and Psychopathology, 21*, 1355-1381.

Fonagy, P. & Target, M. (1997). Attachment and reflective function: Their role in self-organization. *Development and Psychopathology, 9*, 679-700.

Fonagy, P. & Target, M. (2005). Bridging the transmission gap: An end to an important mystery of attachment research? *Attachment & Human Development, 7*(3), 333-343.

Gallagher, S. (2001). The practice of mind. Theory, simulation, or primary interaction? *Journal of Consciousness Studies, 8*(5-7), 83-108.

Gallagher, S. (2004). Understanding Interpersonal Problems in Autism: Interaction Theory as An Alternative to Theory of Mind. *Philosophy, Psychiatry, and Psychology 11*(3), 199-217.

Gallagher, S. (2005). *How the Body Shapes the Mind.* New York: Oxford University Press.

Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition,* 17(2), 535–543.

Gallagher, S. & Hutto, D. D. (2006). Primary interaction and narrative practice. In J. Zlatev, T. Racine, C. Sinha & E. Itkonen (Eds.), *The Shared Mind: Perspectives on Intersubjectivity*. Amsterdam: John Benjamins.

Gallagher, S. & Meltzoff, A. (1996). The earliest sense of self and others: Merleau-Ponty and recent developmental studies. *Philosophical Psychology, 9(2),* 211-233.

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

Gallagher, S. & Povinelli, D. J. (2012). Enactive and behavioral abstraction accounts of social understanding in chimpanzees, infants, and adults. *Review of Philosophy and Psychology.* DOI 10.1007/s13164-012-0093-4

Garnham, W. & Ruffman, T. (2001). Doesn't see, doesn't know: is anticipatory looking really related to understanding of belief? *Developmental Science, 4(1)*, 94-100.

Goldman, A. (1989). Interpretation psychologized. *Mind and Language, 4*, 161-185. (Reprinted 1995 in M. Davies & T. Stone (Eds.) Folk psychology. The theory of mind debate. Oxford: Blackwell Publishers.).

Goldman, A. (2006). *Simulating minds: The philosophy, psychology and neuroscience of mindreading.* Oxford, England: Oxford University Press.

Gomez, J. C. (1996). Second person intentional relations and the evolution of social understanding. *Behavioral and Brain Sciences, 19*(1), 129-130.

Gordon, R. M. (1986). Folk psychology as simulation. *Mind and Language, 1*, 158-171. (Reprinted 1995 in Davies & Stone (Eds.) Folk psychology. The theory of mind debate. Oxford: Blackwell Publishers.

Grienenberger, J., Kelly, K. & Slade, A. (2005). Maternal reflective functioning, mother–infant affective communication, and infant attachment: Exploring the link between mental states and observed caregiving behavior in the intergenerational transmission of attachment. *Attachment & Human Development, 7*(3), 299-311.

Harman, G. (1978). Studying the chimpanzee's theory of mind. *Behavioral and Brain Sciences*, 1, 515-526.

Hauser, M. D. (1996). *The Evolution of Communication.* Cambridge, MA: Bradford book, MIT-Press.

Heal, J. (2005). Joint attention and understanding the mind. In N. Eilan, C. Hoerl, T. McCormack & J. Roessler (Eds.), *Joint Attention: Communication and Other Minds. Issues in Philosophy and Psychology,* (pp. 34-44). Oxford: Clarendon Press.

Hobson, R. P. (1993). The emotional origins of social understanding. *Philosophical Psychology,* 6(3), 227-249.

Hutto, D.D. (2008). *Folk Psychological Narratives: The Sociocultural Basis of Understanding Reasons.* Cambridge, MA: MIT-Press.

Izard, C. (2009). Emotion theory and research: Highlights, unanswered questions, and emerging issues. *Annual Review of Psychology, 60,* 1-25.

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L. & Jasnow, M. D. (2001). Rhythms of dialogue in infancy: coordinated timing in development. *Monographs of the Society for Research in Child Development, 66(2)*, i-viii+1-132.

Jurist, E. (2005). Mentalized affectivity. *Psychoanalytic Psychology, 22(3),* 426-444.

Kahneman, D. & Tversky, A. (1979). Prospect theory: an analysis of decision making under risk. *Econometrica, 47,* 263-291.

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science.* Cambridge, MA: MIT Press.

Kelly, G. A. (1963). *A Theory of Personality: The psychology of personal constructs.* New York: Norton.

Koch, S. (1951). Theoretical psychology, 1950: an overview. *Psychological Review, 58(4),* 295-301. doi:10.1037/h0055768

Koch, S. (1959). *Psychology: A study of science* (Vols. 1-3). New York: McGraw-Hill.

