Chapter 7

Attachment, Mentalization, and the Self

Peter Fonagy and Patrick Luyten

Introduction

Mentalizing is often simplistically understood as synonymous with the capacity of empathy toward other people. In fact, mentalizing comprises a spectrum of capacities that critically involve the ability to see one's own behavior as coherently organized by mental states, and to differentiate oneself psychologically from others. It is these capacities that tend to be noticeably absent in individuals with a personality disorder (PD), particularly at moments of interpersonal stress. In this chapter, we will attempt to demonstrate that such impairments in mentalizing are at the heart of our explanatory framework for conceptualizing PDs. The foundations of our thinking lie in attachment theory, but, according to our most recent formulation, the heart of the relationship between mentalizing and personality pathology lies in the capacity of engaging productively in communication, and more specifically, in the quality of *epistemic trust* the individual possesses in relationships and, formatively, in the relationship between the child and his/her primary caregivers. Epistemic trust is defined in terms of an individual's experience of communication from others, specifically, the ability to receive and treat new knowledge from others as personally relevant and therefore capable of modifying durable representational structures pertaining to self, others, and interpersonal relationships. Underpinning this capability is the consideration of the informant as a "trustworthy" source likely to communicate information that is generalizable and relevant to the self (Fonagy & Luyten, in press).

PDs are often defined in terms of enduring interpersonal difficulties (Higgitt & Fonagy, 1992). We argue that these difficulties are generated by the rigidity of the patient's thinking about their subjective experience and an inability to engage in authentic communication about the causes of his/her own or others' actions, and therefore to modify his/her behavior in response to such communications, in particular. It is this impervious and inflexible quality that has, in the past, made patients with PD so conspicuously "hard to treat". We thus argue that the psychological mechanism behind the rigidity associated with PD is driven by the deep epistemic mistrust that has been generated in the patient—whether through environmental adversity,

genetic vulnerability, or, most likely, a complex combination of these factors—which interferes with social learning and creates an overreliance on unhelpful forms of mentalizing, or outright failures in mentalizing at moments of attachment arousal.

The mentalizing approach was originally formulated in the context of treating patients with borderline personality disorder (BPD) (Bateman & Fonagy, 2004), and has developed into a more comprehensive approach to the understanding and treatment of (severe) PD (Bateman & Fonagy, 2012). More recently, we have moved this thinking a little further, to consider the role of epistemic attitudes and epistemic trust in relation to social learning. These recent formulations have taken us into an even broader integrative theory about normal and disrupted personality development rooted in evolutionary considerations concerning the role of social cognition and the intergenerational transmission of knowledge about the (social) world. Three related features are considered to be central to PD and the typical distortions in self-experience and relationships in patients with PD: (a) disruptions in attachment experiences, (b) associated impairments in mentalizing, and (c) epistemic mistrust and hypervigilance. In this chapter, we direct our argument in relation to attachment, mentalizing, and epistemic trust to focus on the concept of self—a concept that has, historically, been central to thinking about personality pathology, and is notably central to the thinking on PD adopted in proposals to reconceptualize the concept of PD in DSM (Skodol, 2012).

The Mentalizing Approach to Personality Disorder

Attachment Relationships and the Development of Mentalizing

Mentalizing, like language, is a constitutional propensity, largely a developmental achievement made over the early years of life. The optimal development of mentalizing depends on the quality of attachment relationships, and early attachments in particular. The latter reflect the extent to which subjective experiences were adequately mirrored by a trusted other, that is, the extent to which attachment figures were able to respond with contingent and marked affective displays of their own experience in response to the infant's subjective experience, thus enabling the child to develop second-order representations of his/her own subjective experiences. The quality of affect mirroring by attachment figures plays a major role in the development of affect regulative processes and self-control (including attention mechanisms and effortful control), laying the groundwork for mentalizing capacity in the individual. Mentalizing is fundamentally interactive, in that it develops in the context of interactions with others and is

continually influenced by others' levels of mentalizing (i.e., good mentalizing will promote good mentalizing, while poor mentalizing pulls for poor mentalizing).

The Multidimensional Nature of Mentalizing

Mentalizing is not unidimensional. It can be organized around four polarities, each with its own relatively distinct underlying neural circuits (Fonagy & Luyten, 2009): (a) *automatic* versus *controlled* mentalizing, (b) mentalizing with regard to *self* or *others*, (c) mentalizing based on *external* or *internal* features of self and others, and (d) *cognitive* versus *affective* mentalizing. Automatic mentalizing refers to fast, parallel reflective processes that require little consciousness or effort, whereas controlled mentalizing involves more conscious, deliberate and serial reflective processes. The focus in mentalizing can be the self (including one's embodied experiences) or others, and can involve inferences based on external features of others (e.g., facial expressions) or direct assumptions about one's own mind or the mind of others (externally–internally based mentalizing). Finally, full mentalizing involves the integration of both cognitive knowledge (i.e., agent-attitude-propositions) and affective input (i.e., affective-state propositions).

Normally, these polarities are in balance, and can be used flexibly depending on the processing demands of a particular social context or setting. The persistent dominance of one end of a polarity over another signals a potential failure of accurate mental-state understanding. Specific forms of psychopathology can be understood on the basis of different combinations of impairments along the mentalizing dimensions (Bateman, Bolton, & Fonagy, 2013; Fonagy, Bateman, & Bateman, 2011; Fonagy et al., 2010). Different types of psychopathology are characterized by different patterns of inappropriate domination of a polarity, the imbalances generating apparent failures of mentalizing in the individual. Patients with BPD, for example, may typically give a misleading appearance of sophisticated insight or remarkable intuitive empathy based on the dominance of external, automatic, and affective mentalizing over internal, reflective, and cognitive processing of mental states. This imbalanced structure, breaks down at moments of interpersonal stress or attachment arousal (Figure 1). when the capacity for reflective mentalizing and cognitive flexibility about the possible motivation or intentions of other people (or indeed the self) is called for (Figure 2). Understanding mentalizing as being constituted of several sub-processes organized along dimensions rather than a single capacity is essential in preventing the iatrogenic effects of therapy for BPD patients that can arise if the therapist overestimates the patient's overall mentalizing capacity based on having identified conspicuous, isolated strengths.

[insert Figures 1 and 2 about here]

The Context/Relationship-specific Nature of Mentalizing

Furthermore, the capacity to mentalize has both "trait" and "state" aspects that may vary in quality in relation to emotional arousal and interpersonal context (e.g., an individual's mentalizing levels may considerably differ when reflecting on his/her relationship with his/her mother versus father, or when reflecting "off-line" on these relationships versus "on-line" in the course of a real-life interaction). Typically, with increasing arousal, the capacity for controlled mentalizing is likely to become (for the context) inappropriately dominated by automatic and often unreflective mentalizing as the balance between the two polarities is lost.

Attachment Strategies, Mentalizing, and Personality Disorder

Attachment hyperactivating and deactivating strategies play a key role in explaining the relationships among stress/arousal and mentalizing in different interpersonal/arousal contexts. This explains in part some of the heterogeneity observed within BPD patients and among patients with PDs more generally (Fonagy & Luyten, 2009). The relationship between attachment and imbalances in mentalizing in the context of BPD rests upon the way in which attachment hyperactivating or deactivating strategies trigger over-reliance on particular forms of mentalizing, obstructing the ability to call upon a wider and more balanced range of mentalizing skills.

Most patients with BPD, for example, are typically characterized by an excessive use of attachment hyperactivating strategies, often in the context of disorganized attachment. At least two strands of research support the link between hyperactivating strategies and disorganized attachment and mentalizing impairments in BPD. The first strand has provided direct evidence for such a link by showing that BPD is associated with increased levels of insecure attachment styles, by using both interview-based assessment of attachment such as the Adult Attachment Interview (AAI), and self-report measures (Steele & Siever, 2010). The second strand relates to attachment trauma (Allen, 2013), which we will discuss later in this chapter. These views are congruent with general biopsychosocial models of BPD (Oldham, 2009), which assume that adverse childhood experiences and genetic factors interact to create a particular combination of biological factors (neurobiological structures and dysfunctions) and psychosocial factors (personality traits and personality functioning) that underpins BPD pathology (affective and behavioral dysregulation and disturbed relatedness).

Some time ago, Fonagy and colleagues (Fonagy & Higgitt, 1990; Fonagy, Target, & Gergely, 2000) suggested that hyperactivation of the attachment system may be a core aspect of BPD. When exposed to threat and experiencing fear and distress that activates the attachment

system, young children are biologically predisposed to seek proximity to their caregiver. If the caregiver is optimally sensitive and responsive to the distressed child, a down-regulation of the child's emotion and deactivation of his/her attachment system occurs, and a lasting bond is formed to the caregiver who was attentive to the child's need (Bowlby, 1969, 1973). However, in circumstances where early relationships are disrupted, whether for reasons of biology, circumstance, or a combination of the two, the proximity seeking that is triggered by activation of the child's attachment system is anticipated to lead to further adverse emotional experience. This negative experience generates the same emotional response of fear and distress that triggers the attachment system, which then inevitably stimulates further proximity seeking in the increasingly forlorn hope of achieving down-regulation (Main, 2000). Within this model there is considerable room for individual differences to emerge; these differences are affected by genetic predisposition (Fearon, Shmueli-Goetz, Viding, Fonagy, & Plomin, 2013) and formative psychosocial experience (e.g., Fraley, Roisman, Booth-LaForce, Owen, & Holland, 2013; Repetti, Taylor, & Seeman, 2002).

