Chapter 6

The neurobiology of attachment and mentalizing: A neurodevelopmental perspective

Patrick Luyten & Peter Fonagy

To be submitted to Schmahl, Phan, & Friedel. *Neurobiology of personality disorders: The Biological Basis of Personality and Personality Disorders*. Oxford: Oxford University Press.

<1>Abbreviations

ACC anterior cingulate cortex

BPD borderline personality disorder

dACC dorsal anterior cingulate cortex

HPA hypothalamic–pituitary–adrenal

LTC lateral temporal cortex

MPFC medial prefrontal cortex

MSA mental state attribution

PFC prefrontal cortex

SR shared representation

ToM Theory of Mind

VMPFC ventromedial prefrontal cortex

VTA ventral tegmental area

<1>Introduction

This chapter addresses the neurobiology of attachment and mentalizing from a developmental psychopathology perspective. *Attachment* refers to an evolutionarily pre-wired, basic biobehavioral system that is activated in situations of stress and threat (Bowlby, 1973; Ein-Dor, Mikulincer, Doron, & Shaver, 2010; Mikulincer & Shaver, 2007; Panksepp, 1998). Developmentally, the attachment system plays a key role in the modulation of the stress response, another key biobehavioral system, and thus in survival (Gunnar & Quevedo, 2007; Hostinar, Sullivan, & Gunnar, 2014; Lupien, McEwen, Gunnar, & Heim, 2009; Panksepp & Watt, 2011; Watt & Panksepp, 2009). It does so because activation of the attachment system normatively leads to the seeking of proximity of attachment figures, either in reality or by means of activating representations of secure attachment experiences, or both. This leads to (a) the down-regulation of stress at the subjective and neurobiological level and (b) renewed energy and motivation to explore the world, leading to the so-called "broaden-and-build" cycles (Fredrickson, 2001) associated with attachment security (Mikulincer & Shaver, 2007). In individuals with a secure attachment history, this pattern gradually becomes generalized to different situations and circumstances. From a biological perspective, the attachment system is underpinned by a mesocorticolimbic dopaminergic reward system that is involved in the rewarding features of infant–parent, parent–infant, pair-bonding, and other attachment relationships (Insel & Young, 2001; Panksepp & Watt, 2011; Rutherford, Williams, Moy, Mayes, & Johns, 2011; Strathearn, Fonagy, Amico, & Montague, 2009; Swain, Lorberbaum, Kose, & Strathearn, 2007).

The attachment system is also closely related to the *mentalizing* or *social cognition* system, which subserves the human capacity to understand oneself and others in terms of intentional mental states (e.g., feelings, desires, wishes, attitudes, and values). This system has most likely evolved out of the need for human beings to develop the needed

computational power to navigate a complex social world (Fonagy, Luyten, & Allison, 2015; Tomasello & Vaish, 2013). From a biological perspective, as we will describe in more detail in this chapter, different capacities or features of mentalizing are underpinned by relatively different neural circuits (Lieberman, 2007; Luyten & Fonagy, 2015).

Both attachment disruptions and mentalizing impairments, and their links with impairments in stress and arousal regulation, have been amply demonstrated in individuals vulnerable for psychopathology, particularly in individuals with personality disorder (Fonagy & Luyten, 2016; Levy, Meehan, Weber, Reynoso, & Clarkin, 2005). Knowledge of the normal development of neurobiological systems underlying attachment and mentalizing, and disruptions in these normal developmental trajectories, may therefore directly inform our understanding of psychopathology, and of personality disorders specifically.

In this chapter, we first consider the neurobiology of attachment. We then discuss the neurobiological underpinning of mentalizing in relation to attachment and stress regulation. We focus on the early development of both capacities in relation to stress regulation, and discuss the relationship to the development of psychopathology and personality disorder in particular across the lifespan, with a focus on early childhood and adolescence.

<1>Attachment and reward

<2>The roots of attachment in rewarding experiences

That social and attachment relationships should be one of the most rewarding experiences is predicted by both evolutionary (Gilbert, 2006) and psychological (Beck, 2009; Blatt, 2008) theories. Research in both animals and human beings has suggested that attachment, particularly in normal development, is underpinned by a powerful neurobiological reward system (Champagne et al., 2004; Ferris et al., 2005; Insel & Young, 2001; Strathearn, Li, Fonagy, & Montague, 2008). This reward system has been relatively

well described in the literature as comprising various mesolimbic and mesocortical pathways. The ventral tegmental area (VTA) is the origin of mesolimbic pathways, which mainly project to ventral striatal regions, in particular the nucleus accumbens, hippocampus, and amygdala. Meanwhile the mesocortical pathways mainly project to the prefrontal cortex (PFC) and anterior cingulate cortex (ACC) (Nestler & Carlezon, 2006; Pizzagalli, 2014; Russo & Nestler, 2013; Spear, 2000). Dopamine and oxytocin have been regarded as the key biological mediators involved in this system. However, opioid and cannabinoid systems are also of considerable interest because they are associated with the sensation of pain arising from rejection and social loss, a response that is heightened in adolescence, particularly in females (Hsu et al., 2015; Panksepp & Watt, 2011; Spear, 2000).

Studies in animals (including higher primates) and a growing body of research in human beings (Hostinar et al., 2014; Strathearn, 2011; Swain et al., 2014) suggest that this attachment/reward system plays a key developmental role in the emergence and ongoing regulation of the stress system. Typically, secure attachment experiences serve to cushion the effects of stress in early development; this results in "adaptive hypoactivity" of the hypothalamic–pituitary–adrenal (HPA) axis in early development (Gunnar & Quevedo, 2007). For securely attached individuals, relationships become increasingly rewarding, and experiences of effective down-regulation lead to feelings of autonomy and confidence in one's capacity to deal with adversity (i.e., *resilience*). By contrast, insecure attachment experiences can lead to increased susceptibility to stress, indicated by dysfunctions of both the HPA axis and the reward system (Auerbach, Admon, & Pizzagalli, 2014; Pizzagalli, 2014; Strathearn, 2011). In these individuals, attachment experiences become increasingly aversive, which also undermines these individuals' ability to deal with adversity and their confidence in their ability to do so.

Neuropeptides such as oxytocin and vasopressin are key modulators of the relationship between attachment and stress regulation. Oxytocin seems to increase affiliative behavior in the face of distress, particularly in securely attached individuals and in relation to in-group members. Such behavior serves to optimize opportunities for the effective co-regulation of stress with others, reducing behavioral and neuroendocrinological stress responses (Neumann, 2008). Oxytocin also has anxiolytic and anti-stress effects by acting to down-regulate the HPA axis system. Furthermore, it fosters mentalizing and trust in others, again increasing opportunities for the effective down-regulation of distress and the use of exploration (Bartz, Zaki, Bolger, & Ochsner, 2011; Neumann, 2008) and leading to broaden-and-build cycles (Fredrickson, 2001; Mikulincer & Shaver, 2007). However, even in community samples these effects seem mainly limited to in-group members. Research has shown that in relation to out-group members, the administration of oxytocin leads to increased distrust, more bias in attributing intentions, and decreases in cooperative behavior (Bartz, Simeon, et al., 2011).

