Personality Disorders: Theory, Research, and Treatment
2015, Vol. 6, No. 4, 366–379

1949-2715/15/$12.00 http://dx.doi.org/10.1037/per0000117

The Neurobiology of Mentalizing

Patrick Luyten
University of Leuven and University College London

Peter Fonagy
University College London

Mentalizing is the capacity to understand ourselves and others in terms of intentional mental states, such as feelings, desires, wishes, attitudes, and goals. It is a fundamental capacity in our complex social environment. This article reviews our current understanding of the neurobiology of mentalizing. We first summarize the key assumptions of the mentalizing approach to normal and disrupted development. This is followed by discussion of the multiple dimensions of mentalizing and our emerging knowledge of the neural circuits that underlie these dimensions. We then consider the neurobiology of attachment and arousal regulation in relation to mentalizing, and summarize relevant studies in this area. Finally, we discuss the limitations of extant research and outline implications for future research.

Keywords: attachment, borderline personality disorder, Mentalization-Based Treatment, mentalizing, neurobiology

Mentalizing is the capacity to understand ourselves and others in terms of intentional mental states, such as feelings, desires, wishes, attitudes, and goals. It is a fundamental capacity in our social environment: Without this capacity, we would be completely lost in a world that is determined by complex and ever-changing interpersonal relationships that require a high degree of collaboration and mutual understanding (Fonagy, Luyten, & Allison, in press). Although it is to a certain extent "prewired," our capacity for mentalizing is not a given; it is largely a developmental achievement. Research findings suggest that the capacity for mentalizing is first acquired in the context of attachment relationships, and that the extent to which our early and later environment fosters a focus on internal mental states is crucial for its development (Allen, Fonagy, & Bateman, 2008; Fonagy et al., 2010).

The past decade has witnessed a veritable explosion of research on mentalizing. A quick search in Web of Science, for instance, shows an exponential increase in the number of studies in this domain, from only a handful of studies on the topic published in the 1990s to more than 4,000 in 2014. The increasing popularity of the mentalizing approach to understanding both normal and disrupted development is explained not only by the growing realization that this capacity is central in human normative development; it is also explained by the recognition that temporary or stable disruptions in this capacity are one characteristic of almost all forms of psychopathology—ranging from autism and psychosis (Chung, Barch, & Strube, 2014) to major depression (Cusi, Nazarov, Holshausen, Macqueen, & McKinnon, 2012; Ladegaard, Larsen, Videbech, & Lysaker, 2014; Luyten, Fonagy, Lemma, & Target, 2012), eating disorders (Kuipers & Bekker, 2012; Skårderud, 2007), and personality disorders, most notably borderline personality disorder (BPD) (Fonagy & Luyten, in press).

The great upsurge of interest in the role of mentalizing in development is paralleled by an ever-growing interest in the neural underpinnings of this capacity in the field of social and cognitive neuroscience (Herpertz, Jeung, Mancke, & Bertsch, 2014; Lieberman, 2007; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014; Van Overwalle, 2009; Van Overwalle & Baetens, 2009). These studies have fundamentally informed and shaped the mentalization-based approach to psychopathology. Among the many contributions within cognitive and affective neuroscience, two major findings stand out, and these form the backbone of this paper. The first of these is the finding that mentalizing is not a unitary construct, but consists of several dimensions that can be organized along polarities. As we discuss in more detail in this paper, one of the most remarkable and important contributions of neuroscience is that features of mentalizing that were thought to be unitary turned out to be dissociable, which may explain the specific imbalances in mentalizing that can be identified in various psychiatric disorders. The second major finding is from neuroscience studies which suggest that two interacting factors largely determine the quality of mentalizing: (a) stress or arousal and (b) the individual's attachment history. These findings have led to considerable changes in our understanding of the normative development of mentalizing and the mentalizing impairments seen in various types of psychopathology, as well as in the nature of interventions aimed at fostering mentalizing. This also illustrates the growing dialogue between neuroscience and clinical practice (see Fonagy, Luyten, & Bateman, 2015).

In this article we set out where our current understanding of the neurobiology of mentalizing stands. We first briefly outline the core assumptions of the mentalizing approach to normal and

Patrick Luyten, Faculty of Psychology and Educational Sciences, University of Leuven, and Research Department of Clinical, Educational and Health Psychology, University College London; Peter Fonagy, Research Department of Clinical, Educational and Health Psychology, University College London.

Professor Fonagy is one of the original developers of Mentalization-Based Treatment, a treatment that rests on the theories and research evidence discussed in this article. Dr. Luyten is involved in studies evaluating the efficacy of Mentalization-Based Treatment.

Correspondence concerning this article should be addressed to Patrick Luyten, Faculty of Psychology and Educational Sciences, KU Leuven, Tiensestraat 102 - Box 3722, 3000 Leuven, Belgium. E-mail: patrick.luyten@ppw.kuleuven.be

disrupted development. Next, we consider the multiple dimensions of mentalizing and our emerging knowledge of the neural circuits that underlie these dimensions. We then focus on the neurobiology of attachment and arousal regulation in relation to mentalizing and outline implications for future research on the role of mentalizing in normal and disrupted development. Finally, we finish by pointing to where the field may develop.