There is now a rich body of empirical data to support these speculations about the relationship between attachment and BPD. For example, a large investigation with over 1,400 participants compared competing multivariate models of adult attachment patterns, impulsivity, and trait negative affect as these related to borderline features (Scott, Levy, & Pincus, 2009). The relationship between attachment anxiety and features of BPD was most effectively modeled when considered to generate negative affectivity and impulsivity. This provided a better fit than the alternative model of impulsivity or negative affectivity generating attachment anxiety. This study suggested that impulsivity and negative affect can lead to BPD when they occur in the context of high levels of attachment anxiety. A review by Agrawal et al. (2004) considered 13 empirical studies of the adult attachment styles of individuals with BPD; all the studies showed an association between BPD and insecure attachment. These findings have been subsequently confirmed in comparisons with groups of people with other PDs (Aaronson, Bender, Skodol, & Gunderson, 2006; Barone, Fossati, & Guiducci, 2011; Fossati et al., 2005; Meyer, Ajchenbrenner, & Bowles, 2005; Scott et al., 2013).

A further cross-sectional study reported by Choi-Kain, Fitzmaurice, Zanarini, Laverdiere, and Gunderson (2009) differentiated patients with mood disorder from patients with BPD on the basis of attachment style. While both depressed patients and patients with BPD showed greater preoccupation and fearfulness than did control individuals (although BPD patients were more severely affected for both traits), only patients with BPD simultaneously showed preoccupation *and* fearfulness. This study confirmed that, as a number of attachment theorists

have suggested, the key marker of BPD is a lack of a functional regulation strategy to reduce attachment distress (Fonagy & Bateman, 2008; Main, 2000).

Childhood Adversity, Mentalizing, and Personality Disorder

In drawing up a mentalizing, developmental model of PD in relation to childhood adversity, we suggest that two processes unfold, which have a cumulative effect.
(1) We assume that the development of mentalizing depends on the social co-construction of internal states between the child and his/her parents. Following from this, we hypothesize that early neglect and an early emotional environment that is incompatible with the normal acquisition of understanding self and others creates vulnerability to PD.
(2) Furthermore, we hypothesize that subsequent neglect and abuse in an attachment context can disrupt mentalization as part of an adaptive adjustment to adversity when a child whose early experiences of neglect have left him/her less resilient to deal with trauma is in a state of helplessness in relation to the responsible individuals (Fonagy, Steele, Steele, Higgitt, & Target, 1994; Stein, Fonagy, Ferguson, & Wisman, 2000).

In summary, we (Fonagy & Luyten, 2009) propose that early emotional neglect, more so than physical or sexual abuse, predisposes individuals to developing PD, and specifically BPD, by limiting their opportunity to acquire the full spectrum of mentalizing skills, leaving their capacity to mentalize vulnerable to being disrupted under the influence of later stress.

We further contend that childhood environments that are emotionally abusive, or in which the discussion and validation of mental states are absent, lead to the generation of coping strategies that favor rapid, intuitive (as opposed to more deliberate and reflective), emotion-focused (as opposed to cognition-focused) mentalizing—particularly in individuals who have been primed by experiences of early neglect to identify risk rapidly. This intuitive state is dominated by mental-state attributions based on external observable cues (as opposed to the inference of internal states, which may be unreliable in a maltreating environment). In addition, such individuals may need to develop a strategy of being highly sensitive to and readily influenced by (mirroring) mental states in others (at the cost of achieving a clear differentiation of "self" and "other" mental states) (Fonagy, 2000; Fonagy & Luyten, 2009; Luyten, Fonagy, Lowyck, & Vermote, 2012).

We wish to emphasize that while this approach may appear to suggest a deficit theory, our emphasis is upon adaptation. For instance, the specific configuration of mentalizing capacities that characterizes individuals with BPD may be conceived of as that most likely to favor survival under conditions of significant adversity (Frankenhuis, Panchanathan, & Clark

Barrett, 2013). Our model is fully compatible with Linehan's (1993) original biosocial theory of the development of BPD and its later elaboration (Crowell, Beauchaine, & Linehan, 2009), which emphasizes that an invalidating childhood environment can serve to undermine the understanding, regulation, and toleration of affect states, leading the child to adapt by displaying more extreme emotions to achieve a contingent response from carers, that is, the intermittent reinforcement of extreme emotional outbursts.

In line with these assumptions, many studies support the suggestion that secure children perform better than insecure children at mentalization tasks (see, e.g., de Rosnay & Harris, 2002). The first of these, (Fonagy, Steele, Steele, & Holder, 1997), from the London Parent–Child Project, reported that 82% of children who were securely attached to their mother in the Strange Situation passed Harris's Belief-Desire Reasoning Task at age 5.5 years, compared with 50% of those who showed an avoidant attachment style and 33% of the few children with preoccupied attachment. As we will describe in more detail later, findings along these lines are not always consistent (see, e.g., Meins et al., 2002), but in general it appears that secure attachment and mentalizing are affected by similar social influences.

It is necessary to identify what it is about emotional abuse and neglect that can render the child increasingly vulnerable to disrupted mentalizing. We (Fonagy & Luyten, 2009) have stressed the importance of secure attachment in providing a context within which the child can develop the ability to mentalize and regulate his/her own emotions. Clearly, in an environment that is invalidating and emotionally abusive, an insecure and disorganized attachment pattern is likely to develop (Fonagy, 2000; Fonagy & Luyten, 2009). We contend that it is the absence of a "safe base", from which the child can learn the capacity to mentalize and self-regulate, that can create a greater vulnerability for developing PD (and BPD in particular) in later life.

For instance, emotional dysregulation (Gratz, Tull, Baruch, Bornovalova, & Lejuez, 2008), schema modes (particularly disconnection/rejection and impaired limits; Specht, Chapman, & Cellucci, 2009), and distorted self-representation in middle childhood (Carlson, Egeland, & Sroufe, 2009) have all been shown to mediate the relationship between childhood abuse and BPD. However, these studies used a composite score of maltreatment rather than focusing on particular aspects of emotional abuse and/or neglect. Rogosch and Cicchetti (2005) found that attentional networks and processes did not mediate the precursors to BPD in children who had been abused. The formation of schemas and the representation of the self, which are often conspicuously impoverished in individuals with BPD, are themselves mediated by the selective disruption of social cognition (Carlson et al., 2009). These studies point to the importance of these key psychological processes in mediating the impact of childhood

experience of trauma on the development of BPD symptoms in later life. More work is needed in this area, but the research to date suggests that the psychological processes that make the impact of trauma on personality development so harmful and enduring relate to the emergence of the concept of self and identity.

BPD patients have also been found to exhibit high levels of alexithymia (i.e., impairments in mentalizing with regard to the self) and have difficulty in describing their emotions in social situations (New et al., 2012)—features that have both been related to trauma. In samples containing a majority of patients reporting early trauma, patients showed amygdala hyperreactivity in response to viewing emotional or neutral images, combined with blunted subjective appraisals of these same stimuli (Hazlett et al., 2012; Minzenberg, Fan, New, Tang, & Siever, 2007). Considered in relation to attachment, the mentalizing deficits associated with childhood maltreatment may be a form of decoupling, inhibition, or even a phobic reaction to mentalizing: (a) the experience of adversity may undermine cognitive development in general (Cicchetti, Rogosch, & Toth, 2000; Crandell & Hobson, 1999; Stacks, Beeghly, Partridge, & Dexter, 2011); (b) the mentalizing problems may reflect arousal problems associated with exposure to chronic stress (see Cicchetti & Walker, 2001); and (c) the child may avoid mentalizing to avoid perceiving the abuser's hostile and malevolent thoughts and feelings about him/her (e.g., Fonagy, 1991; Goodman, Quas, & Ogle, 2010). These potential problems will all have a significant negative impact on a child's psychological, emotional, social, and educational development.

Maltreatment can contribute to an acquired partial "mind-blindness" by compromising open, reflective communication between parent and child. It may undermine the benefit derived from learning about the links between internal states and actions in attachment relationships (e.g., the child may be told that he/she "deserves", "wants", or even "enjoys" the abuse to which he/she is being subjected). This is more likely to be harmful if the maltreatment is being perpetrated by a family member. In cases when the maltreatment is taking place outside the home, the parents' lack of awareness of it may serve to invalidate the child's communications with the parents about his/her feelings. In such a situation the child finds that reflective discourse does not correspond to his/her feelings—a consistent misunderstanding that could reduce the child's ability to understand and mentalize verbal explanations of other people's actions. In such circumstances, the child is likely to struggle to detect the mental states behind people's actions, and will tend to see these actions as inevitable rather than intended.

