We discuss the neural and neurobiological underpinnings of the core features of borderline personality disorder (BPD), including emotion dysregulation, impulsivity, disturbed interpersonal functioning, identity diffusion, and feelings of inner pain. We review neurobiological research that supports a developmental, biobehavioral switch-model of the relationship between mentalization, stress, and attachment. Although it is likely that there are different developmental pathways to BPD, involving complex interactions between environmental, biological, and psychosocial factors, the final common outcome of these pathways is a low threshold for the activation of the attachment system and for deactivation of controlled mentalization, coupled with impairments in the ability to differentiate mental states of self and other. This leads to hypersensitivity and increased susceptibility to contagion from other people’s mental states, poor integration of cognitive and affective aspects of mentalization, and dysfunctions in stress-regulation systems. The treatment implications of this neurobiologically informed model of BPD are discussed.

Borderline personality disorder (BPD) is a severe condition with a lifetime prevalence that has been estimated to be as much as 6% in the general population (Grant et al., 2008). The prevalence of BPD may be particularly high in outpatient (Korzekwa, Dell, Links, Thabane, & Webb, 2008) and forensic populations (Black et al., 2007), where between one fourth and one third meet criteria for the diagnosis. BPD also shows extensive comorbidity with a range of Axis
I and II disorders, notably mood and anxiety disorders, bipolar disorder, and schizotypal and narcissistic personality disorder (Grant et al., 2008).

BPD probably consists of both stable and unstable features; that is, basic temperamental characteristics that show stability over time and more specific, potentially transient symptoms that are related to these personality traits (Costa, Patriciu, & McCrae, 2005). Moreover, it is highly likely that BPD consists of both transient symptoms and features that are manifestations of an acute illness and symptoms that represent more enduring aspects of the disorder, with many patients experiencing long-term disability (Paris & Zweig-Frank, 2001). There is growing consensus that the core features of the disorder include (a) emotional dysregulation (Linehan, 1993; Reisch, Ebner-Priemer, Tschacher, Bohus, & Linehan, 2008); (b) high levels of impulsivity (Grootens et al., 2008), leading to self-harm and suicidality (Black, Blum, Pfohl, & Hale, 2004); and (c) disturbed interpersonal functioning (Hill et al., 2008), as expressed in high levels of preoccupied and disorganized attachment patterns in BPD patients (Levy, 2005). More specifically, BPD seems to be characterized by a pattern of fearful attachment (attachment-anxiety and relational avoidance), painful intolerance of aloneness, hypersensitivity to social environment, expectation of hostility from others, and greatly reduced positive memories of dyadic interactions (Critchfield, Levy, Clarkin, & Kernberg, 2008; Gunderson & Lyons-Ruth, 2008).

From the perspective of the mentalization-based approach to BPD (Bateman & Fonagy, 2004a, 2006; Fonagy & Luyten, 2009), three other facets of BPD that emerge as aspects of disturbed social relatedness from both factor-analytic and clinical studies deserve special attention (e.g., Sanislow et al., 2002). First, dissociative comorbidity is more frequently reported by patients with BPD than by other groups (Sar, Akyuz, Kugu, Ozturk, & Ertem-Vehid, 2006; Zanarini, Frankenburg, Jager-Hyman, Reich, & Fitzmaurice, 2008), and appears to be closely linked to stress (Stiglmayr et al., 2008), emotional neglect (Sar et al., 2006), and suicidal ideation (Klonsky, 2008). Second, BPD also is characterized by a disturbed sense of identity (e.g., Blatt & Auerbach, 1988; Bradley & Westen, 2005), rooted in a dysfunction or deficit of a sense of agency or self-directedness (e.g., Barnow, Ruge, Spitzer, & Freyberger, 2005; Bender & Skodol, 2007). Finally, BPD seems to be associated with the experience of intense inner pain (Zanarini & Frankenburg, 2007), particularly in relation to disturbed relationships (Gunderson & Lyons-Ruth, 2008), and childhood abuse and neglect (Holm & Severinsson, 2008). More specifically, these feelings are closely related to feelings of rejection, abandonment, and/or isolation (e.g., Stiglmayr et al., 2008), involve an intense sense of shame (Levy, Edell, & McGlashan, 2007; Rusch et al., 2007) and can prompt dissociation (Philipson et al., 2004) and self-harm (Coid, 1993).

BPD remains a somewhat controversial diagnosis in terms of homogeneity, comorbidity between and within axes, arbitrary cutoffs, and poor test-retest reliability (Miller, Muehlenkamp, & Jacobson, 2008). Furthermore, current dimensional models such as the Five-Factor Model of personality may not necessarily be the most suitable for a detailed clinical description of maladaptive features of personality pathology at different ages (Blatt & Luyten, 2009; De Fruyt et al., 2006; Westen, Dutra, & Shedler, 2005). Nevertheless, studies based on dimensional personality models have shed some light on the developmental precursors of BPD. In one study using the Dimensional Personality Symptom item pool, for instance, separation anxiety, depressive traits, insecure attachment, and ineffective stress coping were found to predict trait pathology in childhood (De Clercq, De Fruyt, Van Leeuwen, & Mervielde, 2006). In a 2-year prospective study (Crick, Murray-Close, & Woods, 2005) with a representative sample of 400 children in Grades 4 to 6, it was found that the developmental precursors of BPD features in adolescence showed moderate stability, although “emotion sensitivity” (intense, unstable, inappropriate emotion)
Infancy and Borderline Personality Disorder