Further, in individuals with an insecure attachment history, decreased basal oxytocin levels have been observed, and the administration of oxytocin can have negative effects on social behavior, leading to disruption of the stress response (Bartz, Simeon, et al., 2011; Bertsch, Schmidinger, Neumann, & Herpertz, 2013; Cyranowski et al., 2008; Fries, Hesse, Hellhammer, & Hellhammer, 2005; Heim, Newport, Mletzko, Miller, & Hemeroff, 2008; Meinlschmidt & Heim, 2007; Stanley & Siever, 2010). The effect of oxytocin therefore seems to be that it increases the salience of attachment issues in either a positive or a negative direction. This may be a particular problem in individuals with personality disorders, who often have a marked history of early adversity, particularly attachment trauma. For such individuals, attachment experiences lack rewarding features and are, at best, associated with both reward and anxiety, anger, and/or frustration.

<2>Adolescence and reward: A critical juncture?

During adolescence, the attachment/reward system undergoes marked reorganization (Auerbach et al., 2014; Davey, Yücel, & Allen, 2008; Forbes & Dahl, 2012; Luciana, 2013; Spear, 2000). One of the primary shifts lies in the area of relatedness, which occurs alongside entry into the complex world of peer and romantic relationships (expressed in increased rejection sensitivity) and greater expectations in relation to achievement (reflected in increased sensitivity to failure). At the same time, adolescence is also characterized by the lowest levels of dopamine in striatal regions and the highest levels of dopamine in prefrontal regions; these changes have been suggested to lead to a "mini-reward deficiency syndrome" typical of adolescence (Spear, 2007). This may encourage compensatory behaviors, including risk-taking and drug abuse, as are typically found in individuals with personality disorders (Davey et al., 2008; Spear, 2000).

These findings may also explain why disappointment and/or frustration of needs for relationships and struggles with feelings of belongingness and achievement/status (which are closely intertwined, particularly in adolescence) may lead to a downward spiral characterized by suppression of the reward system, higher levels of stress, and resulting impairments in mentalizing. It has been speculated that the decrease in the incentive value of rewards in adolescence is evolutionarily adaptive, as it serves to encourage novelty and sensation-seeking behavior, which in turn supports adolescents in accomplishing important developmental tasks (e.g., developing feelings of autonomy and achievement, and establishing complex relationships with others).

Two explanations have been formulated to explain the mini-reward deficiency syndrome in adolescence. Low levels of tonic dopamine in combination with high levels of phasic dopamine release in response to rewards might explain why adolescents are

particularly keen to seek out novel and highly rewarding stimuli (Davey et al., 2008; Luciana, 2013). On the other hand, excessive down-regulation of the PFC, as a result of high levels of dopamine in the PFC resulting from increased stress, might lead to impairments in reward sensitivity in adolescence (Pizzagalli, 2014; Spear, 2000). Specifically, high levels of mesocortical dopamine impair mentalizing and representational capacities more generally, and thus may also make the incentive value of rewards seem diminished: there is a heightened perception that the important rewards in adolescence (i.e., love and status) are abstract and temporally distant (Davey et al., 2008).

<1>Mentalizing, attachment and stress regulation

<2>*The origins of the capacity for mentalizing in attachment relationships*

Higher social cognition, in particular the ability to mentalize, is considered to be the factor that underpins humanity's capacity to live in very large social groups. In primate species, including *Homo sapiens*, the size of the social group typically tolerated correlates with the size of the neocortex (the prefrontal and temporoparietal areas that support the large-scale social interactions characteristic of <u>human beings</u>) (Dunbar & Shultz, 2007; Kanai, Bahrami, Roylance, & Rees, 2012; Sallet et al., 2011). With the emergence of this heightened social cognitive capacity in human beings, new and complex ways of collaborating, teaching, and learning – which go far beyond conditioning and emulative learning – were made possible (Csibra & Gergely, 2009; Humphrey, 1988; Tomasello & Vaish, 2013). This form of social cognition made possible (a) the capacity for self-awareness and self-consciousness; (b) the human striving to transcend physical reality; and (c) the human capacity for complex forms of collaboration and relatedness (see Allen, Fonagy, & Bateman, 2008). At the same time, however, these new capacities also resulted in an increased risk for psychopathology (Luyten, Fonagy, Lemma, & Target, 2012).

This vulnerability arising from the capacity for higher social cognition speaks to the fact that mentalizing is not a constitutional given but is largely a developmental achievement. The precise nature of any individual's mentalizing profile – their strengths and weaknesses in relation to mentalizing across the dimensions (see section on *Mentalizing dimensions*), is shaped in the first instance by the cumulative nature of the interactions that take place with the individual's attachment relationships, in particular early attachments during infancy (Fonagy & Luyten, 2016; Kovacs, Teglas, & Endress, 2010). Attachment figures' capacity to respond with contingent and marked affective displays of their own experience in response to the infant's subjective experience positively influences the child's ability to develop mentalizing capacities (see Figure 6.1). Subsequently, contact with other human beings (e.g., peers, teachers, and friends) broadens and strengthens the development of mentalizing (Fonagy & Luyten, 2016).

[INSERT FIGURE 6.1 HERE]

Conversely, consistent or serious failures in marked mirroring by early attachment figures lead to impairments in the capacity to reflect on the self and others, as they lead to unmentalized self-experiences (also called *alien self-experiences*). These experiences are subjectively felt as invalidating the individual's experience, and thus are felt as alien to the self (see Figure 6.2). Such failures are to an extent a part of the fabric of everyday life: some misattunements in marked mirroring are an inevitable experience, as the caregiver may not be constantly available or inclined to engage sensitively with the infant's subjective state at all times. Consequently, all people will have unmentalized self-states. However, in various forms of psychopathology – most paradigmatically, in the case of borderline personality disorder (BPD) and most often as the result of a combination of biological vulnerability and

9

environmental circumstances – these alien self-experiences are so marked that they almost completely dominate the feelings and thoughts of the individual. This leads to a constant pressure to externalize such unmentalized, alien self-experiences, which can manifest itself in the tendency to dominate the mind of others and/or in various types of self-harming behavior (Fonagy & Luyten, 2016).

[INSERT FIGURE 6.2 HERE]

<2>Mentalizing dimensions

Studies concerning the neurobiology of mentalizing have shown that this capacity is organized around at least four dimensions, with each dimension involving relatively distinct neural circuits (see Table 6.1) (Fonagy & Luyten, 2009; Luyten, Fonagy, Lowyck, & Vermote, 2012). These dimensions cover a broad range of related constructs from social cognition research, including empathy, mindfulness, and Theory of Mind (ToM) (Choi-Kain & Gunderson, 2008). Mindfulness, for instance, focuses on a core component of mentalizing about the self (e.g., the ability to attend to one's own internal mental states), while empathy and ToM respectively focus on more affective and more cognitive mentalizing about others.

Solid mentalizing reflects a balance between these four dimensions. Psychopathology (particularly personality disorder) is thought to reflect imbalances between the dimensions. For instance, individuals with BPD are typically overly sensitive to the emotional states of others while showing marked impairments in the capacity to reflect on their own mental states.