Taken together, these views clearly imply that the foundations of subjective selfhood will be less robustly established in individuals who have experienced early neglect. These individuals

will find it more difficult to learn about how subjective experiences inevitably vary between people and the need for flexibility in relation to the relative merits of one's own and alternative perspectives. As we have described, in some longitudinal investigations, low parental affection or nurturing in early childhood appears to be more strongly associated with elevated risk for borderline, antisocial, paranoid, and schizotypal PDs than do physical or sexual abuse in adolescence (Afifi et al., 2011; Gao, Raine, Chan, Venables, & Mednick, 2010; Hengartner, Ajdacic-Gross, Rodgers, Muller, & Rossler, 2013; Johnson, Cohen, Kasen, Ehrensaft, & Crawford, 2006; Powers, Thomas, Ressler, & Bradley, 2011; Tyrka, Wyche, Kelly, Price, & Carpenter, 2009; Widom, Czaja, & Paris, 2009), again pointing to the importance of neglect, low parental involvement, and emotional maltreatment in undermining the normally biologically predetermined learning and communication processes that are involved in the healthy development of perspective taking.

The Re-emergence of Nonmentalizing Modes

Modes of experiencing the self and others that are characteristic of the prementalizing child tend to re-emerge whenever we lose the ability to mentalize, as typically happens in individuals with PD, particularly in high arousal contexts (Fonagy & Target, 1997). According to our conceptualization of mentalizing as consisting of four polarities, we understand the emergence of these nonmentalizing modes as indications of imbalances in mentalizing, in which one extreme form of mentalizing has come to dominate. We have described these nonmentalizing modes as falling into three distinct categories (Fonagy, Gergely, Jurist, & Target, 2002), as a way of understanding the subjective mentalizing experience of the individual and providing a formulation that is useful for clinicians, particularly when they are responding to highly affect driven, nonmentalizing behaviour on the part of the patient.

(a) In the *psychic equivalence mode*, thoughts and feelings become overwhelmingly real, allowing for no alternative perspectives or doubt. This mode reflects the domination of self:affect state thinking with limited internal focus.

(b) In the *teleological mode* there is only a recognition of real, observable goal-directed behavior and objectively discernible events that may potentially constrain these goals. Hence, there is a recognition of the existence and potential role of mental states, but limited to very concrete and observable goals. This mode reflects an extreme exterior focus and the momentary loss of controlled mentalizing.

(c) In the *pretend mode*, thoughts and feelings become severed from reality ("hypermentalizing" or "pseudomentalizing"), which, in the extreme, may lead to feelings of de-realization and dissociation. This reflects controlled mentalizing being dominated by an implicit but inadequately

internal focus, combined with poor belief-desire reasoning and a consequent sense of vulnerability to fusion with others.

Understanding and recognizing the prementalizing modes is important because they often appear alongside a pressure to externalize unmentalized and self-hating aspects of the self (so-called "alien-self" parts). Torturing feelings of badness, possibly linked to experiences of abuse that are felt to be part of the self but are not integrated with it (the "alien-self" parts), can come to dominate self-experience. We assume that these discontinuities in internal experience (when the person feels aspects of their self-experience to be of themselves or their own, and yet also to be alien) generate a sense of incongruence, which is dealt with through externalizing. For example, the individual may attempt to dominate the mind-states of others, through behaving toward others as though the others own the unmentalized self-experiences and on occasions even being successful in generating these experiences in them (Fonagy & Target, 2000). This brings about relief, even if the immediate impact of externalizing a torturing part of the self in this way is to manipulate the other person into punitive persecutory behavior toward the self. Attempts to adapt to these torturing internal experiences can also manifest as attacks upon the self (e.g., self-harming behaviors) or other types of behavior that in the teleological mode are expected to relieve tension and arousal (Fonagy & Target, 2000).

Mentalizing and the Self

How do these considerations help us to understand the subjective experience of patients with PD, and in particular the feelings of identity diffusion that are typical of many of these patients? We approach this question from a developmental psychopathology perspective in line with the more general mentalizing approach. Most modern psychological approaches assume that self-coherence (the sense that one has continuity and consistency in thought and behavior) is something of an illusion (Bargh, 2011, 2014). Thought of in this way, there can be no such thing as self-coherence, identity, or self; these are constructs referring to a self-generated feeling or a subjective state of coherence. This is not to deny the importance of self-coherence. On the contrary, we submit that a feeling or state of self-coherence is at the heart of mental health and is associated with feelings of agency and autonomy.

How then can we understand individuals with PD from this perspective? Identity diffusion, that is, a lack of self-coherence, has notably been central to many theoretical formulations of serious PD. A *disturbed sense of identity*, or "failure to establish stable and integrated representations of self and others" (Livesley, 2003, p. 19), is frequently described as characteristic of BPD (e.g. Blatt & Auerbach, 1988; Bradley & Westen, 2005; Jorgensen, 2010;

Kernberg, 1975, 1984) and was a core part of the reformulation of PD proposed for DSM-5, criticized by many (Shedler et al., 2010; Shedler et al., 2011), which found its way into Section III of DSM-5 (American Psychiatric Association, 2013). Rooted in a dysfunction or deficit of a sense of agency or self-directedness, this characteristic has also been identified in empirical clinical studies (e.g., Adler, Chin, Kolisetty, & Oltmanns, 2012; Barnow, Ruge, Spitzer, & Freyberger, 2005; Bender & Skodol, 2007; Jorgensen et al., 2012).

We argue that it is the capacity for mentalizing that enables us to *create* and *maintain* this consistent sense of the self across different contexts: mental-state language and the experience of being "mind-minded" (Meins et al., 2002) constructs the narrative around one's thoughts and feelings, which in turn helps us continuously to construct a benign and coherent self-structure. When mentalizing impairments weaken this integrative process, an experience of incoherence in self-representation is likely to emerge. Rather than seeing personality impairments as classically defined—as impairments in the self or rigidity in representations—we thus postulate that PD in fact involves failures of this building process and generates a sense of epistemic rigidity, which inhibits the dynamic and ongoing construction of the self. Rigidity is often invoked in discussions of the phenomenology of PD (Beck, Freeman, & Davis, 2004; Caligor, Kernberg, & Clarkin, 2007; McWilliams, 1994; Mikulincer & Shaver, 2007; Rogers & Dymond, 1954). Rigidity, as we use the term here, describes a state of being closed to social learning, which requires being comfortable in taking knowledge from others as relevant to oneself—what we might refer to as a sense of *epistemic trust*. As we will explain in more detail in the next section, we see rigidity as indicating the absence of a willingness to fully engage in social learning, and PD as reflecting impairments in mentalizing that are associated with an exaggeration of *epistemic vigilance* (being on the lookout for the possibility of being misled): it signifies epistemic mistrust of the interpersonal world. If we conceptualize our sense of self as a consequence of a process of mentalizing rather than a fixed representation, then exaggerated vigilance or *epistemic hypervigilance* will inhibit the internalization of social knowledge and the reflection on this knowledge that is necessary for the healthy maintenance of the social learning process.

This more dynamic approach to the ongoing relationship between mentalizing, the social environment, and the construction of the self is at odds with the traditional tendency to reify the notion of representation or internal working models of self and other. This inclination is best exemplified by theoretical writings and empirical studies on the (hierarchical) organization of object or attachment representations, descriptions of splitting of object representations, and even studies on the activation of object representations commonly present in traditional psychodynamic formulations (Fonagy & Target, 2003).

These can be clinically useful descriptions, but they are no longer in line with what we now know about representational processes: representations do not exist; they thus cannot be split or integrated, and they cannot be hierarchically or even nonhierarchically organized (see Fonagy, 1982 for an early exposition of this point of view). A more parsimonious way of describing these processes, which would also be in line with our growing knowledge of representational processes in the brain (Fonagy & Luyten, 2009; Luyten & Blatt, 2013), is to conceptualize them in terms of *states of mind* that, for instance, are created by the capacity for mentalizing, which is itself subserved by different neural circuits (Luyten et al., 2012). It would therefore be more accurate to say that we generate a representation of ourselves and others, rather than that we "have" representations that are stored somewhere, and which are hierarchically organized and can be activated, changed, or structurally modified. What is assumed to be a trait-like quality of representations (e.g., self-representation) is better described as stability in the capacity for generating such representations, or even better as states of mind that bring some stability in our experience of self.

Further, in the context of this notional representation that is stable, differentiated, and integrated, there is little space for the reality that we all sometimes feel inferior, fragmented, or isolated. Traditionally, such states are understood as reflecting the activation of split-off or repressed representations, perhaps because of defensive processes, or as regression. But again, do we really believe that we have stored somewhere various representations of ourselves and others that can be (re-)activated? We can explain these same phenomena much more parsimoniously by arguing that in particular circumstances (i.e., when arousal/conflict is high), we are unable to create a feeling of stability, agency, and positive self-regard; we do not need to assume more.

This has important clinical and research implications. It means that patients do not "change" their self (and object) representations as a result of treatment or experiences (and thus we should not try to assess these changes in their representations); rather, they have acquired the capacity to think and feel differently about themselves and others, and thus to think, reflect, and narrate differently about themselves and others. Hence, the primary focus in treatment (and in outcome studies) should be on these capacities and not on changes in object representations per se, which are a secondary consequence (the fruit on the tree, so to speak) of a growing capacity for social cognition.