and “physical aggression” were no longer significant predictors once level of depression had been controlled for. Social dysfunction (including friend exclusivity; that is, being overly close with friends), relational aggression (impulsivity), and cognitive sensitivity (hostile, untrusting paranoid world view) were the best predictors of borderline pathology in this study. Moreover, specific features of BPD such as self-harm or impulsivity and affective instability during childhood and adolescence have been found to predict adult BPD diagnosis in retrospective studies (Zanarini et al., 2006). Importantly, both clinical- and community-based studies have found only moderate stability of the diagnosis in adolescence (Chabrol & Leichsenring, 2006) and adulthood (Chanen et al., 2004; Cohen, 2008; Deschamps & Vreugdenhil, 2008).

We suggest that genetic and early environmental factors may undermine the development of mentalized affectivity (second order representations of emotional states). The resulting limitations of infant affect regulation will undermine the development of effortful control and the development of a robust understanding of others as motivated by mental states. These are, as with most developmental processes, potentially interactive and bidirectional in terms of causation. Poor affect regulation obviously makes sensitive caregiving more challenging, and the impacts of some environmental influences are evidently exaggerated by certain genetic attributes (e.g., the short allele of the 5-HTT gene; Barry, Kochanska, & Philibert, 2008). Accurate and solid interpersonal relationships, and the ability to direct attention voluntarily, contribute to the emergence of a sound behavioral system that underpins mature attachment relationships. We assume that there are several pathways to the development of BPD depending on the interaction between environmental and biological factors (e.g., Johnson et al., 2002; Torgersen et al., 2000), ranging in severity from individuals who are at increased risk for BPD because of deficits in mentalization (e.g., because they have grown up in a family context characterized by low levels of mentalization and little or no attention to internal mental states) to individuals who are characterized by a defensive decoupling and inhibition of mentalizing because of experiences of abuse and neglect. Moreover, depending on the use of different secondary attachment strategies and contextual factors, some BPD patients will be primarily characterized by preoccupied or avoidant attachment while the attachment system will be disorganized (either from infancy or as a consequence of subsequent stress) in other individuals.

The disorganization of attachment relationships in our view also disorganizes the self-structure, creating incoherence and splitting, which makes stress particularly hard to manage. The key consequence of attachment dysregulation in individuals with BPD is the hyperreactivity of the attachment system, leading to frantic efforts to avoid abandonment, the diagnostic unstable and intense pattern of interpersonal relationships, and a characteristic rapidly escalating tempo moving from acquaintance to great intimacy over extremely brief time periods. We have suggested that the “hyperreactivity” of the attachment system in these patients, possibly linked with traumatic experiences, may be one of the pathways to impairments of mentalization in BPD—intense affect is incompatible with judgments of social trustworthiness (Fonagy & Bateman, 2006, 2008). The vulnerability to an inhibition or decoupling of mentalization may occur for other reasons, such as the understandable reluctance of a maltreated child to contemplate the mental state of adults with frankly destructive thoughts and wishes in relation to him or her. At these times, mature mentalization gives way in these patients to prementalistic modes of subjectivity whereby the thoughts and feelings lose their “as if” quality and become equivalent to physical events, observable physical reality becomes the only criterion for truth, and the internal world genuinely can be separated and be experienced as having no real implications for the world outside as in a very extreme form of pretence. Recently accumulated data suggest
that a further elaboration of this already complex model may be necessary based on improved understanding of the biology of attachment and the neural basis of mentalization (Luyten et al., unpublished manuscript; Luyten, Mayes, Fonagy, & Van Houdenhove, 2009).

THE NEUROBIOLOGY OF ATTACHMENT

Studies on rodent models and human neuroimaging studies have helped clarify the nature of the brain systems underpinning attachment behavior. Two decades ago, MacLean (1990) speculated that substance abuse and drug addiction were attempts to replace opiates or endogenous factors normally provided by social attachments. Ten years later, Jaak Panksepp (1998), the brilliant neurobiologist, identified a common neurobiology to mother–infant, infant–mother, and romantic-partner attachments linked to dependence-inducing narcotic opioid alkaloids. The link of substance dependence to the neural substrate for attachment behaviors was crystallized by the work of Tom Insel and colleagues (see Insel & Young, 2001) at Emory University, who in the title of a major integrative article quipped “Is social attachment an addictive disorder?” (Insel, 2003).

Extant research has suggested that the attachment system incorporating Bowlby’s (1969, 1973) classical secure-base system and the caregiving system as well as pair bonding is associated with the activation of two neural systems that have consistently been shown to play an important role in promoting and maintaining maternal behavior: (a) the dopaminergic reward processing system (Champagne et al., 2004; Ferris et al., 2005; Strathearn, Li, Fonagy, & Montague, 2008) and (b) the oxytocinergic system (Bartels & Zeki, 2004; Champagne, Diorio, Sharma, & Meaney, 2001; Levine, Zagoory-Sharon, Feldman, & Weller, 2007), oxytocin being a neuroactive hormone that is directly synthesized in the hypothalamus and projects to brain areas that are associated with emotions and social behaviors (e.g., striatum, amygdale, and cingulate cortex). Neuropeptides, other than oxytocin, including opioids and vasopressin, also may be involved in mediating some of the dysregulated social experience of disorganized attachment relationships (Stanley et al., 2009; Stanley & Siever, 2009). Vasopressin may play an analogous role to oxytocin, perhaps especially in males (Caldwell, Lee, Macbeth, & Young, 2008; Lim, Murphy, & Young, 2004; Lim, Wang et al., 2004; Lim & Young, 2004; Young & Hammock, 2007).

