

**What is Mentalization? The Concept and its Foundations in Developmental
Research and Social-Cognitive Neuroscience**

Peter Fonagy

Elizabeth Allison

Address for correspondence:

Peter Fonagy PhD FBA

Research Department of Clinical, Educational and Health Psychology

University College London

Gower Street

London WC1E 6BT

E-mail: p.fonagy@ucl.ac.uk

Tel: 44 20 7679 1791

Fax: 44 20 7916 8502

Chapter submitted for Nick Midgley & Ioanna Vrouva (eds) *Keeping Children in Mind:*

Mentalization-based Interventions with Children, Young People and their Families,

January 2011.

What is mentalization?

When we mentalize we are engaged in a form of (mostly preconscious) imaginative mental activity that enables us to perceive and interpret human behavior in terms of intentional mental states (e.g., needs, desires, feelings, beliefs, goals, purposes, and reasons) (Allen, Fonagy, & Bateman, 2008). Mentalizing must be imaginative because we have to imagine what other people might be thinking or feeling. We can never know for sure what is in someone else's mind (Fonagy, Steele, Steele, & Target, 1997). Moreover, perhaps counterintuitively, we suggest that a similar kind of imaginative leap is required to understand our own mental experience, particularly in relation to emotionally charged issues. We shall see that the ability to mentalize is vital for self-organization and affect regulation.

The ability to infer and represent other people's mental states may be uniquely human. It seems to have evolved to enable humans to predict and interpret others' actions quickly and efficiently in a large variety of competitive and cooperative situations. However, the extent to which each of us is able to master this vital capacity is crucially influenced by our early experiences as well as our genetic inheritance. In this chapter, we discuss the evolutionary function of attachment relationships, arguing that their major evolutionary advantage is the opportunity that they give infants to develop social intelligence, as well to acquire the capacity for affect regulation and attentional control. We outline the neurobiological substrates that may link secure attachment and mentalization or social cognition, and describe how secure attachment facilitates both the release of hormones enhancing social sensitivity in mothers and the activation of reward processing regions of the brain in mothers' interactions with their infants, even when the

infant is upset. We review evidence from the developmental literature on the social influences on attachment and mentalization. We then describe how our understanding of ourselves and others as mental agents grows out of interpersonal experience, particularly the child-caregiver relationship (Fonagy, Gergely, Jurist, & Target, 2002), and how the development of the ability to mentalize may be compromised in children who have not benefited from the opportunity to be understood and thought about in this way by a sensitive caregiver. Such individuals are then at greater risk of developing personality pathology, particularly if early neglect is compounded by trauma. Finally we offer some reflections on the challenges of mentalizing in family interactions.

Origins of the concept of mentalization

We first developed our concept of mentalization in the context of a large empirical study, in which security of infant attachment with each parent turned out to be strongly predicted, not only by that parent's security of attachment during the pregnancy (Fonagy, Steele, & Steele, 1991), but even more by the parent's ability to think about and understand their childhood relationship to their own parents in terms of states of mind (Fonagy, Steele, Moran, Steele, & Higgitt, 1991). We proposed that there was a vital synergy between attachment processes and the development of the child's ability to understand interpersonal behaviour in terms of mental states (Fonagy et al., 2002; Fonagy, Redfern, & Charman, 1997).

Alongside this empirical research, inspiration for the development of the concept of mentalization also came from psychoanalytic work with borderline patients. In an early paper effectively co-authored with George Moran, we identified the

repudiation of a concern with mental states as a key aspect of borderline psychopathology (Fonagy, 1991). The first time we used the term mentalization was in 1989 (Fonagy, 1989), influenced by the Ecole Psychosomatique de Paris, but we used the term as operationalized by developmental researchers investigating theory of mind (Leslie, 1987). The failure of mentalizing had of course been apparent to most psychoanalysts working with these patients, particularly Bion, Rosenfeld, Green, Kernberg and the North American object relations theorists. In an early paper reviewing ideas concerning mentalization in relation to classical psychoanalytic concepts, this intellectual indebtedness was carefully documented (Fonagy & Higgitt, 1989). The simple basic suggestion was that representing self and others as thinking, believing, wishing or desiring did not arrive at age 4 as a consequence of maturation; rather, it was a developmental achievement that was profoundly rooted in the quality of early relationships. Its predictable vulnerability to disappearance under stress in borderline conditions was seen as an appropriate focus for psychoanalytically oriented psychological intervention.

A second line of analytic inspiration came from work with children undertaken as part of a project to construct a manual for child analysis and subsequent work in developmental science between Mary Target and Peter Fonagy (Fonagy et al., 2002; Fonagy & Target, 1996, 2000, 2007; Target & Fonagy, 1996). This work helped us to think more deeply about the normal development of thinking or mentalizing capacity and the more primitive modes of thought that precede its emergence. In trying to map the emergence of mentalization on the basis of material from records of child analysis and clinical and research work with children in other contexts, we came up with a heuristic

map of the emergence of mentalization that turned out to be extremely valuable in understanding some qualitative aspects of the thinking of some patients in borderline states. In particular, we noticed that the concrete thinking that many have identified as a hallmark of borderline personality disorder was not dissimilar to the way a two to three year old normally tends to treat their internal experience. The equation of internal and external, treating what is inside my head as equivalent in status to what is there in the physical world, typifies the way toddlers and preschoolers often think until they acquire full mentalizing capacity. We have called this way of thinking *psychic equivalence*. At other times, they appear to be able to use the notion of mental states but paradoxically use it only when they can clearly separate it from physical reality (for example, in play). In this state of mind, which we have called *pretend mode*, thoughts and feelings can be envisioned and talked about but they correspond to nothing real. Finally, the compelling nature of physical reality is also obvious when children only impute intention from what is physically apparent. We noted that this *teleological mode* of thinking was present from a very early stage but is compelling for all of us at moments when mentalizing has ceased, when physical reassurance is demanded and required if emotion regulation is to be reinstated.

Our theory of the developmental emergence of the capacity to mentalize challenges the Cartesian assumption that the mind is transparent to itself and that our ability to reflect on our own minds is innate. We contend that optimal development of the capacity to mentalize depends on interaction with more mature and sensitive minds, and thus a consideration of the role played by attachment in this development is indispensable.

We have come to conceive of mentalization as a multidimensional construct, whose core processing dimensions are underpinned by distinct neural systems. Thus, mentalization involves both a self-reflective and an interpersonal component; it is based both on observing others and reflecting on their mental states, it is both implicit and explicit and concerns both feelings and cognitions (Lieberman, 2007; Saxe, 2006) (Fonagy & Luyten, in press-a; Luyten et al., submitted). When they are working together in optimal combination, the neural systems underpinning these components enable the child to represent causal mental states, distinguish inner from outer reality, infer others' mental states from subtle behavioral and contextual cues, moderate behavior and emotional experience and construct representations of his or her own mental states from perceptible cues (arousal, behavior, context). However, the delicate balance of these systems can be disrupted by neglect and trauma, as we will discuss below.

Attachment and mentalization

Early caregiving relationships are probably key to normal development in all mammals, including humans (Hofer, 1995). John Bowlby, the founder of attachment theory, postulated a universal human need to form close bonds. Bowlby originally proposed that the basic evolutionary function of the attachment instinct was to ensure that infants would be protected from predators (Bowlby, 1969). The baby's attachment behaviors (e.g. proximity seeking, smiling, clinging) are reciprocated by adult attachment behaviors (touching, holding, soothing) and these responses reinforce the baby's attachment behavior toward that particular adult.

However, the evolutionary role of the attachment relationship goes far beyond giving physical protection to the human infant. The infant's attachment behaviors are activated when something about his environment makes him feel insecure. The goal of the attachment system is an experience of security. Thus, the attachment system is first and foremost a regulator of emotional experience (Sroufe, 1996).

None of us are born with the capacity to regulate our own emotional reactions. As the caregiver understands and responds to the newborn infant's signals of moment-to-moment changes in his state, a dyadic regulatory system gradually evolves. The infant learns that while in the caregiver's presence he will not be overwhelmed by his emotional arousal, because the caregiver is there to help him to re-establish equilibrium. Thus, when he starts to feel overwhelmed, he will seek or signal to the caregiver in the hope of soothing and the recovery of homeostasis. By the end of the first year the infant's behavior seems to be based on specific expectations. His past experiences with the caregiver are aggregated into representational systems which Bowlby (1973) termed "internal working models".

Bowlby proposed that the internal working models of the self and others established in infancy provide prototypes for all later relationships. Because internal working models function outside of awareness, they are change-resistant (Crittenden, 1990). The stability of attachment has been demonstrated by longitudinal studies of infants assessed with the Strange Situation and followed up in adolescence or young adulthood with the Adult Attachment Interview (AAI) (George, Kaplan, & Main, 1985). Longitudinal studies have shown a 68-75% correspondence between attachment classifications in infancy and classifications in adulthood (e.g. Main, 1997). This is an

unparalleled level of consistency between behavior observed in infancy and outcomes in adulthood, although it is important to remember that such behavior may be maintained by consistent environments as well as by patterns laid down in the first year of life.