Indeed, when one considers the human ability to reflect on self-experience in relation to others, it naturally follows that this capacity shows both features of stability and change over time, as it is likely to be influenced by context (and particularly attachment contexts) (Luyten et

al., 2012). Similarly, demanding that the self, identity, or personality should be highly stable across the life span makes little sense from this perspective, as people do not *have* a self, identity, or personality; rather, they have the ability, to a greater or lesser extent, to activate a more or less consistent, coherent, and differentiated *feeling* or *experience* of coherence and stability. It is this reflective function that *creates* a feeling of coherence in the moment, rather than it being the case that there *is* a feeling of coherence. This is also congruent with findings from neuroscience suggesting that a cortical midline system is responsible for generating the experience of self as distinct from others (Fonagy & Luyten, in press; Luyten & Blatt, 2013; Northoff, Qin, & Feinberg, 2011).

This idea of the self as a work in progress also allows us to take into account a way of accommodating the role of the social environment in the sense of self. A long-standing critique of psychoanalytic thinking and psychology more broadly is that they both fail to take into account the effects of how the socioeconomic environment might buffet the individual psyche (Fonagy, Target, & Gergely, 2006). Given the evidence accruing that increasing levels of social inequality are connected with an increased prevalence of BPD (Fonagy & Luyten, in press), an approach to PD that understands the mechanism between personality pathology and the social context of the self is even more pertinent (Grant et al., 2008; Wilkinson & Pickett, 2009). An explanation of psychopathology relating to the self that allows for the impact of the environment in an evolutionarily convincing manner is made possible by this dynamic interpretation of our self-representations. If we consider that the evolutionary drive behind mentalizing was to enable human survival in increasingly complex social situations involving matters of hierarchy, cooperation, exclusion, and inclusion, it makes eminent sense that representations of ourselves and those around us should calibrate the extent to which we may be experiencing social isolation, alienation, or inferiority.

Psychological resilience enables an individual to resist these pressures to some degree. By contrast, individuals with PD are often conspicuously reactive to such pressures, while to be wholly impervious to their effects suggests mentalizing impairments of a different nature altogether. Both extremes, however, derive from a lack of capacity to absorb information from the social environment in a way that is compatible with the construction of a normatively coherent sense of self. This ability, or inability, to take in social data, and to be able to reflect upon it with some coherence and pragmatism, is key to understanding PD, with its phenomenology of enduring interpersonal difficulties and troubled sense of self. It is the second-order developmental constructs of flexibility and rigidity that determine how individuals respond to the aspects of their environment or themselves that constitute what we might call

personality—whether we term these cognitive schemas, internal object relationships, interpersonal expectations, or intersubjective concerns. Rigidity points to an individual's inability to progress fluidly, flexibly, and adaptively across the phases of individual development; it is a meta-construct associated with personality functioning. According to this thinking, the concept of rigidity encapsulates an understanding of PD as the failure of appropriate responsiveness to information within a system at the interface of the person and his/her social environment.

The three prementalizing modes, discussed earlier, are significant here; they constitute forms of mentalizing that make the individual appear difficult or even impossible to reach, rendering them potentially impervious to meaningful social influence. Rigidity, or the inability to adjust to other, more reflective forms of mentalizing, breaks down the individual's ability to form a stable self-structure, creating the fragmented sense of self often experientially associated with BPD. Rigidity is perhaps greatest in individuals who show both the high level of avoidance and intense attachment anxiety that is so typical of patients with BPD (e.g., Choi-Kain et al., 2009): An irresolvable dilemma is created by an intense desire for reassurance from an attachment figure if the individual is also unable to fully accept this reassurance owing to a mistrust of the attachment figure's motives. Thus, while security is buttressed by flexibility, which derives from refusing to consider closeness and autonomy as antagonistic and irreconcilable goals, insecurity and partial rigidity arise when an individual is unable to relocate on the closeness–distance dimension without fearing either a permanent loss of autonomy or the loss of affection of his/her attachment figure. The key here is the invalidation of interpersonal information that arises from any encounter, regardless of the nature of the information. Even a positive response from the attachment figure will, in the context of epistemic mistrust, be discounted by assumptions about the person's motives. But dismissal or closing of the flow of information is also unsustainable because of the overriding need for reassurance. Let us now scrutinize further this concept of trust as it applies to socially transmitted information.

Mentalizing and Epistemic Trust

Studies of attachment have confirmed that secure attachment is driven by sensitive responsivity of the caregiver contingent upon the infant's reaction (Belsky & Fearon, 2008; Marvin & Britner, 2008). In addition, we have maintained that attachment experiences also provide the context in which mentalizing by the caregiver can influence the security of attachment and mentalizing in the child (Arnott & Meins, 2007; Fonagy, Steele, Steele, Moran, & Higgitt, 1991; Grienenberger, Kelly, & Slade, 2005; Meins, Fernyhough, Fradley, & Tuckey, 2001; Meins et al., 2002; Slade, Grienenberger, Bernbach, Levy, & Locker, 2005). Recently, we

have added a third consideration: secure attachment is created by a system that simultaneously generates a sense of epistemic trust.

Looked at from a distance, micro-analytic (e.g. Beebe et al., 2010) and more global (e.g., DeWolf & van IJzendoorn, 1997; Isabella, Belsky, & von Eye, 1989; Kiser, Bates, Maslin, & Bayles, 1986; Mills-Koonce et al., 2007) ratings of sensitive caregiving can be seen as in essence recognizing the child's agentive self. We believe that through the down-regulation of affect that results from successful proximity seeking in the distressed infant, attachment not only establishes a lasting bond, but also opens a channel for information that can be used for the transfer of knowledge between the generations. Given that the infant needs to overcome the self-preservative barrier created by natural epistemic vigilance (Sperber et al., 2010; Wilson & Sperber, 2012) and open his/her mind to acquiring the myriad pieces of culturally relevant information on which his/her survival will ultimately depend, it is fortunate that nature (evolution) has provided us with a mechanism of deferential knowledge transmission that can create an "epistemic superhighway" between learners and teachers, who normally share genetic material (Hamilton, 1964). It is this openness to information transfer that we believe offers the cognitive advantage to secure attachment that has been fairly consistently noted, although not to our knowledge systematically studied (e.g., Crandell & Hobson, 1999; Jacobsen & Hofmann, 1997; Moss, Rousseau, Parent, St.-Laurent, & Saintong, 1998). Although still somewhat speculative, evidence for these assumptions comes from two strands of research that are increasingly becoming intertwined: (a) developmental studies and (b) evolutionary approaches concerning the development of social cognition (for a detailed discussion, see Fonagy & Allison, 2014; Fonagy & Luyten, in press; Fonagy, Luyten, & Allison, 2014).

As we suggested elsewhere (Fonagy et al., 2014), building on pioneering work by Dan Sperber (Sperber et al., 2010; Wilson & Sperber, 2012) and accumulating developmental evidence (e.g., Corriveau et al., 2009), secure attachment experiences pave the way for the acquisition of mentalizing at the same time as fostering epistemic trust. In other words, natural selection may have hit upon attachment as a means to mediate the reliable transmission of "memes" from one generation to the next. Secure attachment helps to create a benign condition for the relaxation of epistemic vigilance, and sensitive and appropriate ostensive cueing is a key constituent element of sensitivity on the part of the primary caregiver. Attachment is a much older instinct, in evolutionary terms, than the imperative to generate epistemic trust for the safe transmission of "memes"; in that sense the two processes are distinct. In terms of the phenomenology of child development, however, they are closely interwoven, and it seems likely that in human evolution epistemic trust piggy-backed on to pre-existing attachment processes.

However, we suggest that while attachment may be a key mechanism for mediating epistemic trust, it is secondary to an underlying biological process preserved by evolution. In other words, secure attachment is unlikely to be *necessary* for generating epistemic trust but it may be *sufficient* to do so; furthermore, it is the most *pervasive* mechanism in early childhood because it is a highly evolutionarily effective indicator of trustworthiness.

Implications for the treatment of PD

The Mentalization-Based Treatment Approach

Mentalization-Based Treatment (MBT) is firmly grounded in the theoretical model we have outlined in the early sections of this chapter. The approach regards imbalances in mentalizing as the core of the enduring difficulties of BPD and other types of personality pathology, and the aim of treatment is thus the restoration of more balanced mentalizing (Bateman & Fonagy, 2010). MBT was initially devised for the treatment of patients with BPD in a partial hospital setting. More recently it has developed into a more comprehensive approach to the understanding and treatment of PDs in a range of clinical contexts, including patients with antisocial PD (Bateman & Fonagy, 2008), patients with BPD and marked comorbid eating disorders (Robinson et al., 2014), and patients with less marked personality pathology (Allen, Fonagy, & Bateman, 2008).