[INSERT TABLE 6.1 HERE]

<2>Neurobiology of mentalizing dimensions

Automatic or *implicit* mentalizing involves relatively parallel and fast processing, which requires little effort, focused attention, or intention (Satpute & Lieberman, 2006). Studies suggest that an elementary capacity for implicit mentalizing is present in infants from as young as 7 months of age (Kovacs et al., 2010). Automatic mentalizing clearly facilitates survival (Lieberman, 2007; Mayes, 2006), as the fast processing of social information best serves the fight/flight response in threatening situations. However, in more complex social situations, automatic mentalizing is far less adaptive, as it is typically based on biased assumptions. Hence, human beings need *controlled or explicit* mentalizing to understand both one's own mind and the mind of others, particularly in complex social situations. Controlled mentalizing is verbal, reflective, and conscious, and therefore involves much slower serial processing of social information. Extant research suggests that phylogenetically older brain circuits underlie automatic mentalizing, while controlled mentalizing is underpinned by phylogenetically newer neural circuits. Automatic mentalizing tends to involve the amygdala, basal ganglia, ventromedial prefrontal cortex (VMPFC), lateral temporal cortex (LTC), and dorsal ACC (dACC) (Satpute & Lieberman, 2006). The relation of these brain circuits to threat detection and the fight/flight response is clear. The amygdala, for instance, has been suggested to play a key role in the processing of the biological "value" of information. The VMPFC modulates both the amygdala and basal ganglia, and the VMPFC and basal ganglia have been linked to automatic intuition. Importantly, the basal ganglia have also been linked to reward-related implicit emotion processing, while areas such as the dACC seem to play a central role in the nonreflective processing of emotional distress and pain. The LTC has been linked to the automatic processing of faces, biological motion and attribution of intentions to others.

Controlled mentalizing relies more on the lateral PFC, medial PFC (MPFC), the lateral and medial parietal cortices, medial temporal lobe, and rostral ACC (Lieberman, 2007; Satpute & Lieberman, 2006; Uddin, Iacoboni, Lange, & Keenan, 2007). The lateral PFC and lateral parietal cortex have been related to complex causal reasoning, while the medial parietal cortex has been linked to explicit perspective-taking. The rostral ACC has been implicated in tasks involving explicit conflict processing, and the medial temporal lobe in explicit, declarative memory. Importantly, the MPFC, one of the brain areas that has been most consistently linked to mentalizing, may play a central role in both automatic and controlled mentalizing. Yet, because this structure is larger in human beings than in other primates, and because increasing cognitive load leads to decreasing performance of this structure, it has been suggested that the MPFC is more closely linked to the controlled system (Lieberman, 2007; Satpute & Lieberman, 2006; Uddin et al., 2007).

Mentalizing based on external features of self and others (e.g., facial expressions, posture, movements, and prosody) involves a more lateral frontotemporoparietal network (e.g., the posterior superior temporal sulcus and temporal poles) (indicative of more automatic processing), while mentalizing based on internal features of self and others tends to recruit a medial frontoparietal network (e.g., MPFC) (indicative of more controlled reflective processes) (Lieberman, 2007).

With regard to the self-other dimension, the same core network tends to be activated whenever we reflect on ourselves and others involving the medial prefrontal cortex, temporal poles and the posterior superior temporal sulcus/temporoparietal junction in the LTC (Frith & Frith, 2006; Lieberman, 2007; Uddin et al., 2007; Van Overwalle, 2009; Van Overwalle & Baetens, 2009).

The finding of a common network underlying mentalizing with regard to both self and others sheds light on interesting findings concerning the centrality of both identity problems

and problems with mentalizing about others in most individuals with personality disorders. This may be related to an imbalance in two neural systems involved in self-knowing and knowing others (Dimaggio, Lysaker, Carcione, Nicolo, & Semerari, 2008; Lieberman, 2007; Lombardo, Barnes, Wheelwright, & Baron-Cohen, 2007; Shamay-Tsoory, 2011; Uddin et al., 2007). Ripoll et al. (2013) have called these systems the *shared representation* (SR) system and the *mental state attribution* (MSA) system. The SR system involves a rapid, automatic "visceral recognition" of the experience of others (Lombardo et al., 2010), involving a more body-based, frontoparietal (mirror-neuron) system (Gallese, Keysers, & Rizzolatti, 2004; Rizzolatti & Craighero, 2004; Van Overwalle & Baetens, 2009). Neural areas that have been linked to the SR system include the amygdala, inferior frontal gyrus, inferior parietal lobule, anterior insula, and (dorsal) ACC.

A more controlled, cortical midline system consisting of the ventromedial and dorsomedial PFC, the temporoparietal junction and the medial temporal pole (Lieberman, 2007; Uddin et al., 2007) underlies explicit perspective-taking and both cognitive ToM (dorsomedial PFC) and affective ToM (VMPFC). The MSA plays a central role in the inhibition of the SR system, that is, of automatic mimicry or identification with the mental states of others (Brass & Haggard, 2008; Brass, Ruby, & Spengler, 2009; Brass, Schmitt, Spengler, & Gergely, 2007). For instance, people with BPD seem to be particularly prone to such automatic identification processes, which suggests the existence of impairments in the MSA system (Fonagy & Luyten, 2016). As a result, these individuals are particularly prone to emotional contagion and identity diffusion.

The capacity for mentalizing involves the integration of cognition and affect. Mentalizing has a clear cognitive component, such as perspective-taking and belief-desire reasoning. It depends on several areas of the PFC (Sabbagh, 2004; Shamay-Tsoory & Aharon-Peretz, 2007; Shamay-Tsoory, Aharon-Peretz, & Levkovitz, 2007). The affective

components include affective empathy and mentalized affectivity (Fonagy, Gergely, Jurist, & Target, 2002; Jurist, 2005). The VMPFC appears to play a central role in affective mentalizing, that is, in "marking" mental representations with affect (Rochat & Striano, 1999). This dissociation between the neural systems involved in cognitive and affective mentalizing may also explain, at least in part, the distinction between affective and cognitive empathy. Affective empathy involves a more basic "emotional contagion" system, whereas cognitive empathy involves a cognitive perspective-taking system (Shamay-Tsoory, Aharon-Peretz, & Perry, 2009). Again, these findings concerning the dissociable capacities involved in cognitive versus affective mentalizing have immediate relevance for understanding of personality disorders. Individuals with antisocial personality disorder, particularly those with psychopathic features, show normal cognitive mentalizing but gross impairments in affective mentalizing (Blair, 2013; Viding & McCrory, 2012).

<2>A developmental perspective on the relationships among arousal, attachment and mentalizing

With increasing stress or arousal, controlled, slow and reflective mentalizing is replaced by automatic, fast, and typically biased automatic mentalizing, or *prementalizing* modes of experiencing oneself and others (see Table 6.2 and Figure 6.3).