Moreover, attachment relationships play a key role in the transgenerational transmission of deprivation. Secure adults are three or four times more likely to have children who are securely attached to them (van IJzendoorn, 1995). One might wonder if such powerful intergenerational effects are genetically mediated, but evidence from behavior genetic studies offers no support for genetic transmission (e.g. Fearon et al., 2006). Parental attachment patterns predict unique variance in addition to temperament measures or contextual factors, such as life events, social support and psychopathology (Steele, Steele & Fonagy, unpublished manuscript, University College London). However, the mechanisms which ensure that securely attached mothers and fathers develop secure attachment relationships with their child have been difficult to pin down (van IJzendoorn, 1995).

Insecure, particularly disorganized, infant attachment is a risk factor for sub-optimal emotional and social development (Lyons-Ruth & Jacobvitz, 2008). But accumulating evidence shows that the developmental pathway from disorganized infant attachment to later psychological disorder is complex and sometimes circuitous. Rather than a developmentally reductionist model, moving directly from infancy to adulthood, we must envision a complex series of steps, each involving factors of risk and resilience interacting with past and future developmental phases. However, infant attachment may be a vulnerability factor that can illuminate the entire developmental process

In our view the major evolutionary advantage of attachment in humans is the opportunity it gives the infant to develop social intelligence. Alan Sroufe (1996) and Myron Hofer (2004) played a seminal role in extending the scope of attachment theory from an account of the developmental emergence of a set of social expectations to a far broader conception of attachment as an organiser of physiological and brain regulation. Attachment ensures that the brain processes that serve social cognition are appropriately organized and prepared to enable us to live and work with other people. The brain is experience-expectant (Siegel, 1999) and processes as fundamental as gene expression or changes in receptor densities can be influenced by the infant's environment (e.g. Meaney & Szyf, 2005).

Neurobiological considerations

A consideration of drug addiction can help us to understand the neurobiology of attachment. The mesencephalic dopaminergic reward system has been implicated in the process of drug addiction. Substances which lead to dopamine release in this system are also addictive (e.g. psychostimulants). Since it is unlikely that a brain system would exist specifically to serve drug and alcohol abuse, addiction is probably parasitic on a biological system which plays some other critical evolutionary role. Jaak Panksepp (1998) was the first to delineate a common neurobiology of mother–infant, infant–mother, and romantic attachment relationships, linked to the same mesocorticolimbic dopaminergic reward circuit. MacLean (1990) speculated that substance abuse and drug addiction could be understood as attempts to replace opiates or endogenous factors

normally provided by social attachments. Thomas Insel (2003) has reviewed evidence that suggests that addictive disorder and social attachment share a common neurobiology.

The dopaminergic reward processing system and the oxytocinergic system have been shown to play a key role in promoting and maintaining maternal behavior (see Fonagy & Luyten, in press-b for a detailed review of the evidence). Oxytocin (OT), a neuromodulatory hormone produced in the hypothalamus, has well-described central actions associated with the onset of maternal behavior, as well as peripheral actions in stimulating uterine contraction during labor and milk ejection during lactation. It is released in response to stimuli such as infant suckling, touch, or even the sight or sound of a nursing mother's infant. Oxytocin and vasopressin (AVP) are released by socio-sexual experience. Oxytocin receptors are located in the ventral striatum, a key dopaminergic brain region. There is some evidence to suggest that oxytocin release may facilitate dopamine release. In this way, social and maternally-related cues may be linked with dopamine-associated reinforcement pathways.

Mesocorticolimbic dopamine is an important candidate in the mediation of reward, the capacity for deferred gratification, and addiction, but is also critical for maternal behavior in rats and pair bonding in voles. A circuit linking a vasopressin-sensitive mechanism within the anterior hypothalamus (MPOA) to the VTA and the nucleus accumbens shell may be especially important for mediating the rewarding properties of social interaction. There is good evidence that the mesocorticolimbic pathways mediate pair bonding in rodents (Insel & Young, 2000). Prairie and pine voles form partner preferences and pair bonds after mating but montane and meadow voles do not form selective attachments. This seems to be because in prairie but not montane voles

mating is associated with dopamine release in the nucleus accumbens. Dopamine receptor antagonists infused into the nucleus accumbens prevent partner preferences from being formed (Wang et al., 1999). Voles research also suggests the hypothesis that mating releases OT and AVP, which amplifies the dopamine signal in the nucleus accumbens (Insel, 2003).

These suggestions from animal models have been amply confirmed by fMRI studies that demonstrate an association between functional brain activity related to attachment and cortical and subcortical sites in the human brain that contain a high density of the neurohormones OT and AVP (Swain, Lorberbaum, Kose, & Strathearn, 2007). Imaging shows that the mesocorticolimbic dopaminergic pathway is activated while processing attachment-related stimuli (e.g. Nitschke et al., 2004). The most compelling study using this paradigm comes from Bartels and Zeki (Bartels & Zeki, 2004) who, using the contrast of own vs. other child, controlled for age and familiarity, were able to demonstrate activity in almost all of the brain regions critical for the attachment-mediating neuropeptides in the human brain. As these workers had already reported an fMRI study of romantic love using a similar contrast design (Bartels & Zeki, 2000), they were able to confirm that most of the regions activated by maternal love were the same as those that were associated with romantic love.¹

To what extent can these biological systems explain differences in the quality of human attachment between mothers and infants? Some studies have shown differential brain responses in individuals whose attachment histories and attachment styles differ (e.g. Buchheim et al., 2008; Gillath, Bunge, Shaver, Wendelken, &

¹ These regions were in the striatum (the putamen, globus pallidus, caudate nucleus), the middle insula and the dorsal part of the cingulate cortex.

Mikulincer, 2005). There is evidence for example that those with insecure/dismissing attachment show reduced ventral striatum activation in responses to smiling adult faces and positive task feedback (Vrticka, Andersson, Grandjean, Sander, & Vuilleumier, 2008). In a recent study, Strathearn and colleagues (Strathearn, Fonagy, Amico, & Montague, 2009) aimed to measure differences in maternal brain reward activation and peripheral oxytocin release in response to infant cues, based on the mother's adult attachment classification, and found that first-time mothers with secure patterns of adult attachment compared with those with an insecure/dismissing pattern showed increased activation of mesocorticolimbic reward brain regions, on viewing their own infant's smiling face. This was true on viewing both happy and sad infant face cues. These findings suggest that for secure mothers, their infants' cues have a more reliable and powerful effect in motivating and reinforcing their caregiving responses.

Secure mothers also showed an increased peripheral oxytocin response while playing with their infants, and this was positively correlated with activation of oxytocinergic and dopamine-associated reward processing regions of the brain (hypothalamus/pituitary and ventral striatum) in the fMRI testing session. As the interactive play session occurred several months earlier than the fMRI testing session, it could be strongly argued that the level of oxytocin released when close to the infant is a trait (rather than state) effect. In other words, mothers who enter this new relationship with secure (rather than dismissing) attachment histories are more likely to respond consistently to contact with their child with greater oxytocin release. Because oxytocin is a neuropeptide that enhances social sensitivity (e.g. Domes, Heinrichs, Michel, Berger, & Herpertz, 2007 and see below), these mothers are likely to be relatively more responsive

to their infants' social cues as a consequence of the OT 'surge' and to elicit a positive response from their infant, which may in its turn be felt as rewarding. This means that they are more likely to gain pleasure from their interactions from their babies, potentially yielding a virtuous cycle of increasingly mutually rewarding interactive experiences.

Finally, striking differences in brain activation were seen between secure and insecure mothers in the Strathearn et al. study in response to the mothers' own infant's sad facial affect. This may be important in helping us to understand the intergenerational transmission of attachment security. Securely attached mothers continued to show greater activation in reward processing regions, while insecure/dismissing mothers showed increased activation of the anterior insula, a region associated with feelings of unfairness, pain and disgust (see review, Montague & Lohrenz, 2007). Activation of the anterior insula may signal "norm violations" (Montague & Lohrenz, 2007); insecure/dismissing mothers may cognitively appraise their infant's sadness as a violation of an "expected" neutral or positive affect state. This may lead to avoidance or rejection of negative infant cues (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003), rather than the "approach" responses seen in secure mothers. These results are consistent with a number of previously published models of the cortical organization of the attachment system. For example, a number of authors (Crittenden, 2008; Leckman et al., 2004; Strathearn, 2006) suggest that individuals with insecure/dismissing attachment are biased toward cognitive information processing, and tend to inhibit negative affective responses. Similarly, the findings are consistent with the assumption that insecure mothers experience rather than reflecting on the sadness they see on the faces of their infant and thereby potentially undermine the process of marked mirroring (see below), which we

have suggested is critical in enabling the baby to develop robust emotion understanding and the capacity for affect regulation (Fonagy et al., 2002).