The theoretical model implies that in order to maximize the effect on the patient's ability to think about thoughts and feelings in the context of relationships, especially in the early phases of treatment, the therapist is probably most helpful when his/her 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/her behavior, (e) relate to a current event or activity—that is, whatever is the patient's currently felt mental reality (in working memory), (f) make use of the therapist's mind as a model (by the therapist disclosing his/her expected reaction in response to the event being discussed, by explaining to the patient how he/she anticipates that he or she might react in that situation), and (g) are flexibly adjusted in terms of their complexity and emotional intensity in response to the level of the patient's emotional arousal (i.e., 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 the self and others. MBT therapists achieves this through the judicious use of an *inquisitive stance*, highlighting their own interest in the mental states underpinning behavior, qualifying their own understanding and inferences (in that mental states are opaque),

and showing how such information can help the patient to make sense of his/her own experiences. Pseudomentalization and other fillers to replace genuine mentalization (as described earlier in this chapter) should be explicitly identified by the therapist and the lack of practical success associated with them clearly explained. In this way, therapists can help their patients to learn about how they think and feel about themselves and others, how those thoughts and feelings shape their responses to others, and how "errors" in understanding self and others may lead them to inappropriate actions. Conversely, it is not for the therapist to tell patients about how they feel, what they think, how they should behave, or what the underlying (conscious or unconscious) reasons for their difficulties may be. In fact, any approach which tends toward claiming to "know" how or why patients "are", or to dictate how they should behave and think, is likely to be iatrogenic in individuals whose capacity to mentalize is vulnerable.

While the MBT model has a reasonable evidence base (Bateman & Fonagy, 2009, 2013), it makes the strong and so far almost unwarranted assumption that the increasing capacity to mentalize drives an improvement in BPD symptoms such as self-harming behavior or suicidality. Focusing on the concept of epistemic trust enables us furthermore to reconceptualize the importance of mentalizing as a key part of therapeutic effectiveness.

Reconceptualization of treatment: Three systems

Our thinking in relation to the role of rigidity and epistemic mistrust in PD has led us to reconceptualize treatment and the purported mechanisms of change in the treatment of these patients. We believe there is a case to be made for understanding the underlying processes at work for all therapeutic interventions that have been found to be effective in PDs. In the case of BPD, for instance, a considerable number of different therapies have now been found to be effective (for a review, see Leichsenring, Leibing, Kruse, New, & Leweke, 2011) Such forms of treatment all benefit from a well-articulated theoretical framework and a reliable model for delivery of treatment. However, other than this, it is not possible to isolate a factor common to all these therapies that can explain their effectiveness, and which can be pinpointed as missing from interventions that are less effective. A single model that accounts for how the effective therapies work while accommodating their theoretical specificities is one that accounts for the process that underpins them. In light of our argument, outlined earlier, about epistemic trust, we posit that there are in fact three different processes that are necessarily undergone across successful treatments.

Communication system 1: The teaching and learning of content and the increase of epistemic openness

All evidence-based psychotherapies provide a coherent, consistent, and continuous framework that enables the patient to examine the issues that are deemed to be central to him/her according to a particular theoretical approach (e.g., early schemas, invalidating experiences, object relations, current attachment experiences) in a safe and low-arousal context. These psychotherapies thus provide the patient with helpful skills or knowledge, such as strategies to handle emotional dysregulation or restructured interpersonal relationship schemata. Perhaps more importantly however, all *evidence-based* psychotherapies implicitly provide for the patient a model of mind and an understanding of his/her disorder, as well as a hypothetical appreciation of the process of change, *that are accurate enough to enable the patient to feel recognized as agentive*, empowered to make decisions and alter the course of his/her path through life. The conceptual model of each treatment contains considerable personally relevant information so the patient experiences feeling markedly mirrored or "understood". Helpful, directive approaches may be more likely than a generic exploratory style to communicate a clear recognition of the patient's position (McAleavey & Castonguay, 2013). The idea that psychotherapies have in common the creation of a sense of being understood while differing in the understandings they provide has been part of integrative approaches to psychotherapy since common factor approaches were first proposed (e.g., Frank & Frank, 1991; Prochaska & Norcross, 2013; Rogers, 1951). We know that without a coherent body of knowledge based on a systematically established set of principles, psychological therapy is of little value (Benish, Imel, & Wampold, 2008). Even in large cohort study meta-analyses, therapies without a credible and tight intellectual frame are observed to fail (Abbass, Rabung, Leichsenring, Refseth, & Midgley, 2013).

The fact that so many different therapies, using so many different theoretical models, have been found to be of some benefit indicates that the significance of communication system 1 lies perhaps not only in the essential truth of the "wisdom" of the specific approach, but also in that it causes the patient to give weight to communication from the social world (Ahn & Wampold, 2001; Paris, 2013). This brings us to the second communication system at work in psychotherapy.

Communication system 2: The re-emergence of robust mentalizing

As noted earlier, through passing on knowledge and skills that feel appropriate and helpful to the patient, the therapist implicitly recognizes the patient's agency. The therapist's presentation of information that is relevant to the patient serves as a form of ostensive cue that

conveys the impression that the therapist seeks to understand the patient's perspective; this in turn enables the patient to listen and to hear. In effect, the therapist is modeling how he/she engages in mentalizing in relation to the patient. It is important that in this process both patient and therapist come to see each other more clearly as intentional agents. It is not sufficient for the therapist to present his/her "mentalizing wisdom" to the patient if the therapist is not him/herself clearly seen as an agentive actor whose actions are predictable given the principles of theoretical rationality (Kiraly, Csibra, & Gergely, 2013). The context of an open and trustworthy social situation helps to achieve a better understanding of the beliefs, wishes, and desires underpinning the actions of others and of the self. This in turn allows for a more trusting relationship in the consulting room. Ideally, the patient's feeling of having been sensitively responded to by the therapist opens a second virtuous cycle in interpersonal communication *in which the patient's own capacity to mentalize is regenerated*.

However, the mentalizing of patients—that is, acting in accordance with the patient's perspective—may be a common factor across psychotherapies not because patients need to learn about the contents of their minds or those of others, but because mentalizing may be a generic way of increasing epistemic trust and therefore achieving change in mental *function*.

We would like to underline a point that may seem initially puzzling given our own declared commitment to mentalization-based psychotherapy: *mentalizing in itself is only an intermediate step and is not the ultimate therapeutic objective*. True and lasting improvement, we believe ,rests on a third communication system: learning from experience beyond therapy.

Communication system 3: The re-emergence of social learning with improved mentalizing and epistemic trust

We hypothesize that feeling understood, just as in normal psychological development, opens a key biological route to information transmission and the possibility of taking in knowledge that is felt to be personally relevant and generalizable; this is what brings about change in previously rigidly held beliefs. In essence, the experience of feeling thought about enables us to learn new things about our social world.

We hypothesize that, as the patient's state of epistemic hypervigilance relaxes, he/she develops a greater capacity to trust and begins to discover new ways of learning about others. This facilitates an increase in the patient's willingness to modify his/her cognitive structures for interpreting the behavior of others. Positive social experiences that in the past were discounted as a result of the patient's epistemic hypervigilance now have the potential to have a positive impact. This is the third system of communication, which becomes available once the second system, which is specific to the therapeutic situation, has enhanced the patient's capacity to

mentalize. As the patient begins to experience social interactions in a more benign way and to view his/her social situations more accurately (e.g., not seeing an experience of temporary social disappointment as a complete rejection), he/she can update his/her knowledge of both the self and others.

It is the recovery of capacity for social information exchange that, we feel, may be at the heart of all effective psychotherapies. These therapies impart an ability to benefit from benign social intentions, and to update and build on knowledge about the self and others in social situations. The improved sense of epistemic trust derived from mentalizing enables learning from social experience; in this way the third virtuous cycle is maintained beyond therapy.

As therapists we often assume that the process in the consulting room is the primary driver of change, but experience shows us that change is also brought about by what happens beyond therapy, in the person's social environment. Empirical evidence from studies employing session-by-session monitoring of change suggests that the therapeutic alliance in one session foretells change in the next (Falkenstrom, Granstrom, & Holmqvist, 2013; Tasca & Lampard, 2012). This suggests that the change that occurs is a consequence of changed attitudes to learning, engendered by therapy, modifying behavior *between* sessions. The implication is that the extent to which a patient derives benefit from therapy also depends on what he/she encounters in his/her particular social world during and after treatment, that is, the changes in person–environment exchanges that result from the patient's increased openness to the evolutionary determined and rehabilitated capacity for social learning. We predict that psychotherapy for PD is therefore much more likely to be beneficial if the individual's social environment at the time of treatment is largely benign, or becomes more benign. Although we do not know of any systematic studies that have explored this moderator, clinical experience suggests that there is likely to be some validity to this assertion.

Summary and Conclusions

Patients with PDs are often notoriously difficult to treat. The often paradoxical combination of marked rigidity and the fluidity of their self-experiences and relationships may confuse clinicians, and give them the feeling that they have no grip on what is happening in these patients and in their therapeutic relationships. Patients with PD often have considerable difficulty in developing a working alliance because they distrust what the clinician is offering them; this may even lead to a simple refusal to be treated. Even when these patients accept treatment, they almost invariably face the therapist's suggestions and interpretations with

distrust; this situation frequently persists a very long time, and change typically happens only slowly as they become more trusting of their treating clinician.

