Mentalizing

Competence and process

Clara Möller

Linköping Studies in Arts and Sciences No. 756
Linköping Studies in Behavioural Science No. 211
Faculty of Arts and Sciences

Linköping 2018

Linköping Studies in Arts and Sciences • No. 756
Linköping Studies in Behavioural Science • No. 211

At the Faculty of Arts and Sciences at Linköping University, research and doctoral studies are carried out within broad problem areas. Research is organized in interdisciplinary research environments and doctoral studies mainly in graduate schools. Jointly, they publish the series Linköping Studies in arts and Science. This thesis comes from Division of Psychology at the Department of Behavioural Sciences and Learning.

Distributed by:
Department of Behavioural Sciences and Learning
Linköping University
SE-581 83 Linköping

Clara Möller
Mentalizing
Competence and process

Edition 1:1
ISBN 978-91-7685-189-0
ISSN 0282-9800
ISSN 1654-2029

© Clara Möller
Department of Behavioral Sciences and Learning 2018

Cover image: © The M. C. Escher Company

Printed by: LiU-Tryck, Linköping 2018

ABSTRACT

The concept of mentalizing refers to the process of making sense of oneself and others in terms of mental states, such as thoughts, feelings, wishes, desires and needs. Mentalizing theory as developed by Fonagy and his colleagues has influenced attachment theory by placing the emphasis of attachment development on parents' ability to mentalize their children's minds through the process of marked mirroring (letting the child know in an empathic way that her or his feelings and reactions are comprehensible, valid and possible to regulate). Mentalizing is both an explicit willful action and an implicit nonverbal automatic process.

The golden standard of measurement methods in mentalizing research is the Reflective Functioning (RF) scale, which is usually applied to Adult Attachment Interviews (AAI) and aimed at capturing the level of explicit mentalizing that interviewees express about their attachment relationships. This thesis aimed to explore the concept of mentalizing and its operationalization RF in different contexts. The overarching question guiding the thesis concerns what RF means when it comes to actual human interactions and behavior? To study this question, specific interview guides were developed as well as scoring procedures for therapeutic interactions in order to capture how the mentalizing capacity is manifested when reflecting upon current or recent activity as opposed to reflecting upon attachment relationships from one's childhood. In the first study young criminal offenders incarcerated for a range of different crimes were interviewed. Findings were that their ability to mentalize was significantly lower than the theoretical and/or empirical average, both regarding their attachment relationships and regarding their crimes, suggesting that mentalizing might be relevant as a buffer against the commitment of criminal acts. The results raised questions concerning how the mentalizing ability relates to actual behavior in the prison setting, and what promotes or hampers mentalizing processes in the prison environment. In the second study the relationship between mothers' ability to mentalize about their children and how that related to their emotional availability in interaction with their children was tested. The mothers were also interviewed about limit setting situations and these interviews were rated for RF. The limit setting RF score was a stronger predictor of parental emotional availability than was the general ability to mentalize about the child. The results indicate that mentalizing ability might not be a global capacity but rather a set of skills or, possibly, more context- and relationship-dependent than previously thought. This part of the thesis was directed at capturing the applicability of RF in difficult situations of adulthood and approached the concept of mentalizing as a trait capacity.

In the second part of the thesis, mentalizing was approached as a state that could be seen in, and is dependent upon interactions, by investigating the process of mentalizing established in therapy sessions. In Study III, data came from sessions of mentalization-based therapy (MBT) for patients diagnosed with borderline personality disorder and comorbid substance abuse. The RF scale was applied to relatively unstructured therapy sessions and every statement from the patient was rated for RF. Therapist statements were categorized as either demanding mentalization or not. Sessions were also rated for adherence and competence to the MBT manual. Findings showed that better adherence and competence was related to higher levels of mentalizing from the patient, and that therapist statements demanding mentalizing predicted higher RF scores. Thus, efforts to stimulate mentalizing rendered more explicit mentalizing by the patient, as the MBT manual prescribes. However, pre-treatment RF scores on the AAI were higher than in-session mentalizing, spurring new research questions concerning how the therapeutic interaction shapes the level of mentalizing and whether in-session mentalizing is something qualitatively different from interview-based mentalizing.

In Study IV a newly developed instrument for rating in-session mentalizing was tested by rating the interactional process between patient and therapist. Using a different set of data, consisting of two cases from an randomized controlled trial of short-term therapies for depression, we studied every session from the treatments (16 sessions in each therapy) and scored the level of mentalizing expressed by the therapeutic dyad mutually. In one of the cases the level of in-session dyadic mentalizing was found to be related to the level of depressive symptoms as rated by the patient before the next session, so that higher dyadic mentalizing predicted lower levels of depression, indicating that dyadic mentalizing might be a mechanism of change. In the other case this relationship was not found; despite a low quality of dyadic mentalizing, symptom relief was achieved. In both treatments the interaction could be meaningfully described using a mentalizing framework, suggesting that the dyadic RF measure could be useful in psychotherapeutic process research.

In conclusion, the studies of this thesis suggest that mentalizing can be meaningfully viewed both as an individual competence and as an interactional phenomenon. In addition, RF seems to be highly context- and relationship-specific. In order to gain a wider understanding of the concept and how it relates to human interaction, measures for implicit mentalizing needs to be developed and tested.

ACKNOWLEDGEMENTS

This thesis has been possible due to the never-ending support and encouragement of my supervisors **Rolf Holmqvist** and **Fredrik Falkenström**.

Thank you, **Rolf**, for believing in my ability to plunge into the deep waters of mentalizing theory and encouraging me to go further when I was afraid to drown – you have truly been my secure base in the academic world. Thank you also for being such an inspirational teacher in relational theory and its clinical application.

Thank you, **Fredrik**, for your impressive patience with my statistical ignorance that never seems to improve. Thank you also for inspiring and fun discussions about RF and for excellent supervision and invaluable input in this work.

The research group of clinical psychology at LiU – and especially **Mattias Holmqvist Larsson**, **Anna Gradin Franzén**, **Erika Viklund**, **Börje Lech**, **Anna-Karin Åkerman**, **Ylva Söderberg Gidhagen**, **Lars Back**, **Björn Philips**, **Annika Ekeblad**, **Malin Bäck** and **Anneli Frostell** for inspiration, encouragement and fun times at the pub, in conferences and, not least, in prison! An extra shout-out to **Mattias**, for making me see the therapeutic value of hamsters and to **Björn** for being so generous in sharing your data.

Johanna Kullgard and **Marie Gut** – for being the wisest brains and sources of both clinical and theoretical knowledge in discussing and trying out the RF dyadic instrument. You've made all the work with the endless and sometimes painfully difficult RF ratings a pure pleasure and I really do hope we can continue working together.

A special thank you to all of our informants from prison, both to all of you who were forced stay there and to you who were there by choice – thank you for letting us in and for sharing your stories so generously.

To my mother, **Karin Möller**, for your support and that you always believed in my ability to analyze and to write, despite my inclination for too long sentences. Thanks to you I have gotten a much needed perspective on what is important in my life and you have also taught me that implicit relational knowing carries love and affection at times when words and memories are no longer within reach.

Pappa, Johan, Ebba, Erik and **Hanna**. Family first. Always. I am so lucky to have you!

Jesper, for being the best partner and father to our children I could ever wish for. Without you this simply could never have been done, I would have been stuck changing diapers or crying about not being clever enough to write a thesis. You and our family are the mainstays of my life.

Tage & Sixten, for teaching me about the wonders and challenges of parenthood and about both how difficult and how easy it can be to mentalize in interaction. Jag älskar er båda, alltid, alltid, alltid.

CONTENTS

INTRODUCTION.....	1
DEFINITION	1
Externally and internally oriented mentalizing	2
Mentalizing online and offline	3
Mentalizing as a state or trait capacity	5
Related concepts.....	6
Metacognition.....	6
Mind-mindedness.	7
Implicit relational knowledge.....	8
ASSESSMENT OF MENTALIZING.....	8
Interviews with coding systems	8
RF on the Adult Attachment Interview.	8
Parent Development Interview.....	10
Symptom-specific RF interview.....	10
Metacognition Assessment Scale (MAS).....	10
Self-report questionnaires	11
Reflective functioning questionnaire (RFQ).	11
Mentalization stories test for adolescents.....	11
Experimental/observational tasks.....	12
Reading the mind in the eyes test (RMET).	12
Movie for Assessment of Social Cognition (MASC).....	12
Maternal Mind-Mindedness.	13
DEVELOPMENT OF MENTALIZING	14
Epistemic vigilance and epistemic trust	16
MENTALIZING AND PSYCHOTHERAPY	17
RELATIONAL INTERPRETATION OF REFLECTIVE FUNCTIONING.....	21
 AIM OF THE THESIS.....	 23
 SUMMARY OF INCLUDED STUDIES	 23
ARTICLE 1	23
Mentalizing in young offenders	23
Methods.....	24
Results.....	24
Discussion.....	24
ARTICLE 2	25
Reflective functioning, limit setting, and emotional availability in mother-child dyads	25
Methods.....	25
Results.....	25
Discussion.....	26
ARTICLE 3	26

Mentalization- based therapy adherence and competence stimulates in- session mentalization in psychotherapy for borderline personality disorder with co-morbid substance dependence	26
Methods.....	27
Results.....	27
Discussion.....	27
ARTICLE 4	28
Mentalizing in therapeutic interaction: a relational perspective	28
Methods.....	28
Results.....	28
Discussion.....	29
GENERAL DISCUSSION	29
DISCUSSION OF FINDINGS	29
DISCUSSION OF METHODS.....	31
CLINICAL IMPLICATIONS AND FUTURE RESEARCH	34
REFERENCES	36

MENTALIZING – COMPETENCE AND PROCESS

INTRODUCTION

Definition

The concept of mentalization as it is currently used originates from developmental and cognitive psychology and denotes the individual's ability to attribute hypothetical mental states to explain human behavior and make meaning of the subjective experience of having an inner world, not visible to others. The use of mentalization was prominent within cognitive psychology during the 1990s when Baron-Cohen (1995) presented his research on how Theory of mind/mentalization was central in understanding some of the difficulties associated with autism. Since then, Fonagy and his research group has expanded the theory of mentalization and its application by explaining the development of mentalization from an attachment perspective and emphasizing its significance in affect regulation. They offer a definition of mentalizing as an "imaginative mental activity, namely, perceiving and interpreting human behavior in terms of intentional mental states (e. g. needs, desires, feelings, beliefs, goals, and reasons" (Fonagy, Gergely & Target, 2007, p. 288). There has been a shift towards using the verb mentalizing instead of then noun (mentalization) to underscore the never-ending, constantly shifting activity we engage in rather than associating the concept with a static cognitive skill that we learn, and once learned, can be used whenever needed.

Mentalizing has been described as the process of "holding mind in mind" (Allen, Fonagy & Bateman, 2008), of trying to make oneself and others comprehensible and predictable by imagining an inner reality of thoughts, feelings, wishes and needs that forego or accompany our actions. By attending to our own inner world, we become more familiar with our desires and reactions, which, when we recognize them, can be communicated to others and thus help us sustain and develop close relationships (Fonagy, Gergely, Jurist, & Target, 2002). Mentalizing is an explicit, verbal, willful action (as when we try to understand and set straight a quarrel with a loved one) as well as an implicit, nonverbal, automatic process (as when we hurry to finish what we have to say when noticing the stressed look on our colleague's face). Explicit, or controlled mentalizing has been related to linguistic and symbolic processing, involving various areas of the prefrontal cortex, whereas implicit or automatic mentalizing has been linked to sensory information and found to involve evolutionary

older brain areas such as the amygdala, basal ganglia and dorsal anterior cingulate cortex (Luyten & Fonagy, 2015).

The capacity to mentalize is intrinsically related to the level of emotional arousal experienced (Fonagy & Bateman, 2006). The most challenging for most people is trying to make sense of an emotionally complex or painful event or interaction, i. e. mentalizing under pressure while emotionally unsettled or upset. The difficulties in mentalizing while emotionally upset is explained by Luyten and Fonagy (2015) in that areas of the brain that are active in mentalizing are hampered by high arousal, which is especially pronounced in certain kinds of psychopathology (such as borderline personality disorder). A goal of therapy, according to Bateman & Fonagy (2006), is thus to expand the window of tolerance, to endure emotional pressure while being able to make efforts in teasing out mental states underlying human interactions.