Recent studies have highlighted the role played by oxytocin in facilitating social cognition. In randomized, placebo-controlled trials, intranasal oxytocin produces a broad range of social effects, including enhanced social memory, improved eye gaze when viewing faces, increased recognition and memory of facial expressions and identity, and increased manifestations of trust (e.g. Domes et al., 2007; Guastella, Mitchell, & Mathews, 2008). Evidence suggests that sensitivity to intentional states as inferred from facial expression increases under the influence of OT compared with placebo (Domes et al., 2007). Since, as we have seen, attachment encounters cause the spontaneous release of oxytocin, there is a plausible neurobiological substrate to link attachment and social cognition. Oxytocin levels have been observed to be reduced in populations where social cognition is also observed to be limited, arguably associated with social adversity. Negative early caregiving experiences have an impact on the oxytocin system in monkeys and humans, with lower CSF OT levels reported in nursery than mother-reared monkeys (Winslow, Noble, Lyons, Sterk, & Insel, 2003). Lower CSF oxytocin has been reported in at least one study of women with a history of emotional abuse and/or neglect (Heim et al., 2008). Furthermore, reduced peripheral oxytocin levels have been seen in orphanage-adopted children with histories of early neglect, who also display severe impairments in social reciprocity (Fries, Ziegler, Kurian, Jacoris, & Pollak, 2005).

Thus we can see that the formation of attachment relationships is supported by at least two neurobiological systems: (1) linking attachment experiences to reward and

pleasure, motivating the caregiver (and in all likelihood the infant as well) to seek experiences of closeness; and (2) a neurobiological system linking enhanced social understanding to the attachment context, with closer bonds triggering biological systems that are likely to enhance sensitivity to social cues. Given the availability of a neurobiological pathway, what can the psychological developmental literature tell us about the link between attachment and mentalization?

Understanding the relationship of attachment and mentalization

If attachment underpins the emergence of mentalization we would expect secure children to outperform insecure ones in this domain. Many studies support this hypothesis (measured as passing theory of mind tasks earlier) (see Fonagy & Luyten, in press-b for a review). Generally it seems that secure attachment and mentalization may be subject to similar social influences. We briefly consider some of these influences below.

Mentalizing and parenting

Two decades of research have confirmed that parenting is the key determinant of attachment security. Can aspects of parenting account for the overlap between mentalization and attachment security? In particular, does parental mentalization of the child have an influence? The mother's capacity to think about her child's mind is variously called maternal mind-mindedness, insightfulness and reflective function (RF). These overlapping attributes appear to be associated with both secure attachment and mentalization in the child (see Sharp, Fonagy, & Goodyer, 2006).

Elizabeth Meins (Meins, Ferryhough, Fradley, & Tuckey, 2001), David Oppenheim (Oppenheim & Koren-Karie, 2002) and Arietta Slade (Slade, Grienberger, Bernbach, Levy, & Locker, 2005) have all been able to link parental mentalization of the infant with the development of affect regulation and secure attachment in the child, mostly by analysing interactional narratives between parents and children. Although Meins assessed parents' quality of narrative about their children in real time (i.e. while the parents were playing with their children) and Oppenheim's group did this in a more 'off-line' manner (parent narrating videoed interaction), both concluded that maternal mentalizing was a more significant predictor of security of attachment than, say, global sensitivity. Slade and colleagues (Slade et al., 2005) also observed a strong relationship between attachment in the infant and the quality of the parent's mentalizing about the child. Low mentalizing mothers were more likely to show atypical maternal behaviour on the AMBIANCE system (Atypical Maternal Behavior Instrument for Assessment and Classification) (Bronfman, Parsons, & Lyons-Ruth, 1999), which relates not only to infant attachment disorganization but also to unresolved (disorganized) attachment status in the mother's Adult Attachment Interview (AAI) (Grienberger, Kelly, & Slade, 2005).

Taken together these results suggest that mentalizing parents might well facilitate the development of mentalization in their children. Mindful parenting probably enhances both attachment security and mentalization. However, we should bear in mind that these correlations can be as readily explained as child-to-parent as parent-to-child effects. For example, less power-assertive parenting may be associated with mentalization (Pears & Moses, 2003) not because this may facilitate independent

thinking, but because less mentalizing children may be more likely to elicit controlling parenting behavior. It may also be that the same aspects of family functioning that facilitate secure attachment also facilitate the emergence of mentalizing. .

The process of acquiring mentalization is so ordinary and normal that it may be more appropriate to consider secure attachment as providing an environment that is free of obstacles to its development rather than active and direct facilitation. The key to understanding the interaction of attachment with the development of mentalization may then be to look at instances where normally available catalysts to the development of mentalization are absent.

Family discourse

Exposure to normal family conversations appears to be a precondition of mentalization (Siegal & Patterson, 2008). Nicaraguan deaf adults who grew up without hearing references to beliefs appear to be incapable of passing false belief tests (Pyers, 2003 cited in Siegal & Patterson, 2008). Under normal circumstances, conversations in which adults and children talk about the intentions implied by each others' reasonable comments and link these to each others' appropriately interpreted actions may be the "royal road" to understanding minds. The groundbreaking work of Mary Main (2000) has linked attachment to this kind of verbal communication. Coherent family discourse characteristic of secure attachment (Hill, Fonagy, Safier, & Sargent, 2003) helps to generate explanatory schemas by means of which other people's behaviour can be understood and predicted.

Playfulness

Playfulness is another feature of a secure attachment context. Play may also be important in acquiring mentalizing. The impact of lack of playfulness is most obvious in extreme cases. Blind children's active pretend play is quite limited (Fraiberg, 1977) and they also understand pretend play poorly (Lewis, Norgate, Collis, & Reynolds, 2000). They are delayed on false belief tests and only pass when they reach a verbal mental age of 11 as opposed to the more normal five (McAlpine & Moore, 1985). Blind infants of course also miss out on access to non-verbal information about inner states. They are deprived of cues to internal states such as facial expression, and can experience problems of identity which are perhaps associated with mentalization problems (Hobson & Bishop, 2003).

Maltreatment

Maltreatment disorganizes the attachment system, as we discussed above (see Cicchetti & Valentino, 2006 for a comprehensive review). Does it disrupt mentalization? The evidence for significant developmental delay in these children's understanding of emotions is consistent, if slightly reduced when IQs and SES are controlled for (eg Frodi & Smetana, 1984; Smith & Walden, 1999). As well as problems of emotional understanding, social cognition deficits and delayed theory-of-mind understanding have been reported in maltreated children (e.g. Cicchetti, Rogosch, Maughan, Toth, & Bruce, 2003; Pears & Fisher, 2005).

As reports of maltreatment are commonly associated with diagnoses of personality disorder, let us pause for a moment to consider the apparent deficit in

mentalization in individuals with maltreatment histories. The mentalization deficit associated with childhood maltreatment may be a form of decoupling, inhibition or even a phobic reaction to mentalizing. Maltreatment can contribute to an acquired partial “mind-blindness” by compromising open reflective communication between parent and child. It might undermine the benefit derived from learning about the links between internal states and actions in attachment relationships (e.g. the child may be told that they “deserve”, “want” or even “enjoy” the abuse). This will more obviously be destructive if the maltreatment is perpetrated by a family member. Even where this is not the case, the parents’ lack of awareness that maltreatment is taking place outside the home may invalidate the child’s communications with the parent about how the child is feeling. The child finds that reflective discourse does not correspond to these feelings, a consistent misunderstanding that could reduce the child’s ability to understand/mentalize verbal explanations of other people’s actions. In such circumstances they are likely to struggle to detect mental states behind actions accurately, and will tend to see actions as inevitable rather than intended.

It would be absurd to suggest (either from a scientific or a common-sense perspective) that positive attachment experience is the only relationship influence on the development of mentalization. Negative experiences (e.g. emotionally charged conflict) may as readily facilitate the rapid development of mentalizing as positive emotions linked with secure attachment (Newton, Reddy, & Bull, 2000). The reality is that numerous aspects of relational influence are likely to be involved in the emergence of mentalizing, some of which probably correlate with secure attachment. But studies of social influence

on mentalizing have hitherto mistakenly tended to assume that this social cognitive capacity is unimodal.