This chapter reflects our evolving attempt to understand these phenomena from a mentalizing perspective. Our previous views primarily focused on disruptions in attachment relationships and associated mentalizing impairments in explaining the typical features of patients with (severe) PD. Specifically, we have consistently argued that early attachment disruptions, likely often in combination with biological vulnerability, give rise to often severe mentalizing impairments, which lead to serious discontinuity in the self and associated relational problems. This inconsistency in the self-experience, which we conceptualize as resulting from an incapacity to generate a more coherent sense of self because of impairments in mentalizing capacity, leads to a constant pressure to externalize "alien self" parts, which may be expressed in a tendency to dominate the mental states of others and/or (particularly in BPD, but also, for instance, in paranoid PD and antisocial PD) self-harm. More recently, we have drawn attention to a third, closely related factor, the *epistemic mistrust* or *epistemic hypervigilance* that results from attachment disruptions. This inhibits openness to social knowledge and the reflection on this knowledge that is necessary for the healthy maintenance of the evolutionary rooted social learning process that is typical of human beings.

These formulations are likely to have important implications for the treatment of patients with PDs and particularly in relation to the discussion on the role of so-called specific versus common factors in the treatment of these patients, and our continuous efforts to develop interventions for these "hard to reach" patients.

References

Aaronson, C. J., Bender, D. S., Skodol, A. E., & Gunderson, J. G. (2006). Comparison of attachment styles in borderline personality disorder and obsessive-compulsive personality disorder. *Psychiatric Quarterly, 77*, 69-80.

Abbass, A. A., Rabung, S., Leichsenring, F., Refseth, J. S., & Midgley, N. (2013). Psychodynamic psychotherapy for children and adolescents: A meta-analysis of short-term psychodynamic models. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*, 863-875.

Adler, J. M., Chin, E. D., Kolisetty, A. P., & Oltmanns, T. F. (2012). The distinguishing characteristics of narrative identity in adults with features of borderline personality disorder: An empirical investigation. *Journal of Personality Disorders, 26*, 498-512.

Afifi, T. O., Mather, A., Boman, J., Fleisher, W., Enns, M. W., Macmillan, H., & Sareen, J. (2011). Childhood adversity and personality disorders: Results from a nationally representative population-based study. *Journal of Psychiatric Research, 45*, 814-822.

Agrawal, H. R., Gunderson, J., Holmes, B. M., & Lyons-Ruth, K. (2004). Attachment studies with borderline patients: A review. *Harvard Review of Psychiatry, 12*, 94-104.

Ahn, H., & Wampold, B. E. (2001). Where oh where are the specific ingredients? A meta-analysis of component studies in counseling and psychotherapy. *Journal of Counseling Psychology, 48*, 251−257.

Allen, J. G. (2013). *Mentalizing in the development and treatment of attachment trauma*. London, UK: Karnac Books.

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

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association.

Arnott, B., & Meins, E. (2007). Links among antenatal attachment representations, postnatal mind-mindedness, and infant attachment security: A preliminary study of mothers and fathers. *Bulletin of the Menninger Clinic, 71*, 132-149.

Bargh, J. A. (2011). Unconscious thought theory and its discontents: A critique of the critiques. *Social Cognition, 29*, 629-647.

Bargh, J. A. (2014). Our unconscious mind. *Scientific American, 310*, 30-37.

Barnow, S., Ruge, J., Spitzer, C., & Freyberger, H. J. (2005). Temperament und Charakter bei Personen mit Borderline-Personlichkeitsstorung [Temperament and character in persons with borderline personality disorder]. *Nervenarzt, 76*, 839-840, 842-834, 846-838.

Barone, L., Fossati, A., & Guiducci, V. (2011). Attachment mental states and inferred pathways of development in borderline personality disorder: A study using the Adult Attachment Interview. *Attachment and Human Development, 13*, 451-469.

Bateman, A., Bolton, R., & Fonagy, P. (2013). Antisocial personality disorder: A mentalizing framework. *FOCUS: The Journal of Lifelong Learning in Psychiatry, 11*, 178-186.

Bateman, A., & Fonagy, P. (2008). Comorbid antisocial and borderline personality disorders: Mentalization-based treatment. *Journal of Clinical Psychology, 64*, 181-194.

Bateman, A., & Fonagy, P. (2009). Randomized controlled trial of outpatient mentalization-based treatment versus structured clinical management for borderline personality disorder. *American Journal of Psychiatry, 166*, 1355-1364.

Bateman, A., & Fonagy, P. (2010). Mentalization based treatment for borderline personality disorder. *World Psychiatry, 9*, 11-15.

Bateman, A., & Fonagy, P. (2013). Impact of clinical severity on outcomes of mentalisation-based treatment for borderline personality disorder. *British Journal of Psychiatry, 203*, 221-227.

Bateman, A. W., & Fonagy, P. (2004). Mentalization-based treatment of BPD. *Journal of Personality Disorders, 18*, 36-51.

Bateman, A. W., & Fonagy, P. (Eds.). (2012). *Handbook of mentalizing in mental health practice*. Washington, DC: American Psychiatric Publishing.

Beck, A. T., Freeman, A., & Davis, D. D. (2004). *Cognitive therapy of personality disorders*. New York, NY: Guilford Press.

Beebe, B., Jaffe, J., Markese, S., Buck, K., Chen, H., Cohen, P., . . . Feldstein, S. (2010). The origins of 12-month attachment: A microanalysis of 4-month mother-infant interaction. *Attachment and Human Development, 12*, 3-141.

Belsky, J., & Fearon, P. R. M. (2008). Precursors of attachment security. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment theory and research* (2nd ed., pp. 295-316). New York, NY: Guilford Press.

Bender, D. S., & Skodol, A. E. (2007). Borderline personality as a self-other representational disturbance. *Journal of Personality Disorders, 21*, 500-517.

Benish, S. G., Imel, Z. E., & Wampold, B. E. (2008). The relative efficacy of bona fide psychotherapies for treating post-traumatic stress disorder: A meta-analysis of direct comparisons. *Clinical Psychology Review, 28*, 746-758.

Blatt, S. J., & Auerbach, J. S. (1988). Differential cogntive disturbances in three types of borderline patients. *Journal of Personality Disorders, 2*, 198-211.

Bowlby, J. (1969). *Attachment and loss, Vol. 1: Attachment*. London, UK: Hogarth Press and Institute of Psycho-Analysis.

Bowlby, J. (1973). *Attachment and loss, Vol. 2: Separation: Anxiety and anger*. London, UK: Hogarth Press and Institute of Psycho-Analysis.

Bradley, R., & Westen, D. (2005). The psychodynamics of borderline personality disorder: A view from developmental psychopathology. *Development and Psychopathology, 17*, 927-957.

Caligor, E., Kernberg, O. F., & Clarkin, J. F. (2007). *Handbook of dynamic psychotherapy for higher level personality pathology*. Washington, DC: American Psychiatric Press.

Carlson, E. A., Egeland, B., & Sroufe, L. A. (2009). A prospective investigation of the development of borderline personality symptoms. *Development and Psychopathology, 21*, 1311-1334.

Choi-Kain, L. W., Fitzmaurice, G. M., Zanarini, M. C., Laverdiere, O., & Gunderson, J. G. (2009). The relationship between self-reported attachment styles, interpersonal dysfunction, and borderline personality disorder. *Journal of Nervous and Mental Disease, 197*, 816-821.

Cicchetti, D., Rogosch, F. A., & Toth, S. L. (2000). The efficacy of toddler-parent psychotherapy for fostering cognitive development in offspring of depressed mothers. *Journal of Abnormal Child Psychology, 28*, 135-148.

Cicchetti, D., & Walker, E. F. (2001). Editorial: Stress and development: Biological and psychological consequences. *Development and Psychopathology, 13*, 413-418.

Corriveau, K. H., Harris, P. L., Meins, E., Fernyhough, C., Arnott, B., Elliott, L., . . . de Rosnay, M. (2009). Young children's trust in their mother's claims: Longitudinal links with attachment security in infancy. *Child Development, 80*, 750-761.

Crandell, L. E., & Hobson, R. P. (1999). Individual differences in young children's IQ: A social-developmental perspective. *Journal of Child Psychology and Psychiatry, 40*, 455-464.

Crowell, S. E., Beauchaine, T. P., & Linehan, M. M. (2009). A biosocial developmental model of borderline personality: Elaborating and extending Linehan's theory. *Psychological Bulletin, 135*, 495-510.

de Rosnay, M., & Harris, P. L. (2002). Individual differences in children's understanding of emotion: The roles of attachment and language. *Attachment and Human Development, 4*, 39-54.

DeWolf, M. S., & van IJzendoorn, M. H. (1997). Sensitivity and attachment: A meta-analysis on parental antecedents of infant attachment. *Journal of Marriage and the Family, 68*, 571-591.

Falkenstrom, F., Granstrom, F., & Holmqvist, R. (2013). Therapeutic alliance predicts symptomatic improvement session by session. *Journal of Counseling Psychology, 60*, 317-328.

Fearon, P., Shmueli-Goetz, Y., Viding, E., Fonagy, P., & Plomin, R. (2013). Genetic and environmental influences on adolescent attachment. *Journal of Child Psychology and Psychiatry*. doi: 10.1111/jcpp.12171

Fonagy, P. (1982). The integration of psychoanalysis and empirical science: A review. *International Review of Psychoanalysis, 9*, 125-145.