Koriat, A. (2000). The feeling of knowing: some metatheoretical implications for consciousness and control. *Consciousness and Cognition, 9*, 149-171.

Kovács, A. M., Téglás, E. & Endress, A. D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science, 330*, 1830-1834.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Kukla, A. (1989). Nonempirical issues in psychology. *American Psychologist*, 44(5), 785-794.

Kukla, A. (1995). Amplification and simplification as modes of theoretical analysis in psychology. *New Ideas in Psychology, 13(3),* 201-217.

Kukla, A. (2001). *Methods of Theoretical Psychology.* Cambridge, MA: MIT-Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the Growth of Knowledge* (pp. 91-196). Cambridge: Cambridge University Press.

Laranjo, J., Bernier, A., Meins, E. & Carlson, S. M. (2010). Early manifestations of children's theory of mind: the roles of maternal mind-mindedness and infant security of attachment. *Infancy, 15(3)*, 300-323.

Legerstee, M. (2005). *Infants' sense of other people. Precursors to a Theory of Mind.* Cambridge: Cambridge University Press.

Legerstee, M., Varghese, J. & van Beek, Y. (2002). Effects of maintaining and redirecting infant attention on the production of referential communication in infants with and without Down syndrome *Journal of Child Language, 29*(1), 23-48.

Leslie, A. M. (1994). ToMM, ToBy, and agency: Core architechture and domain specificity. In: L. Hirschfeld and S. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 119-148). New York: Cambridge University Press.

Lewin, K. (1951). Formalization and progress in psychology. In D. Cartwright (Ed.), *Field Theory in Social Science* (pp. 1-29). New York: Harper & Brothers. (Reprinted from *University of Iowa Studies in Child Welfare,* 1940, 16(3), 9-42.)

Lieberman,M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology*, 58, 259–289.

Looren De Jong, H. (2010). From theory construction to deconstruction. The many modalities of theorizing in psychology. *Theory & Psychology, 20(6)*, 745-763.

Looren De Jong, H., Bem, S. & Schouten, M. (2004). Theory in psychology: a review essay of Andre Kukla's *Methods of theoretical psychology. Philosophical Psychology, 17(2)*, 275-295.

Low, J. & Perner, J. (2012). Implicit and explicit theory of mind: State of the art. Editorial. *British Journal of Developmental Psychology, 30,* 1-13. DOI:10.1111/j.2044-835X.2011.02074.x

Luyten, P., Fonagy, P., Lowyck, B. & Vermote, R. (2012). Assessment of Mentalizing. In A.Bateman, & P. Fonagy (Eds.), *Handbook of Mentalizing in Mental Health Practice* (pp. 3-41). Arlington, VA: American Psychiatric Publishing.

Main, M. (1991). Metacognitive knowledge, metacognitive monitoring, and singular (coherent) vs. multiple (incoherent) models of attachment: Findings and directions for future research. In C. M. Parkes, J. Stevenson-Hinde, & P. Marris

(Eds.), *Attachment across the life cycle* (pp. 127–159). London: Tavistock-Routledge.

Malle, B. F. (2005). Folk theory of mind: conceptual foundations of human social cognition. In J. S. Uleman, R. R. Hassin & J. A. Bargh (Eds.), *The new unconscious* (pp. 225-255). Oxford: Oxford University Press.

Meins, E., Fernyhough, C., Russell, J., & Clark-Carter, D. (1998). Security of attachment as a predictor of symbolic and mentalising abilities: a longitudinal study. *Social Development, 7*(1), 1-24.

Meins, E., Fernyhough, C., Wainwright, R., Das Gupta, M., Fradley, E., & Tuckey, M. (2002). Maternal Mind-Mindedness and attachment security as predictors of theory of mind understanding. *Child Development, 73*(6), 1715-1726.

Meins, E., Fernyhough, C., Wainwright, R., Clark-Carter, D., Das Gupta, M., Fradley, E., & Tuckey, M. (2003). Pathways to understanding mind: construct validity and predictive validity of maternal mind-mindedness. *Child Development, 74(4),* 1194-1211.

Meltzoff, A. N. & Moore, K. M. (1998). Infant intersubjectivity: broadening the dialogue to include imitation, identity and intention. In S. Bråten (Ed.), *Intersubjective Communication and Emotion in Early Ontogeny* (pp. 47-62). Cambridge, UK.: Cambridge University Press.

Michael, J. (2011). Interactionism and Mindreading. *Review of Philosophy and Psychology* 2, 559-578.