As we suggested above, mentalization is probably better considered as a complex multi-component capacity with a variety of determinants, some of which are genetic while others are more influenced by environmental interference and facilitation. Each of the correlates of secure attachment we considered above may interface with one or more of a series of a range of neuropsychologically defined components of mentalizing (see Fonagy & Luyten, 2009 for a review)

The development of an agentive self: The social acquisition of social cognition

If children's caregiving environments are key to their development as social beings, how do these environmental influences have their effect? Our model relies on the child's innate capacity to detect aspects of his world that react contingently to his own actions. In his first months of life, the baby begins to understand that he is a physical agent whose actions can bring about changes in bodies with which he has immediate physical contact (Leslie, 1994). At the same time, he begins to understand that he is a social agent, as he learns that his behaviour affects his caregiver's behaviour and emotions (Neisser, 1988). Both these early forms of self-awareness probably evolve through the workings of an innate contingency detection mechanism that enables the infant to analyse the probability of causal links between his actions and stimulus events (Watson, 1994). The child's initial preoccupation with perfectly response-contingent stimulation (provided by the proprioceptive sensory feedback generated by his own

actions) allows him to differentiate himself from his environment and to construct a primary representation of his bodily self.

At about 3-4 months, infants' preference appears to change. They begin to be drawn to high-but-imperfect contingencies rather than perfect contingency (Bahrck & Watson, 1985) – the level of contingency that characterises an attuned caregiver's empathic mirroring responses to a baby's emotional displays. Repeated experience of these responses enables the baby to begin to differentiate his internal self-states: a process we have termed 'social biofeedback' (Gergely & Watson, 1996). A congenial and secure attachment relationship vitally contributes to the emergence of early mentalizing capacities by allowing the infant to 'discover' or 'find' his/her psychological self in the social world (Gergely, 2001).

At first, infants are not introspectively aware of different emotion states. Rather, their representations of these emotions are primarily based on stimuli received from the external world. Babies learn to differentiate the internal patterns of physiological and visceral stimulation that accompany different emotions by observing their caregivers' facial or vocal mirroring responses to these (e.g. Legerstee & Varghese, 2001; Mitchell, 1993). The baby comes to associate his control over the parents' mirroring displays with the resulting improvement in his emotional state, and this lays the foundations for his eventual development of the capacity for emotional self-regulation. The establishment of a second-order representation of affect states creates the basis for affect regulation and impulse control: affects can be manipulated and discharged internally as well as through action, they can also be experienced as something recognizable and hence shared.

Two conditions need to be met if the capacity to understand and regulate emotions is to develop: (a) reasonable congruency of mirroring whereby the caregiver accurately matches the infant's mental state and (b) 'markedness' of the mirroring, whereby the caregiver is able to express an affect while indicating that she is not expressing her own feelings (Gergely & Watson, 1999). If the caregiver's mirroring is incongruent, the resulting representation of the infant's internal state will not correspond to a constitutional self state, which might predispose the infant to develop a narcissistic personality structure (perhaps analogous to Winnicott's notion of a 'false self', Winnicott, 1965). If the mirroring is unmarked, the caregiver's expression may seem to externalize the infant's experience and may overwhelm the infant, making his experience seem contagious and escalating rather than regulating his state. A predisposition to experiencing emotion through other people (as in a borderline personality structure) might be established (Fonagy et al., 2002).

Affect regulation, attentional control and mentalization

The child is thought to internalize his experience of well-regulated affect in the infant parent couple to form the basis of a secure attachment bond and internal working model (Sroufe, 1996). In this account, affect regulation is a prelude to mentalization; yet, once mentalization occurs, the nature of affect regulation is transformed. Not only does mentalization allow adjustment of affect states, but more fundamentally it is used to regulate the self. The emergence of mentalizing function follows a well-researched developmental line that identifies 'fixation points':

(a) During the second half of the first year of life, the child begins to be able to grasp the causal relations between actions, their agents and the environment. At around nine months, infants begin to look at actions in terms of the actor's underlying intentions (Baldwin, Baird, Saylor, & Clark, 2001) and they begin to understand themselves as teleological agents who can choose the most efficient way to bring about a goal from a range of alternatives (Csibra & Gergely, 1998). However, at this stage agency is understood purely in terms of physical actions and constraints. Infants expect actors to behave rationally, given physically apparent goal states and the physical constraints of the situation (Gergely & Csibra, 2003). The infant does not yet have any idea about the agent's mental state. We have suggested that there is a connection between this focus on understanding actions in terms of their physical as opposed to mental outcomes (a teleological stance) and the mode of experience of agency that we often see in the self-destructive acts of individuals with borderline personality disorder (see below). For these individuals, slight changes in the physical world can trigger elaborate conclusions about states of mind and only modifications in the realm of the physical can convince them as to the intentions of the other.

(b) During the second year, children begin to understand that they and others are intentional agents whose actions are caused by prior states of mind such as desires (Wellman & Phillips, 2000) and that their actions can bring about changes in minds as well as bodies (e.g. by pointing, Corkum & Moore, 1995). At this stage the capacity for emotion regulation comes to reflect the prior and current relationship with the primary caregiver (Calkins & Johnson, 1998). Most importantly, children begin to acquire an internal state language and the ability to reason non-egocentrically about feelings and

desires in others (Repacholi & Gopnik, 1997). Paradoxically, this becomes evident not only through the increase in joint goal directed activity but also through teasing and provocation of younger siblings (Dunn, 1988). However, the child is not yet able to represent mental states independently of physical reality and therefore the distinction between internal and external, appearance and reality is not yet fully achieved (Flavell & Miller, 1998). This means that internal reality is sometimes experienced as far more compelling and at other times seems inconsequential relative to the child's awareness of the physical world. We have referred to these modes of experiencing internal reality as psychic equivalence and pretend modes respectively (see below).

(c) Around three to four years of age, the child begins to grasp that people's actions are caused by their beliefs. A meta-analytic review of over 500 tests showed that by and large children younger than three fail the false-belief task and as the child's age increases they are increasingly likely to pass (Wellman, Cross, & Watson, 2001), suggesting that mentalizing abilities take a quantum leap forward around age four. From this point, the child can understand both himself and others as representational agents. He knows that people do not always feel what they appear to feel, and that their emotional reactions to an event are influenced by their current mood or even by earlier emotional experiences which were linked to similar events (Flavell & Miller, 1998). Reaching this milestone transforms his social interactions. His understanding of emotions comes to be associated with empathic behaviour (Zahn-Waxler, Radke-Yarrow, Wagner, & Chapman, 1992) and more positive peer relations (Dunn & Cutting, 1999). His understanding that human behaviour can be influenced by transient mental states (such as thoughts and feelings) as well as by stable characteristics (such as personality or

capability) creates the basis for a structure to underpin an emerging self-concept (Flavell, 1999). His newfound ability to attribute mistaken beliefs to himself and others enriches his repertoire of social interaction with tricks, jokes and deception (Sodian & Frith, 1992; Sodian, Taylor, Harris, & Perner, 1992). Notably, at this time the child also begins to prefer playing with peers to playing with adults (Dunn, 1994). This shift brings to a close the time when mentalization was acquired through the agency of an adult mind and opens a lifelong phase of seeking to enhance the capacity to understand self and others in mental state terms through bonds with individuals who share one's interest and humour.

(d) In the sixth year, we see related advances such as the child's ability to relate memories of his intentional activities and experiences into a coherent causal-temporal organisation, leading to the establishment of the temporally extended self (Povinelli & Eddy, 1995). Full experience of agency in social interaction can emerge only when actions of the self and other can be understood as initiated and guided by assumptions concerning the emotions, desires and beliefs of both. Further theory of mind skills that become part of the child's repertoire at this stage include second order theory of mind (the capacity to understand mistaken beliefs about beliefs), mixed emotions (e.g. understanding being in a conflict), the way expectations or biases might influence the interpretation of ambiguous events, and the capacity for subtle forms of social deceptions (e.g. white lies). As these skills are acquired the need for physical violence begins to decline (Tremblay, 2000) and relational aggression increases (Cote, Tremblay, Nagin, Zoccolillo, & Vitaro, 2002).

Subjectivity before mentalization

In order to make use of this model of the emergence of mentalizing function in understanding children's subjective experience, we also need a clear conception of what the non-mentalizing mind might be like before the child fully recognizes that his internal states are mere representations of reality. It is important to grasp that at first the small child assumes that what he knows is known by others and vice versa. Our sense of the uniqueness of our own perspective develops only slowly. While infants already possess a distinct sense of their physical integrity by three months or so at the latest, we start with the assumption that knowledge is common and that our own thoughts or feelings are shared by others. Young children assume that other children will know facts that they themselves have just learned (Taylor, Esbensen, & Bennett, 1994). One reason why toddlers are so prone to outbursts of rage and frustration may be that as the world and individual minds are not yet clearly demarcated, they expect other people to know what they are thinking and feeling, and to see situations in the same way they do. Thus frustration of their wishes seems malign or wilfully obtuse, rather than the result of a different point of view or alternative priorities.