Fonagy, P. (1991). Thinking about thinking: Some clinical and theoretical considerations in the treatment of a borderline patient. *International Journal of Psycho-Analysis, 72 (Pt 4)*, 639-656.

Fonagy, P. (2000). Attachment and borderline personality disorder. *Journal of the American Psychoanalytic Association, 48*, 1129-1146.

Fonagy, P., & Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy*. doi: 10.1037/a0036505

Fonagy, P., & Bateman, A. (2008). The development of borderline personality disorder: A mentalizing model. *Journal of Personality Disorders, 22*, 4-21.

Fonagy, P., Bateman, A., & Bateman, A. (2011). The widening scope of mentalizing: A discussion. *Psychology and Psychotherapy: Theory, Research and Practice, 84*, 98-110.

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., & Higgitt, A. (1990). A developmental perspective on borderline personality disorder. *Revue Internationale de Psychopathologie, 1*, 125-159.

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

Fonagy, P., & Luyten, P. (in press). A multilevel perspective on the development of borderline personality disorder. In D. Cicchetti (Ed.), *Development and psychopathology* (3rd ed.). New York, NY: John Wiley & Sons.

Fonagy, P., Luyten, P., & Allison, E. (2014). Epistemic petrifaction and the restoration of epistemic trust: A new conceptualization of borderline personality disorder and its psychosocial treatment. *Manuscript in preparation*.

Fonagy, P., Luyten, P., Bateman, A., Gergely, G., Strathearn, L., Target, M., & Allison, E. (2010). Attachment and personality pathology. In J. F. Clarkin, P. Fonagy & G. O. Gabbard (Eds.), *Psychodynamic psychotherapy for personality disorders: A clinical handbook* (1st ed., pp. 37-88). Washington, DC: American Psychiatric Publishing.

Fonagy, P., Steele, H., Steele, M., & Holder, J. (1997). Attachment and theory of mind: Overlapping constructs? *Association for Child Psychology and Psychiatry Occasional Papers, 14*, 31-40.

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.

Fonagy, P., Steele, M., Steele, H., Moran, G. S., & Higgitt, A. C. (1991). The capacity for understanding mental states: The reflective self in parent and child and its significance for security of attachment. *Infant Mental Health Journal, 12*, 201-218.

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

Fonagy, P., & Target, M. (2000). Playing with reality: III. The persistence of dual psychic reality in borderline patients. *International Journal of Psychoanalysis, 81*, 853-874.

Fonagy, P., & Target, M. (2003). *Psychoanalytic theories: Perspectives from developmental psychopathology*. London: Whurr.

Fonagy, P., Target, M., & Gergely, G. (2000). Attachment and borderline personality disorder: A theory and some evidence. *Psychiatric Clinics of North America, 23*, 103-122.

Fonagy, P., Target, M., & Gergely, G. (2006). Psychoanalytic perspectives on developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology. Volume 1: Theory and method* (2nd ed., pp. 701-749). Hoboken, NJ: John Wiley & Sons.

Fossati, A., Feeney, J. A., Carretta, I., Grazioli, F., Milesi, R., Leonardi, B., & Maffei, C. (2005). Modeling the relationships between adult attachment patterns and borderline personality disorder: The role of impulsivity and aggressiveness. *Journal of Social and Clinical Psychology, 24*, 520-537.

Fraley, R. C., Roisman, G. I., Booth-LaForce, C., Owen, M. T., & Holland, A. S. (2013). Interpersonal and genetic origins of adult attachment styles: A longitudinal study from infancy to early adulthood. *Journal of Personality and Social Psychology, 104*, 817-838.

Frank, J. D., & Frank, J. B. (1991). *Persuasion and healing: A compatative study of psychotherapy*. Baltimore, MD: Johns Hopkins University Press.

Frankenhuis, W. E., Panchanathan, K., & Clark Barrett, H. (2013). Bridging developmental systems theory and evolutionary psychology using dynamic optimization. *Developmental Science, 16*, 584-598.

Gao, Y., Raine, A., Chan, F., Venables, P. H., & Mednick, S. A. (2010). Early maternal and paternal bonding, childhood physical abuse and adult psychopathic personality. *Psychological Medicine, 40*, 1007-1016.

Goodman, G. S., Quas, J. A., & Ogle, C. M. (2010). Child maltreatment and memory. *Annual Review of Psychology, 61*, 325-351.

Grant, B. F., Chou, S. P., Goldstein, R. B., Huang, B., Stinson, F. S., Saha, T. D., . . . Ruan, W. J. (2008). Prevalence, correlates, disability, and comorbidity of DSM-IV borderline personality disorder: Results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Clinical Psychiatry, 69*, 533-545.

Gratz, K. L., Tull, M. T., Baruch, D. E., Bornovalova, M. A., & Lejuez, C. W. (2008). Factors associated with co-occurring borderline personality disorder among inner-city substance

users: The roles of childhood maltreatment, negative affect intensity/ reactivity, and emotion dysregulation. *Comprehensive Psychiatry, 49*, 603-615.

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

Hamilton, W. D. (1964). The genetic evolution of social behaviour. *Journal of Theoretical Biology, 7*, 1-52.

Hazlett, E. A., Zhang, J., New, A. S., Zelmanova, Y., Goldstein, K. E., Haznedar, M. M., . . . Chu, K. W. (2012). Potentiated amygdala response to repeated emotional pictures in borderline personality disorder. *Biological Psychiatry, 72*, 448-456.

Hengartner, M. P., Ajdacic-Gross, V., Rodgers, S., Muller, M., & Rossler, W. (2013). Childhood adversity in association with personality disorder dimensions: New findings in an old debate. *European Psychiatry*.

Higgitt, A., & Fonagy, P. (1992). Psychotherapy in borderline and narcissistic personality disorder. *British Journal of Psychiatry, 161*, 23-43.

Isabella, R. A., Belsky, J., & von Eye, A. (1989). Origins of infant-mother attachment: An examination of interactional synchrony during the infant's first year. *Developmental Psychology, 25*, 12-21.

Jacobsen, T., & Hofmann, V. (1997). Children's attachment representations: Longitudinal relations to school behavior and academic competency in middle childhood and adolescence. *Developmental Psychology, 33*, 703-710.

Johnson, J. G., Cohen, P., Kasen, S., Ehrensaft, M. K., & Crawford, T. N. (2006). Associations of parental personality disorders and axis I disorders with childrearing behavior. *Psychiatry (Edgmont), 69*, 336-350.

Jorgensen, C. R. (2010). Invited essay: Identity and borderline personality disorder. *Journal of Personality Disorders, 24*, 344-364.

Jorgensen, C. R., Berntsen, D., Bech, M., Kjolbye, M., Bennedsen, B. E., & Ramsgaard, S. B. (2012). Identity-related autobiographical memories and cultural life scripts in patients with borderline personality disorder. *Consciousness and Cognition, 21*, 788-798.

Kernberg, O. F. (1975). *Borderline conditions and pathological narcissism*. New York, NY: Jason Aronson.

Kernberg, O. F. (1984). *Severe personality disorders: Psychotherapeutic strategies*. New Haven, CT: Yale University Press.

Kiraly, I., Csibra, G., & Gergely, G. (2013). Beyond rational imitation: Learning arbitrary means actions from communicative demonstrations. *Journal of Experimental Child Psychology, 116*, 471-486.

Kiser, L. J., Bates, J. E., Maslin, C. A., & Bayles, K. (1986). Mother-infant play at six months as a predictor of attachment security of thirteen months. *Journal of the American Academy of Child Psychiatry, 25*, 68-75.

Leichsenring, F., Leibing, E., Kruse, J., New, A. S., & Leweke, F. (2011). Borderline personality disorder. *Lancet, 377*, 74-84.

Linehan, M. M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. New York, NY: Guilford Press.

Livesley, W. J. (2003). *Practical management of personality disorder*. New York, NY: Guilford Press.

Luyten, P., & Blatt, S. J. (2013). Interpersonal relatedness and self-definition in normal and disrupted personality development: Retrospect and prospect. *American Psychologist, 68*, 172-183.

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.

Main, M. (2000). The organized categories of infant, child, and adult attachment: Flexible vs. inflexible attention under attachment-related stress. *Journal of the American Psychoanalytic Association, 48*, 1055-1096; discussion 1175-1087.

Marvin, R. S., & Britner, P. A. (2008). Normal development: The ontogeny of attachment. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment theory and research* (2nd ed., pp. 269-294). New York, NY: Guilford Press.

McAleavey, A. A., & Castonguay, L. G. (2013). Insight as a common and specific impact of psychotherapy: Therapist-reported exploratory, directive, and common factor interventions. *Psychotherapy*.

McWilliams, N. (1994). *Psychoanalytic diagnosis*. New York: The Guilford Press.

Meins, E., Fernyhough, C., Fradley, E., & Tuckey, M. (2001). Rethinking maternal sensitivity: Mothers' comments on infants' mental processes predict security of attachment at 12 months. *Journal of Child Psychology and Psychiatry, 42*, 637-648.

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

Meyer, B., Ajchenbrenner, M., & Bowles, D. P. (2005). Sensory sensitivity, attachment experiences, and rejection responses among adults with borderline and avoidant features. *Journal of Personality Disorders, 19*, 641-658.