[INSERT TABLE 6.2 HERE]

[INSERT FIGURE 6.3 HERE]

Both noradrenergic and dopaminergic systems have been shown to be involved in this switch from controlled to automatic mentalizing, which has been hypothesized to protect the PFC from excessive stimulation and to facilitate coordination among the attentional,

executive, and sensory systems in response to threat (Arnsten, Mathew, Ubriani, Taylor, & Li, 1999). Norepinephrine enhances the activation of the PFC, while α_1 postsynaptic receptor stimulation impairs its functioning. The D1 dopamine receptor family also enhances PFC functioning, but when amygdala activation leads to catecholamine release, D1 impairs PFC functioning.

Individuals' attachment history plays an important role in modulating three key parameters involved in this switch: (a) the point at which an individual switches from controlled to automatic mentalizing, (b) the extent of the loss of the capacity for controlled mentalizing, and (c) the duration of the loss of controlled mentalizing until it is re-established (Fonagy & Luyten, 2009; Fonagy & Luyten, 2016; Luyten, Fonagy, Lowyck, et al., 2012).

<3>Secure attachment

Individuals who predominantly use *secure attachment strategies* in response to stress seem to be able to keep controlled mentalizing on-line longer and more effectively compared with individuals who tend to rely on insecure attachment strategies. They also seem to be able to recover from losses of controlled mentalizing faster and more effectively. In securely attached individuals, studies suggest that activation of the attachment system actually fosters controlled mentalizing in combination with a relaxation of epistemic hypervigilance, leading to effective down-regulation of stress and the development of "broaden and build" cycles (Fredrickson, 2001).

The neuropeptides oxytocin, vasopressin, and opioids appear to play an important role here, leading to deactivation of behavioral mechanisms involved in social avoidance and attenuating both behavioral and neuroendocrine stress responses (Heinrichs & Domes, 2008; Insel & Young, 2001; Panksepp & Watt, 2011). Oxytocin release has also been found to foster mentalizing in securely attached individuals, as apparent in improvements in social memory

(and memory of facial expressions in particular) and the recognition of mental states based on facial expressions, as well as increasing trust in others (Bartz, Zaki, et al., 2011; Neumann, 2008). These findings may have important implications for our understanding of the neurobiology of resilience (Fonagy, Steele, Steele, Higgitt, & Target, 1994). Yet, even in securely attached individuals, mentalizing is not always solid, as contextual factors play an important role in determining both the quality and quantity of mentalizing (Bartz, Zaki, et al., 2011). With increasing arousal, the capacity for controlled mentalizing is easily lost, particularly in relation to out-group members (Bartz, Zaki, et al., 2011). For instance, research has consistently shown that, even in normal community samples, the majority of whose members are securely attached, oxytocin administration leads to increased distrust, increased bias in attributing intentions to others, and decreases in cooperative behavior with regard to out-group members (Bartz, Zaki, et al., 2011). From a neurobiological perspective, increasing arousal and the neurobiological cascade of events that follows seems to make attachment issues more salient, which increases the likelihood of a deactivation of controlled mentalizing.

<3>*Attachment hyperactivating strategies*

Some individuals tend to use attachment hyperactivating strategies in response to stress because of their developmental history or because of contextual factors. These strategies reflect attempts to find security and to co-regulate stress with others, based on the underlying belief that others are not able to provide security and support, despite these individuals' strong wishes for care, support and love. Attachment hyperactivating strategies have been shown to be characterized by a low threshold for switching to nonmentalizing modes, more extensive lapses in controlled mentalizing, and a longer time to recovery from such lapses, compared with individuals whose attachment strategies are secure. This may lead, in the extreme, to hyperactivation of the attachment system, with resulting impairments

in stress regulation and mentalizing as a result of a failure to benefit from broaden-and-build experiences. This is particularly the case in traumatized individuals, who often show marked hypersensitivity to stress, hyperactivation of the attachment system, and profound impairments in mentalizing. For example, studies have shown that early trauma is associated with kindling of the amygdala (Botterill et al., 2014), leading to an exaggerated response to threats. Similarly, research clearly suggests both structural and functional changes in the amygdala (and the stress response more generally) in individuals with a history of early adversity. Dysfunction of the HPA axis has been demonstrated in a wide variety of conditions characterized by marked early adversity, ranging from depression to functional somatic disorders to personality disorders (Jogems-Kosterman, de Knijff, Kusters, & van Hoof, 2007; Nater et al., 2010; Scott, Levy, & Granger, 2013; Wingenfeld, Spitzer, Rullkotter, & Lowe, 2010). In BPD patients with an explicit trauma history, for instance, a reduction in pituitary volume (Garner et al., 2007), elevated levels of corticotropin-releasing factor in cerebrospinal fluid (Lee, Geracioti, Kasckow, & Coccaro, 2005), dysfunctions of cortisol responsivity (Jogems-Kosterman et al., 2007; Minzenberg et al., 2006; Walter et al., 2008), and disturbed dexamethasone suppression test responses (Wingenfeld et al., 2007) have been observed.

Impaired stress regulation as a result of continued stress has been shown to negatively influence brain areas involved in mentalizing, as is, for instance, demonstrated by findings that chronic stress disrupts amygdala–VMPFC connectivity (Tottenham & Sheridan, 2009).

Attachment deactivating strategies

Attachment deactivating strategies involve denying attachment needs and asserting one's own autonomy and independence when faced with adversity, in an attempt to down-regulate stress. Developmentally, attachment deactivating strategies develop on the basis of repeated experiences that others are unavailable to provide security, support, and comfort. In

response, these individuals tend to de-emphasize the importance of attachment relationships, which becomes a habitual response associated with rapid deactivation of the attachment system and social information processing of threat cues. Yet, at the same time, these individuals are often able to keep the neural systems involved in controlled mentalizing on-line even when under considerable stress because they have had to learn to rely on their own capacity for affect regulation (Vrticka, Andersson, Grandjean, Sander, & Vuilleumier, 2008). Experimental studies have shown that these deactivating strategies are likely to fail under increasing stress, leading to a "rebound" of suppressed feelings of insecurity and lack of self-worth (Mikulincer & Shaver, 2007; Vrticka et al., 2008).

A series of studies by Strathearn and colleagues (Strathearn et al., 2009; Strathearn et al., 2008) provides a good example of the influence of attachment deactivating strategies on stress regulation and mentalizing. Strathearn and colleagues first assessed attachment security in 30 first-time mothers using the Adult Attachment Interview before the birth of their child. About 10 months after birth of their child, the same mothers were asked to view images of their own or other infants' smiling and crying faces. Mothers who were classified as securely attached showed greater activation in brain regions of the reward system, such as the ventral striatum, and the oxytocin-associated hypothalamus/pituitary region. They also showed higher and increasing peripheral oxytocin release while playing with their infant, which was positively correlated with brain activation to their own infant in the reward system. By contrast, mothers who were categorized as having insecure/dismissing attachment (who would tend to rely predominantly on attachment deactivating) showed less activation of the reward system and greater activation of the insula in response to seeing their own infant's sad face. Studies have suggested that the insula plays a key role in generating feelings of unfairness, pain, and disgust (see review by Montague & Lohrenz, 2007), and is involved in automatic mentalizing. Thus, these mothers appeared to be less able to regulate the sad

feelings evoked in them by viewing their infant's crying face. These findings are consistent with findings in adults showing that attachment deactivating strategies are associated with down-regulation of activity in reward-related brain regions and, at the same time, activation of the MPFC and the ventral ACC, brain areas that have been implicated in controlled mentalizing as well as social rejection and emotion suppression (Vrticka et al., 2008).