The ability to identify the needs of infants from visual and other sensory cues, and to differentially respond to these, is thought to be the basis for establishing secure mother–infant attachment (Fonagy & Target, 2005). Thus, a mother’s behavioral and brain response to her infant’s cues may be an important predictor of infant development. Several research groups have sought to understand how a mother’s brain responds to her child’s auditory or visual cues, using functional MRI (fMRI) (Leibenluft, Gobbini, Harrison, & Haxby, 2004; Lorberbaum et al., 2002; Ranote et al., 2004; Seifritz et al., 2003; Swain, 2008; Swain et al., 2008). One common theme emerging from these studies has been the possible role of the mesocorticolimbic dopamine system in processing reward-based signals and motivating maternal care, as seen in animal models (see comprehensive review in Swain, Lorberbaum, Kose, & Strathearn, 2007). Several studies have shown that the striatum, a key projection of midbrain dopamine neurons which includes the putamen and caudate head, is activated in response to face images of a mother’s own child compared to unknown (or familiar, but unrelated) children (Bartels & Zeki, 2004; Leibenluft et al., 2004) as well as to infant cry stimuli (Lorberbaum et al., 2002; Swain et al., 2008). The observation of comparable activation patterns in response to pictures of romantic partners (Bartels & Zeki, 2000; Zeki, 2007) and beautiful faces (Aharon et al., 2001) also implies a possible link between brain reward circuits and the human attachment system.
Strathearn et al. (2008) addressed important inconsistencies in relation to maternal response studies that have failed to show striatal activation (e.g., Seifritz et al., 2003) by specifically controlling for variation in facial affect, comparing maternal brain responses to happy, neutral, and sad baby faces. Their study showed that all of the key dopaminergic regions and cortical regions which project to and from these striatal regions were activated when mothers viewed their own babies’ faces compared to unknown baby faces. Own versus unknown baby face contrasts revealed specific hemodynamic responses in prominent dopaminergic brain regions involved in cognitive and affective information processing: (a) the limbic circuits in mothers’ brains responding preferentially to their own infant, including the midbrain ventral tegmental area, ventral striatum, ventral anterior cingulate cortex, insula, and medial prefrontal cortex; and (b) a cognitive associative network including the substantia nigra, precommissural dorsal caudate, precommissural dorsal putamen, ventral anterior thalamic nucleus, and the dorsolateral prefrontal cortex. The first-time mothers in this study showed the strongest responses when viewing their own baby’s happy face, and the weakest responses when the baby’s face was sad. There are feed-forward loops between the striatum and the substantia nigra, suggesting that these striatonigrostrial circuits funnel information between ventromedial (limbic), central (associative), and dorsolateral (motor) striatal regions (Haber, Fudge, & McFarland, 2000). Each striatal region is integrally connected to a corresponding region of the midbrain’s ventral tegmental area and substantia nigra via ascending and descending dopaminergic neurons. Likewise, there are corresponding connections between the striatum and the forebrain. Thus, the striatum is believed to be an important relay station between the limbic and motor systems, integrating affective information from limbic regions with cognitive information from the prefrontal cortex, in shaping motor/behavioral responses.

In responding to infant social cues, whether positive or negative, mothers need to integrate both affective and cognitive information about their baby as well as evaluate competing demands before executing the most appropriate behavior. For example, a distressed baby usually evokes an empathic emotional response from a mother as well as cognitive processes to determine, based on past experience and knowledge, possible causes and remedies for her baby’s distress. Likewise, a smiling baby face usually leads to positive affective arousal in a mother, associations with other rewarding experiences, and contingent behavioral responses such as smiling, caressing, or playing. Cocaine, a common drug of abuse which activates both mesocorticolimbic and nigrostriatal dopamine systems, appears to compete with natural infant-related reward signals (Molitor, Mayes, & Ward, 2003), which may help to explain relatively high rates of child maltreatment in cocaine-exposed mothers.

**EVIDENCE LINKING OXYTOCIN TO THE ATTACHMENT SYSTEM**

The neuropeptides oxytocin and vasopressin play key roles in two aspects of creating attachment relationships: (a) by activating the reward/attachment system (the “push” mechanism involved in attachment) (Heinrichs & Domes, 2008) and (b) by deactivating neurobehavioral systems that are involved in mediating social avoidance (the “pull” mechanism involved in attachment). Rodent research has demonstrated that oxytocin plays a role in attachment and prosocial behavior (Insel & Young, 2001). For instance, oxytocin and vasopressin have been shown to inhibit aversion of both female and male rodents to infant pups as well as lead to a number of affiliative behaviors, including caregiving behavior (Insel & Young, 2001).

Oxytocin also has been structurally linked with brain regions involved in attachment. Oxytocin receptors are located in the ventral striatum, a key dopaminergic brain region, and
receptor binding is linked functionally to maternal behavior in the rat (Olazabal & Young, 2006).