Basic Assumptions of the Mentalizing Approach

The basic assumptions of the mentalizing approach to normal and disrupted development are depicted in Figures 1 to 3 and Table 1. Research findings suggest that the capacity for mentalizing is not a constitutional given but is largely a developmental achievement that depends initially on the quality of the individual's attachment relationships, in particular early attachments during infancy (Fonagy & Luyten, in press; Kovács, Téglás, & Endress, 2010). Specifically, the extent to which attachment figures have been able to respond with contingent and marked affective displays of their own experience in response to the infant's subjective experience is thought to be positively associated with the child's ability to develop mentalizing capacities, that is, second-order representations of his or her own subjective experiences (see Figure 1). This in turn positively influences affect-regulative processes and self-control (including attentional mechanisms and effortful control), as the development of the capacity to reflect on internal mental states represents a major leap in the individual's capacity to regulate his or her affect. Later in life, exposure to a wider environment (e.g., peers, teachers, and friends) which fosters a focus on internal mental states is thought to broaden and strengthen the development of mentalizing (Fonagy & Luyten, in press). Conversely, failures in the process of marked mirroring from early attachment figures lead to impairments in the capacity to reflect on the self and others, as they lead to unmentalized self-experiences, also called "alien-self" experiences, which do not validate the individual's experience and thus are felt as alien to the self (see Figure 2). Such failures in marked mirroring are to a certain extent inevitable, and thus we all have unmentalized mental states. In various types of psychopathology, however, most often as the result of a combination of biological vulnerability and environmental circumstances, these alien self-experiences are so pronounced that they dominate the individual's subjectivity. This leads to a constant pressure to externalize these unmentalized self-experiences—which may be expressed, for instance, in a tendency to dominate the mind of others and/or in various types of self-harming behavior (Fonagy & Luyten, in press).

The capacity to mentalize is therefore only in part a trait-like capacity. It is always to a certain extent relationship- and context-specific (e.g., mentalizing levels may differ considerably among relationships, or between when reflecting "off-line" on a past event vs. "online" in a real-life interaction). Mentalizing, therefore, is a fundamentally bidirectional or transactional social process (Fonagy & Target, 1997): It is thought to develop in the context of interactions with others, and its quality in relation to understanding others is assumed to be influenced by the mentalizing capacities of those with whom we interact. Mentalizing is also distinct from attentional processes and general (cognitive) reasoning, although it partly relies on these capacities and in turn fosters them. Neuroimaging studies clearly demonstrate the existence of distinct neural circuits involved in these capacities (Van Overwalle, 2011).

Furthermore, mentalizing is not a unitary, unidimensional capacity. Here, neuroscience findings have been particularly instrumental in defining mentalizing as being organized around four dimensions or polarities, with each polarity having relatively distinct underlying neural circuits (see Table 1). These four polarities are (a) automatic versus controlled mentalizing, (b) mentalizing with regard to self and to others, (c) mentalizing based on external or internal features of self and others, and (d) cognitive versus affective mentalizing (see Table 1, and below; Fonagy & Luyten, 2009; Luyten, Fonagy, Lowyck, & Vermote, 2012). On the basis of this view, different types of psychopathology can be characterized by their different combinations of impairments along these polarities (i.e., different *mentalizing profiles*; see Fonagy et al., 2015).

The concept of mentalizing is thus an umbrella concept, which encompasses and covers related constructs from social cognition research such as empathy, mindfulness, and Theory of Mind (ToM; Choi-Kain & Gunderson, 2008). Empathy and ToM, for instance, tap into features of mentalizing about others, whereas mindfulness primarily involves a core component of mentalizing about the self (e.g., the ability to attend to one's own internal mental states). Both empathy and mindfulness focus on affective components of mentalizing, whereas ToM is more about cognitive

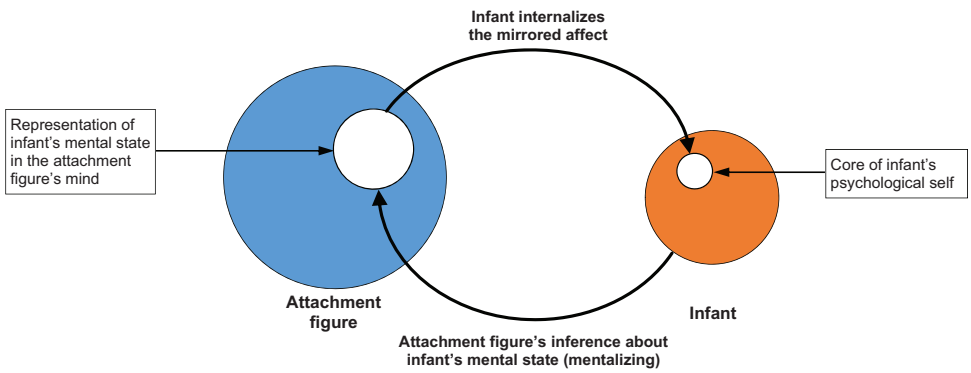

Figure 1. The role of marked mirroring in the development of mentalizing. See the online article for the color version of this figure.

Table 1
Four Dimensions of Mentalizing: Distinguishing Features and Hypothesized Underlying Neural Circuits

Polarity	Features	Neural circuits
Automatic	Unconscious, parallel, fast processing of social information that is reflexive and requires little effort, focused attention, or intention; therefore prone to bias and distortions, particularly in complex interpersonal interactions (i.e. when arousal is high)	Amygdala Basal ganglia Ventromedial prefrontal cortex (VMPFC) Lateral temporal cortex (LTC) Dorsal anterior cingulate cortex (dACC)
Controlled	Conscious, verbal, and reflective processing of social information that requires the capacity to reflect consciously and deliberately on and make accurate attributions about the emotions, thoughts, and intentions of self and others. Relies heavily on effortful control and language	Lateral prefrontal cortex (LPFC) Medial prefrontal cortex (MPFC) Lateral parietal cortex (LPAC) Medial parietal cortex (MPAC) Medial temporal lobe (MTL) Rostral anterior cingulate cortex (rACC)
Internal	Understanding one's own mind and that of others through a direct focus on the mental interiors of both the self and others	Medial frontoparietal network (more controlled)
External	Understanding one's own mind and that of others based on external features (such as facial expressions, posture, and prosody)	Lateral frontotemporoparietal network (more automatic)
Self–Other	Shared networks underpin the capacity to mentalize about the self and others	Shared representation system (more automatic) versus mental state attribution system (more controlled)
Cognitive–Affective	Mentalizing may focus on more cognitive features (more controlled), such as belief-desire reasoning and perspective-taking, versus more affective features (more automatic), including affective empathy and mentalized affectivity (the feeling and thinking-about-the-feeling)	Cognitive mentalizing involves several areas in prefrontal cortex; affectively oriented mentalizing seems particularly related to the VMPFC

features of mentalizing (e.g., belief-desire reasoning; although this concept has broadened considerably in recent years to include affect). Mentalizing is broader than any of these concepts: It focuses on both self and other, and on both cognition and affect. Furthermore, mentalizing also encompasses processes involved in interpreting one's own mind and that of others based on external features (such as facial expressions, posture, and prosody) and balancing this sensitivity with knowledge about the mental interiors of both the self and others. Mentalizing is thus all about the *balance* between the systems underlying these four dimensions and potential *imbalances* (e.g., being overly sensitive to the emotional states of others at the expense of reflective awareness of one's own state of mind). Good mentalizing thus balances the various systems that are responsible for being aware of how one feels oneself, what one thinks, and what others feel and think.