Nelson, T. O., & Narens, L. (1990). Metamemory: A Theoretical Framework and New Findings. In A. E. Graesser & G. H. Bower (Eds.), *The Psychology of Learning and Motivation: Inferences in Text Comprehension* (Vol. 25, pp. 125-173). New York: Academic Press.

Nöe, A. (2004). *Action in Perception*. Cambridge, MA: MIT-Press.

Onishi, K. H. & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255-258.

Perner, J. (1991). *Understanding the Representational Mind.* Cambridge, MA: MIT-Press.

Perner, J. (1996). Simulation as explicitation of predication-implicit knowledge about the mind: arguments for a simulation-theory mix. In P. Carruthers & P. Smith

(Eds.), *Theories of theories of mind* (pp. 90-104). Cambridge, UK: Cambridge University Press.

Perner, J. (2010). Who took the cog out of cog*nitive* science? Mentalism in an era of anti-cognitivism. In P. A. Frensch & R. Schwarzer (Eds.), *Cognition and Neuropsychology: International Perspectives on Psychological Science* (Vol. 1, pp. 241-261). Hove, UK: Psychology Press.

Perner, J. & Roessler, J. (2010). Teleology and causal understanding in children's theory of mind. In J. Aguilar & A. A. Buckareff (Eds.), *Causing Human Action: New Perspectives on the Causal Theory of Action* (pp. 199-228). Cambridge, MA: Bradford Book, MIT-Press.

Peterson, C. & Slaughter, V. (2003). Opening windows into the mind: mothers' preferences for mental state explanations and children's theory of mind. *Cognitive Development, 18*, 399-429.

Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition*, 36, 1-16.

Premack, D. & Woodruff, G. (1978). Does the Chimpanzee have a Theory of Mind? *Behavioral and Brain Sciences, 4*, 515-526.

Proust, J. (2003a). Does metacognition necessarily involve metarepresentation? *Behavioral and Brain Sciences, 26*(3), 352.

Proust, J. (2003b). Thinking of oneself as the same. *Consciousness and Cognition, 12*, 495-509.

Proust, J. (2006). Rationality and metacognition in non-human animals. In S. Hurley, & M. Nudds (Eds.), Rational animals? (pp. 247–274). Oxford: Oxford University Press.

Proust, J. (2007). Metacognition and metarepresentation: is a self-directed theory of mind a precondition for metacognition? *Synthese, 159*(2), 271-295.

Reddy, V. (1991). Playing with others' expectations: teasing, joking and mucking about in the first year. In A. Whiten (Ed.), *Natural Theories of Mind: Evolution, Development and Simulation of Everyday Mindreading* (pp. 143-158). Cambridge, MA: Blackwell.

Reddy, V. (1996). Omitting the second person in social understanding. *Behavioral and Brain Sciences, 19*(1), 140-141.

Reddy, V. (2000). Coyness in early infancy. *Developmental Science, 3*, 186-192.

Reddy, V. (2001). Infant clowns: the interpersonal creation of humour in infancy. *Enfance, 3*, 247-256.

Reddy, V. (2003). On being an object of attention: Implications for self-other-consciousness. *Trends in Cognitive Sciences, 7(9)*, 397-402.

Reddy, V. (2005). Before the 'Third element': Understanding Attention to Self. In N. Eilan, C. Hoerl, T. McCormack & J. Roessler (Eds.), *Joint Attention: Communication and Other Minds. Issues in Philosophy and Psychology* (pp. 85-109). Oxford: Oxford University Press.

Reddy, V. (2008). *How Infants Know Minds*. Cambridge, MA: Harvard University Press.

Reddy, V. (2010). Engaging Minds in the First Year: The Developing Awareness of Attention and Intention. In J.G. Bremner & T.D. Wachs (Eds.), *The Wiley-Blackwell Handbook of Infant Development* (2nd ed., vol. 1, pp. 365-393). Wiley-Blackwell.

Reddy, V. & Morris, P. (2004). Participants don't need theories. Knowing minds in engagement. *Theory & Psychology, 14*(4), 647-665.

Ruffman, T., Garnham, W., Import, A. & Connolly, D. (2001). Does Eye Gaze Indicate Implicit Knowledge of False Belief? Charting Transitions in Knowledge. *Journal of experimental child psychology, 80*, 201-224.

Ruffman, T. & Perner, J. (2005). Do infants really understand false belief? *Trends in Cognitive Sciences, 9(10),* 462-463.

Ryder, A. G., Yang, J., Zhu, X., Yao, S., Yi, J., Heine, S. J., et al. (2008). The cultural shaping of depression: Somatic symptoms in China, psychological symptoms in North America? *Journal of Abnormal Psychology, 117*(2), 300-313.