Oxytocin receptors are abundant in areas involved in attachment and other social behaviors (the bed nucleus of the stria terminalis, the hypothalamic paraventricular nucleus, the central nucleus of the amygdala, the ventral tegmental area, and the lateral septum) (Francis, Champagne, & Meaney, 2000; Stanley & Siever, 2009). The remarkable work of Young, Lim, Gingrich, and Insel (2001) showed that oxytocin receptors are more numerous in the ventral pallidum subspecies of voles (the prairie or pine voles), which have monogamous attachments, when compared with the promiscuous meadow or montane voles. This applies to vasopressin (V1a) receptor distribution as well. Remarkably, partner-preference formation in the socially promiscuous meadow vole can be increased by using viral vector V1aR gene transfer into the ventral forebrain, which leads to an increase in partner preference (in terms of time spent huddling), thus providing a potential molecular mechanism for the rapid evolution of a complex social behavior (Lim, Wang et al., 2004).

Increasing the expression of oxytocin receptors in these regions enhances maternal care behavior in rodents (Francis et al., 2000). The same process, enhanced oxytocin activity, may play a part in human offspring’s attachment feelings toward their parents (Gordon et al., 2008). Oxytocin may link sensitivity to social cues, such as infant facial expressions, with dopamine-associated reinforcement pathways (Strathearn, Fonagy, Amico, & Montague, 2009). Oxytocin is a facilitator of social connectedness at a number of levels. It has been found to facilitate social function in experimental studies, including improving social memory, and memory of facial expressions and identity (Baumgartner, Heinrichs, Vonlanthen, Fischbacher, & Fehr, 2008; Guastella, Mitchell, & Dadds, 2008; Guastella, Mitchell, & Mathews, 2008). Oxytocin appears to selectively facilitate positive, happy social memories (Guastella, Mitchell, & Mathews, 2008). It creates or activates a positive emotional experience in relation to other people [e.g., helping the person overcome negative (conditioned) associations in relation to someone by attenuating activity in the extended amygdala] (Petrovic, Kalisch, Singer, & Dolan, 2008). It also seems to generate a sense of trust in relation to the putative intentions of the other (Bartz & Hollander, 2006; Theodoridou, Rowe, Penton-Voak, & Rogers, 2009), and trust is the aspect of attachment that is considered by a number of us to be one of its key evolutionarily conserved facets in humans (Fonagy, Gergely, & Target, 2007; Gergely, 2007). Affective states associated with secure (or even insecure organized) attachment are the emotional reactions indicating to human infants that the information that is being presented to them by an adult may be relied on (or be epistemically trusted) to be incorporated into semantic memory (Gergely, Kiraly, & Egyed, 2007). These states seem to be a mediator of the potential to experience interpersonal connectedness, the human equivalent of licking and grooming (Francis et al., 2000). In a fascinating study, Gordon et al. (2008) demonstrated that in a sample of young adults without serious romantic attachments, serum oxytocin was reasonably strongly correlated with the self-reported quality of bonding with parents ($r = .42$) and predicted relatively lower levels of depression. In a study of cohabiting couples, greater partner support was found to be associated with higher plasma oxytocin after a period of warm partner contact (Grewen, Girdler, Amico, & Light, 2005).

**Oxytocin and Stress**

Oxytocin reduces behavioral and neuroendocrine responses to social stress and seems both to enable animals to overcome their natural avoidance of proximity and to inhibit defensive behavior, thereby facilitating approach behavior (Heinrichs & Gaab, 2007). Vasopressin has primarily
been implicated in male-typical social behaviors, including aggression and pair-bond formation, and mediates anxiogenic effects (Heinrichs & Domes, 2008). Experimentally administered oxytocin reduces the stress response in a laboratory stress test to a similar extent as the presence of active social support (Heinrichs, Baumgartner, Kirschbaum, & Ehler, 2003). Women who are breastfeeding appear to be less physiologically reactive to stress, which may be accounted for by the stimulation of the nipple as well as the close social contact, both of which increase oxytocin release (Amico, Johnston, & Vagnucci, 1994). This increased resilience to stress may account for the epidemiologic observation of reduced likelihood of child maltreatment in mothers who breastfeed (Strathearn, Mamun, Najman, & O’Callaghan, 2009).

Oxytocin can inhibit hypothalamicpituitary stress axis activity, suggesting that activation of the attachment system may generate increased experience of reward, increased sensitivity to social cues, and decreased social avoidance, but also the potential for the reward to override lack of trust (Luyten, Mayes, Fonagy, & Van Houdenhove, 2009). This complex set of associations with social behavior may help us to account for the puzzling combination of facilitative and inhibitory associations between attachment history and social cognition.