This balance is thought to depend on the interaction between two determining factors: (a) stress or arousal and (b) the use of attachment strategies in response to arousal (see Figure 3). As explained in more detail below, as stress or arousal increases, there is a tendency to switch from slow and reflective mentalizing to fast, automatic and so-called *prementalizing* modes of experiencing oneself and others (see Table 2 and Figure 3). Automatic mentalizing tends to be rigid and typically involves biased assumptions about the self and others. Individual differences in the use of attachment strategies are thought to influence three key parameters related to the switch from controlled to automatic mentalizing:

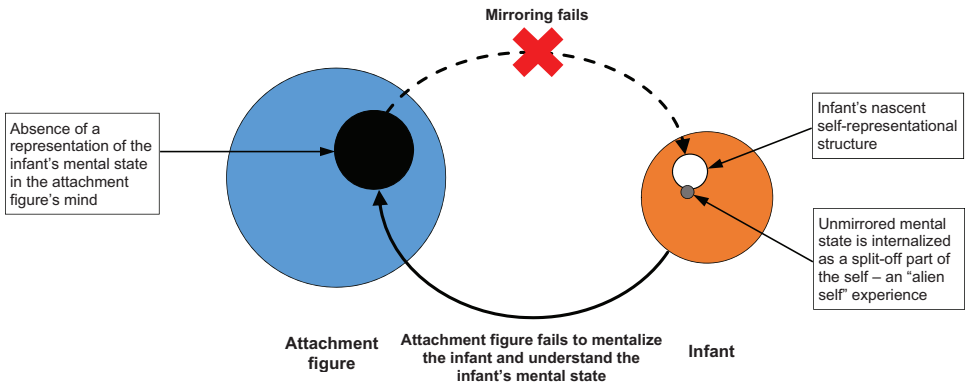

Figure 2. Failure to adequately mirror mental states, problems with mentalizing, and the emergence of alien self-parts. See the online article for the color version of this figure.

(a) how readily individuals switch to nonmentalizing modes, (b) the extent to which the individual loses the capacity for more controlled mentalizing, and (c) the duration of the loss of controlled mentalizing. As we will discuss, a neurobiological understanding of the relationship between arousal, attachment strategies, and mentalizing is particularly pertinent to our understanding of disorders that are characterized by extreme impairments and imbalances in mentalizing, such as BPD.

The Neurobiology of the Mentalizing Polarities

Automatic and Controlled Mentalizing

Findings from both behavioral science and neuroscience support a distinction between two different types of mentalizing underpinned by relatively distinct neural circuits. *Automatic* or *implicit* mentalizing presumes the use of parallel and therefore much faster processing; it is reflexive and requires little effort, focused attention, or intention (Satpute & Lieberman, 2006). Automatic mentalizing seems to be our default position: We constantly tend to automatically "read" the mind of others. This capacity is already present in a rudimentary form in infants as young as 7 months of age (Kovács et al., 2010). From an evolutionary perspective, automatic mentalizing has clear value for survival (Lieberman, 2007; Mayes, 2006): Typical fight/flight responses are best subserved by fast (and thus automatic) processing of social information (e.g., when, at night in a dark alley, we see a man with a gun in his hand approaching us). Yet, in many circumstances, and particularly in our complex interpersonal world, the switch under high arousal conditions from controlled, explicit, and reflective mentalizing to automatic or implicit mentalizing may not always be that adaptive—and particularly not in individuals who have a low threshold for such a switch, as it hampers their ability to pause and reflect, and so to develop appropriate models of their mind and that of others. As we will discuss in more detail below, particularly in situations of increasing arousal (see Figure 3), automatic mentalizing is likely to be based on simple heuristics that may work well under some conditions, but utterly fail to capture the complexity of human motivations in more complex interpersonal situations (e.g., when we find ourselves or one of our loved ones in a difficult love relationship, or when we are involved in a conflict at work). The (often much-needed) correction of biased assumptions associated with automatic mentalizing is exerted by *controlled* or *explicit* mentalizing, which is typically conscious, verbal, and reflective. Mentalizing in real time under realistic contextual demands requires the capacity to reflect consciously and deliberately on and make accurate attributions about the emotions, thoughts, and intentions of others, and to display an accurate and balanced appreciation of a social situation—which relies heavily on the capacity for effortful control and the subtle distinctions language allows us to make.

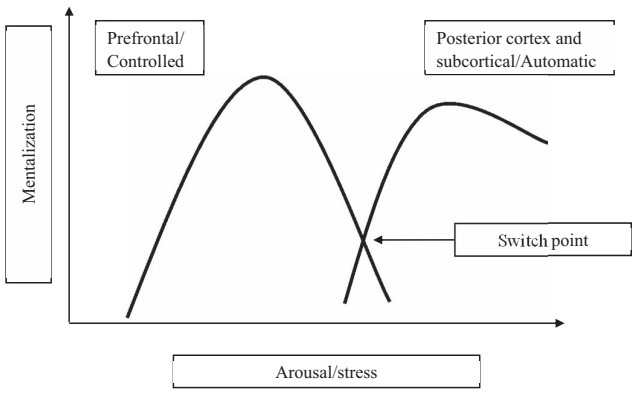

Figure 3. A biobehavioral switch model of the relationship between arousal/stress and controlled versus automatic mentalizing.