In describing the normal development of mentalizing in the child of two to five years (Fonagy & Target, 1996; Target & Fonagy, 1996), we have suggested that there is a transition from a split mode of experience to mentalization. We hypothesize that the very young child equates the internal world with the external. What exists in the mind must exist out there and what exists out there must also exist in the mind. At this stage there is no room yet for alternative perspectives: "How I see it is how it is". The toddler's or young pre-school child's insistence that "there is a tiger under the bed" is not allayed by parental reassurance. This 'psychic equivalence', as a mode of experiencing the internal

world, can cause intense distress, since the experience of a fantasy as potentially real can be terrifying. The acquisition of a sense of pretend in relation to mental states is therefore essential. While playing, the child knows that internal experience may not reflect external reality, but then the internal state is thought to have no implications for the outside world (“pretend mode”).

Normally, at around four years old, the child integrates these modes to arrive at mentalization, in which mental states can be experienced as representations. Inner and outer reality can then be seen as linked, yet differing in important ways, and no longer have to be either equated or dissociated from each other (Gopnik, 1993). However, under certain circumstances pre-mentalistic forms of subjectivity may still re-emerge to dominate social cognition years after the acquisition of full mentalization.

The impact of maltreatment on attachment and mentalization

The impact of maltreatment has had particular importance in our model. Above we outlined how *experienced* maltreatment in the context of an attachment relationship might create a vicious cycle of fearful attachment, where fear triggers proximity seeking which, once achieved, triggers further anxiety in its turn (see e.g. Bartholomew, Kwong, & Hart, 2001). In our model of intersubjective self development we have also stressed (following Winnicott) that when children are regularly unable to “find” themselves as intentional beings in the reflective stance of the caregiver, either because of neglect or difficulties in infant temperament, they have no choice but to internalize a representation of their caregiver into their self-representation. This inevitably disorganizes the self structure, creating splits, and gives rise to the well recognized phenomena associated with

disorganization of attachment, such as the manipulateness in middle childhood characteristic of individuals whose attachment was disorganized in infancy (George & Solomon, 1999). Manipulateness (and oppositional behavior as a consequence of attachment disorganization) occurs because the child prefers to externalize the alien part of the self by nudging the caregiver (gently and sometimes not so gently) to experience the angry or anxious states of mind that they have internalized, but experience as alien (Fonagy et al., 2002). Creating a situation where their own unwanted mental states can be felt to belong to somebody else enables the individual with disorganized attachment history to experience a measure of self-coherence.

Maltreatment creates a further profound complication within this mechanism, when the child uses split-off parts of the self to gain illusory control over the abuser, a process that has been richly described in the psychoanalytic literature as identification with the aggressor (Freud, 1936). When the child internalizes the mental state of the victimizer into the alien part of the self, they experience a part of their own mind as torturing, bent on the destruction of the ego. This leads to an unbearably painful emotional state where the self is experienced as evil and hateful. In these circumstances it may feel as though the only solution is to turn the attack from within the mind against the body by self-harming. Alternatively the individual with a history of maltreatment may resort to constantly externalizing the alien, torturing parts of the self-structure into a person close by. Through projective identification (as best described by Rosenfeld, 1971) the persecutory parts of the self are located in someone else. In this way the need for the alien experiences to be owned by another mind may lead to their involvement in a sequence of further abusive relationships.

The need for projective identification is a matter of life and death for those with a traumatising part of the self-structure, but the constellation creates a dependence on the object that has many of the features of addiction. This is not surprising given the overlap of the neural structures responsible for mediating intense attachment and addiction discussed earlier (Insel, 2003).

Mentalizing in families

While on the one hand we have contended that the infant's early relationship to its caregivers is the cradle of mentalizing, on the other hand there is no context more likely to induce a loss of mentalizing than family interactions. It is here that relationships tend to be at their most fraught, their most loving, and their most intense emotionally and so the stage is set on a daily basis for interactions that potentially stimulate a loss of mentalizing in one or more members of the family.

Mentalizing problems may emerge in a variety of contexts and with differing severity and presentations. These occur along a spectrum from relatively mild and specific difficulties to highly destructive non-mentalizing attitudes that may have long-term effects on a child's mentalizing capacity and well-being. Mentalizing strategies may also be under-used or applied erratically because of other demands or high levels of perceived stress, or if a family member or a relationship has a circumscribed 'blind spot'. At the more extreme end of the spectrum, one or more family members may consciously or unconsciously misuse mentalization in their dealings with others.

Sometimes specific problems with mentalizing may occur. For example, in an acrimonious parental separation, one parent who is otherwise highly sensitive to the

child's feeling states may find it particularly hard to tune into the child's feelings and thoughts about loss of the parental couple, perhaps because of ongoing hatred of the other parent. As a result the parent is unable to mentalize that aspect of his or her child's life.

Such specific loss of mentalization may be associated with stress: when exposed to great pressure, most people tend to lose their capacity to think about the thoughts and feelings of others. For example, quite dramatic temporary failures of mentalization can arise in individuals and families during emotionally intense interchanges. This can also happen merely in response to thoughts and feelings that trigger high arousal and anti-mentalizing reactions. Under such circumstances grossly inaccurate or even seemingly malevolent feelings can be attributed to others and feelings of resentment and mistrust grow in the relationship context. The representation of the mind of others can literally be obliterated and replaced by an empty or hostile image. For example, when a parent becomes convinced that their child is deliberately and maliciously provoking them, his/her mind becomes closed to seeing the child in alternative ways. Or a parent who suffered physical or sexual abuse may temporarily lose the capacity to mentalize, when faced with a reminder of their own (past) states of helplessness, anger or shame. Their child's distressed response may act as an additional reminder.

In other situations a parent may be temporarily preoccupied with other important concrete issues in their life, such as a crisis at work, and this may propel them into a non-mentalizing frame that gets carried into the family life. This state may fluctuate, preventing the parent from being attuned to the child's feeling states at certain times. In such a scenario, the child, who usually has had good experiences of feeling

thought about and understood, is suddenly confused by the parent's apparent emotional unavailability. The problem can be compounded further if the parent is incapable of appreciating the child's disappointment and confusion. Other specific family problems with mentalizing can arise from the child obscuring his/her own mental states, making the parent's task of 'mind-reading' difficult. This can happen for a wide range of reasons, such as the arrival of a new step-parent, or in the course of limited visitation rights, when the non-resident parent simply lacks the contextual information necessary to make sense of the child's state of mind. In both these cases, the parent's understandable inability to mentalize the child can nevertheless leave the child feeling that they are not understood, limiting their motivation for making themselves available to be understood.

There are background conditions which increase the frequency with which non-mentalising family interactions arise. Long standing mental health problems can compromise mentalizing in families in a number of ways. A parent with schizophrenia, particularly with repeated episodes of the illness, will find it difficult to take perspectives, develop and model a trusting attitude, or be able to take turns, and will have strong unshakable beliefs which will impede curiosity and reflective contemplation. A child in such a family may respond to this from early on by 'hyper-mentalising' – being a precocious mentalizer - as part of his/her development into a 'young carer'. Other children appear to disengage from the mental state of adults. In both scenarios, an interest in the child's own mental state decreases as a consequence of the parents impaired mentalizing.

Where a parent suffers from major depression, the child may be overactive in stimulating the parent, not into a mentalizing attitude but rather into action in order to

break through oppositional behaviours as ways of making the parent connect, even if only via disciplining and other non-mentalizing actions. Some children may adopt a stance analogous to that of their parents, shutting down and opting for not thinking as the least painful way of coping with emotional neglect. Parents with high levels of arousal, such as those with chronic anxiety states, can find themselves excessively engaged with the child's mental world, anxiously loading the child with their own anxious preoccupations. The child, who does not understand the source of severe anxiety, can be perturbed by it and search for an explanation in their actions and thoughts, engaging in turn in excessive mentalizing. In a sense, similar processes appear to take place in the parent and child, almost in parallel but failing to inform each other directly.

When these dyadic processes take place in a family context, it is inevitable that it will affect everyone, with others attempting to mentalize the relationship in question. Another family member, faced with a dyadic non-mentalizing interaction and attempting to understand the two parties, will run the risk of being only partially understood by one or both parties. Each non-mentalizing person is likely to understand only some aspects of the onlooker's stance: those which correspond to their own perspective. At the same time, each person in the dyad will feel invalidated by those aspects of the onlooker's stance which describe the mind state of the other, as though the onlooker has taken sides in opposition to him or her. In that way the onlooker is recruited to the non-mentalizing interaction as they themselves feel that they have only partially been heard by each of the protagonists. In this way, a non-mentalizing dyad becomes a triad. Gradually, the system can recruit other members of the family as well as professionals.