<1>Genetics of attachment and mentalizing: Are attachment styles evolutionary and culturally determined adaptive strategies?

Although the genetic loading of attachment in childhood is probably very small, this may change over the course of life. Indeed, a recent twin study found that by adolescence approximately 40% of individual differences in attachment security may be genetically determined (Fearon, Shmueli-Goetz, Viding, Fonagy, & Plomin, 2014), again pointing to adolescence as a critical juncture in the development of attachment and associated capacities such as mentalizing and stress/affect regulation. Specifically, adolescence may involve a resetting of the attachment system, for better or for worse, that is determined in part by genetic factors.

If further replicated, these findings shed an important light on the potential evolutionary functions of insecure attachment. Indeed, if attachment is an evolutionarily rooted capacity that enhances the survival of the human species, why would insecure attachment strategies still exist in the human behavioral repertoire? Building on the work of Ein-Dor and colleagues (Ein-Dor et al., 2010), we argue that insecure attachment styles represent a strategy to adapt to the environment, but one that may come with a high cost, as is evidenced most clearly by individuals with personality disorder (Fonagy & Luyten, 2016; Tottenham & Sheridan, 2009). From this perspective, secure attachment is accompanied by a relative relaxation of threat processing as a result of repeated experiences of security in

19

relation to attachment figures, which leads to a relaxation of interpersonal distrust and avoidance. These individuals show appropriate openness and flexibility in new relationships and novel situations, which increases opportunities for salutogenesis (Antonovsky, 1987).

By contrast, insecure attachment experiences give rise to hypersensitivity to threat. A pattern of attachment hyperactivation may follow, characterized by hypersensitivity to threat that is expressed as an emphasis on externally focused mentalizing, to the neglect of more internally focused mentalizing. Although this is adaptive to some extent in an environment characterized by attachment figures who are unreliable, the cost associated with this adaptation strategy is that these individuals show constant hypervigilance to threat, leading to chronic stress and, as a consequence of the wear and tear of chronic stress, dysregulation of the reward and stress regulation systems. Hypervigilance is also related to problems with epistemic trust, that is, the capacity to trust others as a source of knowledge about the world; impairments in epistemic trust seriously limit opportunities for salutogenesis (Fonagy et al., 2015). BPD might be considered the disorder par excellence of this pattern. However, insecure attachment experiences have also been shown to give rise to the excessive use of attachment deactivating strategies, downplaying the importance of attachment relationships and subjective distress. Although these strategies are also adaptive in the short term, the compulsive autonomy and often marked distrust of others (as is, for instance, observed in paranoid personality disorder) that they entail is associated with considerable intra- and interpersonal costs.

<1>Conclusions and directions for future research

This chapter summarizes extant research and thinking concerning the relationships among attachment, stress/affect regulation, and social cognition/mentalizing in relation to the development of personality disorders.

Although research in these areas has considerably improved our insight into the nature of personality disorders, and their underlying neurobiology in particular, much remains to be explored. Many of the research findings summarized in this chapter rest on cross-sectional and small sample size studies. The generalization of findings from research in animals to humans continues to be a major challenge, although research over the past two decades has generally confirmed findings from studies of animals in humans. Future research should employ more ecologically valid paradigms to study the interplay between attachment, stress regulation, and mentalizing over time in both normative development and clinical populations. In addition, rather than relying on consensus-based descriptive diagnoses as the basis for research, future research should investigate the neurobiology of attachment-related processes from a developmental psychopathology perspective, focusing on the role of these systems across different types of psychopathology. The study of major transitions in life, such as from childhood to adolescence and from adolescence to adulthood, as illustrated in this chapter, may be particularly productive.

References

Allen, J. G., Fonagy, P., & Bateman, A. W. (2008). *Mentalizing in clinical practice*. Washington, DC: American Psychiatric Publishing.

Antonovsky, A. (1987). *Unraveling the mystery of health: How people manage stress and stay well*. San Francisco, CA: Jossey-Bass.

Arnsten, A. F., Mathew, R., Ubriani, R., Taylor, J. R., & Li, B. M. (1999). α-1 noradrenergic receptor stimulation impairs prefrontal cortical cognitive function. *Biological Psychiatry, 45*, 26-31. doi: 10.1016/S0006-3223(98)00296-0

Auerbach, R. P., Admon, R., & Pizzagalli, D. A. (2014). Adolescent depression: Stress and reward dysfunction. *Harvard Review of Psychiatry, 22*, 139-148. doi:

10.1097/hrp.0000000000000034

Bartz, J., Simeon, D., Hamilton, H., Kim, S., Crystal, S., Braun, A., . . . Hollander, E. (2011). Oxytocin can hinder trust and cooperation in borderline personality disorder. *Social Cognitive and Affective Neuroscience, 6*, 556-563. doi: 10.1093/scan/nsq085

Bartz, J. A., Zaki, J., Bolger, N., & Ochsner, K. N. (2011). Social effects of oxytocin in humans: Context and person matter. *Trends in Cognitive Sciences, 15*, 301-309. doi: 10.1016/j.tics.2011.05.002

Beck, A. T. (2009). Cognitive aspects of personality disorders and their relation to syndromal disorders: A psychoevolutionary approach. In C. R. Cloninger (Ed.), *Personality and psychopathology* (pp. 411-429). Washington, DC: American Psychiatric Press.

Bertsch, K., Schmidinger, I., Neumann, I. D., & Herpertz, S. C. (2013). Reduced plasma oxytocin levels in female patients with borderline personality disorder. *Hormones and Behavior, 63*, 424-429. doi: 10.1016/j.yhbeh.2012.11.013

Blair, R. J. R. (2013). The neurobiology of psychopathic traits in youths. *Nature Reviews. Neuroscience, 14*, 786-799. doi: 10.1038/nrn3577

Blatt, S. J. (2008). *Polarities of experience: Relatedness and self definition in personality development, psychopathology, and the therapeutic process*. Washington, DC: American Psychological Association.

Botterill, J. J., Fournier, N. M., Guskjolen, A. J., Lussier, A. L., Marks, W. N., & Kalynchuk, L. E. (2014). Amygdala kindling disrupts trace and delay fear conditioning with parallel changes in Fos protein expression throughout the limbic brain. *Neuroscience, 265*, 158-171. doi: 10.1016/j.neuroscience.2014.01.040

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

Brass, M., & Haggard, P. (2008). The what, when, whether model of intentional action.