Table 2
Automatic, Nonmentalizing Modes That Re-Emerge With the Loss of Controlled Mentalizing

Psychic equivalence mode
Individuals equate inner (mental) reality with outer reality ("mind–world isomorphism"). Because of this, the internal has the same power as the external
Intolerance of alternative perspectives – leads to "concrete" understanding
Teleological mode
Extreme exterior focus
Only observable change or action is considered a true indicator of the intentions of the other
Pretend mode
Ideas form no bridge between inner and outer reality; thoughts and feelings are decoupled from external reality
In extreme, may manifest as "dissociation" of thought (hypermentalizing or pseudomentalizing)

In our complex contemporary social world, which demands increasingly sophisticated collaboration with others, considerable "computational power" is needed to develop models of the minds of ourselves and others (Fonagy et al., in press). In this context, relying on automatic mentalizing is not always adaptive, leading to an evolutionary "friction-rub" of the neural systems involved in mentalizing, particularly when automatic mentalizing is dominated by nonreflexive and biased assumptions about the self and others. In the absence of such biases, however, automatic mentalizing may provide a very effective and efficient way of processing social contexts. Yet, studies tend to suggest that, even in normative development, automatic mentalizing is often biased toward nonreflective assumptions about the self and others—particularly when it is the result of high arousal (i.e., when feeling ashamed, embarrassed, or threatened, or when confronted with out-group members). For example, this is illustrated by studies demonstrating the rapid activation of biased views toward people of another race in priming studies (Knutson, Mah, Manly, & Grafman, 2007).

Many types of psychopathology, and serious personality disorder pathology in particular, seems to be characterized by temporary or permanent impairments in the capacity for controlled mentalizing (Fonagy et al., 2015). Neuroscience findings have begun to shed more light on the neural circuits underpinning this capacity and therefore on the neurobiological basis of these disorders.

Automatic and controlled mentalizing seem to be subserved by two relatively different neural circuits. Phylogenetically older brain circuits that rely primarily on sensory information appear to underlie automatic mentalizing, whereas controlled mentalizing involves phylogenetically newer brain circuits that rely more on

linguistic/symbolic processing. Although the assignment of particular brain regions to each of these circuits currently is at best tentative, neural circuits underlying automatic mentalizing probably include the amygdala, basal ganglia, ventromedial prefrontal cortex (VMPFC), lateral temporal cortex (LTC), and dorsal anterior cingulate cortex (dACC) (Satpute & Lieberman, 2006). These brain areas are primarily involved in the rapid detection of threat and the fast and automatic modulation and processing of (social) information. The amygdala, for instance, has been consistently linked to processing of the biological "value" of information, and is particularly reactive to facial emotional expressions; this highlights its central role in the rapid processing of social information in the context of the fight/flight response. The VMPFC plays a key role in the modulation of the amygdala and basal ganglia, and both the VMPFC and basal ganglia are involved in automatic intuition. The basal ganglia have also been shown to be involved in reward-related implicit emotion processing, to which we will return in detail below. The dACC has been implicated in nonreflective emotional distress related to both physical and social (i.e., exclusion) pain. The LTC—in particular the superior temporal sulcus region—plays a role in fast and automatic processing of biological motion, face recognition, and attribution of intentions. Hence, all these regions are involved in fast and implicit processing of social information.

Controlled mentalizing involves the lateral prefrontal cortex (LPFC), medial prefrontal cortex (MPFC), lateral parietal cortex (LPAC), medial parietal cortex (MPAC), medial temporal lobe (MTL), and rostral anterior cingulate cortex (rACC) (Lieberman, 2007; Satpute & Lieberman, 2006; Uddin, Iacoboni, Lange, & Keenan, 2007). The LPFC has been most consistently shown to be activated by tasks requiring asymmetrical reasoning (e.g., X causes Y, but this does not imply that Y causes X), requiring effortful control and involving considerable computational resources. The LPAC is similarly involved in tasks that require reasoning, and the MPAC is involved in explicit perspective-taking. The rACC seems to be involved in explicit, reflected-upon conflict processing; the MTL has been implicated in explicit, declarative memory. The MPFC seems to be one of the core structures involved in mentalizing, but it is not clear whether this structure primarily belongs to the automatic or the controlled circuit, or both. Because the MPFC is larger in humans than in other primates, and because cognitive load decreases its performance, it is considered to belong to the controlled system (Lieberman, 2007; Satpute & Lieberman, 2006; Uddin et al., 2007).

Given its potential evolutionary function, the capacity for automatic mentalizing appears to be neurally "prewired." Developmental studies suggest that automatic or implicit mentalizing is robust early in the second year of life (Baillargeon, Scott, & He, 2010) or perhaps even earlier, as indicated by a study that showed that babies as young as 7 months of age automatically encode other's beliefs (Kovács et al., 2010). Verbal recognition of another's perspective, by contrast, is reliable only in the fourth year (Carpendale & Lewis, 2006) or even later, perhaps after age 8 (Gweon, Dodell-Feder, Bedny, & Saxe, 2012), which is probably related to language acquisition (Beeghly & Cicchetti, 1994) and the development of effortful control (Fonagy & Luyten, 2009). Congruent with these speculations about the potential evolutionary process underpinning the capacity for controlled mentalizing, there is a strong positive association between the mean social group size and the size of the neocortex for different species of primates (Dunbar, 2008); this is particularly notable for brain areas that support the large-scale social interactions characteristic of *Homo sapiens*. Additionally, a positive correlation has been found between individual explicit mentalizing competences, gray matter volume of mentalizing areas as measured with voxel-based morphometry of magnetic resonance images, and social group size (Lewis, Rezaie, Brown, Roberts, & Dunbar, 2011).