Oxytocin May Mediate the Relationship of Attachment and Mentalizing

Oxytocin is, of course, best known for its roles in female reproduction: It is released in large amounts after distension of the cervix and vagina during labor and after stimulation of the nipples, facilitating childbirth and breastfeeding, respectively. Its mentalization-enhancing function presents a powerful opportunity for the mother to optimize her understanding of and focus on the mental state of her infant by close observation of the infant’s expression ("minding the eyes"). Feldman, Weller, Zagoory-Sharon, and Levine (2007) demonstrated that oxytocin concentrations during early pregnancy and during the postpartum period are strongly associated with maternal bonding as suggested by indicators such as positive affect, gaze, affectionate touch, and frequency of monitoring the infant as well as thoughts related to attachment. The functioning of the oxytocin mechanism provides an obvious and potentially powerful account for the association we have reported linking attachment and mentalization processes. As we have seen, studies with intranasally administered oxytocin have shown oxytocin to improve social memory (Guastella, Mitchell, & Mathews, 2008; Savaskan, Ehrhardt, Schulz, Walter, & SchSchinger, 2008), and increase trust, generosity, and the experience of empathy associated with watching emotional videos in humans (Barraza & Zak, 2009; Baumgartner et al., 2008; Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005). More pertinently in this context, oxytocin enhances the recognition of mental states revealed by facial expressions (Domes, Heinrichs, Michel, Berger, & Herpertz, 2007), probably by causing selective fixation on the eye region when viewing faces (Guastella, Mitchell, & Dadds, 2008).

This is at least suggestive evidence that the engagement of the caregiver’s attachment system, which increases the level of oxytocin, may make the caregiver more sensitive to the infant’s mental state, which in turn may generate (a) a greater likelihood of secure attachment between the infant and that parent and (b) a more rapidly developing understanding of minds on the part of the child. If oxytocin indeed plays a critical role in attachment-related mentalization, we would expect that conditions which are known to interfere with mentalization to be linked with oxytocin as well. For example, the finding of lower urinary oxytocin concentrations in maltreated children is consistent with the assumption that oxytocin mediates human attachment behavior (Fries, Ziegler, Kurian, Jacoris, & Pollak, 2005). Similarly, in adult males with a history
of early separation, intranasal challenge with oxytocin resulted in attenuated cortisol decreases compared with control subjects, suggesting altered central sensitivity to the effects of oxytocin in this group (Meinlschmidt & Heim, 2007). These findings echo primate research, which also has found lower cerebrospinal-fluid oxytocin levels in nursery-reared monkeys compared to those of mother-reared monkeys (Winslow, Noble, Lyons, Sterk, & Insel, 2003). Perhaps even more persuasive is a unique study which has found lower cerebrospinal fluid oxytocin in women who experienced maltreatment, particularly emotional abuse and/or neglect (Heim et al., 2008).

There also is reasonable evidence linking the maltreatment of young children with problems in mentalization. Maltreated children engage in less symbolic and dyadic play (Alessandri, 1991); they may fail to show typical empathic responses to distress in other children (Howes & Espinosa, 1985; Klimes-Dougan & Kistner, 1990; Main & George, 1985); they make fewer references to internal states, and maltreating mother–child dyads discuss emotions less frequently than do nonmaltreating dyads (Beeghly & Cicchetti, 1994; Shipman & Zeman, 1999); and they manifest delayed theory-of-mind understanding (Cicchetti, Rogosch, Maughan, Toth, & Bruce, 2003; Pears & Fisher, 2005). Maltreated preschoolers have poorer understanding of universal adult (During & McMahon, 1991) and child facial expressions of emotion (Camras, Grow, & Ribordy, 1983; Camras et al., 1990) and show clear limitations of social cognition in story-stem completion tasks (Macfie, Cicchetti, & Toth, 2001; Macfie et al., 1999). They also manifest a delay in emotion-focused mentalization (Cicchetti, Rogosch, Maughan, Toth, & Bruce, 2003; Pears & Fisher, 2005; Rogosch, Cicchetti, & Aber, 1995; Smith & Walden, 1999), which was found to predict social competence at 8 years of age (Rogosch et al., 1995). Lower oxytocin levels may mediate the apparent undermining of the robust establishment of representations of mental states in maltreated individuals.

Attachment History and Individual Differences in Stress Responsivity

In the context of secure attachment, the activation of the attachment system predictably involves a relaxation of normal strategies of interpersonal caution. Congruent with this assumption, expressions in most languages associate love with various severe forms of sensory handicap, particularly blindness. There is good evidence that intense activation of the neurobehavioral system underpinning attachment is associated with deactivation of arousal and affect regulation systems (Luyten et al., 2009) as well as deactivation of neurocognitive systems likely to generate interpersonal suspicion; that is, those involved in social cognition or mentalization, including the lateral prefrontal cortex, medial prefrontal cortex (MPFC), lateral parietal cortex, medial parietal cortex, medial temporal lobe, and rostral anterior cingulated cortex (Bartels & Zeki, 2000, 2004; Lieberman, 2007; Mayes, 2000, 2006; Satpute & Lieberman, 2006).

The activation and deactivation of the attachment system appears to be closely linked to arousal and stress regulation (Heinrichs & Domes, 2008; Lieberman, 2007; Mayes, 2006). Following the model outlined by Mayes (2000, 2006), we suggest that with increased arousal, there is a switch from cortical to subcortical systems, from controlled to automatic mentalizing and subsequently to nonmentalizing modes. Based on Arnsten’s ((1998) dual-process model, Mayes (2000, 2006) proposed that stress regulation is not a generalized state of activation/deactivation but a differential balance of excitation and inhibition involving multiple, interactive neural systems with different neurochemical substrates regulating specific and different aspects of prefrontal, posterior cortical, and subcortical functions. For instance, as the level of cortical activation increases through mutually interactive norepinephrine alpha 2 and dopamine D1
systems, prefrontal cortical function improves, including the capacity for attentional control, planning/organization, and explicit mentalization. However, with further increases of stimulation, norepinephrine alpha 1 and dopamine D1 inhibitory activity increases to the point that the prefrontal cortex goes “offline,” and posterior cortical and subcortical functions (e.g., more automatic, implicit, affect-focused forms of mentalization) are enhanced and finally take over. Increasing levels of norepinephrine and dopamine interact such that above a certain threshold, the balance shifts from prefrontal executive functioning to amygdala-mediated memory encoding and posterior-subcortical automatic responding (fight–flight–freeze).