Another systemic perspective on mentalization failure is the experience of the individual faced with non-responsive minds. When faced with family members who cannot respond to a member's inquiring or curious mind, the person will give up, reinforcing the hopelessness of all concerned, resulting in a 'circular' or cyclic hopelessness. For example, a child who is depressed may experience her thoughts and feelings about herself as entirely real, and will be deprived of the perspective that would allow her to think differently about herself or others, as a result of the lack of interest in her state of mind that she experiences from her carers. In the absence of relational mentalizing strengths, such as curiosity or reflective contemplation and perspective taking, pessimism about the possibility of feelings changing takes over. A feeling of hopelessness is taken to be 'physical reality' by the child the moment it is experienced and it cannot be treated as 'just a thought' which could then be challenged cognitively. The parents of a depressed child or adolescent may resonate with their child's predicament (e.g. they all may feel that having few friends is a hopeless situation) or because the child's behaviour may be experienced by them as an expression of their own failure or incompetence. Just as mentalizing engenders more mentalizing, so non-mentalizing can engender further non-mentalizing in a family context.

In families with poor boundaries between the generations (often described as enmeshed – e.g. Minuchin, 1974) forms of intrusive mentalizing can take place. Here the separateness of minds is not respected within the family: a family member strongly believes s/he knows what another member thinks and feels. In such cases, the family discourse may sound as if everyone is mentalizing well, but, paradoxically, this does not have the usual consequence of people feeling understood. We have described such

interactions, which are marked by their ‘pretend’ quality, as pseudomentalization. Family members’ narratives fail to connect with each other, which may incite each family member to redouble his/her efforts to have his/her view accepted by the rest of the family. In consequence, more and more unjustified assumptions are made about other people’s mental states: family members invest a lot of energy in thinking or talking about how other family members think or feel, but their interpretations bear little or no relationship to other people’s reality. The result is that mentalization is experienced as being obstructive and confusing and it can lead to certain members of the family avoiding further mentalization efforts altogether.

If a member of the family ‘leaves the field’, becoming unavailable for mentalizing within the family system, other family members may respond in kind, demonstrating more extreme non-mentalizing by taking on a stance that directly attacks mentalization. Statements indicating this kind of extreme non-mentalizing could include: “you are trying to drive me crazy”; “your grandma is in league with your father against us”; “you provoked me”; “you don’t care about whether your Dad is here or not”; “you don’t care about me”; “you would be glad if I was dead”. Such statements generate further arousal that is incompatible with mentalization and can lead to nothing but further non-mentalizing cycles. Any attempts to discuss the meaning of such statements are almost certain to fail, as such statements only make sense in a non-mentalizing world. A therapist who attempts to question the meaning of such statements is therefore inadvertently contributing to the non-mentalizing cycle and will at best achieve pseudo-mentalization.

At the extreme end of the non-mentalizing spectrum is the misuse of mentalization. Here, understanding of mental states of self and others is not directly impaired, but is misused to further an individual's interest at the expense of the well-being of the family or one of its members. For example, a parent might use a child's current mental state (e.g. sadness) as ammunition in a marital battle (e.g. "whenever you visit your father you feel so sad afterwards, don't you think you should stop seeing him?"). In these situations the child might experience mentalization as aversive because being understood occurs in the context of being manipulated. In such cases, children's feelings are typically exaggerated or distorted in the interest of the parent's unspoken intention or attitude. Another example might be a father who claims that he objects to his wife working because it makes the children feel neglected, whereas the true cause of his objection is that his wife's work requires him to be more involved at home, leaving him less time for himself.

Another misuse of mentalization is coercion against the child's thoughts. This involves the parent undermining the child's capacity to think by deliberately humiliating the child for her or his thoughts and feelings. For example, the parent exposes the child's sexual feelings in a family gathering in a belittling and insensitive manner, disclosing what the child might have confided in private. These phenomena are most pernicious in the context of abuse, where the abusive party may falsely maintain that the child "fell down the stairs, I never hit you" or that the child "enjoyed it when I touched you like that", for instance. This kind of misuse of mentalization may undermine the child's capacity to mentalize, not simply because they directly contradict the child's own reality,

but because the child may be unable to construct a bearable image of the thoughts that the parent must have had in order to make such confusing statements.

Conclusion

In this chapter we have briefly summarized the relationship between attachment and mentalizing, suggesting that the process of mentalizing should be given central importance in child development. We suggest that a link between abnormal development of social cognition during childhood and adult psychopathology may be partially mediated through mentalizing. This suggestion implies that a focus on mentalizing process could enhance clinical practice. Unlike other integrative approaches such as interpersonal psychotherapy, mentalization based treatment has a theoretical frame of reference which includes a developmental model, a theory of psychopathology and a hypothesis about the mechanism of therapeutic action. We suggest that a focus on enhancing mentalizing may be an important factor distinguishing mentalizing therapies from other psychotherapies. The therapist is focused on mental processes and is not engaged in cognitive restructuring, he is not working to provide insight and he does not attempt to alter behavior directly. Cognitive and behavioral changes or patients coming to recognise underlying meanings or identify reasons why they are as they are occur in MBT as consequences of the change in mentalizing, rather like positive side-effects, and also because the concept is broad, almost an umbrella term for a group of basic psychological processes.

Mentalizing could be seen as one of many common factors in psychotherapy. All psychotherapies, whatever their focus, share the potential to recreate an interactional

matrix of attachment in which mentalization develops and sometimes flourishes. This perhaps acts as a catalyst for further change in cognitions, emotions and behavior, irrespective of the therapeutic target. Having a patient's mind in mind will make any therapeutic effort more efficient.

Mentalizing is a developmental construct. This raises questions about the variability not simply of mother-child interaction but of families and the significance of developmental milestones, particularly the importance of the move from childhood to adolescence. Distortions in the development of mentalizing are therefore likely to go beyond the diagnostic group of individuals with personality disorder and there may be other individuals who can benefit from having their mentalizing problems addressed directly. This opens up the possibility of preventive work during childhood. As mentalizing is a fundamental psychological process and so interfaces with all major mental disorders, mentalizing-focused approaches may have the potential to improve well-being across a range of disorders. So whilst on the one hand, our claims for mentalizing continue to be modest, on the other hand, we see it as a unifying mental process that interfaces with and can therefore interfere with a wide range of psychological functions. This suggests that regardless of an approach focusing on mentalizing being adopted in a treatment, there is a need for any practitioner to see the world from the patient's perspective, and that whenever that focus on the patient's internal mental process is dominant there is intrinsic value because of the powerful commitment to the patient's subjectivity.

References

- Allen, J., Fonagy, P., & Bateman, A. (2008). *Mentalizing in clinical practice*. Washington: American Psychiatric Press.
- Bahrnick, L. R., & Watson, J. S. (1985). Detection of intermodal proprioceptive-visual contingency as a potential basis of self-perception in infancy. *Developmental Psychology, 21*, 963-973.
- Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Dev, 72*(3), 708-717.
- Bartels, A., & Zeki, S. (2000). The neural basis of romantic love. *Neuroreport, 11*(17), 3829-3834.
- Bartels, A., & Zeki, S. (2004). The neural correlates of maternal and romantic love. *Neuroimage, 21*(3), 1155-1166.
- Bartholomew, K., Kwong, M. J., & Hart, S. D. (2001). Attachment. In W. J. Livesley (Ed.), *Handbook of Personality Disorders: Theory, Research and Treatment* (pp. 196-230). New York: Guilford Press.
- Bowlby, J. (1969). *Attachment and Loss, Vol. 1: Attachment*. London: Hogarth Press and the Institute of Psycho-Analysis.
- Bowlby, J. (1973). *Attachment and Loss, Vol. 2: Separation: Anxiety and Anger*. London: Hogarth Press and Institute of Psycho-Analysis.
- Bronfman, E., Parsons, E., & Lyons-Ruth, K. (1999). *Atypical Maternal Behavior Instrument for Assessment and Classification (AMBIANCE): Manual for coding disrupted affective communication, version 2. Unpublished manuscript*. Cambridge, MA: Harvard Medical School.
- Buchheim, A., Erk, S., George, C., Kachele, H., Kircher, T., Martius, P., et al. (2008). Neural correlates of attachment trauma in borderline personality disorder: a functional magnetic resonance imaging study. *Psychiatry Res, 163*(3), 223-235.
- Calkins, S., & Johnson, M. (1998). Toddler regulation of distress to frustrating events: Temperamental and maternal correlates. *Infant Behavior and Development, 21*, 379-395.
- Cicchetti, D., Rogosch, F. A., Maughan, A., Toth, S. L., & Bruce, J. (2003). False belief understanding in maltreated children. *Dev Psychopathol, 15*(4), 1067-1091.
- Cicchetti, D., & Valentino, K. (2006). An ecological-transactional perspective on child maltreatment: Failure of the average expectable environment and its influence on child development. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental Psychopathology*. (2nd ed., Vol. 3, pp. 129-201). New York: John Wiley & Sons.
- Corkum, V., & Moore, C. (1995). Development of joint visual attention in infants. In C. Moore & P. Dunham (Eds.), *Joint Attention: Its Origins and Role in Development* (pp. 61-83). New York: Erlbaum.
- Cote, S., Tremblay, R. E., Nagin, D., Zoccolillo, M., & Vitaro, F. (2002). The development of impulsivity, fearfulness, and helpfulness during childhood: patterns of consistency and change in the trajectories of boys and girls. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 43*(5), 609-618.
- Crittenden, P. (2008). *Raising Parents. Attachment, parenting and child safety*. Devon: Willan Publishing.