Externally and internally oriented mentalizing

Mentalizing concerns both cognitive and emotional aspects and can be oriented towards both oneself and others (Choi-Kain & Gunderson, 2008). Another important distinction with clinical implications has been made between internally and externally oriented mentalizing. The latter refers to reliance on external features such as body language, facial expressions and tone of voice and activates parts of the brain (among which the temporal lobes) that are less willfully controlled.

Externally oriented mentalizing is closely linked to, but not equal to, implicit mentalizing, in that we rarely think consciously of how we interpret others' facial expressions, for example, but the information is stored in memory like a picture, available to be processed, and reprocessed, explicitly when one tries to understand why a misunderstanding has taken place. Of course, one's thoughts and feelings might distort the picture, but external cues, like frowning or yawning, are often easier to describe more objectively than the other's mental state which one has inferred and reacted to. External cues can also be helpful in facilitating discussion of misunderstandings based on different interpretations of such cues in terms of mental states. Individuals with autistic disorders have been found to have decreased ability to recognize and interpret facial expressions (Harms, Martin & Wallace, 2010) whereas patients with borderline personality disorder tend to be hypersensitive to facial expressions but have bias in interpreting ambiguous affect expressions (Domes et al., 2008). Problems with internally oriented mentalizing seems to be more characteristic for the borderline group (King-Casas et al., 2008).

Internal mentalizing is a conscious process that requires deliberate action to imagine the inner world of oneself or others and using that imagination to understand actions and mental states (Luyten & Fonagy, 2015). Internal mentalizing is rooted in the intentional stance described by philosopher Dennett (1987), which refers to the strategy of prediction that is adopted by humans, helping us interpret the behaviors of ourselves and others as intentional, i. e. *about* something, motivated by internal states such as thoughts and wants. Internal mentalizing relies upon the ability to imagine and envision such interior states that cannot be detected by, but may be inferred by, visible cues. Being able to oscillate between external and internal mentalizing is crucial to make sense of one's own reactions as well as making plausible interpretations of others.

Mentalizing online and offline

In the context of psychotherapy, a distinction between online versus offline mentalizing has been made. Offline mentalizing refers to the process of reflecting about past events or experiences, whereas online mentalizing concerns trying to mentalize about experiences in the here and now, according to Luyten and Fonagy (2015). The latter is often considered to be the most therapeutically challenging intervention for patients with attachment-related difficulties (Bateman & Fonagy, 2006; Morken, Karterud & Arefjord, 2014; Safran & Muran, 2000). Offline mentalizing tends to be more cognitive and explicitly oriented in trying to understand past events and creating a comprehensible narrative on how these events relate to oneself. Online mentalizing requires both cognitive and emotional aspects of mentalizing and, in particular, being able to abide with the frustration that often accompanies the uncertainty of here and now interactions. There is accumulating evidence that working with transference and countertransference in psychotherapy is beneficial (Norcross, 2011). Barreto and Matos (2018) offer an interesting perspective on how therapists by mentalizing the countertransference can stimulate mentalizing in the therapist-patient relationship. Offline mentalizing can be done both within a therapeutic relationship and on one's own, whereas online mentalizing is more intertwined with the current interpersonal context and depends upon the mentalizing capacity of the other party (Luyten & Fonagy, 2015). Online mentalizing bears a resemblance to the concept of *mindfulness in action* as described by Safran and Muran (2000), where they outline meta-communicative strategies to "articulate one's *implicit* or intuitive sense of something that is taking place in the therapeutic relationship in

order to initiate an *explicit* exploration of that which is being unwittingly enacted.” (p. 110, italics in original).

However, the literature on online versus offline mentalizing is scarce and differences between online and offline mentalizing are not clear-cut. When describing the mechanisms of change in mentalization-based treatment for borderline personality disorder, Fonagy and Bateman (2006) describe that too high levels of affect disrupt the ability to mentalize, explaining borderline symptomatology in the words of high arousal and activation of the attachment system take the mentalizing parts of the brain “offline”. Allen (2003) uses the term online mentalizing to describe mentalizing implicitly, automatically, without words about something that goes on in the present moment, which he distinguishes from offline mentalizing that he equates with explicit reflection about past experiences. Taken together, it is hard to disentangle whether it is the time frame, level of emotional arousal or the degree of conscious processing that is the dividing line between the two. But if we stay with the time frame perspective, since Fonagy and Luyten (2015) also emphasized that aspect, blurry points remain. For example, when a patient discusses events from the past with her therapist and the dyad struggle together to find new ways of understanding what happened and links to what the patient is experiencing in the present moment, is that online mentalizing in the here and now or is it offline mentalizing about there and then? Or, when a patient felt neglected by his therapist in the waiting room before the session started and manages to bring that to his therapist’s attention by the end of a 45-minute session, is that reflecting about the past or in the present?

In their article on bridging the transmission gap, Fonagy and Target (2005) mention the term off-line mentalizing capacity in one sentence, referring to parent’s ability to reflect about their relationship with their child when interviewed about it, thus implicitly distinguishing this off-line mentalizing capacity from thinking about the relationship “‘in real time’ in the course of an interaction” (Fonagy & Target, 2005, p. 335), something that could be referred to as online mentalizing. For the purpose of the present thesis, this distinction will be applied since I find it the most useful in the context of psychotherapy and process research thereof. That is, the difference between online and offline mentalizing refers to whether the activity takes place within a relationship or within oneself. Online mentalizing refers to the action of mentalizing (explicitly and implicitly) while interacting with someone else, such as exploring a misunderstanding that has taken place between the two interacting parties. Online mentalizing thus also incorporates a therapeutic dyad who is trying to make sense of the patient’s history or emotional problems together (i. e. the topic of

conversation does not have to concern the dyadic relationship), but when the patient leaves the room and goes home reflecting upon the content of the session he just had, he then engages in offline mentalizing.

Mentalizing as a state or trait capacity

Fonagy, Bateman and Luyten (2012) hold that mentalizing capacity has “both ‘trait’ and ‘state’ aspects that vary in quality in relation to emotional arousal and interpersonal context” (p. 4). According to mentalizing theory, infants are predisposed to develop mentalizing skills, but depending on their attachment context, the propensity and ability to understand oneself and others as intentional beings will be more or less pronounced (Fonagy et al., 2002). Fonagy et al. (2002) discard the idea of a relational prototype that is learned early on in life and through which every subsequent relationship is sculptured, as “naive” (p. 98), but nevertheless stress the importance of attachment relationships for the child’s ability to process psychosocial experiences, endure negative affect and develop and sustain intimate relationships. Mentalizing is thus established in childhood and has been found to be protective in the event of potentially traumatic experiences (Allen, 2013) as well as protective against aggressive behavior (Taubner, White, Zimmermann, Fonagy & Nolte, 2013), suggesting a trait-like component of mentalizing ability.

On the other hand, research findings indicate context-specific differences in ability to mentalize about attachment relationships versus mentalizing about one’s psychiatric symptoms (Kullgard et al., 2013; Rudden, Milrod, Target, Ackerman & Graf, 2006), suggesting state-like aspects of mentalizing. In addition, mentalizing may differ significantly among relationships, and seems to be “strongest in relation to conflict, distress, and confusing behavior in emotionally important relationships” according to Target (2008).

The notion of whether mentalizing should be viewed as primarily a trait or a state phenomenon has bearing on how to investigate the significance of mentalizing in interpersonal interaction. For example, is mentalizing ability a relatively stable personality trait that can be captured in an interview and predict one’s actual behavior or ability to make use of psychotherapeutic treatment? Is mentalizing ability enhanced by psychotherapeutic treatment or is it too stable a trait to be usable as an outcome measure? Is one’s ability to mentalize dependent upon what kind of therapist one meets? Or should mentalizing rather be viewed as a “process of joint attention” (Bateman & Fonagy, 2006, p. 93) within the realms of a psychotherapeutic

relationship and something that should be considered a mechanism of change in order to establish symptom relief? In that case, a process measure of mentalizing is needed which makes it possible to investigate whether interventions aimed at enhancing mentalizing in the present moment are effective in reducing psychological distress.

Related concepts

Metacognition.

The concept of metacognition has considerable theoretical overlap with mentalizing, in that it refers to the ability to identify and reason about mental states in oneself and others (Maillard et al., 2017). According to Maillard et al. (2017) metacognition is an umbrella term which incorporates different cognitive capacities that people use to handle the nature of mental states (for example, that different people have different perspectives, or that one's thoughts are not necessarily true).

However, there are also important differences between metacognition and mentalizing, of which the most significant concerns their relations to attachment theory. Metacognitive disturbances are seen as caused by a range of reasons of which attachment related difficulties can be one, and a broader view of causes (among which maladaptive interpersonal schema is one; Dimaggio, 2015) is stressed as a difference compared to the mentalizing concept as defined by the Fonagy group (Maillard et al., 2017). Another theoretical difference is that metacognition is thought of as a set of subsystems that includes knowledge about cognition and cognitive strategies, monitoring of current mental content and control of cognition (such as inhibition of impulses, attention shifting and cognitive flexibility), whereas the concept of mentalizing presented by Fonagy and colleagues is a global capacity underlying affect regulation and ability to mentalize about oneself and others. The operationalizations, the Metacognitive Assessment Scale and Reflective functioning scale respectively, reflect these conceptual differences (for further details about the instruments, see Section "Assessment of mentalizing" below).

In the description of metacognition, emphasis is put on problem solving and cognitively oriented mastery of difficult situations without being overwhelmed or clouded by emotions. Mentalizing, on the other hand, equally emphasizes cognitive and affective mentalizing and, in its therapeutic application, stresses the difficulties of both too little and too much emotional arousal. What is considered effective mentalizing always contains affectivity, and being able to be aware of and identify one's affects is considered the most fundamental aspect of mentalizing (Jurist, 2010).

Another important distinction between the two concepts seems to be that mentalizing includes and relies heavily on implicit activity and knowledge (Fonagy, Target, Steele & Steele, 1998), whereas metacognition is presented as and exemplified with only explicit efforts in understanding human behavior. Research using the Metacognitive Assessment Scale has shown improvements in metacognitive abilities in patients who undertook cognitive behavioral therapy or metacognitive interpersonal therapy (Dimaggio, Semerari, Carcione, Niccolò, & Procacci, 2007) and it might be that the concept of metacognition is closer to the cognitive behavioral tradition.

Mind-mindedness.

The research of Meins and colleagues about mother-infant interactions, has shed light on how the interplay is more or less influenced by the caregiver's inclination to treat her or his child as an intentional being with mental states (Meins, 2013). By studying mother-child dyads that had been categorized as either sensitive or non-sensitive (according to Ainsworth's notion of sensitivity), Meins, Fernyhough, Fradley and Tuckey (2001) found that sensitive mothers tune in to their children and interpret their child's behavior from the child's perspective. By contrast, less sensitive mothers tend to either misinterpret their child's behavior or fail to respond empathically but instead tease or ignore their child. The level of maternal mind-mindedness in interaction with her 6 months old infants, i. e. her tendency to treat her child as having an own mind and to tune into the mental states of her child, has been found to predict infant attachment security at 12 months of age (Meins, 1999). Bernier and Dozier (2003) found, in their study of mediation of attachment transmission, that mind-mindedness accounted for part of the predictive power of the parental state of mind on infant attachment behavior. What still remains unclear, they argue, is by what mechanisms the parental representations of the child influence attachment behavior. They contend that previous research on parental sensitivity indicates that further investigation of the connection between parental mind-mindedness and sensitivity in actual interaction is needed (Bernier & Dozier, 2003).

The concept of mind-mindedness has resemblance to the concept of parental reflective functioning (i. e. parents' capacity to mentalize about their children) presented by Slade in 2005. Slade holds that parental RF is a specific type of mentalizing, as described by Fonagy. It differs from mind-mindedness as it, beyond parental use of mental state words in interaction with their children, also embraces how the parent links mental states with behavior or other mental states (such as "oh,

you want to be held now because you're sad and afraid" as opposed to merely stating "you're sad"; Slade, 2005).

Implicit relational knowledge.

The concept of implicit relational knowing presented by the Boston Process of Change Study Group in the 1990s refers to non-verbal, procedural and non-reflective knowledge of being with another person, of how to interact and attune to the other (Stern, 2004). Implicit relational knowledge incorporates body language, gestures and vocal prosody but also mental nonvisible activity such as affective attunement, expectations and motivation (Stern, 2004). Studies of moment-to-moment interaction of mothers and their infants have demonstrated how even the dyad attune to one another in facial expressions, gestures etc. and coordinate their actions clearly visible to the observer (Beebe & Lachman, 1988). The Boston group argues that implicit relational knowledge is crucial to understanding how therapeutic change is brought about (Lyons-Ruth et al., 1998) and that it "operates outside focal attention and conscious experience, without benefit of translation into language" (Lyons-Ruth et al., 1998, p. 285). Within the realms of psychotherapy, the Boston group argues, both the implicit relational knowing of the patient and the therapist contribute to and define the relationship between them. The patient's experience of being with another person can possibly be rearranged by what they call "moments of meeting" (Lyons-Ruth et al., 1998, p. 286), referring to a dyadic state of consciousness where both parties are recognized in their own subjectivity.