There are good reasons to suppose that different attachment histories are associated with attachment styles that differ in terms of the associated background level of activation of the attachment system, and the point at which the switch from more prefrontal, controlled to more automatic mentalizing occurs (Luyten et al., unpublished manuscript). Dismissing individuals tend to deny attachment needs, asserting autonomy, independence, and strength in the face of stress, using attachment-deactivation strategies. In contrast, a preoccupied attachment classification or an anxious attachment style are generally thought to be linked with the use of attachment hyperactivating strategies (Cassidy & Berlin, 1994; Cassidy & Kobak, 1988; Mikulincer & Shaver, 2003, 2007, 2008). Attachment hyperactivating strategies have been consistently associated with the tendency to exaggerate both the presence and seriousness of threats, and frantic efforts to find support and relief, often expressed in demanding, clinging behavior. In the context of this article, note that both Adult Attachment Interview studies and self-report studies have found a predominance of anxious-preoccupied attachment strategies in BPD patients (Agrawal, Gunderson, Holmes, & Lyons-Ruth, 2004; Fonagy et al., 1996; Gunderson, 1996; Levy, 2005; Levy, Meehan, Weber, Reynoso, & Clarkin, 2005; Minzenberg, Poole, & Vinogradov, 2006; Patrick, Hobson, Castle, Howard, & Maughan, 1994; Rosenstein & Horowitz, 1996; Stovall-McClough & Clotire, 2003), although there is every indication that the two instruments are sensitive to different forms of psychological dysfunction (Riggs et al., 2007). In one study, 75% of personality disorder patients meeting criteria for BPD fell into the rarely used subgroup of the AAI (E3): “fearfully preoccupied with respect to trauma” (Fonagy et al., 1996). In BPD patients, we and others noted a characteristic pattern of fearful attachment (attachment-anxiety and relational avoidance), painful intolerance of aloneness, hypersensitivity to social environment, expectation of hostility from others, and greatly reduced positive memories of dyadic interactions (e.g., Critchfield et al., 2008; Gunderson & Lyons-Ruth, 2008).1

Neuroimaging studies have supported the notion that attachment history affects the setting of the “switch,” which turns the mentalizing system from planned, controlled, and organized cognition to automatic processing with narrowed, poorly sustained attention, and increased vigilance for attachment disruptions such as rejection and abandonment. For example, Gillath, Bunge, Shaver, Wendelken, and Mikulincer (2005) reported an fMRI study of women with high and low scores on attachment anxiety and avoidance who were asked to think about or stop thinking about various relationship scenarios. When thinking about negative scenarios, women

---

1Recently, in a compelling formulation of the relationship of attachment, maltreatment, and cognitive deficits that overlap with the current proposals in many ways, Minzenberg, Poole, and Vinogradov (2008) also suggested that BPD patients’ reactivity in attachment relationships was related to temporal-limbic dysfunction and avoidance within attachment relationships, which could be a relational strategy that serves to compensate for the emotional consequences of frontal-executive dysregulation, while childhood adversity contributed both independently and in interaction with neurocognitive dysfunction.
with high levels of attachment anxiety showed more activation in the hippocampus (memory retrieval) and in emotion-related brain regions (the anterior temporal pole and the dorsal anterior cingulate), and less activation in the orbitofrontal cortex. It seems that in contemplating negative scenarios, those with an anxious attachment style may underrecruit the brain areas associated with emotion regulation and show enhanced retrieval of negative memories. In a prospective study, Strathearn, Fonagy, et al. (2009) examined whether differences in attachment security of 30 first-time mothers, assessed using the AAI before the birth of their child, were related to brain reward and peripheral oxytocin response to infant cues. About 10 months after the birth of their child, mothers viewed their own or others’ infants’ smiling and crying faces during fMRI scanning. Mothers with secure attachment showed greater activation of brain reward regions, including the ventral striatum, and the oxytocin-associated hypothalamopituitary region. Peripheral oxytocin response during infant contact also was significantly higher in secure mothers, and the size of change from baseline oxytocin levels was positively correlated with brain activation to their own infants in both regions. The important differences based on attachment history emerged when the mothers viewed their infants’ sad faces. Securely attached mothers continued to show greater activation in reward-processing regions while insecure/dismissing mothers, congruent with findings described earlier, showed reward system deactivation and insular activation in response to seeing their own infants’ sad faces. The insula may be a region associated with feelings of unfairness, pain, and disgust (see review by Montague & Lohrenz, 2007). Mothers with insecure/dismissing attachment histories appeared less able to downregulate the sad feelings evoked in them by their infants’ sad faces, possibly because they felt overwhelmed by sad memories of their own past. We have argued that the insular activation signals that the infant’s sad expression elicits sadness in insecure mothers, to the extent that they might not be able to provide “marked contingent mirroring” to their infants. This mirroring is necessary for infants to internalize the mother’s expression as a representation of their own experience of sadness and thus assist in the organization of their emotional responses. If confirmed by subsequent investigations, the difference in hemodynamic response between mothers of secure and insecure attachment histories may be part of the answer to the puzzle of the transmission gap identified by researchers who have been unable to identify a mediator that could account for the concordance of maternal and infant attachment security (Fonagy, Steele, & Steele, 1991; van IJzendoorn, 1995). These findings have suggested that for securely attached mothers, infant cues (whether positive or negative in affect) may act as an important affective signal of “incentive salience” (Berridge, 2007), reinforcing and motivating responsive maternal care. However, insecure mothers may experience a negative subjective reaction that would make them reflect (i.e., mirror) their infants’ sadness without being able to create a symbolic/mentalizing distance between their infants’ and their own states of mind.