Internal and External Mentalizing

As for the automatic versus controlled distinction, neuroimaging studies have identified two relatively different neural networks that underlie the capacities for *internally focused* and *externally focused* social cognition. Relatively speaking, mentalizing based on external features of self and others (such as facial expressions, posture, and prosody) tends to recruit a lateral frontotemporoparietal network (e.g., posterior superior temporal sulcus (pSTS) and temporal poles), which essentially is involved in less controlled and reflective processes. Mentalization focused on internal features (which requires the intention to represent the internal mental states of self and others), on the other hand, activates a medial frontoparietal network (e.g., MPFC), which is involved in more active and controlled reflection (Lieberman, 2007). Beer, John, Scabini, and Knight (2006), for instance, investigated patients with extensive damage to the MPFC and VMPFC and who showed low levels of self-consciousness when they behaved inappropriately. However, when these patients were shown a video recording of their behavior, they did show self-consciousness, and realized the inappropriateness of their behavior. In these patients, mentalizing based on internal self-monitoring did not elicit embarrassment, but watching the video recruited their intact externally focused self-reflection.

This distinction also seems to be reflected in the ontogenesis of the capacity for mentalizing. Given the nonverbal nature of infants under the age of 24 months, the only way parents can obtain information about their babies' internal mental states is by relying on external features such as the infant's behavior and facial expression (Beebe et al., 2007, 2008). Hence, in early development, attachment figures are almost completely dependent on external cues to develop a model of the mind of their child. Some parents seem to have considerable difficulty with this process, but may do much better at reflecting on the internal mental states of the child once he or she is older—that is, once they can rely more on internally directed mentalizing processes (Sharp & Fonagy, 2008). Other parents seem to have the opposite problem. As an example, Sleed and Fonagy (2010) found that some mothers who appeared highly attuned to their infants when their interaction with their child was rated from videotapes of the interaction scored low on measures of reflective function based on a representation of the child's internal state assessed by Slade's (2005) Parent Development Interview. This may also explain in part the effects of parent–infant intervention programs that use video feedback (Beebe et al., 2008; Slade, 2005). In such interventions, parents are invited to reflect together with a therapist about the possible meanings of their infant's behavior and expressions, thus developing their ability to read others' minds based on external features as well as linking this ability to their capacity to reflect on the minds of others based on internal features.

Hence, even within normative development, individuals may show considerable differences in the capacity for internally versus externally focused mentalizing. These differences are frequently more pronounced in psychopathology. For instance, studies have amply demonstrated major impairments in internally focused mentalizing in individuals with BPD, whereas these individuals appear to show little or no deficit in externally focused mentalizing and may even be hypersensitive toward external social cues (see Fonagy et al., 2015).

Self and Other Mentalizing

Neuroimaging studies have identified a core network of neural systems that is activated whenever individuals reflect on the self and others. This core network consists of the medial prefrontal cortex and temporal poles and the pSTS/temporoparietal junction (TPJ) in the LTC (Frith & Frith, 2006; Lieberman, 2007; Uddin et al., 2007; Van Overwalle, 2009; Van Overwalle & Baetens, 2009). Hence, a shared network seems to underpin the capacity to mentalize about the self and others. Interestingly, the overlapping brain circuitry used in mentalizing about self and others may explain the difficulty of normally developing children to acquire a sense of selfhood, which in the extreme may give rise to serious difficulties with identity integration, as is observed in many types of psychopathology of the self. Patients with BPD in particular seem to constantly struggle to free themselves from the undue influence of others' mental states (which is termed *identity diffusion*).

The neuroimaging literature may help us to understand this phenomenon better: It suggests that two distinct neural networks are involved in self-knowing and knowing others (Lieberman, 2007; Uddin et al., 2007) and that patients with serious pathology of the self may show a marked imbalance between these two systems (Fonagy & Luyten, 2009). Ripoll et al. (2013) suggested distinguishing these two systems as the *shared representation* (SR) system, in which empathic processing relies on shared representations of others' mental states, and the *mental state attribution* (MSA) system, which relies more on symbolic and abstract processing. This distinction overlaps with those suggested by Shamay-Tsoory (2011) and others (Dimaggio, Lysaker, Carcione, Nicolo, & Semerari, 2008; Lieberman, 2007; Lombardo, Barnes, Wheelwright, & Baron-Cohen, 2007; Uddin et al., 2007).

The SR system entails a "visceral recognition" of the experience of others without high-level cognitive processing, based on a similarity of neural activation while experiencing and observing others experiencing states of mind (Lombardo et al., 2010). It is assumed to be a more body-based, frontoparietal (mirror-neuron) system that is involved in understanding the multimodal embodied self (e.g., face and body recognition) and understanding others through motor-simulation mechanisms (Gallese, Keysers, & Rizzolatti, 2004; Rizzolatti & Craighero, 2004; Van Overwalle & Baetens, 2009). This suggests that a fundamental process allowing us to appreciate the actions and emotions of others involves the activation of the mirror neuron system for actions and the activation of visceromotor centers for the understanding of affect (Lombardo et al., 2010). This is thought to be one of the key evolutionary mechanisms underpinning social empathy—knowing from the inside, as it were, how another feels. Hence, this is an implicit, automatic system, providing physical other-to-self and self-to-other mapping, allowing the immediate understanding (but also misunderstanding, as we shall see) of self and others. SR processing is present from infancy and phylogenetically dates back to rodents. Neuroanatomically, it may engage the amygdala, inferior frontal gyrus, inferior parietal lobule (both of these zones are rich in mirror neurons; Bernhardt & Singer, 2012; Van Overwalle & Baetens, 2009), anterior insula, and (dorsal) ACC (both of which are involved in observed and felt pain). Congruent with these assumptions, Seyfarth and Cheney (2013) argue that trust, empathy, and sensitivity to others' emotional states develop out of a largely unconscious mimicking tendency, which we share with primates and which plays a key role in affiliative behavior (van Baaren, Holland, Kawakami, & van Knippenberg, 2004). The SR system provides for motor empathy, underpins shared pain, and explains emotion contagion, as well as rudimentary recognition of intention and emotional states (Shamay-Tsoory, 2011).