Assessment of mentalizing

There are three main ways of assessing mentalizing: interviews with coding systems, self-report questionnaires, and experimental/observational tasks. In addition, a few projective tests have been proposed to measure mentalizing ability but these tests were not developed to be measures of mentalizing and will thus be excluded from this overview.

Interviews with coding systems

RF on the Adult Attachment Interview.

The golden standard of mentalizing assessment is the operationalization called the Reflective Functioning (RF) Scale. The RF scale was developed by Fonagy et al. in 1998, to be used on Adult Attachment Interviews (AAI; George, Kaplan & Main, 1985). The AAI consists of questions concerning interactions with attachment figures

and potentially attachment activating situations (such as separations). The questions are aimed at “surprising the unconscious”, thus eliciting the attachment style of the interviewee. The RF coding procedure (which is entirely separate from the attachment classification) is complex and involves looking for explicit indications of mentalization by the interviewee in response to interview questions. The questions are divided into *demand* and *permit* questions, i. e. questions that either demand reflection on human interaction or mental states from the interviewee or questions where such reflection is not necessary. Demand questions that are answered without any indication of (a) awareness of the nature of mental states, (b) the explicit effort to tease out mental states’ underlying behavior, (c) recognition of developmental aspects of mental states or (d) recognition of mental states in relation to the interviewer are scored low in RF (RF 3 or below). The scale runs from -1 (active rejection of mentalizing) to +9 (exceptional mentalizing). Each question is given a single score that the rater aggregates and uses together with a qualitative impression of the interview as a whole to decide upon a global RF score for the interview. In a study of 200 parents to be (Fonagy, Steele, Steele, Moran & Higgitt, 1991), the most common RF level in a non-clinical sample was found to be 5, which in the RF manual has been labeled *Ordinary RF*. Such a score is applicable for interviewees who make behavior meaningful to themselves and others as motivated by mental states that can be envisioned and reflected upon. These interviewees are able to maintain a relatively stable and coherent model of the mind of their attachment figures, as long as the topic is not too complex or emotionally difficult. Ratings below this theoretical median indicate that the interviewee is either unable or unwilling to reflect upon their own minds or on those of other people. The lowest level, *Negative RF* (-1 or 0), is characterized by refusing and actively avoiding demands of mentalizing, and hostile, bizarre, or clearly unintegrated answers to the interviewer’s questions. Ratings of -1 or 0 are very unusual in non-clinical samples, according to Fonagy and colleagues (1998), but have been found in violent offenders (Levinson & Fonagy, 2004). Ratings of RF 1 or 2, *Lacking in RF*, are more common and refer to an absence of mentalizing statements in the interview. Such interviews have no, or very little, indication of the interviewee taking mental states into account or of having an understanding of the minds of her/himself or of the attachment figures. Commonly in these interviews, demand questions are answered by “I don’t know”. The next defined level of the scale is called *Questionable or Low RF* (RF 3), which means that the interviewee uses mental state words and shows a partial, yet simplistic and superficial, understanding of the thoughts, feelings, and intentions of others, for example in citing others without contributing with her or his own

reflections on the statements. These interviews give the rater a sort of screenshot of the interviewee's experiences, without the narrator's voice commenting and explaining the events that occurred. Transcripts falling in the upper part of the RF scale, *Marked RF* (6–7) or *Exceptional RF* (8–9), are characterized by numerous instances of full reflective functioning, an ability to maintain a detailed and complex understanding of what goes on in the minds of different protagonists. One of the hallmarks of full RF is the occurrence of newly formed understandings of the events being narrated that gives a fresh and vibrant touch to the story. In these interviews, mental states are causally related to one another so that new reflection and integration might be achieved, with the interviewee maintaining a tentative way of expressing her/himself, without losing awareness of the limitations of her/his insight.

Parent Development Interview.

(PDI; Slade, Aber, Berger, Bresgi & Kaplan, 2010). PDI is a semi-structured interview aimed at capturing a parent's ability to mentalize her/his child and their relationship. The interview resembles the AAI in that it consists of permit and demand questions concerning an attachment relationship. Unlike the AAI, the PDI is focused on a current, constantly evolving relationship, which lessens the burden on autobiographical memory and often makes it easier for the interviewee to come up with emotionally salient examples of interactions that call upon mentalizing the child and oneself as a parent. The PDI is scored for RF by applying the original RF scale (Fonagy et al., 1998) together with an addendum created by Slade, Bernbach, Grienberger, Levy and Locker (2005).

Symptom-specific RF interview.

In their research on psychodynamic therapy for panic disorder Milrod and colleagues developed a panic-specific RF interview designed at capturing whether patients with panic disorder have difficulties in mentalizing their symptoms independent of their ability to mentalize their attachment relationships (Rudden, Milrod, Aronson & Target, 2008). Their research is highly interesting since the results showed that there may be mentalizing difficulties within a specific domain, indicating the context-specificity of mentalizing (Rudden et al., 2006). Since their studies, several other specific RF interviews have been developed (such as depression-specific RF; Ekeblad, Falkenström & Holmqvist, 2016).

Metacognition Assessment Scale (MAS).

The MAS is another instrument for assessing mentalization based on narratives, developed by Carcione and colleagues (described in Dimaggio, Procacci et al., 2007).

The MAS consists of three subscales measuring an individual's ability to 1) reason about her/his own mind, 2) others' minds or 3) master increasingly complex psychological problems. The MAS is to be applied to therapy sessions (or specialized interviews) and a scoring unit is identified as a patient's speech turn framed by two therapist statements. Ratings yield one global score and three subscale scores (one of each of the above presented subscales) indicating frequency and level of sophistication in using each skill, for example the extent to which a patient is able to represent her or his emotional state with regard to psychological conflicts and soothe suffering (Maillard et al., 2017).

Self-report questionnaires

Self-reported capacity for mentalization is a challenge because of the immanent paradox of needing the capacity for mentalization in order to notice a possible lack thereof. However, the time-consuming efforts of interviewing and scoring RF on the AAI spurs the development of other instruments to capture mentalizing capacity.

Reflective functioning questionnaire (RFQ).

In 2016, Fonagy et al. published an article on the development and validation of the Reflective functioning questionnaire (RFQ). The RFQ originally consisted of 46 items. After several data collections, a factor analysis yielded a two-factor structure, and the six items that loaded highest on each factor were chosen to make up the final RFQ with the two subscales Certainty of mental states of self and others and Uncertainty of mental states of self and others. According to the authors, the results were promising and indicated that the RFQ discriminated between non-clinical adults and patients with borderline personality disorder or eating disorder and was correlated with measures of mindfulness, empathy and perspective taking in theoretically expected ways. The RFQ was related to parental RF (measured with a self-report questionnaire) but only vaguely to external mentalizing as measured by the Reading the mind in the eyes test (see further description below). The relationship between RFQ and observer rated measures of mentalization such as RF on the AAI or PDI has not yet been studied, although such a relationship seems essential to establish before concluding that self-reported RF measures the same concept as observer measures.

Mentalization stories test for adolescents.

This questionnaire primarily aimed at capturing ability to mentalize internally about others, consists of 21 real-life scenarios of adolescents (Vrouva & Fonagy, 2009). Each story includes a protagonist in an emotionally charged interaction with someone

else. The scenarios are intended to elicit negative feelings such as anger, guilt or jealousy in the protagonist and incorporate what she or he does or says in response to that mental state. Following the scenario are questions about the reasons for the protagonist's actions and three choices to select from. Each answer is scored according to the degree of mentalizing it reflects (accurate, excessive or distorted mentalizing) and a total score is calculated for the test as a whole (Rutherford et al., 2012). The test has not been widely used, so its psychometrics properties are not known to me.

Experimental/observational tasks

Reading the mind in the eyes test (RMET).

This is a test of externally based mentalizing, developed by Baron-Cohen to discriminate individuals with high functioning autism or Asperger syndrome from individuals without function variations (Baron-Cohen, Wheelwright, Hill, Raste & Plumb, 2001). The test consists of 36 photographs of human faces or parts of human faces (the eye-region). With every photograph, four words denoting different mental states are presented and the task is to choose which word most accurately describes the mental state of the person in the picture. A lexicon with the semantics of each word is available in order to eliminate differences in verbal ability among test persons. In the study of Baron-Cohen et al. (2001) the test was found to discriminate between adults with Asperger or high functioning autism from normal controls and the Eyes test was inversely correlated in both groups with results on the Autism Spectrum Quotient.

Movie for Assessment of Social Cognition (MASC).

The MASC (Fossati, Borroni, Dziobek, Fonagy & Somma, 2018) is a 15-minute-long movie of four people who interact with each other in common social situations, each displaying stable personality characteristics and experiencing different situations that elicit mental states such as joy, jealousy and anger. The movie is used to test non-attachment related mentalizing. The test person is instructed to watch the movie closely and try to imagine what the protagonists feel and think. The movie is cut into 43 segments with increasing complexity and in each one the test person is asked why the protagonists acted the way they did. The test person may choose from four alternative answers and is instructed to choose the one that is most likely (despite that one cannot be certain of the right answer.) Each alternative represents a qualitatively different level of mentalizing (i. e. non-mentalizing, hyper-mentalizing, hypo-mentalizing or adequate mentalizing) and the test person's choice is thought to give information as to which kind of mentalizing difficulty is most prominent, if any. In

their overview, Luyten, Fonagy, Lowyck and Vermote (2012) describes the MASC as tapping on explicit (rather than implicit) mentalizing, as well as requiring both internally and externally focused mentalizing. All questions concern the protagonists of the film, thus naturally testing mentalizing about others (and not self). The instrument has been found to be significantly and meaningfully correlated with the Reading the mind in the eyes test in both clinical and nonclinical groups (Fossati et al., 2018).

Maternal Mind-Mindedness.

Based on observations of mother-infant interactions, Meins et al. (2001) developed the measure of mind-mindedness in order to quantify the extent to which mothers express interpretations of their infants' behaviors as motivated by intentional states. The rater observes if and how the mother interprets her baby's gaze direction, playing with objects and how she imitates her child, creating a dialogue between herself and the baby and how she encourages autonomy in the child as an own individual with distinct thoughts, feelings and wishes. Using the measure of mind-mindedness also includes determining the appropriateness of the mothers' comments, since only comments referring to mental states that appear to be in accordance with the child's behavior qualifies as mind-minded. Mind-related comments can also be classified as non-attuned if it seemed at odds with the infant's behavior or did not fit the interaction (for example stating that the infant liked a particular food when the baby was clearly engaged in playing with a toy). Mind-mindedness have been found to be related to maternal sensitivity (Meins et al., 2001) and, of particular interest, are the findings of misattunement where mothers have been found to act sensitively to their child's mental state but misattribute the reason for the child's feelings or conversely, display an actual understanding of the child's inner state but doing so in a tone of voice that conveys irritation or discontent with the child's state or behavior which has been seen in mothers suffering from mental illness (Meins, 2013).

To bear in mind is that most of the above listed tests are developed primarily to assess mentalizing ability for the purpose of research. When assessing mentalizing clinically there are a few tests aimed at capturing the mentalizing deficits associated with autistic disorders (for example, RMET). A comprehensive assessment of mentalizing ability and deficits associated with psychiatric conditions such as borderline personality disorder or eating disorders should, according to Luyten et al. (2012), include several clinical interviews focusing on attachment history, current and

past relationships and demand questions concerning the patient and her/his attachment figures as well as the psychiatric symptoms experienced.

The research on mentalizing can be divided into two main areas, i. e. 1) development of mentalizing and its role in the transmission of attachment patterns within families and 2) the application of mentalizing in psychotherapy, which started with the understanding and treatment of borderline personality disorder from a mentalizing perspective.