- Crittenden, P. M. (1990). Internal representational models of attachment relationships. *Infant Mental Health Journal, 11*, 259-277.
- Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science, 1*(2), 255-259.
- Domes, G., Heinrichs, M., Michel, A., Berger, C., & Herpertz, S. C. (2007). Oxytocin improves "mind-reading" in humans. *Biol Psychiatry, 61*(6), 731-733.
- Dunn, J. (1988). *The Beginnings of Social Understanding*. Oxford: Basil Blackwell Ltd and Cambridge, MA: Harvard University Press.
- Dunn, J. (1994). Changing minds and changing relationships. In C. Lewis & P. Mitchell (Eds.), *Children's Early Understanding of Mind: Origins and Development* (pp. 297-310). Hove, UK: Lawrence Erlbaum.
- Dunn, J., & Cutting, A. L. (1999). Understanding others, and individual differences in friendship interactions in young children. *Social Development, 8*(2), 202-219.
- Fearon, R. M., Van Ijzendoorn, M. H., Fonagy, P., Bakermans-Kranenburg, M. J., Schuengel, C., & Bokhorst, C. L. (2006). In search of shared and nonshared environmental factors in security of attachment: a behavior-genetic study of the association between sensitivity and attachment security. *Dev Psychol, 42*(6), 1026-1040.
- Flavell, J. H. (1999). Cognitive development: children's knowledge about the mind. *Annu Rev Psychol, 50*, 21-45.
- Flavell, J. H., & Miller, P. H. (1998). Social cognition. In W. Damon, D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology, 5th ed.* (pp. 851-898). New York: Wiley.
- Fonagy, P. (1989). On tolerating mental states: Theory of mind in borderline patients. *Bulletin of the Anna Freud Centre, 12*, 91-115.
- Fonagy, P. (1991). Thinking about thinking: Some clinical and theoretical considerations in the treatment of a borderline patient. *International Journal of Psycho-Analysis, 72*, 1-18.
- Fonagy, P., Gergely, G., Jurist, E., & Target, M. (2002). *Affect Regulation, Mentalization and the Development of the Self*. New York: Other Press.
- Fonagy, P., & Higgitt, A. (1989). A developmental perspective on borderline personality disorder. *Revue Internationale de Psychopathologie, 1*, 125-159.
- Fonagy, P., & Luyten, P. (2009). A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Development and Psychopathology, 21*(4), 1355-1381.
- Fonagy, P., & Luyten, P. (in press-a). A developmental, mentalization based approach to the understanding and treatment of borderline personality disorder. *Development and Psychopathology*.
- Fonagy, P., & Luyten, P. (in press-b). A mentalization-based approach to the understanding and treatment of borderline personality disorder. *Development and Psychopathology*.
- Fonagy, P., Redfern, S., & Charman, T. (1997). The relationship between belief-desire reasoning and a projective measure of attachment security (SAT). *British Journal of Developmental Psychology, 15*, 51-61.

- Fonagy, P., Steele, H., & Steele, M. (1991). Maternal representations of attachment during pregnancy predict the organization of infant-mother attachment at one year of age. *Child Development, 62*, 891-905.
- Fonagy, P., Steele, M., Moran, G. S., Steele, H., & Higgitt, A. C. (1991). Measuring the ghost in the nursery: a summary of the main findings of the Anna Freud Centre/University College London parent-child study. *Bulletin of the Anna Freud Centre, 14*, 115-131.
- Fonagy, P., Steele, M., Steele, H., & Target, M. (1997). *Reflective-Functioning Manual, version 4.1, for Application to Adult Attachment Interviews*. London: University College London.
- Fonagy, P., & Target, M. (1996). Playing with reality: I. Theory of mind and the normal development of psychic reality. *International Journal of Psycho-Analysis, 77*, 217-233.
- Fonagy, P., & Target, M. (2000). Playing with reality III: The persistence of dual psychic reality in borderline patients. *International Journal of Psychoanalysis, 81*(5), 853-874.
- Fonagy, P., & Target, M. (2007). Playing with reality: IV. A theory of external reality rooted in intersubjectivity. *Int J Psychoanal, 88*(Pt 4), 917-937.
- Fraiberg, S. (1977). *Insights from the blind*. London: Souvenir Press.
- Freud, A. (1936). *The Ego and the Mechanisms of Defence*. New York: International Universities Press, 1946.
- Fries, A. B., Ziegler, T. E., Kurian, J. R., Jacoris, S., & Pollak, S. D. (2005). Early experience in humans is associated with changes in neuropeptides critical for regulating social behavior. *Proc Natl Acad Sci U S A, 102*(47), 17237-17240.
- Frodi, A., & Smetana, J. (1984). Abused, neglected, and nonmaltreated preschoolers' ability to discriminate emotions in others: The effects of IQ. *Child Abuse & Neglect, 8*(4), 459-465.
- George, C., Kaplan, N., & Main, M. (1985). The Adult Attachment Interview. Unpublished manuscript, Department of Psychology, University of California at Berkeley.
- George, C., & Solomon, J. (1999). A comparison of attachment theory and psychoanalytic approaches to mothering. *Psychoanalytic Inquiry, 19*, 618-646.
- Gergely, G. (2001). The obscure object of desire: 'Nearly, but clearly not, like me. Contingency preference in normal children versus children with autism. In J. Allen, P. Fonagy & G. Gergely (Eds.), *Contingency Perception and Attachment in Infancy, Special Issue of the Bulletin of the Menninger Clinic* (pp. 411-426). New York: Guilford Press.
- Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences, 7*, 287-292.
- Gergely, G., & Watson, J. (1996). The social biofeedback model of parental affect-mirroring. *International Journal of Psycho-Analysis, 77*, 1181-1212.
- Gergely, G., & Watson, J. (1999). Early social-emotional development: Contingency perception and the social biofeedback model. In P. Rochat (Ed.), *Early social cognition: Understanding others in the first months of life* (pp. 101-137). Hillsdale, NJ: Erlbaum.

- Gillath, O., Bunge, S. A., Shaver, P. R., Wendelken, C., & Mikulincer, M. (2005). Attachment-style differences in the ability to suppress negative thoughts: exploring the neural correlates. *Neuroimage*, 28(4), 835-847.
- Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, 16, 1-14, 29-113.
- Grienenberger, 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. *Attach Hum Dev*, 7(3), 299-311.
- Guastella, A. J., Mitchell, P. B., & Mathews, F. (2008). Oxytocin Enhances the Encoding of Positive Social Memories in Humans. *Biological Psychiatry*, 64, 256-258.
- Heim, C., Young, L. J., Newport, D. J., Mletzko, T., Miller, A. H., & Nemeroff, C. B. (2008). Lower CSF oxytocin concentrations in women with a history of childhood abuse. *Mol Psychiatry*, 14(10), 954-958.
- Hill, J., Fonagy, P., Safier, E., & Sargent, J. (2003). The ecology of attachment in the family. *Fam Process*, 42(2), 205-221.
- Hobson, R. P., & Bishop, M. (2003). The pathogenesis of autism: insights from congenital blindness. *Philos Trans R Soc Lond B Biol Sci*, 358(1430), 335-344.
- Hofer, M. A. (1995). Hidden regulators: Implications for a new understanding of attachment, separation and loss. In S. Goldberg, R. Muir & J. Kerr (Eds.), *Attachment Theory: Social, Developmental, and Clinical Perspectives* (pp. 203-230). Hillsdale, N.J.: The Analytic Press, Inc.
- Hofer, M. A. (2004). The Emerging Neurobiology of Attachment and Separation: How Parents Shape Their Infant's Brain and Behavior. In S. W. Coates & J. L. Rosenthal (Eds.), *September 11- "When the Bough Broke", Attachment Theory, Psychobiology, and Social Policy: An Integrated Approach to Trauma*. New York: Analytic Press.
- Insel, T. R. (2003). Is social attachment an addictive disorder? *Physiol Behav*, 79(3), 351-357.
- Insel, T. R., & Young, L. J. (2000). Neuropeptides and the evolution of social behavior. *Curr Opin Neurobiol*, 10(6), 784-789.
- Leckman, J. F., Feldman, R., Swain, J. E., Eicher, V., Thompson, N., & Mayes, L. C. (2004). Primary parental preoccupation: circuits, genes, and the crucial role of the environment. *J Neural Transm*, 111(7), 753-771.
- Legerstee, M., & Varghese, J. (2001). The role of maternal affect mirroring on social expectancies in 2-3 month-old infants. *Child Development*, 72, 1301-1313.
- Leslie, A. M. (1987). Pretense and representation: The origins of 'Theory of Mind'. *Psychological Review*, 94, 412-426.
- Leslie, A. M. (1994). TOMM, ToBy, and agency: core architecture and domain specificity. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119-148). New York: Cambridge University Press.
- Lewis, V., Norgate, S., Collis, G., & Reynolds, R. (2000). The consequences of visual impairment for children's symbolic and functional play. *British Journal of Developmental Psychology*, 18, 449-464.