Neuroscientist, 14, 319-325. doi: 10.1177/1073858408317417

Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 364*, 2359-2367. doi: 10.1098/rstb.2009.0066

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: Inferential processes versus action simulation. *Current Biology, 17*, 2117-2121. doi: 10.1016/j.cub.2007.11.057

Champagne, F. A., Chretien, P., Stevenson, C. W., Zhang, T. Y., Gratton, A., & Meaney, M. J. (2004). Variations in nucleus accumbens dopamine associated with individual differences in maternal behavior in the rat. *Journal of Neuroscience, 24*, 4113-4123. doi: 10.1523/JNEUROSCI.5322-03.2004

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. doi: 10.1176/appi.ajp.2008.07081360

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Science, 13*, 148-153. doi: 10.1016/j.tics.2009.01.005

Cyranowski, J. M., Hofkens, T. L., Frank, E., Seltman, H., Cai, H. M., & Amico, J. A. (2008). Evidence of dysregulated peripheral oxytocin release among depressed women. *Psychosomatic Medicine, 70*, 967-975. doi: 10.1097/PSY.0b013e318188ade4

Davey, C. G., Yücel, M., & Allen, N. B. (2008). The emergence of depression in adolescence: Development of the prefrontal cortex and the representation of reward. *Neuroscience & Biobehavioral Reviews, 32*, 1-19. doi: 10.1016/j.neubiorev.2007.04.016

Dimaggio, G., Lysaker, P. H., Carcione, A., Nicolo, G., & Semerari, A. (2008). Know yourself and you shall know the other... to a certain extent: Multiple paths of

influence of self-reflection on mindreading. *Consciousness and Cognition, 17*, 778-789. doi: 10.1016/j.concog.2008.02.005

Dunbar, R. I. M., & Shultz, S. (2007). Evolution in the social brain. *Science, 317*, 1344-1347. doi: 10.1126/science.1145463

Ein-Dor, T., Mikulincer, M., Doron, G., & Shaver, P. R. (2010). The attachment paradox: How can so many of us (the insecure ones) have no adaptive advantages? *Perspectives on Psychological Science, 5*, 123-141. doi: 10.1177/1745691610362349

Fearon, P., Shmueli-Goetz, Y., Viding, E., Fonagy, P., & Plomin, R. (2014). Genetic and environmental influences on adolescent attachment. *Journal of Child Psychology and Psychiatry, 55*, 1033-1041. doi: 10.1111/jcpp.12171

Ferris, C. F., Kulkarni, P., Sullivan, J. M., Harder, J. A., Messenger, T. L., & Febo, M. (2005). Pup suckling is more rewarding than cocaine: Evidence from functional magnetic resonance imaging and three-dimensional computational analysis. *Journal of Neuroscience, 25*, 149-156. doi: 10.1523/Jneurosci.3156-04.2005

Fonagy, P., Gergely, G., Jurist, E., & Target, M. (2002). *Affect regulation, mentalization, and the development of the self*. New York, NY: Other Press.

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. doi: 10.1017/s0954579409990198

Fonagy, P., & Luyten, P. (2016). A multilevel perspective on the development of borderline personality disorder. In D. Cicchetti (Ed.), *Developmental Psychopathology* (3rd ed.). New York, NY: Wiley.

Fonagy, P., Luyten, P., & Allison, E. (2015). Epistemic petrification and the restoration of epistemic trust: A new conceptualization of borderline personality disorder and its psychosocial treatment. *Journal of Personality Disorders, 29*, 575-609. doi:

10.1521/pedi.2015.29.5.575

Fonagy, P., Steele, M., Steele, H., Higgitt, A., & Target, M. (1994). The Emanuel Miller Memorial Lecture 1992. The theory and practice of resilience. *Journal of Child Psychology and Psychiatry, 35*, 231-257.

Forbes, E. E., & Dahl, R. E. (2012). Research Review: Altered reward function in adolescent depression: What, when and how? *Journal of Child Psychology and Psychiatry, 53*, 3-15. doi: 10.1111/j.1469-7610.2011.02477.x

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology. The broaden-and-build theory of positive emotions. *American Psychologist, 56*, 218-226.

Fries, E., Hesse, J., Hellhammer, J., & Hellhammer, D. H. (2005). A new view on hypocortisolism. *Psychoneuroendocrinology, 30*, 1010-1016. doi: 10.1016/j.psyneuen.2005.04.006

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*, 531-534. doi: 10.1016/j.neuron.2006.05.001

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences, 8*, 396-403. doi: 10.1016/j.tics.2004.07.002

Garner, B., Chanen, A. M., Phillips, L., Velakoulis, D., Wood, S. J., Jackson, H. J., . . . McGorry, P. D. (2007). Pituitary volume in teenagers with first-presentation borderline personality disorder. *Psychiatry Research, 156*, 257-261. doi: 10.1016/j.pscychresns.2007.05.001

Gilbert, P. (2006). Evolution and depression: Issues and implications. *Psychological Medicine, 36*, 287-297. doi: 10.1017/s0033291705006112

Gunnar, M., & Quevedo, K. (2007). The neurobiology of stress and development. *Annual Review of Psychology, 58*, 145-173. doi: 10.1146/annurev.psych.58.110405.085605

Heim, C., Newport, D. J., Mletzko, T., Miller, A. H., & Hemeroff, C. B. (2008). The link

between childhood trauma and depression: Insights from HPA axis studies in humans. *Psychoneuroendocrinology, 33*, 693-710. doi: 10.1016/j.psyneuen.2008.03.008

Heinrichs, M., & Domes, G. (2008). Neuropeptides and social behaviour: Effects of oxytocin and vasopressin in humans. *Progress in Brain Research, 170*, 337-350. doi: 10.1016/S0079-6123(08)00428-7

Hostinar, C. E., Sullivan, R. M., & Gunnar, M. R. (2014). Psychobiological mechanisms underlying the social buffering of the hypothalamic-pituitary-adrenocortical axis: A review of animal models and human studies across development. *Psychological Bulletin, 140*, 256-282. doi: 10.1037/a0032671

Hsu, D. T., Sanford, B. J., Meyers, K. K., Love, T. M., Hazlett, K. E., Walker, S. J., . . . Zubieta, J. K. (2015). It still hurts: Altered endogenous opioid activity in the brain during social rejection and acceptance in major depressive disorder. *Molecular Psychiatry, 20*, 193-200. doi: 10.1038/mp.2014.185

Humphrey, N. K. (1988). The social function of intellect. In R. W. Byrne & A. Whiten (Eds.), *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans* (pp. 13-26). New York, NY: Oxford University Press.