Development of mentalizing

The essence of the mentalizing theory presented by Fonagy et al. (2002) is that we are born predisposed to develop mentalizing skills, but in order for this development to flourish babies need secure attachment relationships in which their affect displays are mirrored and their behaviors mentalized by attachment figures. This idea, that the quality of the parent-child-relationship has impact on how the brain areas activated in mentalizing develop, is supported by research findings presented by Luyten and Fonagy (2015). The parent-child interactions need to be characterized by *contingent* and *marked* mirroring, meaning that the affect that is mirrored by the parent is congruent with the child's experience and clearly processed by the parent. The markedness of mirroring is central in that an exact mirroring (parent's expression is identical to the child's) would be overwhelming and terrifying for the child, whereas markedness conveys that the parent has understood what the child is feeling and empathizes with that feeling. This process is the vehicle by which the child learns how to recognize her/his own feelings, how they are experienced in the body and how to build affect representations. In addition, it is the beginning of affect regulation.

The mechanisms of attachment are crucial foundations of mentalization theory. According to Bowlby's attachment theory (Bowlby, 1988), fear (stress arousal) activates the attachment system in the child and when fear is experienced all other behavioral systems (including areas of the brain that are active in mentalizing processes) are downplayed in favor of seeking closeness with the parent. This mechanism has the obvious evolutionary advantage of increasing the chances of the child's survival in the face of dangers. When the attachment system is activated, the child approaches his parent, and closeness is in itself usually affect regulating (when the attachment relationship is secure), which calms the attachment system (Bowlby, 1988). However, research by Lyons-Ruth and colleagues suggests that it is not primarily physical closeness to attachment figures that regulates the child's level of

arousal, but rather continuous positive intersubjective communication between infant and attachment figure (Lyons-Ruth, 2006). Lyons-Ruth (1999) stresses the importance of parents maintaining a psychological connection and intersubjective communication with their child for the child to be able to develop affect regulation strategies and mentalizing ability of her own. Lyons-Ruth's longitudinal studies provide empirical support for the notion that children of parents who are psychologically withdrawn and withhold information about the mind of the child and important others, tend to develop psychopathology later in life (Lyons-Ruth, 2006). In order to maintain a psychological connection and continue to envision the child's inner world, parents need to be able to mentalize their own inner experiences and regulate their feelings and reactions that arise in interaction with the child. Affect regulation is set in motion by the parent's mirroring and mentalizing of the child, i. e. the child whose own mentalizing is deactivated needs to benefit from his mother's ability to mentalize both his and her own feelings in order to establish the experience of being secure in the world again and not overwhelmed by whatever caused the unpleasant feeling in the first place. When the attachment system is deactivated, other behavioral systems such as exploring (and mentalizing) can be activated again. Securely attached children display social skills earlier than children with insecure or disorganized attachment and securely attached children have better affect regulation strategies and self-inhibition (Sroufe, Egeland, Carlson & Collins, 2005).

According to Fonagy and colleagues (2002), it is in interactions with primary caregivers that children goes from only experiencing themselves as physical entities to also learning to understand themselves as social and intentional (as in having intentions that forego behaviors) agents in the world. Gradually the child creates second order representations for emotional experiences and begins to understand others and their observable behaviors as being causally linked to mental states (Fonagy et al., 2002). Several phases of pre-mentalizing modes precede the acquiring of mentalizing ability in children (Allen et al., 2008). These phases are not mutually exclusive but are seen as evolving ways of understanding oneself and others, layers that remain within us and can become activated throughout life when emotional arousal is too high. As stated by Allen et al. (2008) infants around the age of nine months have been found to expect rational and goal-directed actions of others. They are surprised by actions that seem non-rational from the perspective of physically observable circumstances. This is seen as evidence of a *teleological stance* in mentalization theory, i. e. a concrete non-mentalistic way of understanding people and their behaviors as if there were no inner world or mental states. Empirical evidence

suggests that infants as young as 7 months are beginning to take others' mental states and beliefs into account (Kovács, Téglás, & Endress, 2010), indicating that the teleological stance is followed by the next developmental step earlier than previously stated. The next step is described by Allen et al. (2008) when small children start to envision mental states as underlying behavior, and demonstrate an implicit understanding of false beliefs.

Still, children in this stage cannot distinguish reality from mental states and the inner and outer worlds are confused, a mode termed *psychic equivalence*. This confusing of fantasies and thoughts with reality can be rather overwhelming and frightening, since fantasies can be experienced as reality. This is thought to be the catalyst by which mentalizing skills develop further, into the next phase called *pretend mode*. In this mode, inner and outer worlds are separated from each other, and the child learns that his own experience does not mirror the external reality. However, in this phase the child does not see any connection between the inner and outer realities. They have no impact on one another, which by extension makes feelings, desires and thoughts meaningless and unreal. It can be thought of as internal mentalizing gone awry, that there is too little connection between the fantasized interior of oneself/others and reality, making the fantasies very vulnerable to any challenge from the outside.

The parent's mentalizing capacity has been found to predict secure attachment in her or his child (Fonagy et al., 1991; Lyons-Ruth, 1999). More specifically, Slade and colleagues investigated parents' ability to understand their children in a mentalizing way and found that parental mentalization mediates generational transmission of attachment patterns in families (Slade, Grienemberger, Bernbach, Levy & Locker, 2005).

Epistemic vigilance and epistemic trust

According to Sperber et al. (2010), humans are born with *epistemic vigilance*, which refers to a cognitive, self-protective mechanism predisposing us to be vigilant to signs of threat and potential disinformation coming from others. We scrutinize those surrounding us in order to detect who might harm us and whose information is not to be trusted. Empirical evidence from a study of Corriveau et al. (2009) suggests that secure attachment experiences lay the foundation for forming *epistemic trust*, which is defined as a person's "willingness to consider new knowledge from another person as trustworthy, generalizable, and relevant to the self" (Fonagy & Allison, 2014, p 373).

Securely attached children were inclined to trust information coming from their attachment figure, but could also question and discard that information when it seemed implausible (Corriveau et al., 2009). Conversely, results from the same study showed that children with insecure-disorganized attachment appeared to suffer from chronic epistemic hyper-vigilance, implying that neither information from an attachment figure nor a stranger was to be trusted. When epistemic trust is lost, or never adequately built up, the individual tends to be more rigid and defensive in his knowledge structures, and limited in his capacity to internalize new relational information that could challenge the epistemic vigilance. By extension, this may lead to the experience of being totally isolated and without communicative bridges to others (Fonagy & Allison, 2014). Attachment traumas may lead to distrust in others and tend to limit ability to explore solutions to problems and autonomy (Allen, 2005). Patients coming to therapy need not only learn how to trust the therapist, but also to regain reasonable trust in others, i. e. manage to balance epistemic vigilance with trust outside the treatment setting in order to benefit from treatment and be able to generalize knowledge gained in therapy into their real lives. Fonagy and Allison (2014) hold that mentalizing work in therapy is a vehicle by which epistemic trust can be established. By mentalizing in their real life outside therapy, patients can gain new understandings and experiences of their relationships, evaluate information in a nuanced manner and thus function better in the social world and with themselves (Fonagy & Allison, 2014). So far, the empirical support for this notion is scarce, but in a small study of MBT group therapy for adolescents with borderline personality disorder, post-treatment improvements in borderline symptoms were associated with enhanced trust in peers and parents (Bo et al., 2017). However, the study lacked a control group, so the improvements cannot be said to be caused by the treatment but may have been due to other factors such as spontaneous recovery.

Mentalizing and psychotherapy

An important turning point in the history of mentalization theory was the publication of the first study of successful mentalization-based treatment (MBT) for borderline personality disorder (Bateman & Fonagy, 2008; Bateman & Fonagy, 1999). The symptoms and features of borderline personality disorder (BPD) were conceptualized as deficits in mentalizing (Fonagy & Luyten, 2009) and successfully treated by therapists aiming to foster mentalizing capacity in these, by many considered as very difficult to treat, patients.

Since then mentalization-based treatment has expanded its applications to other clinical domains, such as eating disorders, family therapy and psychotic disorders (Bateman & Fonagy, 2012) and there has been an increasing interest in mentalizing as a mechanism of change in therapy. In theory, the change that enhanced mentalizing brings about lies in the patient experiencing being mentalized about by her therapist, thus learning to think about herself and her inner world in relation to others and gain new knowledge about how to interact with others and negotiate social relationships (Fonagy & Allison, 2014). Empirically, this remains to be investigated.

The dissertation by Bernbach (2002) was the pioneering study of mentalizing in therapy sessions. She investigated the mentalizing process in brief relational therapy (BRT) sessions of patients with personality disorders. In BRT occurrences of alliance ruptures are central to treatment, so her study focused on sessions that contained evidence of alliance tension and ruptures. She rated patient statements in sessions from beginning, middle and end phase of treatment and found that the patterns of how the level of mentalizing fluctuated throughout treatment differed between good outcome therapies and bad outcome therapies. Interestingly, there was no significant relationship between the level of RF and outcome (that is, the magnitude of RF was not higher in the good outcome cases, contrary to expectations). However, the good outcome therapies had greater variation of RF and decreased in RF at the end of treatment and Bernbach argued that the patterns of mentalizing that she found may be due to the relationship between the patient and the therapist. She addressed the work of relational theorists Benjamin and Aron who hold that “intersubjectivity is a mental state and a development achievement that occurs in relation to another individual” (Bernbach, 2002, p. 26), suggesting that a high level of RF in therapy sessions indicates an intersubjective process. She then proposed that the stress of experiencing a rupture close to the end in a treatment that is successful, is more challenging and therefore harder to process and explore, as compared to handling ruptures in a treatment which has not been helpful in reducing patient symptoms.

Since Bernbach’s dissertation several other studies have been conducted where the RF scale has been applied to therapy processes. Goldstein (2015) present an excellent and detailed overview of the studies (and of those where RF has been applied to interview transcripts before and after treatment) and the findings are mixed, to say the least. Josephs, Anderson, Bernard, Fatzer and Streich (2003) found that RF increased over the course of analytic treatment and correlated with symptom reduction. Karlsson and Kermott (2006) found that patient RF decreased during interpersonal therapy and remained low and stable in cognitive behavioral therapy as

well as in brief psychodynamic therapy. Hörz-Sagstetter, Mertens, Isphording, Buchheim and Taubner (2015) found in their case study of two patients who undertook psychoanalysis, that RF fluctuated during sessions as well as during the years of psychotherapy and that session-RF increased during treatment and corresponded to pre- and posttreatment measurements of RF.

With regards to RF as an outcome measure, RF scores increased and correlated with symptom reduction in transference-focused therapy (Levy et al., 2006; Fischer-Kern et al., 2015). Similar results were found in the Middleby-Clements unpublished study (referred in Goldstein, 2015) of cannabis dependent patients, where RF increased only in patients with good outcome. In a case study of a patient with borderline personality disorder, a significant increase in RF was reported after three years of mentalization-based treatment (Gullestad & Wilberg, 2012). Vermote et al. (2010) on the contrary, found that inpatients with personality disorders did not change their RF after treatment, despite significant symptom reduction and Boldrini et al. (2018) reported no change in RF during treatment (as measured by a computerized method of analyzing transcripts of therapy sessions) in their sample of 27 patients in psychoanalytic treatments. Boldrini et al. (2018) found, however, that RF measured in the first month of treatment predicted changes in two outcome measures of personality functioning post therapy. In Goldstein's own study, no change in RF was found over the 30 session course of treatment in her sample of 30 therapies. Further investigation revealed that attachment insecurity was negatively correlated with RF so that patients who were low in attachment insecurity did show a correlation with increasing RF over the course of treatment (Goldstein, 2015).

Concluding those studies, one can say that the level of RF can vary substantially depending on the way it is measured, depending on patient characteristics and treatment conditions, and it remains difficult to understand what reflective functioning really signifies.

Almost all studies of mentalizing focus on the explicit aspects of mentalizing and the implicit domain is often left aside. Davidsen and Fosgerau (2015) elucidate the relatively undertheorized domain of implicit mentalizing in mentalizing theory as well as the lack of agreement in how occurrence of implicit mentalizing can be detected. Existing instruments to assess implicit mentalizing are very limited and in the overview by Luyten et al. (2012) only three test of implicit mentalizing are presented, all of which are experimental. In addition, Davidsen and Fosgerau (2015) argue that there is a need to further study how implicit mentalizing take place in interpersonal interaction since that is the context in which implicit mentalizing develops and evolves

according to the theory of Fonagy et al. (2002). They criticize the existing operationalization of mentalizing, the RF scale, for not taking the dialogical nature and context-dependent aspects that are inherent in the mentalizing theory itself, into account when trying to measure mentalizing ability: “If the claim that mentalizing processes are enacted intersubjectively and are contextually dependent is to be taken seriously, there is a need to focus on the interactional behavior of participants” (Davidsen & Fosgerau, 2015, p. 439). Thus, both the implicit domain and the significance of interpersonal context are in need of further investigation.