- Luyten, P., Fonagy, P., Mayes, L., Vermote, R., Lowyck, B., Bateman, A., et al. (submitted). Broadening the scope of the mentalization based approach to psychopathology.
- Lyons-Ruth, K., & Jacobvitz, D. (2008). Attachment disorganization: Unresolved loss, relational violence, and lapses in behavioral and attentional strategies. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of Attachment: Theory, Research and Clinical Applications* (pp. 520-554). New York: Guilford.
- MacLean, P. (1990). *The triune brain in evolution: role in paleocerebral functions*. New York: Plenum.
- Main, M. (1997). *Attachment narratives and attachment across the lifespan*. Paper presented at the Fall Meeting of the American Psychoanalytic Association, New York.
- Main, M., & Hesse, E. (2000). The organized categories of infant, child, and adult attachment: flexible vs. inflexible attention under attachment-related stress. *Journal of the American Psychoanalytic Association*, 48, 1055-1096.
- McAlpine, L. M., & Moore, C. L. (1985). The development of social understanding in children with visual impairments. *Journal of Visual Impairment and Blindness*, 89, 349-358.
- Meaney, M. J., & Szyf, M. (2005). Environmental programming of stress responses through DNA methylation: life at the interface between a dynamic environment and a fixed genome. *Dialogues Clin Neurosci*, 7(2), 103-123.
- Meins, E., Ferryhough, 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.
- Minuchin, S. (1974). *Families and family therapy*. Cambridge, Mass: Harvard University Press.
- Mitchell, R. W. (1993). Mental models of mirror self-recognition: Two theories. *New Ideas in Psychology*, 11, 295-325.
- Montague, P. R., & Lohrenz, T. (2007). To detect and correct: norm violations and their enforcement. *Neuron*, 56, 14-18.
- Neisser, U. (1988). Five kinds of self-knowledge. *Philosophical Psychology*, 1, 35-59.
- Newton, P., Reddy, V., & Bull, R. (2000). Children's everyday deception and performance on false-belief tasks. *British Journal of Developmental Psychology*, 18, 297-317.
- Nitschke, J. B., Nelson, E. E., Rusch, B. D., Fox, A. S., Oakes, T. R., & Davidson, R. J. (2004). Orbitofrontal cortex tracks positive mood in mothers viewing pictures of their newborn infants. *Neuroimage*, 21(2), 583-592.
- Oppenheim, D., & Koren-Karie, N. (2002). Mothers' insightfulness regarding their children's internal worlds: The capacity underlying secure child-mother relationships. *Infant-Mental-Health-Journal*, 23, 593-605.
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford: Oxford University Press.
- Pears, K. C., & Fisher, P. A. (2005). Emotion understanding and theory of mind among maltreated children in foster care. *Development and Psychopathology*, 17(1), 47-65.

- Pears, K. C., & Moses, L. J. (2003). Demographics, parenting, and theory of mind in preschool children. *Social Development, 12*, 1-20.
- Povinelli, D. J., & Eddy, T. J. (1995). The unduplicated self. In P. Rochat (Ed.), *The Self in Infancy: Theory and Research* (pp. 161-192). Amsterdam: Elsevier.
- Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology, 33*, 12-21.
- Rosenfeld, H. (1971). Contribution to the psychopathology of psychotic states: The importance of projective identification in the ego structure and object relations of the psychotic patient. In E. B. Spillius (Ed.), *Melanie Klein today* (pp. 117-137). London: Routledge, 1988.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science, 300*(5626), 1755-1758.
- Sharp, C., Fonagy, P., & Goodyer, I. (2006). Imagining your child's mind: Psychosocial adjustment and mothers' ability to predict their children's attributional response styles. *British Journal of Developmental Psychology, 24*(1), 197-214.
- Siegal, M., & Patterson, C. C. (2008). Language and theory of mind in atypically developing children: Evidence from studies of deafness, blindness, and autism. In C. Sharp, P. Fonagy & I. Goodyer (Eds.), *Social cognition and developmental psychology*. Oxford: Oxford University Press.
- Siegel, D. J. (1999). *The developing mind: Toward a neurobiology of interpersonal experience*. New York: Guilford.
- Slade, A., Grienenberger, J., Bernbach, E., Levy, D., & Locker, A. (2005). Maternal reflective functioning, attachment and the transmission gap: A preliminary study. *Attachment and Human Development, 7*(3), 283-298.
- Smith, M., & Walden, T. (1999). Understanding feelings and coping with emotional situations: A comparison of maltreated and nonmaltreated preschoolers. *Social Development, 8*(1), 93-116.
- Sodian, B., & Frith, U. (1992). Deception and sabotage in autistic, retarded and normal children. *J Child Psychol Psychiatry, 33*(3), 591-605.
- Sodian, B., Taylor, C., Harris, P. L., & Perner, J. (1992). Early deception and the child's theory of mind: false trails and genuine markers. *Child Development, 62*, 468-483.
- Sroufe, L. A. (1996). *Emotional development: The organization of emotional life in the early years*. New York: Cambridge University Press.
- Strathearn, L. (2006). Exploring the neurobiology of attachment. In L. C. Mayes, P. Fonagy & M. Target (Eds.), *Developmental Science and Psychoanalysis: Integration and Innovation*. London: Karnac.
- Strathearn, L., Fonagy, P., Amico, J., & Montague, R. (2009). Adult attachment predicts maternal brain and oxytocin response to infant cues. *Neuropsychopharmacology, 34*(13), 2655-2666.
- Swain, J. E., Lorberbaum, J. P., Kose, S., & Strathearn, L. (2007). Brain basis of early parent-infant interactions: psychology, physiology, and in vivo functional neuroimaging studies. *Journal of Child Psychology and Psychiatry, 48*, 262-287.
- Target, M., & Fonagy, P. (1996). Playing with reality II: The development of psychic reality from a theoretical perspective. *International Journal of Psycho-Analysis, 77*, 459-479.

- Taylor, M., Esbensen, B. M., & Bennett, R. T. (1994). Children's understanding of knowledge acquisition: the tendency for children to report that they have always known what they have just learned. *Child Dev*, 65(6), 1581-1604.
- Tremblay, R. E. (2000). The origins of violence. *ISUMA*(Autumn), 19-24.
- van IJzendoorn, M. H. (1995). Adult attachment representations, parental responsiveness, and infant attachment: A meta-analysis on the predictive validity of the Adult Attachment Interview. *Psychological Bulletin*, 117, 387-403.
- Vrticka, P., Andersson, F., Grandjean, D., Sander, D., & Vuilleumier, P. (2008). Individual attachment style modulates human amygdala and striatum activation during social appraisal. *PLoS ONE*, 3, e2868.
- Wang, Z., Yu, G., Cascio, C., Liu, Y., Gingrich, B., & Insel, T. R. (1999). Dopamine D2 receptor-mediated regulation of partner preferences in female prairie voles (*Microtus ochrogaster*): a mechanism for pair bonding? *Behav Neurosci*, 113(3), 602-611.
- Watson, J. S. (1994). Detection of self: The perfect algorithm. In S. Parker, R. Mitchell & M. Boccia (Eds.), *Self-Awareness in Animals and Humans: Developmental Perspectives* (pp. 131-149): Cambridge University Press.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Dev*, 72(3), 655-684.
- Wellman, H. M., & Phillips, A. T. (2000). Developing intentional understandings. In L. Moses, B. Male & D. Baldwin (Eds.), *Intentionality: A Key to Human Understanding*. Cambridge, MA: MIT Press.
- Winnicott, D. W. (1965). Ego distortion in terms of true and false self. In *The maturational process and the facilitating environment* (pp. 140-152). New York: International Universities Press.
- Winslow, J. T., Noble, P. L., Lyons, C. K., Sterk, S. M., & Insel, T. R. (2003). Rearing effects on cerebrospinal fluid oxytocin concentration and social buffering in rhesus monkeys. *Neuropsychopharmacology*, 28(5), 910-918.
- Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E., & Chapman, M. (1992). Development of concern for others. *Developmental Psychology*, 28, 126-136.