The finding of reduced “reward” activation in mothers with insecure/dismissing attachment also is consistent with a recent study of responses to smiling adult faces and positive task feedback (Vrticka, Andersson, Grandjean, Sander, & Vuilleumier, 2008). This study reported a negative correlation between dismissing attachment scores and activation of the ventral striatum. This is certainly congruent with the typical distant and detached relational style of avoidant individuals, and their tendency to deny the importance of attachment relationships. In both studies, challenges to the attachment system yielded greater ventral striatal activation in securely attached individuals while in avoidant/dismissing individuals, it was associated with a relative downregulation of reward-related activity. Consistent with the earlier claims about the complementary activation of the attachment and mentalizing systems, in the Vrticka et al. (2008) study,
avoidant attachment was positively related to activation in the MPFC and the Ventral Anterior Cingulate Cortex, areas that have been implicated in controlled social cognition and mentalization, social rejection, and emotion suppression. Moreover, Vrticka et al. also reported that secure attachment was positively related to the activation of the ventral striatum in response to positive reinforcement, but negatively with activation of the amygdala to negative reinforcement. This is in line with previous suggestions that we have made that secure attachment requires the simultaneous (paradoxical) activation of components that are normally reciprocally activated, mentalizing, and reward-salience associated regions of the brain (Fonagy & Bateman, 2006). These observations also are congruent with the assumption that secure attachment consists of a combination of low anxiety and low avoidance.

TREATMENT IMPLICATIONS

The therapeutic implications of the model outlined earlier are summarized in greater detail elsewhere (Allen, Fonagy, & Bateman, 2008; Bateman & Fonagy, 2004b, 2006). Yet, what must be clear from our formulations is that they imply the need to abandon the overvaluation of specific techniques in favor of a generic therapeutic stance that cuts across theoretical modalities. More specifically, given the generic nature of mentalization as a mental function, most treatments of BPD are expected to be effective to the extent that they include important components facilitating mentalization, even though this capacity may be addressed using different languages by various models of therapy, such as “mindfulness,” “validation,” “self-states,” and so on. Randomized clinical trials have shown a number of psychological treatments to be moderately effective when compared to either routine care or other therapies (Bateman & Fonagy, 2008; Blum et al., 2008; Clarkin, Levy, Lenzenweger, & Kernberg, 2007; Davidson et al., 2006; Giesen-Bloo et al., 2006; Linehan, Comtois, Brown, Heard, & Wagner, 2006), and most authors now suggest that an integrative framework should be adopted to explain treatment effects (Clarkin, 2006; de Groot, Verheul, & Trijsburg, 2008; Levy et al., 2006; Lynch, Chapman, Rosenthal, Kuo, & Linehan, 2006; Paris, 2008).

From the perspective developed in this article, the overall aim of treatment should be to simultaneously stimulate a patient’s attachment and involvement with treatment while also helping him or her maintain mentalization. A titrated, but more or less exclusive, focus on the BPD patient’s current mental state while activating the attachment relationship is expected to enhance the patient’s mentalizing capacities without generating iatrogenic effects, as it inevitably activates the attachment system. Hence, treatment should avoid situations where patients are expected to talk of mental states that they cannot link to subjectively felt reality. Thus, with regard to dynamic therapies, this implies that there should be (a) a de-emphasis of “deep,” unconscious interpretations in favor of conscious or near-conscious content; (b) a modification of therapeutic aim, especially with severely disturbed patients, from insight to recovery of mentalization (i.e., achieving representational coherence and integration); (c) careful eschewing of descriptions of complex mental states (conflict, ambivalence, unconscious) that are incomprehensible to a person whose mentalizing is vulnerable; (d) avoidance of extensive discussion of past trauma, except in the context of reflecting on current perceptions of mental states of maltreating figures and changes in mental state from being a victim in the past versus one’s experiences now.

Our theoretical model also implies that to maximize impact on the patient’s ability to think about thoughts and feelings in relationship contexts, especially in the early phases of treatment, the therapist is probably most helpful when interventions (a) are simple and easy to understand,
(b) are affect-focused, (c) actively engage the patient, (d) focus on the patient’s mind rather than on his or her behavior, (e) relate to current event or activity—whatever the patient’s currently felt mental reality (in working memory), (f) make use of the therapist’s mind as a model (by therapists disclosing their anticipated reaction in response to the event being discussed (i.e., talking to the patient about how the therapist anticipates that he or she might react in that situation), and (g) are adjusted flexibly regarding their complexity and emotional intensity in response to the intensity of the patient’s emotional arousal (withdrawing when arousal and attachment are strongly activated).