As Keselman, Kullgard, Holmqvist and Osvaldsson Cromdal (2016) demonstrated in their conversation analysis study of therapy sessions, a patient who resists invitations to mentalize (i. e. demand questions) from her therapist does not necessarily show a lack of mentalizing ability, but the resistance can actually be interpreted as a linguistic way of displaying disagreement with the anticipated aim of the therapist, which in effect is a sign of mentalizing ability. Thus, the significance of each patient statement needs to be interpreted in relation to the therapist’s actions and the context of the specific therapy session. By this line of thought, a statement cannot in itself be indicative of mentalizing or not, but its function can only be said to exist in relation to how it is linked to what has been said before and how it is interpreted and responded to by the other party. An authentic, anonymized example would be as follows:

Therapist: “what is your feeling towards him for letting you down?”

Patient: “I don’t know.”

T: “When you think about the situation, what do you experience in your body?”

P: “I feel tense.”

T: “That’s a sign of anxiety, if you don’t block your feeling with anxiety, what is your feeling towards him for letting you down?”

P shrugs his shoulders: “I don’t feel angry”.

From a mentalizing perspective this patient’s utterances would grant him the rating of absent or questionable mentalizing, but from a relational interactional perspective he refuses the therapist’s probing, anticipating that there is a preferred answer that he does not want to give and which could be interpreted as a sign of an ongoing, non-verbal mentalizing process on behalf of the patient. If the dyad does not address the

potential disagreement we will never know whether the patient's utterances reflect non-mentalizing or non-verbalized mentalizing.

Studies of psychotherapy sessions using the method of conversation analysis have shown that therapists edit their patients' speech by for example summaries, conclusions, interpretations or noticing, i. e. therapist directing attention to the patient's recent action (Vehviläinen, 2008). The editing or tailoring indicates what the therapist considers to be relevant and can have the intention of increasing the patient's awareness of something that she or he, supposedly, had not been aware of before (for example, an observation of the patient's tendency to talk about someone else but herself or that she tells an emotionally difficult story without showing any signs of emotional arousal). By the nature of the therapeutic relationship, therapists have the power to direct the topic of conversation and decide what is relevant and not and can also, as illustrated by Vehviläinen (2008), have the power to decide whether there are underlying motives to a patient's action or not, which in terms of mentalizing could be construed as exerting a knowing stance (as opposed to the not knowing stance that is proscribed by Bateman and Fonagy, 2006). With this in mind, it seems plausible to assume that patients can react differently to whatever stance the therapist exerts, and inevitable that the dynamics of each therapeutic dyad and unfolding of their interactions in every single session are unique.

Relational interpretation of reflective functioning

Relational therapy, represented by Mitchell and colleagues, is grounded in a two-person psychodynamic psychology perspective where primary motivation is rooted in real relationships, as opposed to drive/conflict theories that see people as motivated by innate drives and inner conflicts (Greenberg & Mitchell, 1983). The self is seen as inherently related to others, striving to be recognized. Intersubjectivity is a developmental achievement that includes mutual recognition, where the other is seen as a subject in herself and not an object to my needs (Benjamin, 1990). Behaviors, affects and subjective experiences are to be understood, not in themselves, but as relational acts in a relational matrix. An implication of the two-person psychology perspective is, for example, that every therapeutic relationship is seen as unique and constructed in the present moment and that both parties have their own subjective understanding of what goes on in the interaction. Therapeutic change incorporates not only explicitly formulated insights but also a new, implicitly experienced way of being together with the other person that is qualitatively different from how the therapist and

patient related to each other beforehand (Lyons-Ruth, 1999). There are many points of contact between relational theory and mentalizing theory, for example when it comes to what is considered therapeutically important. For example, being able to endure conflict within the therapeutic relationship and to handle different perspectives on oneself and the current interaction could be thought of as reaching the mode of intersubjectivity within relational theory (Mitchell, 2000) or, in Fonagy's words, as an advanced level of mentalizing.

In therapeutic interaction, the patient's statements are related to the therapist's and vice versa and to disentangle them and sift out only the mentalizing ability of the patient has proven to be difficult empirically (Karlsson & Kermott, 2006). Theoretically, one may argue that it is not even possible to categorize statements as representative of one of the participants' ability, because of the constructivist nature of the interaction. That is, both the therapist and the patient may be seen as bringing into the room their respective ways of functioning and could also be understood as mutually influencing each other so that a unique way of interacting is constructed. Benjamin (1988) uses the image of Escher's birds to describe the mutuality aspect of self-other relationships. In his work *Day and Night* (see image on cover), Escher depicted black and white birds that appear to fly in both directions depending on which color that is considered the main focus and which is considered the backdrop. Whichever color is main, the reciprocal need for the other remains. As Benjamin describes, we find it harder to conceptualize the birds' movement as a whole and tend to orient our impression sequentially, to give priority to one direction and downplay the other. The challenge of the Escher birds is to look at the two directions, the white and the black birds, simultaneously, and recognize the mutual dependency.

Aron (1996) holds that interventions should not be isolated and thought of as the therapist's own contribution, for example as in the analytic idea of interpretations as conveying information to the patient from the therapist. Instead, he argues, the therapist statements should be regarded as co-constructed with the patient and that the fruitfulness of the interventions is whether they stimulate further inquiry and exploration of the patient's mind and relational inclinations. Translated into research methodology this suggests that the object of scrutiny should be the dyadic interaction in itself, rather than the patient's and the therapist's respective statements.

AIM OF THE THESIS

The overall aim of the thesis was to investigate the concept of mentalizing and how it corresponds to the operationalization of explicit mentalizing (RF). What do the RF scores mean when it comes to human interactions and actual behavior? We developed specific interview guides in order to capture how mentalizing capacity is manifested when reflecting upon something that took place recently (i. e. concerning limit setting situations with one's child as in article no 2 or concerning crimes committed as in article no 1) as compared to reflecting upon attachment relationships from one's childhood. This part of the thesis was directed at capturing the applicableness of RF in difficult situations of adulthood. We then went into the realm of therapy and aimed at finding out what role mentalizing plays in therapy sessions and whether it is fruitful to try to stimulate mentalizing in therapeutic interaction. During the work with the third study (when the RF scale was applied to therapy sessions) questions concerning ways of operationalizing mentalizing arose, of whether there was something to gain by looking at the therapeutic relationship from a relational and interactional perspective instead of looking at therapist statements and patient responses separately.

SUMMARY OF INCLUDED STUDIES

Article 1

Mentalizing in young offenders

Lowered mentalizing ability increases the risk of impulsive behaviors and relational problems and in a study by Levinson and Fonagy (2004) criminal offenders were found to be rated low in RF as compared with non-criminal controls. Another study indicated that RF mediated expressions of aggression (Taubner et al., 2012), with higher RF being associated with less proactive (i. e. deliberate) aggression. Very few studies had been conducted and we aimed to investigate mentalizing ability in young offenders imprisoned in Sweden. We hypothesized that their RF would be lowered (i. e. below RF 5) and that RF would be correlated with degree of psychopathy and alexithymia. A second aim was to investigate whether mentalizing ability differed between those who had committed interpersonal crimes and those who had committed non-interpersonal crimes and whether there were any differences in ability to mentalize around attachment relationships versus ability to mentalize about their criminal acts.

Methods.

We interviewed 42 male prison interns (aged 18-21) who volunteered to participate in our project with the AAI and a new interview consisting of demand and permit questions concerning the principal crime they were imprisoned for (Reflective Functioning Crime-Specific Interview; RF CS). Both interviews were rated for RF (Fonagy et al., 1998). The interns also completed the Toronto Alexithymia Scale (TAS-20; Parker, Bagby, Taylor, Endler & Schmitz, 1993) which is a self-report inventory to assess alexithymia. Based on interviews held with prison officers who were familiar with the interns, scores of psychopathy were made by a clinical psychologist using the Psychopathy Checklist (PCL:SV, Hare, 1991).

Results.

RF scores were low both on the AAI ($M = 2.71$, $SD = 1.7$) and on the RF CS ($M = 2.20$, $SD = 1.17$) and moderately correlated ($r = .56$). None of the subjects were classified above cutoff for psychopathy (i. e. all were scored below PCL = 18). None exceeded cutoff for alexithymia on TAS, but the scores were higher (TAS Total $M = 52.59$, $SD = 12.92$) than the scores of non-clinical adults ($M = 42.51$, $SD = 9.09$) in the study of Berthoz and Hill (2005). There were no correlations between RF and psychopathy and only one correlation was found between mentalizing and alexithymia, namely a positive, moderate correlation between RF CS and the TAS subscale called *Difficulty in identifying feelings and distinguishing them from the bodily sensations of emotion*. RF did not differ between interpersonal and non-interpersonal offenders, however the scores of antisocial behaviors on PCL were significantly higher in the group of interpersonal offenders. Different levels of crime-specific RF were illustrated by interview excerpts.

Discussion.

The results indicated low mentalizing in the sample (75% of the AAIs were rated RF 0-3) but, contrary to expectations, no systematic relationship between RF and psychopathy. One possible explanation is the problem of restriction of range, since no one in the sample exceeded the PCL cutoff score for psychopathy. The excerpts illustrated how two interviewees with the same PCL score answered very differently when asked to reflect about their crimes, suggesting that the crime specific interview can yield qualitatively different yet important information about the minds of the offenders as compared to previously established measures such as the PCL. The study gave rise to further research questions concerning the relationship between mentalizing ability and actual behavior in the prison setting, how interview-based RF

relates to online mentalizing and what promotes or hampers mentalizing processes in the prison environment. Study limitations were the absence of a control group of non-criminal young men and the lack of a measure of verbal ability which may be related to the explicit mentalization ability according to findings of Rutherford et al. (2012).

Article 2

Reflective functioning, limit setting, and emotional availability in mother-child dyads

In previous research, the importance of parents' mentalizing capacity for their children's attachment security has been established (Grienenberger, Kelly & Slade, 2005; Slade, Grienenberger et al., 2005). Parents' ability to be emotionally available and sensitive to their children's needs and emotional signals in interactions has also been found to be linked to children's attachment security and exploratory behavior (Sorce & Emde, 1981; Ziv, Aviezer, Gini, Sagi & Koren-Karie, 2000). The results of two studies of depressed mothers and their children suggested that mothers' ability to represent their child's mental states and be mind-minded may serve as a protective factor in mother-child interactions since depressed mothers with intact mind-mindedness were unimpaired in emotional availability whereas depressed mothers with low mind-mindedness also had low emotional availability (Lok & McMahon, 2006; Trapolini, Ungerer & McMahon, 2008). No previous studies have investigated the relation between parental mentalizing (operationalized as parental reflective functioning) and emotional availability in parent-child interaction. We wanted to investigate how parental mentalizing regarding their children was related to behavior, more specifically emotional availability in actual interactions between parent and child.

Methods.

Forty mother-child dyads were observed in a play setting. Children were aged 3-10 years. Mothers were interviewed with Parent Development Interview (PDI) and a new interview concerning limit setting (RF LS). Interviews were rated for RF and observations were rated with Emotional Availability Scales (EAS) by independent raters.

Results.

Levels of RF and EAS were in line with expectations in a non-clinical sample (RF on PDI $M = 5.20$, $SD = 1.49$) and all EAS subscales were in the Non-risk or Optimal zones). The mean RF LS was 4.26 ($SD = 1.25$) which was approximately 1 point lower

than RF on PDI. Moderate correlations between the EAS subscale Sensitivity and RF on PDI ($r = .38$) and between EAS Sensitivity and RF on Limit-Setting ($r = .47$) were found. Regression analyses showed that on average 13% of the variance in EA could be accounted for by the RF ratings and that RF LS was far more predictive for EAS than was RF PDI.

Discussion.

Mothers with higher RF demonstrated greater sensitivity to their children's emotional cues, which was consistent with expectations. Interestingly the RF concerning limit setting was more predictive of mothers' behavior in the play situations than was the RF on PDI. This may be due to the similarities between the play situation and the content of the limit setting interview. Another possible interpretation is that the questions about limit setting were more demanding for the mothers than the PDI questions (as indicated by the lower RF mean) and thus a better indicator of how mothers mentalize when they actually interact with their children. The difference between RF PDI and RF LS in predictability of the emotional availability supports the idea of mentalizing as a set of abilities rather than a homogenous skill. Limitations of the study were the small sample size, the large range of age among the children and the fact that only mothers were included, which restricts the possibility for generalization of the results.