Although SR processing dominates early development (Decety & Michalska, 2010), because automatic mirroring inevitably generates distress in response to others' distress, a further neural development is needed to supplement it, based on a cortical midline system consisting of the VMPFC and DMPFC, the TPJ and the medial temporal pole (Lieberman, 2007; Uddin et al., 2007). This system is less bodily based, and processes information about the self and others in more abstract and symbolic ways, as we have seen (Frith & Frith, 2006; Uddin et al., 2007). It also appears to be mainly shaped across development by interpersonal relationships, is phylogenetically initially found in primates, emerges fully in adolescence, and is neurochemically strongly linked to dopaminergic functioning (Lackner, Bowman, & Sabbagh, 2010). Behaviorally, this explicit MSA network underpins perspective-taking and both cognitive ToM (involving the DMPFC) and affective ToM (underpinned by the activity of the VMPFC). It is important to note that cognitive inference of affect is an act of imagination that is not the same as "feeling another's feelings," which SR processing entails (Gweon et al., 2012).

The SR and MSA systems may be mutually inhibitory (Brass & Haggard, 2008; Brass, Ruby, & Spengler, 2009; Brass, Schmitt, Spengler, & Gergely, 2007), which in our opinion further elucidates typical features of BPD and other types of psychopathology of the self. Neuroimaging studies indicate that the neural regions most often recruited in the inhibition of imitative behavior are those involved in explicit mental state attributions. Studies suggest that patients with BPD have (often serious) impairments in more controlled, explicit mentalizing (Fonagy & Luyten, in press), suggesting the existence of impairments in the MSA system. Hence, these individuals may experience an excessive and developmentally inappropriate activation of the SR system, leaving them with difficulties in decoupling their representation of another person's experience from their self-representations (Fonagy & Luyten, 2009; Ripoll et al., 2013). This leads to an overemphasis on others' feelings and emotions (*emotional contagion*) and, in turn, to confusion about the self (identity diffusion and feelings of inner emptiness), as well as an undue emphasis on externally oriented mentalizing ("jumping to conclusions") that is difficult to modulate.

Although these assumptions are still somewhat speculative, decreased activity has been noted in the STS of patients with BPD during deliberate empathic processing (Dziobek et al., 2011; Mier et al., 2013). Decreased activation of the MPFC was noted during the regulation of provoked aggression (New et al., 2009), which

may indicate an inadequate perspective-taking stance on the part of these patients. Similarly, a task calling for the use of psychological distancing from affective stimuli failed to demonstrate higher activation in the posterior cingulate cortex for BPD patients (Koenigsberg et al., 2009). Dysfunction may also be indicated by hyperactivation of the MPFC (leading to hypermentalizing) in response to an experience of rejection (Ruocco et al., 2010), an important finding, as BPD has been associated with both hypomentalizing and hypermentalizing, particularly in the context of tasks involving complex interpersonal relationships (Sharp et al., 2011). In turn, research findings support the assumption that BPD is associated with inappropriate activation of the SR system, perhaps in part because of a heightened response to stress, which inhibits the capacity for systematic mental state attributions (i.e., controlled mentalizing; see Figure 3). Amygdala hyperreactivity has been shown in individuals with BPD in response to both positive and negative stimuli (Hazlett et al., 2012; Mier et al., 2013; Minzenberg, Fan, New, Tang, & Siever, 2007), as well as in attempts by patients to regulate their response to negative social cues (Koenigsberg et al., 2009). Ripoll et al. (2013) cite unpublished data suggesting that the lack of habituation indicated by amygdala activity is associated with limitations in subjective perception of social support. The anterior insula has also been shown to be hyperreactive during affective empathy tasks. Dziobek et al. (2011) reported that during emotional empathy the right midinsula was more strongly activated in individuals with BPD than in nonclinical controls. This anatomical region is associated with bodily arousal (Simmons et al., 2013), suggesting that an emotional empathy task engages BPD patients more than controls. In support of this speculation, a positive association between right midinsula activation and skin conductance was shown for this patient group. King-Casas et al. (2008) reported that BPD patients' mistrustful reactions to fair offers in a multiround social exchange task with a partner were attributable to insula hyperreactivity, hindering more controlled reflective functioning about the intentions of their partner in the task.

Cognitive and Affective Features of Mentalizing

Full mentalizing involves the integration of cognition and affect, yet, again, both capacities can be relatively dissociated. The cognitive features of mentalizing include belief-desire reasoning and perspective-taking, and affective features include affective empathy and mentalized affectivity (Fonagy, Gergely, Jurist, & Target, 2002; Jurist, 2005). The former aspect is typically emphasized in ToM research, and seems to overlap to a large extent with more controlled mentalizing, whereas the latter is associated with affective empathy, and is largely automatic and embodied (Sabbagh, 2004).

As with the other three dimensions of mentalizing discussed earlier, there is increasing evidence that distinct, though somewhat more overlapping, neurocognitive systems are involved in these two capacities (Sabbagh, 2004; Sebastian et al., 2012; Shamay-Tsoory & Aharon-Peretz, 2007; Shamay-Tsoory, Aharon-Peretz, & Levkovitz, 2007). Whereas cognitively oriented mentalization depends on several areas in the prefrontal cortex, affectively oriented mentalizing seems particularly related to the VMPFC. This suggests that the VMPFC may play an important role in "marking" mental representations of self and others with affective information that can subsequently be integrated with cognitive knowledge such as belief-desire reasoning (Rochat & Striano, 1999). Again, a more automatic, embodied, and lateralized system is distinguished from a cortical midline structure that is more based on abstract and linguistic processing. This has also led to speculations about two possible systems underlying empathy; these are a more basic "emotional contagion" system and a more advanced cognitive perspective-taking system, as expressed in notable behavioral and anatomic dissociations between deficits in cognitive empathy associated with the VMPFC and emotional empathy associated with the inferior frontal gyrus (Shamay-Tsoory, Aharon-Peretz, & Perry, 2009).