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

Mills-Koonce, W. R., Gariepy, J. L., Propper, C., Sutton, K., Calkins, S., Moore, G., & Cox, M. (2007). Infant and parent factors associated with early maternal sensitivity: A caregiver-attachment systems approach. *Infant Behavior & Development, 30*, 114-126.

Minzenberg, M. J., Fan, J., New, A. S., Tang, C. Y., & Siever, L. J. (2007). Fronto-limbic dysfunction in response to facial emotion in borderline personality disorder: An event-related fMRI study. *Psychiatry Research, 155*, 231-243.

Moss, E., Rousseau, D., Parent, S., St.-Laurent, D., & Saintong, J. (1998). Correlates of attachment at school-age: Maternal reported stress, mother-child interaction and behavior problems. *Child Development, 69*, 1390-1405.

New, A. S., aan het Rot, M., Ripoll, L. H., Perez-Rodriguez, M. M., Lazarus, S., Zipursky, E., . . . Siever, L. J. (2012). Empathy and alexithymia in borderline personality disorder: Clinical and laboratory measures. *Journal of Personality Disorders, 26*, 660-675.

Northoff, G., Qin, P., & Feinberg, T. E. (2011). Brain imaging of the self--conceptual, anatomical and methodological issues. *Consciousness and Cognition, 20*, 52-63.

Oldham, J. M. (2009). Borderline personality disorder comes of age. *American Journal of Psychiatry, 166*, 509-511.

Paris, J. (2013). How the history of psychotherapy interferes with integration. *Journal of Psychotherapy Integration, 23*, 99-106.

Powers, A. D., Thomas, K. M., Ressler, K. J., & Bradley, B. (2011). The differential effects of child abuse and posttraumatic stress disorder on schizotypal personality disorder. *Comprehensive Psychiatry, 52*, 438-445.

Prochaska, J. O., & Norcross, J. C. (2013). *Systems of psychotherapy: A transtheoretical analysis* (8th ed.). Belmont, CA: Brooks/Cole Cengage Advantage Books.

Repetti, R. L., Taylor, S. E., & Seeman, T. E. (2002). Risky families: Family social environments and the mental and physical health of offspring. *Psychological Bulletin, 128*, 330-366.

Robinson, P., Barrett, B., Bateman, A., Hakeem, A., Hellier, J., Lemonsky, F., . . . Fonagy, P. (2014). Study protocol for a randomized controlled trial of mentalization based therapy against specialist supportive clinical management in patients with both eating disorders and symptoms of borderline personality disorder. *BMC Psychiatry, 14*, 51.

Rogers, C. R. (1951). *Client-centered therapy*. Boston, MA: Houghton Mifflin.

Rogers, C. R., & Dymond, R. F. (1954). *Psychotherapy and personality change: Coordinated research studies in the client-centered approach*. Chicago, IL: University of Chicago Press.

Rogosch, F. A., & Cicchetti, D. (2005). Child maltreatment, attention networks, and potential precursors to borderline personality disorder. *Development and Psychopathology, 17*, 1071-1089.

Scott, L. N., Kim, Y., Nolf, K. A., Hallquist, M. N., Wright, A. G., Stepp, S. D., . . . Pilkonis, P. A. (2013). Preoccupied attachment and emotional dysregulation: Specific aspects of borderline personality disorder or general dimensions of personality pathology? *Journal of Personality Disorders, 27*, 473-495.

Scott, L. N., Levy, K. N., & Pincus, A. L. (2009). Adult attachment, personality traits, and borderline personality disorder features in young adults. *Journal of Personality Disorders, 23*, 258-280.

Shedler, J., Beck, A., Fonagy, P., Gabbard, G. O., Gunderson, J., Kernberg, O., . . . Westen, D. (2010). Personality disorders in DSM-5. *American Journal of Psychiatry, 167*, 1026-1028.

Shedler, J., Beck, A. T., Fonagy, P., Gabbard, G. O., Kernberg, O., Michels, R., & Westen, D. (2011). Revision of the Personality Disorder Model for DSM-5: Response to Skodol letter. *American Journal of Psychiatry, 168*, 97-98.

Skodol, A. E. (2012). Personality disorders in DSM-5. *Annual Review of Clinical Psychology, 8*, 317-344.

Slade, A., Grienenberger, J., Bernbach, E., Levy, D., & Locker, A. (2005). Maternal reflective functioning, attachment, and the transmission gap: A preliminary study. *Attachment and Human Development, 7*, 283-298.

Specht, M. W., Chapman, A., & Cellucci, T. (2009). Schemas and borderline personality disorder symptoms in incarcerated women. *Journal of Behavior Therapy and Experimental Psychiatry, 40*, 256-264.

Sperber, D., Clement, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language, 25*, 359-393.

Stacks, A. M., Beeghly, M., Partridge, T., & Dexter, C. (2011). Effects of placement type on the language developmental trajectories of maltreated children from infancy to early childhood. *Child Maltreatment, 16*, 287-299.

Steele, H., & Siever, L. (2010). An attachment perspective on borderline personality disorder: Advances in gene-environment considerations. *Current Psychiatry Reports, 12*, 61-67.

Stein, H., Fonagy, P., Ferguson, K. S., & Wisman, M. (2000). Lives through time: An ideographic approach to the study of resilience. *Bulletin of the Menninger Clinic, 64*, 281-305.

Tasca, G. A., & Lampard, A. M. (2012). Reciprocal influence of alliance to the group and outcome in day treatment for eating disorders. *Journal of Counseling Psychology, 59*, 507-517.

Tyrka, A. R., Wyche, M. C., Kelly, M. M., Price, L. H., & Carpenter, L. L. (2009). Childhood maltreatment and adult personality disorder symptoms: Influence of maltreatment type. *Psychiatry Research, 165*, 281-287.

Widom, C. S., Czaja, S. J., & Paris, J. (2009). A prospective investigation of borderline personality disorder in abused and neglected children followed up into adulthood. *Journal of Personality Disorders, 23*, 433-446.

Wilkinson, R., & Pickett, K. (2009). *The spirit level: Why equality is better for everyone*. London, UK: Penguin Books.

Wilson, D., & Sperber, D. (2012). *Meaning and relevance*. Cambridge, UK: Cambridge University Press.

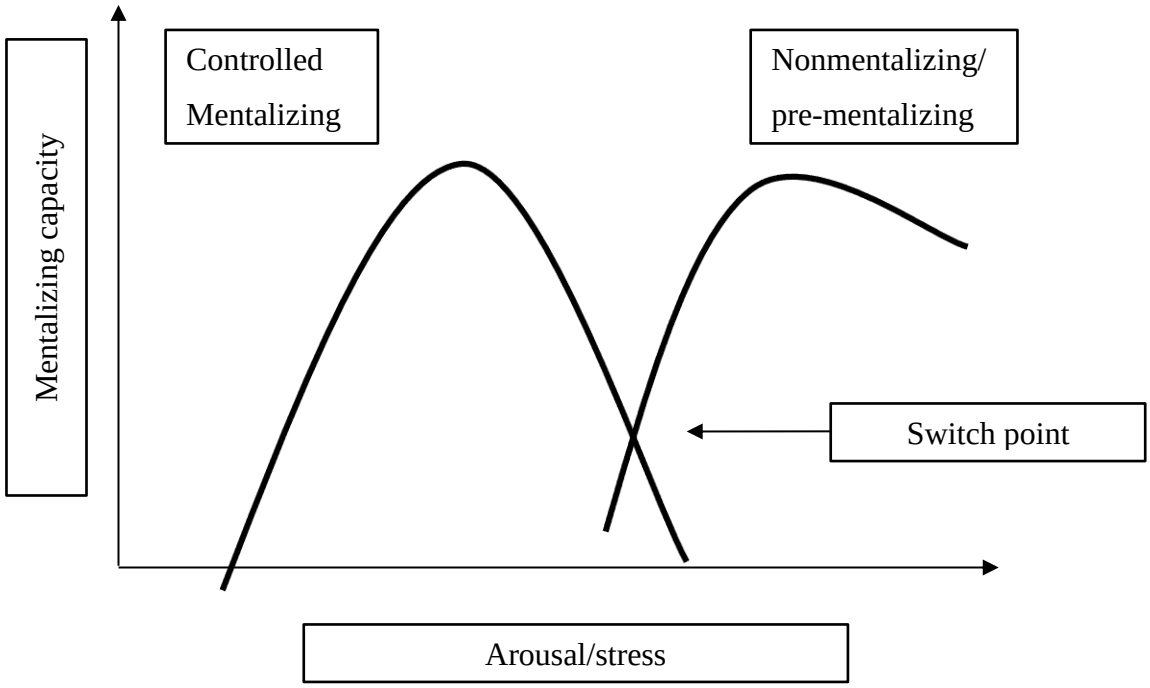

Figure 1. Switch from Controlled Mentalizing to Nonmentalizing/Pre-mentalizing Modes under Conditions of High Arousal.

Figure 2. Mentalizing Profile of Borderline Personality Disorder Across the Four Polarities that Underlie Mentalizing.