Insel, T. R., & Young, L. J. (2001). The neurobiology of attachment. *Nature Reviews. Neuroscience 2*, 129-136. doi: 10.1038/35053579

Jogems-Kosterman, B. J., de Knijff, D. W., Kusters, R., & van Hoof, J. J. (2007). Basal cortisol and DHEA levels in women with borderline personality disorder. *Journal of Psychiatric Research, 41*, 1019-1026. doi: 10.1016/j.jpsychires.2006.07.019

Jurist, E. L. (2005). Mentalized affectivity. *Psychoanalytic Psychology, 22*, 426-444. doi: 10.1037/0736-9735.22.3.426

Kanai, R., Bahrami, B., Roylance, R., & Rees, G. (2012). Online social network size is reflected in human brain structure. *Proceedings of the Royal Society B: Biological*

Sciences, 279, 1327-1334. doi: 10.1098/rspb.2011.1959

Kovacs, A. M., Teglas, E., & Endress, A. D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science, 330*, 1830-1834. doi: 10.1126/science.1190792

Lee, R., Geracioti, T. D., Jr., Kasckow, J. W., & Coccaro, E. F. (2005). Childhood trauma and personality disorder: Positive correlation with adult CSF corticotropin-releasing factor concentrations. *American Journal of Psychiatry, 162*, 995-997. doi: 10.1176/appi.ajp.162.5.995

Levy, K. N., Meehan, K. B., Weber, M., Reynoso, J., & Clarkin, J. F. (2005). Attachment and borderline personality disorder: Implications for psychotherapy. *Psychopathology, 38*, 64-74. doi: 10.1159/000084813

Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology, 58*, 259-289. doi: 10.1146/annurev.psych.58.110405.085654

Lombardo, M. V., Barnes, J. L., Wheelwright, S. J., & Baron-Cohen, S. (2007). Self-referential cognition and empathy in autism. *PLOS ONE, 2*, e883. doi: 10.1371/journal.pone.0000883

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Wheelwright, S. J., Sadek, S. A., Suckling, J., . . . Baron-Cohen, S. (2010). Shared neural circuits for mentalizing about the self and others. *Journal of Cognitive Neuroscience, 22*, 1623-1635. doi: 10.1162/jocn.2009.21287

Luciana, M. (2013). Adolescent brain development in normality and psychopathology. *Development and Psychopathology, 25*, 1325-1345. doi: 10.1017/s0954579413000643

Lupien, S. J., McEwen, B. S., Gunnar, M. R., & Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews. Neuroscience, 10*,

434-445. doi: 10.1038/nrn2639

Luyten, P., & Fonagy, P. (2015). The neurobiology of mentalizing. *Personality Disorders: Theory, Research, and Treatment, 6*, 366-379. doi: 10.1037/per0000117

Luyten, P., Fonagy, P., Lemma, A., & Target, M. (2012). Depression. In A. Bateman & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice* (pp. 385-417). Washington, DC: American Psychiatric Association.

Luyten, P., Fonagy, P., Lowyck, B., & Vermote, R. (2012). Assessment of mentalization. In A. W. Bateman & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice* (pp. 43-65). Washington, DC: American Psychiatric Publishing.

Mayes, L. C. (2006). Arousal regulation, emotional flexibility, medial amygdala function, and the impact of early experience: comments on the paper of Lewis et al. *Annals of the New York Academy of Sciences, 1094*, 178-192. doi: 10.1196/annals.1376.018

Meinlschmidt, G., & Heim, C. (2007). Sensitivity to intranasal oxytocin in adult men with early parental separation. *Biological Psychiatry, 61*, 1109-1111. doi: 10.1016/j.biopsych.2006.09.007

Mikulincer, M., & Shaver, P. R. (2007). *Attachment in adulthood: Structure, dynamics and change*. New York, NY: Guilford Press.

Minzenberg, M. J., Grossman, R., New, A. S., Mitropoulou, V., Yehuda, R., Goodman, M., . . . Siever, L. J. (2006). Blunted hormone responses to ipsapirone are associated with trait impulsivity in personality disorder patients. *Neuropsychopharmacology, 31*, 197-203. doi: 10.1038/sj.npp.1300853

Montague, P. R., & Lohrenz, T. (2007). To detect and correct: Norm violations and their enforcement. *Neuron, 56*, 14-18. doi: 10.1016/j.neuron.2007.09.020

Nater, U. M., Bohus, M., Abbruzzese, E., Ditzen, B., Gaab, J., Kleindienst, N., . . . Ehlert, U. (2010). Increased psychological and attenuated cortisol and alpha-amylase responses

to acute psychosocial stress in female patients with borderline personality disorder. *Psychoneuroendocrinology, 35*, 1565-1572. doi: 10.1016/j.psyneuen.2010.06.002

Nestler, E. J., & Carlezon, W. A., Jr. (2006). The mesolimbic dopamine reward circuit in depression. *Biological Psychiatry, 59*, 1151-1159. doi: 10.1016/j.biopsych.2005.09.018

Neumann, I. D. (2008). Brain oxytocin: A key regulator of emotional and social behaviours in both females and males. *Journal of Neuroendocrinology, 20*, 858-865. doi: 10.1111/j.1365-2826.2008.01726.x

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. Oxford, UK: Oxford University Press.

Panksepp, J., & Watt, D. (2011). Why does depression hurt? Ancestral primary-process separation-distress (PANIC/GRIEF) and diminished brain reward (SEEKING) processes in the genesis of depressive affect. *Psychiatry, 74*, 5-13. doi: 10.1521/psyc.2011.74.1.5

Pizzagalli, D. A. (2014). Depression, stress, and anhedonia: Toward a synthesis and integrated model. *Annual Review of Clinical Psychology, 10*, 393-423. doi: 10.1146/annurev-clinpsy-050212-185606

Ripoll, L. H., Snyder, R., Steele, H., & Siever, L. J. (2013). The neurobiology of empathy in borderline personality disorder. *Current Psychiatry Reports, 15*, 344. doi: 10.1007/s11920-012-0344-1

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169-192. doi: 10.1146/annurev.neuro.27.070203.144230

Rochat, P., & Striano, T. (1999). Social-cognitive development in the first year. In P. Rochat (Ed.), *Early social cognition* (pp. 3-34). Mahwah, NJ: Lawrence Erlbaum Associates.

Russo, S. J., & Nestler, E. J. (2013). The brain reward circuitry in mood disorders. *Nature*

Reviews. Neuroscience, 14, 609-625. doi: 10.1038/nrn3381

Rutherford, H. J., Williams, S. K., Moy, S., Mayes, L. C., & Johns, J. M. (2011). Disruption of maternal parenting circuitry by addictive process: Rewiring of reward and stress systems. *Frontiers in Psychiatry, 2*, 37. doi: 10.3389/fpsyt.2011.00037

Sabbagh, M. A. (2004). Understanding orbitofrontal contributions to theory-of-mind reasoning: Implications for autism. *Brain and Cognition, 55*, 209-219. doi: 10.1016/j.bandc.2003.04.002

Sallet, J., Mars, R. B., Noonan, M. P., Andersson, J. L., O'Reilly, J. X., Jbabdi, S., . . . Rushworth, M. F. (2011). Social network size affects neural circuits in macaques. *Science, 334*, 697-700. doi: 10.1126/science.1210027

Satpute, A. B., & Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research, 1079*, 86-97. doi: 10.1016/j.brainres.2006.01.005

Scott, L. N., Levy, K. N., & Granger, D. A. (2013). Biobehavioral reactivity to social evaluative stress in women with borderline personality disorder. *Personality Disorders, 4*, 91-100. doi: 10.1037/a0030117

Shamay-Tsoory, S. G. (2011). The neural bases for empathy. *Neuroscientist, 17*, 18-24. doi: 10.1177/1073858410379268

Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: A lesion study. *Neuropsychologia, 45*, 3054-3067. doi: 10.1016/j.neuropsychologia.2007.05.021

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Levkovitz, Y. (2007). The neuroanatomical basis of affective mentalizing in schizophrenia: Comparison of patients with schizophrenia and patients with localized prefrontal lesions. *Schizophrenia Research, 90*, 274-283. doi: 10.1016/j.schres.2006.09.020

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain, 132*, 617-627. doi: 10.1093/brain/awn279

Spear, L. (2007). The developing brain and adolescent-typical behavior patterns: An evolutionary approach. In D. Romer & E. F. Walker (Eds.), *Adolescent psychopathology and the adolescent brain* (pp. 9-30). New York, NY: Oxford University Press.