Arousal, Attachment, and Mentalizing

A Developmental Psychobiological Perspective

Based on Arnsten, Mathew, Ubriani, Taylor, and Li (1999) and Mayes (2000, 2006), we have proposed a biobehavioral model which suggests that with increasing arousal there is a switch from controlled to automatic mentalizing (see Figure 3). As we have discussed above, this switch serves a clear evolutionary function: the emergence in situations of threat of a fight/flight/freeze response has clear survival value. Both noradrenergic and dopaminergic systems seem to be involved in this switch, which is hypothesized to protect the prefrontal cortex from excessive stimulation as well as facilitate coordination among the attentional, executive and sensory systems (Arnsten et al., 1999). For instance, norepinephrine enhances the activation of the prefrontal cortex, but α_1 postsynaptic receptor stimulation impairs functioning, which results in turning the prefrontal cortex "off-line" and facilitating subcortical functioning. Similarly, the D1 dopamine receptor family enhances prefrontal functioning, but under excessive catecholamine release (partly mediated by amygdala activation), D1 impairs functioning.

Yet, there are important differences between individuals in the switch point, as increasing stress activates not only a fight/flight/freeze response but also the attachment system, a behavioral system that modulates threat by prompting the individual to seek proximity to real or internalized attachment figures.

From a neurobiological perspective, increasing stress or arousal is thus associated with a complex set of coordinated responses involving (a) stress regulation systems involving the detection and processing of stress (i.e., involving the amygdala and hypothalamic–pituitary–adrenal (HPA) axis), (b) the mesocorticolimbic dopaminergic system, which has also been described as the brain's reward circuitry, which underlies attachment behavior (Champagne et al., 2004; Ferris et al., 2005; Insel & Young, 2001; Strathearn, Li, Fonagy, & Montague, 2008), and (c) neural circuits involved in mentalizing (Bartels & Zeki, 2000, 2004; Bull, Phillips, & Conway, 2008; Hurlemann, Hawellek, Maier, & Dolan, 2007; Lieberman, 2007; Mayes, 2000, 2006; Satpute & Lieberman, 2006).

Individuals' attachment history seems to be crucially important in understanding variations in these responses. Individual differences in the use and strength of attachment hyperactivation and deactivation strategies in response to stress in particular appear to determine three essential parameters in the switch from prefrontal to posterior cortical systems, or from controlled to automatic

mentalizing: (a) the threshold (intercept) at which the switch happens, (b) the strength or slope of the relationship between stress and the activation of neural circuits involved in controlled versus automatic mentalizing, and (c) the time to recovery from stress (Fonagy & Luyten, 2009; Fonagy & Luyten, in press; Luyten, Fonagy, Lowyck, et al., 2012). In the following sections, we discuss each of these factors and summarize the evidence that relates to them.

Secure Attachment Strategies, Arousal, and Mentalizing

In individuals who predominantly use *secure attachment strategies* in response to stress, the activation of the attachment system seems to foster controlled mentalizing, in combination with a relaxation of epistemic hypervigilance, leading to an effective down-regulation of stress and so-called "broaden and build" cycles (Fredrickson, 2001) that are typically associated with attachment experiences. Activation of the attachment system predictably seems to involve a relaxation of normal strategies of interpersonal caution. There is good evidence that intense activation of the neurobehavioral system underpinning attachment is associated with the deactivation of arousal and affect-regulation systems (Fonagy & Luyten, 2009; Luyten, Fonagy, Lowyck, et al., 2012), as well as the deactivation of neurocognitive systems likely to generate interpersonal suspicion—that is, those systems and brain regions involved in social cognition or mentalizing, including the LPFC, MPFC, LPAC, MPAC, MTL, and rACC (Bartels & Zeki, 2000, 2004; Lieberman, 2007; Satpute & Lieberman, 2006; Van Overwalle, 2009). For example, with increased intimacy, regions of the brain associated with reflective mentalizing will be deactivated.

Studies suggest that neuropeptides such as opioids, oxytocin, and vasopressin play an important role in this process. This role is both in activating the reward/attachment system and in deactivating the behavioral mechanisms involved in social avoidance and in attenuating both behavioral and endocrine stress responses (Heinrichs & Domes, 2008; Insel & Young, 2001; Panksepp & Watt, 2011). This explains, at the neurobiological level, the downregulation of arousal that is typically associated with secure attachment. Furthermore, oxytocin has been found to facilitate mentalizing in these individuals, as expressed in improvements in social memory, memory of facial expressions and identity, enhancements of the recognition of mental states based on facial expressions, probably by causing selective fixation on the eye region when viewing faces, and increasing trust (Bartz, Zaki, Bolger, & Ochsner, 2011; Neumann, 2008). Thus, the activation of the attachment system generates increased experience of reward, increased sensitivity to social cues, decreased stress levels, and decreased social avoidance, leading to "broaden and build" cycles (Fredrickson, 2001). These findings thus shed more light on the neurobiology of resilience (Fonagy, Steele, Steele, Higgitt, & Target, 1994): Individuals who predominantly use secure attachment strategies when faced with adversity have the ability to turn to (internalized) secure attachment figures in times of need, they find interpersonal contacts rewarding, and they have the capacity to keep controlled mentalizing "online" even when faced with considerable stress.

Yet, contextual factors should not be forgotten (Bartz, Zaki, et al., 2011), and mentalizing is not always consistently solid, even in predominantly securely attached individuals. For instance, studies clearly suggest that with increasing arousal, particularly in relation to out-group members, the likelihood of a switch to automatic mentalizing increases in everyone, even those who are securely attached (Bartz, Zaki, et al., 2011). In line with these findings, studies have reported that oxytocin administration leads to increased distrust, more bias in attributing intentions, and decreases in cooperative behavior with regard to out-group members even in normal community samples (Bartz, Zaki, et al., 2011). Hence, the increase in mentalizing and relaxation of interpersonal distrust and the fight/flight response associated with the use of secure attachment strategies is clearly limited to close attachment figures, or at best to a relatively small number of people who are seen as belonging to the in-group. Increasing stress may simply make attachment issues more salient, which may increase the likelihood of a deactivation of controlled mentalizing. This was also shown in a direct investigation of the neural phenomena underlying the switch model in community adults, which reported that exposure to idiosyncratic scripts eliciting attachment-related stress resulted in reduced controlled mentalizing-related activation in the left pSTS, left inferior frontal gyrus and left TPJ. Moreover, the left middle frontal gyrus and left anterior insula showed greater functional connectivity to the left pSTS after attachment stress (Nolte et al., 2013).