Article 3

Mentalization- based therapy adherence and competence stimulates in- session mentalization in psychotherapy for borderline personality disorder with co-morbid substance dependence

Several studies of mentalizing in therapy sessions (Karlsson & Kermott, 2006; Bernbach, 2002) have suggested that the patient adapts to therapist style and interventions, thus not necessarily displaying her or his highest ability to mentalize during therapy. However, in the manual of mentalization-based therapy (Bateman & Fonagy, 2006), it is clearly stated that the therapist should aim at regulating patient arousal in order to create the best possible conditions for the patient to mentalize as well as aiming at stimulating an increasing ability to explore and reflect upon mental states in self and others. This study aimed at exploring whether adherence to and competence in the MBT method promote mentalizing in therapy sessions, how mentalizing in a therapy session is related to mentalizing in an interview, and how a therapist's interventions may facilitate patient mentalizing in a therapy session.

Methods.

In order to study these research questions, we applied the RF manual to therapy sessions from MBT treatment with borderline patients with comorbid substance abuse (Philips, Wennberg, Konradsson & Franck, 2018). Two sessions from 15 unique therapies, one from the beginning (2 months) and one from the middle of treatment (6 months) were used. Every therapist statement was classified as either a demand or permit statement, i.e. whether the therapist asked the patient to reflect on mental states or not, and every patient statement was rated for RF. Interrater reliability for session RF was excellent ($ICC = .86$). Adherence and competence to the MBT manual was rated, using the MBT Adherence and Competence Scale, rendering one score of adherence and one of competence on a scale between 1-7 for each session (Karterud et al., 2013). Interrater reliability for adherence and competence was excellent ($ICC = .87$ and $.89$ respectively).

Results.

Adherence and competence in MBT predicted higher levels of patient mentalizing in session, even when controlling for pre-treatment RF. Demand questions (i.e. questions directed at exploring mental states) from the therapist also predicted higher RF levels in subsequent patient statements compared to responses to permit questions. Mean session-RF at two months was strongly correlated with pre-treatment RF ($r = .64$) and with mean session-RF at six months ($r = .60$). Mean session-RF in session was lower both at two months ($M = 1.81$, $SD = 1.15$) and six months ($M = 1.73$, $SD = 1.14$) than pre-treatment RF ($M = 3.13$, $SD = 1.19$). Adherence and competence levels were low on average ($M = 3.17$, $SD = 1.05$ for adherence and $M = 3.53$, $SD = .90$ for competence). Cutoff for adequate adherence and competence is 4 (Karterud & Bateman, 2011).

Discussion.

The results provide empirical support for the notion of MBT principles, that work directed towards stimulating mentalizing actually has an effect on the level of mentalizing. The strong correlation between RF in session at 2 and 6 months and pre-treatment RF suggests that a component of RF is stable over time. The higher level of pre-treatment RF (compared to session-RF) indicates that the patients do not access or make full use of their capacity for mentalizing. The reason may be that that patient arousal level was too high and that dysregulation hampered explicit mentalizing, that implicit mentalizing is more profoundly relied upon in therapy than explicit mentalizing or perhaps that the patients adapted to the way their therapists conducted

treatment and did not get enough opportunity to enhance their mentalizing in session. Yet another reason for the lower in session-RF scores might be that in session-mentalizing is something qualitatively different from mentalizing in an interview setting. The RF interview is designed to demand explicit mentalizing from the patient and therefore is composed of many demand questions, whereas the therapy session has both multiple aims as well as several ways (in addition to demand questions) to stimulate mentalizing. Limitations of the study were the small sample and the suboptimal level of MBT adherence and competence in the sessions.

Article 4

Mentalizing in therapeutic interaction: a relational perspective

The key question studied here concerned whether mentalizing should be viewed as an individual competence or whether it in nature is an interactional phenomenon, since this question has a bearing on the methodology of studying mentalizing in therapy sessions. A new instrument for rating dyadic RF was piloted, in which the level of explicit mentalizing within the therapeutic dyad was rated.

Methods.

Two brief relational therapies for depression were selected, both conducted by the same therapist. Each session of each therapy (in total 31 sessions) was divided into 20 episodes and each episode was rated by two raters together who were blind to outcome. Both raters had previously trained in rating dyadic RF and their interrater reliability was calculated on two different sets of therapy sessions not included in the study and found to be moderate to good ($ICC^1 = .71$, $ICC^2 = .77$).

Results.

It was possible to rate dyadic RF. The main finding was that a significant, lagged correlation between mentalizing and depressive symptoms was found in one of the therapies. Higher dyadic RF in one session was followed by reduced level of depressive symptoms in the next session and lowered RF was followed by increasing levels of depression. This correlation was not found in the other therapy, where RF levels were low and rather stable and rated as overanalytic/hyperactive. Despite the poor quality of RF in this therapy, the treatment was successful in significantly reducing depressive symptoms.

Discussion.

The attempt to rate dyadic RF was partly successful. It was possible to attain acceptable inter-rater reliability. The treatment process in two cases could be understood in a meaningful way by applying the mentalization framework. Restricted generalizability is an obvious study limitation, given the use of only two cases. Still, in one of the cases an interesting correlation between dyadic mentalizing and levels of depressive symptoms was found, which implies the need for further investigation of this possible mechanism of change in therapy.

GENERAL DISCUSSION

Discussion of findings

The aims of this thesis were to investigate the concept of mentalizing as theorized by the Fonagy group and to analyze its operationalization in different forms of reflective functioning scales. The four included studies approached the concept of mentalizing from different angles. One question concerned the utility of studying mentalizing as a trait capacity and as a state in interactions (i. e. offline and online mentalizing). In the first study of RF in relation to criminal behavior, mentalizing was treated as a trait and investigated as to how it correlated with other traits such as psychopathy, empathy and alexithymia. The study used two different interview-based measures of RF, both investigating offline mentalizing. The findings were mixed. As expected, a low level of RF was found in the sample. But, contrary to expectations, there was no correlation between the nature of the crime and the level of RF. The only correlation with the other measures was in the opposite direction of what was expected; higher levels of alexithymia correlated with higher levels of RF. In the article, different levels of mentalizing about one's own criminal behavior are exemplified, showing that the crime-specific RF interview indeed yielded a diversity in levels of explicit mentalizing. The results raised questions concerning the built-in difficulties of measuring concepts that intend to capture aspects of the person that are opaque to conscious reflection like alexithymia and empathy with self-assessment questionnaires as well as questions about what the RF scores signify, and how the ability to mentalize translates into actual behavior. Would it be possible to influence mentalizing ability in criminal offenders and thus reduce the risk for impulsive behavior and increase the ability to form and sustain close relationships? Several important research steps are needed before that question can be answered, for example, investigating the link

between mentalizing ability and behavior as well as studies concerning the link between interventions aimed at enhancing mentalizing ability and outcome in terms of prosocial behavior.

The second study had a similar approach with two different measures of RF, aimed at capturing offline mentalizing. Mothers were interviewed about their relationship with their child as well as about limit-setting situations in parenthood. Their RF scores were compared to ratings of their emotional availability in actual interactions with their child, a behavioral indicator of online mentalizing. In this study, mentalizing was considered an ability (or abilities) that influences (or possibly gets influenced by) interactions in attachment relationships, thus more trait than state but still a context-dependent set of skills. The shorter limit setting interview scores showed stronger correlations with emotional availability than did RF on the PDI scores. The reason may be that a more focused reflection task better captures mentalizing that is associated with the quality of the interaction with the child than the general parental reflective functioning rated on PDI. Another possible explanation for the stronger association between LS RF and EAS is that limit setting situations might constitute stronger tests of parental mentalizing ability than general RF about the child, since limit setting situations involve more of the parent's own inner states and potentially conflicting feelings than do demands to mentalize a child's inner world in general. Studies of intervention programs for substance-dependent mothers and their babies have shown that improving mothers' ability to mentalize their own feelings and thoughts has greater impact on their caregiving behavior than interventions aimed at stimulating mentalizing about the child (Suchman et al., 2011; Suchman et al., 2010) which may be interpreted as support for the notion that parents who have conflicting or difficult feelings about themselves and their relations to others need to regulate them before being able to care for their child. Attachment theory and research is focused mostly on a child in need of comfort and support, which may be a situation that is not experienced as distressing to the parent her/himself as much as a limit setting situation may be. Thus, research on mentalizing in limit setting situations may extend and enrich attachment theory.

The two last studies focus on the possibility to rate the state quality of mentalizing in therapeutic interaction, online mentalizing. As Davidsen and Fosgerau (2015) point out, almost all empirical studies of mentalizing focus on verbal behavior, downplaying the importance of the implicit domain. Interestingly, when revisiting the manual for rating RF, it is clearly stated that RF refers to "an automatic procedure, unconsciously invoked in interpreting human action...an overlearned skill, which may

be systematically misleading in a way much more difficult to detect and correct than mistakes in conscious attributions would be... [that] lends a shape and coherence to self-organisation which is entirely outside awareness” (Fonagy et al., 1998, p. 10). This description of RF in the manual is at odds with the fact that almost all studies of mentalizing leave the implicit processes aside. Or rather, the idea of what RF ought to signify does not match the way the results of the studies are often discussed.

Discussion of methods

RF is described as an individual capacity that has extensive impact on communication and relational abilities as well as profound significance in self-organization and affect regulation (Fonagy et al., 1998). At the same time, its development and advancement is dependent on the mentalization of a significant other. Mentalization is fundamental to the capacity for intersubjective experiencing, since a crucial aspect of psychological well-being is to be able to appreciate one’s own feelings, experiences and wishes as valid even though others might not agree or behave in the way one desires, and to be able to reflect on them in interaction with other persons. The parent-child relationship incorporates numerous instances of wishes not realized or needs not met, where the frustrated child gets disappointed with her/his parent. The hallmarks of a secure attachment relationship are that these inner experiences are recognized and validated, are given space and meaning psychologically, in for example a mirroring and comforting interactive process. In insecure attachment relationships, negative feelings such as frustration, anger or sadness may be experienced as threatening the closeness to the parent. The child may learn that the regulation of such feelings is not (entirely) possible within parent-child interaction, leaving the child to manage too much on her/his own. Theoretically this can be translated to the therapeutic relationship, where the therapist’s ability to function as a secure attachment figure lay the foundation for developing mentalizing skills (Safran, 2012). The way Safran describes the therapy process as exploring transference and countertransference as well as identifying ruptures and managing these, is the arena where intersubjectivity can be experienced and both explicit and implicit relational knowledge may develop. Naturally, the roles of therapist and patient differ, and there are at least two different angles to investigate the therapeutic interaction from. Either the object of study is the patient and how s/he reacts to interventions made by the therapist in order to encourage and stimulate mentalization growth, or the object of study is the interaction itself and how

intersubjectivity is achieved, and how that relates to symptom relief. I would say there are strengths and weaknesses to each approach, as discussed below.

In the third study, we studied how patients responded to therapist interventions. An obvious advantage with this procedure is that it may be easier to get good interrater reliability. The rating of every patient statement is a distinct and systematic procedure that is easy to replicate and evidently can be done with high interrater reliability. This way of applying the RF scale “makes sense”, in that it mimics the way the AAI interviews are rated for mentalizing and the same categories can be applied without stretching the principles of the original RF manual. The analyses in this paper are fairly strong for causal inferences concerning the effect of therapist interventions. There are disadvantages to this methodology, however. One of them is the heavy reliance upon the verbal statements of the patients, downplaying the implicit aspects of the therapeutic interaction. For example, when a patient responds to a demand question with a simple “I don’t know”, it would be rated as an RF 1 reply, suggesting lack of mentalizing. As mentioned above, this might also be an example of a situation where the patient anticipates what the therapist aims at and resists to go there, which in fact would indicate at least moderately advanced mentalizing skills (albeit, perhaps less advanced interpersonal skills). Unless this is explored by the therapist in a way that allows the patient safety enough to explicitly criticize, disagree or resist the therapist’s idea, the RF rater runs the risk of simplifying a complex interpersonal situation. The results showed low levels of RF in session, lower than the reflective capacity the patients displayed on the interviews conducted before treatment, which left us with questions concerning to what degree the patient statements reflected adjusting themselves to their therapists. Another possibility is that the different measurement methods (RF on an interview versus RF in session) reflect the difference between online and offline mentalizing, which may be rather different skills. The rating procedure in this paper takes the impact of the therapist into consideration, in the sense that interventions are categorized as either demand or permit (other ways of categorizing therapist statements could of course also be possible) and their role in stimulating mentalizing is investigated. The study showed that the therapist impact is important and that the therapists varied a lot in adherence to the treatment manual. However, this approach leaves little room for understanding why therapists did not follow the manual better or why their interventions sometimes appeared to hamper mentalizing rather than stimulate it.