Spear, L. P. (2000). The adolescent brain and age-related behavioral manifestations. *Neuroscience and Biobehavioral Reviews, 24*, 417-463. doi: 10.1016/S0149-7634(00)00014-2

Stanley, B., & Siever, L. J. (2010). The interpersonal dimension of borderline personality disorder: Toward a neuropeptide model. *American Journal of Psychiatry, 167*, 24-39. doi: 10.1176/appi.ajp.2009.09050744

Strathearn, L. (2011). Maternal neglect: Oxytocin, dopamine and the neurobiology of attachment. *Journal of Neuroendocrinology, 23*, 1054-1065. doi: 10.1111/j.1365-2826.2011.02228.x

Strathearn, L., Fonagy, P., Amico, J., & Montague, P. R. (2009). Adult attachment predicts maternal brain and oxytocin response to infant cues. *Neuropsychopharmacology, 34*, 2655-2666. doi: 10.1038/npp.2009.103

Strathearn, L., Li, J., Fonagy, P., & Montague, P. R. (2008). What's in a smile? Maternal brain responses to infant facial cues. *Pediatrics, 122*, 40-51. doi: 10.1542/peds.2007-1566

Swain, J. E., Kim, P., Spicer, J., Ho, S. S., Dayton, C. J., Elmadih, A., & Abel, K. M. (2014). Approaching the biology of human parental attachment: Brain imaging, oxytocin and coordinated assessments of mothers and fathers. *Brain Research, 1580*, 78-101. doi:

10.1016/j.brainres.2014.03.007

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. doi: 10.1111/j.1469-7610.2007.01731.x

Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Review of Psychology, 64*, 231-255. doi: 10.1146/annurev-psych-113011-143812

Tottenham, N., & Sheridan, M. A. (2009). A review of adversity, the amygdala and the hippocampus: A consideration of developmental timing. *Frontiers in Human Neuroscience, 3*, 68. doi: 10.3389/neuro.09.068.2009

Uddin, L. Q., Iacoboni, M., Lange, C., & Keenan, J. P. (2007). The self and social cognition: The role of cortical midline structures and mirror neurons. *Trends in Cognitive Sciences, 11*, 153-157. doi: 10.1016/j.tics.2007.01.001

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping, 30*, 829-858. doi: 10.1002/hbm.20547

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *Neuroimage, 48*, 564-584. doi: 10.1016/j.neuroimage.2009.06.009

Viding, E., & McCrory, E. J. (2012). Genetic and neurocognitive contributions to the development of psychopathy. *Development and Psychopathology, 24*, 969-983. doi: 10.1017/s095457941200048x

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. doi: 10.1371/journal.pone.0002868

Walter, M., Bureau, J. F., Holmes, B. M., Bertha, E. A., Hollander, M., Wheelis, J., . . .

Lyons-Ruth, K. (2008). Cortisol response to interpersonal stress in young adults with borderline personality disorder: A pilot study. *European Psychiatry, 23*, 201-204. doi: 10.1016/j.eurpsy.2007.12.003

Watt, D. F., & Panksepp, J. (2009). Depression: An evolutionarily conserved mechanism to terminate separation distress? A review of aminergic, peptidergic, and neural network perspectives *Neuro-Psychoanalysis, 11*, 7-51.

Wingenfeld, K., Lange, W., Wulff, H., Berea, C., Beblo, T., Saavedra, A. S., . . . Driessen, M. (2007). Stability of the dexamethasone suppression test in borderline personality disorder with and without comorbid PTSD: A one-year follow-up study. *Journal of Clinial Psychology, 63*, 843-850. doi: 10.1002/jclp.20396

Wingenfeld, K., Spitzer, C., Rullkotter, N., & Lowe, B. (2010). Borderline personality disorder: Hypothalamus pituitary adrenal axis and findings from neuroimaging studies. *Psychoneuroendocrinology, 35*, 154-170. doi: 10.1016/j.psyneuen.2009.09.014

Figure 6.1. The role of marked mirroring in the development of mentalizing

Figure 6.2. Failure to adequately mirror mental states, problems with mentalizing, and the emergence of alien self-parts

Figure 6.3. A biobehavioral switch model of the relationship between arousal/stress and controlled versus automatic mentalizing

Table 1. Dimensions of mentalizing

Dimension	Defining features	Hypothesized neural circuits
Automatic–	Unconscious, parallel, fast processing of social information versus	Amygdala, basal ganglia, ventromedial prefrontal cortex (VMPFC), lateral temporal cortex (LTC), dorsal anterior cingulate cortex (dACC) versus
Controlled	Conscious, verbal, and reflective processing of social information that relies on effortful control and language	Lateral prefrontal cortex (LPFC), medial prefrontal cortex (MPFC), lateral parietal cortex (LPAC), medial parietal cortex (MPAC), medial temporal lobe (MTL), rostral anterior cingulate cortex (rACC)
Internal–	Understanding one's own mind and that of others through direct focus on the mental interiors	Medial frontoparietal network (more controlled)
External	or through a focus on external features (such as facial expressions, posture, and prosody)	Lateral frontotemporoparietal (more automatic)
Self–Other	Capacity to reflect about both the self and others in terms of inner mental states	Shared representation system (more automatic) versus mental state attribution system (more controlled)
Cognitive–Affective	Focus on cognitive (more controlled), such as belief-desire reasoning and perspective-taking, versus affective features (more automatic), including affective empathy and mentalized affectivity (the feeling and thinking-about-the-feeling), of mental states of self and others	Prefrontal cortex (cognitive mentalizing) versus VMPFC (affectively oriented mentalizing)

Table 2. Automatic nonmentalizing modes that emerge with the loss of controlled mentalizing

Nonmentalizing mode	Features
Psychic equivalence mode	• Equation of inner reality with outer reality: "What I think is real" • Overly concrete/literal understanding
Teleological mode	• Focus on directly observable goals or actions • Only observable changes or actions can be true indicators of the intentions of others
Pretend mode	• Thoughts and feelings are decoupled from external reality • May lead to "dissociation" of thought (hypermentalizing or pseudomentalizing)

Infant internalizes
the mirrored affect

Representation of
infant's mental state
in the attachment
figure's mind

Core of infant's
psychological self

Attachment
figure

Infant

Attachment figure's inference about
infant's mental state (mentalizing)

Mirroring fails

Absence of a representation of the infant's mental state in the attachment figure's mind

Infant's nascent self-representational structure

Unmirrored mental state is internalized as a split-off part of the self – an "alien self" experience

Attachment figure

Attachment figure fails to mentalize the infant and understand the infant's mental state

Infant