The key task of therapy is thus to promote curiosity about the way mental states motivate and explain the actions of self and others. Therapists achieve this through the judicious use of “the inquisitive stance,” highlighting their own interest in the mental states underpinning behavior, qualifying their own understanding and inferences (i.e., showing respect for the opaqueness in mental states), and demonstrating how such information can help the patient to make sense of his or her experiences. Pseudomentalization and other fillers to replace genuine mentalization must be explicitly identified by therapists, and the lack of practical success associated with them should be clearly highlighted. In this way, therapists can help their patients to learn about how they think and feel about themselves and others, how that shapes their responses to others, and how “errors” in understanding self and others may lead to inappropriate actions. Put simply, it is not for therapists to “tell” patients about how they feel, what they think, or how they should behave, or what the underlying reasons may be, conscious or unconscious, for their difficulties. Any therapy approach which moves toward claiming to “know” how patients “are,” how they should behave and think, and “why they are the way they are,” is likely to be harmful to patients with a vulnerable capacity to mentalize. This principle applies to cognitive behavioral therapy as much as it does to psychodynamic psychotherapy. For example, Davidson, Livingstone, McArthur, Dickson, and Gumley (2007) demonstrated that high levels of therapists’ integrative complexity (an indication of the number of ideas being combined in a single statement) was associated with poor outcome in cognitive behavioral therapy rather than with good outcome while the patients’ increase in integrative complexity marked improvement in social functioning.

From a practical standpoint, an effective mentalization-based intervention may unfold along the following lines: (a) The therapist identifies a break in mentalizing (described earlier as psychic equivalence, pretend, or teleological modes of thought), (b) patient and therapist rewind to the moment before the break in subjective continuity, (c) the current emotional context for the break in the session is explored by identifying the momentary affective state between patient and therapist, (d) the therapist explicitly identifies and owns up to his or her own contribution to the break in mentalizing, and (e) the therapist seeks to help the patient understand the mental states implicit in the current state of the patient–therapist relationship (i.e., mentalize the transference).

The therapist’s mentalizing therapeutic stance should include (a) humility deriving from a sense of “not knowing;” (b) whenever possible, taking time to identify differences in perspectives; (c) legitimizing and accepting different perspectives; (d) active questioning of the patient in relation to his or her experience—asking for detailed descriptions of experience (“what” questions) rather than explanations (“why” questions); and (e) eschewing the need to understand what makes no sense (i.e., saying explicitly that something is unclear). An important component of this stance is monitoring one’s own mistakes as a therapist. This not only models honesty and courage through such acknowledgments and tends to lower arousal through the therapist taking responsibility but also offers invaluable opportunities to explore how mistakes can arise out of mistaken assumptions about opaque mental states and how misunderstanding can lead
Infancy and Borderline Personality Disorder

...to massively aversive experiences. In this context, it is important to be aware that the therapist is constantly at risk of losing the capacity to mentalize in the face of a nonmentalizing patient. Consequently, we consider therapists’ occasional enactments as an acceptable concomitant of the therapeutic alliance, something that simply has to be owned up to. As with other instances of breaks in mentalizing, such incidents require that the process is “rewound and the incident explored.” Hence, in this collaborative patient–therapist relationship, the two partners involved have a joint responsibility to understand enactments.

So far, mentalization as a therapeutic approach has been evaluated in randomized controlled trials for BPD (Bateman & Fonagy, 1999, 2001, 2003, 2008) and a cluster randomized trial of a school-wide antiviolence intervention based on mentalization-based therapy principles (Fonagy et al., 2009). Ongoing applications include a treatment protocol for family therapy, a specific modification of the treatment for adolescents who self-harm and individuals with restricting eating disorders, a complex outreach intervention for adolescents who misuse substances or have other severe conduct problems, and a parenting intervention for enhancing mother–infant relationships in high-risk mothers.

Conclusion

Although developmentally it is highly likely that different pathways to BPD exist, we suggest that they all result in a low threshold for activation of the attachment system under stress. In combination with low thresholds for deactivation of the capacity for controlled mentalization, particularly the ability to differentiate mental states of self from those of others, this renders both the interpersonal and the internal worlds of individuals with BPD incomprehensible to them and leads to a cascade of impairments in other aspects of mentalization. This explains their propensity to become involved in vicious interpersonal cycles, characterized by marked affective dysregulation. Hence, disruptions of the attachment system, and identity diffusion closely linked to such disruptions, are seen as the core features of BPD. To deal with the consequent interpersonal dysfunction and distress, impulsivity, affect dysregulation, and feelings of inner pain, BPD patients rely on a number of maladaptive affect regulation strategies such as self-harm, substance abuse, or hypersexuality. Treatment should aim to simultaneously stimulate patients’ attachment and involvement with treatment while also helping them recover and maintain their ability to mentalize. Mentalization-based treatment for BPD focuses on specific instances of mentalization failure and uses a limited set of techniques to help patients recover mentalization when reconsidering experiences where they lost the capacity to think about the feelings and thoughts of others.

REFERENCES


Infancy and Borderline Personality Disorder


Infancy and Borderline Personality Disorder