It was striking that therapists did not follow the treatment manual very closely. How can that be? Does the proscribed way to intervene interfere with something else

that the therapists feel more natural to engage in? Perhaps there are other interactional cues that allure therapists to less adherence with the manual despite all the circumstances pointing them in the direction of close adherence (being videotaped, being invested in the therapy method, being part of a research project, receiving supervision etc.). What, then, are they doing instead? And is that helpful to the patient? In our third study we only classified all such observations as non-adherent due to the need to keep to our research questions, but closer study of what therapists actually do when they do not follow the manual can render answers to these questions.

In the fourth study, we tried another approach, treating mentalizing as an interactive phenomenon that could be investigated in the therapeutic dyad. It aligns with relational psychoanalytic lines of thought about what impacts the patient-therapist relationship and therapeutic change. This procedure could be seen as improving the validity of measuring reflective functioning on non-structured clinical material. However, there were difficulties in achieving high levels of interrater reliability. This may be a consequence of the attempt to operationalize aspects of intersubjectivity. It was found difficult to pinpoint observational behaviors that reflect partially implicit interactions. How do the observer know when the patient and the therapist experience an intersubjective mode, of mutual recognition while at the same time feeling validated and understood as individual subjects, when there are “feelings in the air” to phrase it in Mitchell’s words (Mitchell, 2000, p. 66)? There is a built-in paradox in such a pursuit, since an outside observer cannot participate in the felt experience of the interactional field between the two parties s/he observes. She or he can only detect signs on the outside and needs to infer the inner states in a reliable manner. The attempt taps into the humbling reality of the therapy profession itself, where therapists can never know what a patient experiences or what is going to be helpful, but still needs to strive towards the goal of helping her or him.

Another aspect of implicit interaction is that of the narrative structure of the attachment interviews that are used to determine the interviewee’s ability to mentalize. The interview questions are designed to elicit attachment related affects and reflections about the impact of parents’ behaviors on one’s own personality. For the interviewees in prison, in our first study, the context was that they were incarcerated for their actions and the interviewer asked them to provide reasons for their behavior. Remarkably often the invitation to discuss connections between childhood and their current situation was rejected. When using the RF scale, this is interpreted as avoiding to mentalize due to hostility to the act of mentalizing itself, but may of course be due to other reasons, such as wanting to avoid a “blame the parents”-discourse, or wanting

to position oneself as certain type of character who is resistant to influences from others (in criminology research “the bogeyman” is someone who is essentially evil and stands beyond impact of people trying to save or change him). A structured interview like AAI or the crime-focused interview directs the interviewee explicitly and implicitly towards causally connecting different aspects of her or his history. There is a need for studies to investigate the structure of the conversation and how the interviewees handle implicit demands of preferred answers in order to discover other possible ways to interpret and understand the answers beyond the rules set by the rating system (in this case the RF scale and mentalization theory). The same can be said about the PDI and Limit setting interviews, to what degree do the interview questions allow for parents to position themselves outside the preferred parenthood of attachment theory and society’s norms of what a good parent is? It would be very interesting to see if there are differences in how men and women respond to the questions of separation from their child and how (in?)-sensitive the RF scale and the LS scale are to different ways of handling limit setting in different cultural contexts.

Clinical implications and future research

The findings in this thesis have implications for what therapists should strive for in their training and professional development. Safran’s (2012) ideas of skillfulness in therapy are interesting to consider in this context. Safran refers to D. Schön’s studies on different types of professional rationality (Schön, 1983). He argues that “technical rationality” (professional knowledge as a sort of scientific problem solving) is used by experts in a range of professional fields. However, many therapists argue that *reflection in action*, rather than technical problem solving skills may differentiate skilled professionals from novice ones. Safran describes reflection in action as “an ongoing appraisal of the evolving situation in a rapid, holistic, and (at least partially) tacit fashion. It involves a reflective conversation with the relevant situation that allows for modification of one’s understanding and actions in response to ongoing feedback” (Safran, 2012, p. 53). Within psychotherapy sessions, this calls for the ability to read the patient and attend to both implicit and explicit feedback on what impact interventions have as well as having a flexible, self-reflexive therapeutic stance in order to adjust to the details and subtle changes in the therapeutic relationship, which by the terminology of Fonagy and colleagues might be described as mentalizing the therapeutic relationship.

If the notion of mentalizing as a dyadic phenomenon is to be pursued, there is a need to further develop the instrument of RF in dyadic interaction and use it on clinical material to better assess both validity and reliability. Another interesting possibility would be to operationalize therapeutic dialogues in terms of what type of process that unfolds, for example by classifying interactions as non-explorative, validating, explorative, soothing or defense-oriented (L-G., Lundh, personal communication, June 1st, 2018). Yet another angle that needs further exploration is that of investigating therapeutic interaction using conversation analysis, to answer questions of what kind of interventions actually promote or stimulate mentalizing and what hampers it from the empirical bottom-up-perspective rather than to try to match interactions with existing theory of how mentalization-stimulating therapy ought to be conducted.

On a more clinically relevant note, there is a clearly under-investigated area of research relating to criminal behavior and mentalizing skills, both in terms of whether improved mentalizing has implications for relapsing in criminal acts and how propensity to mentalize can be stimulated within treatment programs conducted during imprisonment.

Both the first and second study showed stronger correlations between context-specific mentalizing and behavior than between general mentalizing ability and specific behaviors (committing offences or interacting with one's child). This implicates that mentalizing is more of a complex set of skills than a general ability or perhaps that implicit and explicit mentalizing are two different capacities rather than two sides of the same coin. It might be that explicit mentalizing has less impact on actual behavior (and is more relevant in making sense of and reflecting upon past experiences) and that implicit mentalizing and the preconscious interpretation of interaction that implicit mentalizing amounts to, is more influential in real life situations. In order to investigate these ideas further there is an obvious need for operationalizations of implicit mentalizing.

References

- Allen, J. G. (2003). Mentalizing. *Bulletin of the Menninger Clinic*, 67, 91-112.
- Allen, J. G. (2005). *Coping with trauma. Hope through understanding* (2nd ed.). Washington/DC: American Psychiatric Publishing.
- Allen, J. G. (2013). *Mentalizing in the development and treatment of attachment trauma*. London: Karnac.
- Allen, J. G., Fonagy, P., & Bateman, A. (2008). *Mentalizing in clinical practice*. Washington/DC: American Psychiatric Publishing.
- Aron, L. (1996). *A meeting of minds: Mutuality in psychoanalysis*. Hillsdale, NJ: Analytic Press.
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Boston: MIT Press/Bradford Books.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the mind in the eyes” test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology & Psychiatry*, 42, 241-251.
- Barreto, J. F., & Matos, P. M. (2018). Mentalizing countertransference? A model for research on the elaboration of countertransference experience in psychotherapy. *Clinical Psychology & Psychotherapy*, 1-13. <https://doi.org/10.1002/cpp.2177>
- Bateman, A. W., & Fonagy, P. (1999). Effectiveness of partial hospitalization in the treatment of borderline personality disorder: A randomized controlled trial. *The American Journal of Psychiatry*, 156(10), 1563–1569.
- Bateman, A., W., & Fonagy, P. (2006). *Mentalization-based treatment for borderline personality disorder: A practical guide*. Oxford/UK: Oxford University Press.
- Bateman, A., W., & Fonagy, P. (2008). 8-Year follow-up of patients treated for borderline personality disorder: Mentalization-based treatment versus treatment as usual. *The American Journal of Psychiatry*, 165(5), 631–8.
- Bateman, A. W., & Fonagy, P. (2012). *Handbook of mentalizing in mental health practice*. Washington/DC: American Psychiatric Publishing.
- Beebe, B., & Lachmann, F. (1988). The contribution of mother-infant mutual influence to the origins of self and object representations. *Psychoanalytic Psychology*, 5, 304–337. doi: 10.1037/0736-9735.5.4.305
- Benjamin, J. (1988). *The bonds of love. Psychoanalysis, feminism, and the problem of domination*. New York: Pantheon Books.
- Benjamin, J. (1990). Recognition and destruction: An outline of intersubjectivity. I S. A. Mitchell, & L. Aron (Eds.) *Relational Psychoanalysis: The emergence of a tradition* (pp. 181-210). New York: The Analytic Press.
- Bernbach, E. (2002). *Reflective functioning and the therapeutic relationship: Understanding*

change in brief relational therapy. Dissertation Abstracts International: Section B: The Sciences and Engineering (Vol. 62). Ann Arbor, MI: ProQuest Information & Learning.

- Bernier, A. & Dozier, M. (2003). Bridging the transmission gap: The role of maternal mind-mindedness. *International Journal of Behavioral Development*, 27, 355-365.
- Berthoz, S., & Hill, E. L. (2005). The validity of using self-reports to assess emotion regulation abilities in adults with autism spectrum disorders. *European Psychiatry*, 20, 291-198.
- Bo, S., Sharp, C., Beck, E., Pedersen, J., Gondan, M., & Simonsen, E. (2017). First empirical study of outcomes for mentalization-based group therapy for adolescents with BPD. *Journal of Personality Disorders: Theory, Research, and Treatment*, 8, 396-401.
- Boldrini, T., Nazzaro, M. P., Damiani, R., Genova, F., Gazzillo, F., & Lingardi, V. (2018). Mentalization as a predictor of outcome of psychoanalytic outcome: An empirical study of transcribed psychoanalytic sessions through the lenses of a computerized text analysis measure of reflective functioning. *Psychoanalytic Psychology*, 35, 196-204.
- Bowlby, J. (1988). *A secure base: Clinical applications of attachment theory*. London/UK: Routledge.
- 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, 1127-1135.
- Corriveau, K. H., Harris, P. L., Meins, E., Fernyhough, C., Arnott, B., Elliot, L., Liddle, B., Hearn, A., Vittorini, L., & de Rosnay, M. (2009). Young children's trust in their mother's claims: Longitudinal links with attachment security in infancy. *Child Development*, 80, 750-761.
- Davidsen, A. S., & Fosgerau, C. F. (2015). Grasping the process of implicit mentalization. *Theory & Psychology*, 25, 434-454.
- Dennett, D. (1987). *The intentional stance*. Cambridge/MA: The MIT Press.
- Dimaggio, G. (2015). Awareness of maladaptive interpersonal schemas as a core element of change in psychotherapy for personality disorders. *Journal of Psychotherapy Integration*, 25, 39-44.
- Dimaggio, G., Semerari, A., Caricone, A., Nicolò, G., & Procacci, M. (2007). *Psychotherapy of personality disorders: Megacognition, states of mind and interpersonal cycles*. London, UK: Routledge.
- Dimaggio, G., Procacci, M., Nicolò, G., Popolo, R., Semerari, A., Carcione, A., & Lysaker, P. H. (2007). Poor metacognition in narcissistic and avoidant personality disorders: Four psychotherapy patients analysed using the metacognition assessment scale. *Clinical Psychology and Psychotherapy*, 14, 386-401.

- Domes, G., Czeschnek, D., Weidler, F., Berger, C., Fast, K., & Herpertz, S. C. (2008). Recognition of facial affect in borderline personality disorder. *Journal of Personality Disorders*, 22, 135-147.
- Ekeblad, A., Falkenström, F., & Holmqvist, R. (2016). Reflective functioning as a predictor of working alliance and outcome in the treatment of depression. *Journal of Consulting and Clinical Psychology*, 84, 67-78.
- Fischer-Kern, M., Doering, S., Taubner, S., Hörz, S., Zimmermann, J., Rentrop, M., Schuster, P., Buchheim, P., & Buchheim, A. (2015). Transference-focused psychotherapy for borderline personality disorder: Change in reflective function. *The British Journal of Psychiatry*, 207, 173-174.
- Fonagy, P., & Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy*, 51, 372-380.
- Fonagy, P., & Bateman, A. W. (2006). Mechanisms of change in mentalization-based treatment of BPD. *Journal of Clinical Psychology*, 62, 411-430.
- Fonagy, P., Bateman, A. W., & Luyten, P. (2012). Introduction and overview. In A. Bateman & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice*, (pp. 3-42). Washington, DC: American Psychiatric Association.
- Fonagy, P., Gergely, G., Jurist, E., & Target, M. (2002). Affect regulation, mentalization, and the development of the self. New York: Other Press.
- 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, 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., Luyten, P., Moulton-Perkins, M., Lee, Y-W., Warren, F., Howard, S., Ghinal, R., Fearon, P., & Lowyck, B. (2016). Development and validation of a self-report measure of mentalizing: The reflective functioning questionnaire. *PLoS One* 11(7): e0158678. Doi: 10.1371/journal.pone.0158678
- Fonagy, P., Steele, M., Steele, H., Moran, G. S., & Higgitt, A. C. (1991). The capacity for understanding mental states: The reflective self in parent and child and its significance for security of attachment. *Infant Mental Health Journal*, 12, 201-218.
- Fonagy, P., & Target, M. (2005). Bridging the transmission gap: An end to an important mystery of attachment research? *Attachment & Human Development*, 7, 333-343.
- Fonagy, P., Target, M., Steele, H., & Steele, M. (1998). *Reflective-functioning manual Version 5. For application to Adult Attachment Interviews*. London: University College London.