Satpute, A. B. & Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research*, 1079, 86–97.

Savage, L. (1954). *The Foundations of Statistics*. New York: Wiley.

Seligman, S. (2000). Clinical implications of current attachment theory. *Journal of the American Psychoanalytic Association,* 48, 1189-1194.

Shai, D. & Belsky, J. (2011). When words just won’t do: introducing parental embodied mentalizing. *Child Development Perspectives, 5(3),* 173-180. DOI: 10.1111/j.1750-8606.2011.00181.x

Slade, A. (2005). Parental reflective functioning: An introduction. *Attachment & Human Development, 7*(3), 269-281.

Slife, B. D. & Williams, R. N. (1997). Toward a theoretical psychology. Should a subdiscipline be formally recognized? *American Psychologist, 52(2),* 117-129.

Smedslund, J. (2002). From hypothesis-testing psychology to procedure-testing psychologic. *Review of General Psychology, 6(1),* 51-72.

Sroufe, L. A., Egeland, B., Carlson, E., & Collins, W. A. (2005). Placing early attachment experiences in developmental context. In K. E. Grossmann, K. Grossmann, & E. Waters (Eds.), *Attachment from Infancy to Adulthood: The Major Longitudinal Studies* (pp. 48-70). New York: Guilford Press.

Stern, D. N. (1985). *The interpersonal world of the infant: A view from psychoanalysis and developmental psychology.* New York: Basic Books.

Strawson, P. (1985). *Skepticism and naturalism: Some varieties.* New York: Columbia University Press.

Thoermer, C., Sodian, B., Vuori, M., Perst, H. & Kristen, S. (2012). Continuity from an implicit to an explicit understanding of false belief from infancy to preschool age. *British Journal of Developmental Psychology, 30,* 172-187. DOI:10.1111/j.2044-835X.2011.02067.x

Tomasello, M. (1999). *The Cultural Origins of Human Cognition.* Cambridge, MA: Harvard University Press.

Tooby, J. & Cosmides, L. (1995). Foreword. In S. Baron-Cohen*, Mindblindness: An Essay on Autism and Theory of Mind* (pp. xi–xviii). Cambridge, MA: MIT Press.

Trevarthen, C. (1974). Conversations with a two-month old. *New Scientist, 62*(896), 230-235.

Trevarthen, C. (1979). Communication and cooperation in early infancy. In M. Bullowa (Ed.), *Before speech: The beginnings of human communication* (pp. 321-347). Cambridge, MA: Cambridge University Press.

Trevarthen, C. (1992). An infant´s motive for speaking and thinking in the culture. In A. Heen Wold (Ed.), *The Dialogical Alternative, Towards a Theory of Language and Mind.* Oslo: Scandinavian University Press.

Trevarthen, C. (1998a). The concept and foundation of infant intersubjectivity. In S. Bråten (Ed.), *Intersubjective Communication and Emotion in Early Ontogeny* (pp. 15-46). Cambridge, UK: Cambridge University Press.

Trevarthen, C. (1998b). Explaining emotions in attachment. [Review of the book *Emotional Development: The Organisation of Emotional Life in the Early Years*, by A. Sroufe]. *Social Development, 7(2),* 269-272.

Trevarthen, C. & Aitken, K. J. (2001). Infant intersubjectivity: research, theory and clinical applications. *Journal of Child Psychology and Psychiatry, 42*(1), 3-48.

Trevarthen, C., Aitken, K., Papoudi, D. & Robarts, J. (1998). *Children with Autism: diagnosis an interventions to meet their needs* (2nd ed.). London: Jessica Kingsley Publishers.

Trevarthen, C. & Hubley, P. (1978). Secondary intersubjectivity: Confidence, confiding, and acts of meaning in the first year. In A. Lock (Ed.), *Action, Gesture, and Symbol: The Emergence of Language* (pp. 183-229). New York: Academic Press.

von Neumann, J. & Morgenstern, O. (1944). *Theory of Games and Economic Behaviour.* Princeton, NJ: Princeton University Press.

Wellman, H. M., Cross, D. & Watson, J. (2001). Meta-analysis of Theory-of-Mind development: The truth about false-belief. *Child development, 72*(3), 655-684.

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103-128.

Wittgenstein, L. (1980). *Remarks on the philosophy of psychology* (vol. 2, G. H. von Wright & H. Nyman, eds.). Oxford, UK: Basil Blackwell.

von Wright, G. H. (1971). *Explanation and Understanding.* London: Routledge & Kegan, Paul.