- Fossati, A., Borroni, S., Dziobek, I., Fonagy, P., & Somma, A. (2018). Thinking about assessment: Further evidence of the validity of the movie for assessment of social cognition as a measure of mentalistic abilities. *Psychoanalytic Psychology*, 35, 127-141.
- George, C., Kaplan, N., & Main, M. (1985). *The Adult Attachment Interview, unpublished protocol* (3rd ed.). Berkeley: Department of Psychology, University of California.
- Goldstein, R. E. (2015). *Promoting reflection in psychotherapy: Patient attachment, patient and trainee therapist reflective functioning, affective word use, and symptom change in psychodynamic therapy sessions* (Unpublished doctoral dissertation). Department of psychology, Long Island University: New York.
- Greenberg, J. R., & Mitchell, S. A. (1983). *Object relations in psychoanalytic therapy*. Cambridge, MA: Harvard University Press.
- Grienberger, J. F., 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.
- Gullestad, F. S., & Wilberg, T. (2012). Change in reflective functioning during psychotherapy: A single-case study. *Psychotherapy Research*, 21, 97-111.
- Hare, R. D. (1991). *The Hare psychopathy checklist, revised*. Toronto, Canada: Multi-Health Systems.
- Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review*, 20, 290-322.
- Hörz-Sagstetter, S., Mertens, W., Isphording, S., Buchheim, S., & Taubner, S. (2015). Changes in reflective functioning during psychoanalytic psychotherapies. *Journal of the American Psychoanalytic Association*, 63, 481-509.
- Josephs, L., Anderson, E., Bernard, A., Fatzer, K., & Streich, J. (2003). Assessing progress in analysis interminable. *Journal of American Psychoanalytic Association*, 52, 1185-1215.
- Jurist, E. (2010). Mentalizing minds. *Psychoanalytic Inquiry*, 30, 289-300.
- Karlsson, R., & Kermott, A. (2006). Reflective-functioning during the process in brief psychotherapies. *Psychotherapy: Theory, Research, Practice, Training*, 43, 65–84. <http://doi.org/10.1037/0033-3204.43.1.65>
- Karterud, S., & Bateman, A. (2011). *Mentaliseringsbaserad terapi – manual och bedömnings skala*. Stockholm: Liber.
- Karterud, S., Pedersen, G., Engen, M., Selsbakk Johansen, M., Johansson, P. N., Schlüter, C., Urnes, O., Wilberg, T., & Bateman, A. W. (2013). The MBT adherence and

- competence scale (MBT-ACS): Development, structure and reliability. *Psychotherapy Research*, 23, 705-717.
- Keselman, H., Kullgard, N., Holmqvist, R., & Osvaldsson Cromdal, K. (2016). Responding to mentalization invitations in psychotherapy session: A conversation analysis approach. *Psychotherapy Research*, 1, 1-16. Doi: 10.1080/10503307.2016.1219422
- King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., & Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science*, 321, 806-810.
- 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.
- Kullgard, N., Persson, P., Möller, C., Falkenström, F., & Holmqvist, R. (2013). Reflective functioning in patients with obsessive-compulsive disorder (OCD): Preliminary findings of a comparison between reflective functioning (RF) in general and OCD-specific reflective functioning. *Psychoanalytic Psychotherapy*, 27, 154-169.
- Levinson, A., & Fonagy, P. (2004). Offending and attachment: The relationship between interpersonal awareness and offending in a prison population with psychiatric disorder. *Canadian Journal of Psychoanalysis*, 12, 225-251.
- Levy, K., Meehan, K. B., Kelly, K. M., Reynoso, J. S., Weber, M., Clarkin, J. F., & Kernberg, O. F. (2006). Change in attachment patterns and reflective function in a randomized control trial of transference-focused psychotherapy for borderline personality disorder. *Journal of Consulting and Clinical Psychology*, 74, 1027-1040.
- Lok, S. M., & McMahon, C. A. (2006). Mothers' thoughts about their children: Links between mind-mindedness and emotional availability. *British Journal of Developmental Psychology*, 24(3), 477-488.
- Luyten, P., & Fonagy, P. (2015). The neurobiology of mentalizing. *Personality Disorders: Theory, Research, and Treatment*, 6, 366-379.
- Luyten, P., Fonagy, P., Lowyck, B., & Vermote, R. (2012). The assessment of mentalization. In A. Bateman & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice*, (pp. 43-65). Washington, DC: American Psychiatric Association.
- Lyons-Ruth, K. (1999). The two-person unconscious: Intersubjective dialogue, enactive relational representation, and the emergence of new forms of relational organization. *Psychoanalytic Inquiry*, 19, 576-617.
- Lyons-Ruth, K. (2006). The interface between attachment and intersubjectivity: Perspective from the longitudinal study of disorganized attachment. *Psychoanalytic Inquiry*, 26, 595-616.

- Lyons-Ruth, K., Bruschweiler-Stern, N., Harrison, A. M., Morgan, A. C., Nahum, J. P., Sander, L., Stern, D. N., & Tronick, E. Z. (1998). Implicit relational knowing: Its role in development and psychoanalytic treatment. *Infant Mental Health Journal*, 19, 282-289.
- Maillard, P., Dimaggio, G., de Roten, Y., Berthoud, L., Despland, J-N., & Kramer, U. (2017). Metacognition as a predictor of change in the treatment for borderline personality disorder: A preliminary pilot study. *Journal of Psychotherapy Integration*, 27, 445-459.
- Meins, E. (1999). Sensitivity, security and internal working models: bridging the transmission gap. *Attachment & Human Development*, 1, 325-342.
- Meins, E. (2013). Sensitive attunement to infants' internal states: operationalizing the construct of mind-mindedness. *Attachment & Human Development*, 15, 524-544.
- Meins, E., Fernyhough, C., Fradley, E., & Tuckey, M. (2001). Rethinking maternal sensitivity: Mothers' comments on infants' mental processes predict security of attachment at 12 months. *Journal of Child Psychology and Psychiatry*, 42, 637-648.
- Mitchell, S. A. (2000). *Relationality: From attachment to intersubjectivity*. Hillsdale, NJ: Analytic Press.
- Morken, K., Karterud, S., & Arefjord, N. (2014). Transforming disorganized attachment through mentalization-based treatment. *Journal of Contemporary Psychotherapy*, 44, 117-126.
- Norcross, J. C. (2011). *Psychotherapy relationships that work. Evidence based responsiveness*. New York: Oxford University Press.
- Parker, J. D. A., Bagby, R. M., Taylor, G. J., Endler, N. S., & Schmitz, P. (1993). Factorial validity of the 20-item Toronto Alexitymia Scale. *European Journal of Personality*, 7, 221-232.
- Philips, B., Wennberg, P., Konradsson, P., & Franck, J. (2018). Mentalization-based treatment for concurrent borderline personality disorder and substance use disorder: A randomized controlled feasibility study. *European Addiction Research*, 24, 1-8.
- Rudden, M., Milrod, B., Aronson, A., & Target, M. (2008). Reflective functioning in panic disorder patients. Clinical observations and research design. In F. N. Busch (Ed.), *Mentalization. Theoretical Considerations, Research Findings, and Clinical Implications* (pp. 185-205). New York, NY: Analytic Press.
- Rudden, M., Milrod, B., Target, M., Ackerman, S., & Graf, E. (2006). Reflective functioning in panic disorder patients: A pilot study. *Journal of the American Psychoanalytic Association*, 54, 1339-1343. doi:10.1177/00030651060540040109

- Rutherford, H. J. V., Wareham, J. D., Vrouva, I., Mayes, L. C., Fonagy, P., & Potenza, M. N. (2012). Sex differences moderate the relationship between adolescent language and mentalization. *Personality Disorders: Theory, Research, and Treatment*. Advance online publication. doi: 10.1037/a0028938
- Safran, J. D. (2012). *Psychoanalysis and Psychoanalytic Therapies*. Washington, DC: American Psychological Association.
- Safran, J. D., & Muran, J. C. (2000). *Negotiating the therapeutic alliance. A relational treatment guide*. New York: The Guilford Press.
- Schön, D. (1983). *The reflective practitioner*, New York, NY: Basic Books.
- Slade, A. (2005). Parental reflective functioning: An introduction. *Attachment & Human Development*, 7, 269-281.
- Slade, A., Aber, J. L., Berger, B., Bresgi, I., & Kaplan, M. (2010). *Parent Development Interview Revised Short Version, PDI-S*. New York, NY: The City College of New York.
- Slade, A., Bernbach, E., Grienemberger, J., Levy, D., & Locker, A. (2005). *Addendum to reflective functioning scoring manual. For use with the parent development interview*. New York, NY: The City College of New York.
- Slade, A., Grienemberger, J., Bernbach, E., Levy, D., & Locker, A. (2005). Maternal reflective functioning, attachment, and the transmission gap: a preliminary study. *Attachment & Human Development*, 7(3), 283–98.
- Sorce, J. F., & Emde, R. N. (1981). Mother's presence is not enough: Effect of emotional availability on infant exploration. *Developmental Psychology*, 17(6), 737–745.
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, 25, 359-393.
- Sroufe, L. A., Egeland, B., Carlson, E., & Collins, W. A. (2005). *The development of the person: The Minnesota study of risk and adaption from birth to adulthood*. New York: Guilford.
- Stern, D. N. (2004). *The present moment in psychotherapy and everyday life*. New York: W. W. Norton.
- Suchman, N., DeCoste, C., Leigh, D., & Borelli, J. (2010). Reflective functioning in mothers with drug use disorders: Implications for dyadic interactions with infants and toddlers. *Attachment and Human Development*, 12, 567-585.
- Suchman, N. E., DeCoste, C., McMahon, T. J., Rounsaville, B., & Mayes, L. (2011). The mothers and toddlers program, an attachment-based parenting intervention for substance-using women: Results at 6-week follow-up in a randomized clinical pilot. *Infant Mental Health Journal*, 32, 427-449.

- Target, M. (2008). Commentary. In F. N. Busch (Ed.), *Mentalization. Theoretical considerations, research findings, and clinical applications* (pp. 261–279). New York: The Analytic Press.
- Taubner, S., White, L. O., Zimmermann, J., Fonagy, P., & Nolte, T. (2013). Attachment-related mentalization moderates the relationship between psychopathic traits and proactive aggression in adolescence. *Journal of Abnormal Child Psychology*, 41, 929-938.
- Trapolini, T., Ungerer, J. a, & McMahon, C. a. (2008). Maternal depression: relations with maternal caregiving representations and emotional availability during the preschool years. *Attachment & Human Development*, 10 (1), 73–90.
- Vehviläinen, S. (2008). Identifying and managing resistance. In A. Peräkylä, C. Antaki, S. Vehviläinen, & I. Leudar (Eds.), *Conversation analysis and psychotherapy* (pp. 120-138). Cambridge/UK: Cambridge University Press.
- Vermote, R, Lowyck, B., Luyten, P. Vertommen, H., Corveleyn, J., Verhaest, Y., Stroobants, R., Vandeneede, B., Vansteelandt, K., & Peuskens, J. (2010). Process and outcome in psychodynamic hospitalization-based treatment for patients with a personality disorder. *The Journal of Nervous and Mental Disease*, 198, 110-115.
- Vrouva, I., & Fonagy, P. (2009). The development of the mentalizing stories for adolescents (MSA). *Journal of American Psychoanalytic Association*, 57, 1174-1179.
- Ziv, Y., Aviezer, O., Gini, M., Sagi, A., & Koren-Karie, N. (2000). Emotional availability in the mother-infant dyad as related to the quality of infant-mother attachment relationship. *Attachment & Human Development*, 2 (2), 149-169.

Papers

The papers associated with this thesis have been removed for copyright reasons. For more details about these see:

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-152253>