Attachment Hyperactivating Strategies

Individuals who primarily use *attachment hyperactivating strategies* (strategies that reflect desperate attempts to find security based on the conviction that others are not there to provide security and support, correlating with the anxious and preoccupied attachment styles) seem to be characterized by a relatively low threshold for switching to nonmentalizing modes, more extensive lapses in controlled mentalizing, and a relatively longer time to recovery compared to secure individuals. The threshold for deactivation of brain areas involved in controlled mentalizing seems to be relatively low, and more automatic, subcortical systems, including the amygdala, have a low threshold for responding to stress.

In these individuals, stress seems to readily activate the attachment system (seeking for protection), and attachment trauma may lead to chronic activation of the attachment system. In the situation where a child is seeking proximity to a traumatizing attachment figure (e.g., an abusive or neglectful parent) as a consequence of trauma, he or she is, naturally, likely to be further traumatized. Prolonged activation of the attachment system may create further difficulty resulting from increased emotional arousal. Many patients with BPD, for example, present with these features, which is unsurprising given the high prevalence of preoccupied and disorganized attachment as well as severe developmental trauma in this group (see Fonagy et al., 2015).

Attachment Deactivation Strategies

Individuals who primarily rely on *attachment deactivating strategies* (i.e., individuals with anxious-avoidant and dismissive attachment, which involve denying attachment needs and asserting one's own autonomy, independence, and strength in an attempt to

downregulate stress, based on the belief that others cannot provide support and comfort) tend to demonstrate fast deactivation of the attachment system and social information processing of threat cues. Attachment deactivating strategies have been shown to keep the neural systems involved in controlled mentalizing "online" for longer (Vrtička, Andersson, Grandjean, Sander, & Vuilleumier, 2008). Hence, these individuals often resemble those who predominantly use secure attachment strategies. Yet, this deactivating strategy is likely to fail under increasing stress. If securely attached individuals are those who are able to retain a relatively high activation of prefrontal areas in the presence of activation of the dopaminergic mesolimbic pathways (the attachment/reward system), then differences in mentalizing between securely attached individuals and individuals who primarily rely on attachment deactivating strategies may become apparent only under increasing stress—an assumption that is consistent with the findings of both experimental (Mikulincer & Shaver, 2007) and neuroimaging (Vrtička et al., 2008) studies.

Neurobiological Research on Arousal, Attachment, and Mentalizing

Neuroscience findings converge to suggest that attachment history is indeed crucial in understanding the relationship between arousal and mentalizing. Following Arnsten et al. (1999) and Mayes (2000), studies suggest that the threshold for switching from controlled to automatic mentalizing can be lowered as a result of exposure to early stress and attachment trauma. There is a close relationship between stress/arousal regulation through the HPA axis and the amygdala, a core structure within the neural circuits that subserve automatic mentalizing (see above), as is, for instance, evidenced in the high prevalence of corticotropin-releasing hormone (CRH)-expressing neurons and receptors in the amygdala (Tottenham & Sheridan, 2009). Early adversity has been shown to lead to kindling of the amygdala (Botterill et al., 2014), again supporting the role of the amygdala in potentiating fear and the stress response more generally. Research also clearly suggests the presence of both structural and functional changes in the amygdala in individuals with impairments in mentalizing, who typically have a history of early adversity. For instance, in BPD—a condition whose sufferers are commonly characterized by histories of high levels of early adversity—stress regulation, mediated by the HPA axis, is disturbed (Jogems-Kosterman, de Knijff, Kusters, & van Hoof, 2007; Nater et al., 2010; Scott, Levy, & Granger, 2013; Wingenfeld, Spitzer, Rullkotter, & Löwe, 2010). Functional MRI (fMRI) studies of BPD patients in which the background level of stress and/or attachment system activation was manipulated (e.g., Minzenberg et al., 2007) confirm the abnormal pattern of frontal deactivation and associated hyperresponsiveness of the limbic system in a range of contexts, using situational induced stress (Kraus et al., 2010) and in studies of the moderating influence of trait arousal (Holtmann et al., 2013) (see reviews by Mier et al., 2013; Salavert et al., 2011). For example, Silbersweig et al. (2007) reported that under conditions of negative emotion and behavioral inhibition, BPD patients showed relatively decreased VMPFC activity (including the medial orbitofrontal and subgenual ACC) and increased amygdalar–ventral striatal activity, correlating with decreased constraint and increased negative emotion, respectively. Furthermore, BPD patients with an explicit trauma history show a reduction in pituitary size (Garner et al., 2007), elevated levels of CRH in cerebrospinal fluid (Lee, Geracioti, Kasckow, & Coccaro, 2005), dysfunctions of cortisol responsivity (Jogems-Kosterman et al., 2007; Minzenberg et al., 2006; Walter et al., 2008), and disturbed dexamethasone suppression test responses (Wingenfeld et al., 2007). These dysfunctions cascade into other brain areas involved in automatic mentalizing. For instance, chronic stress has been shown to disrupt amygdala–VMPFC connectivity (Tottenham & Sheridan, 2009).

Although these studies require further replication because of several methodological limitations, including small sample sizes, disparate experimental paradigms, and considerable heterogeneity in sample selection (e.g., comorbidity with depression, childhood abuse, PTSD, and coping styles; Fertuck et al., 2006; Kahl et al., 2006), there is also more direct evidence concerning the neurobiological basis of the influence of individual differences in attachment on the relationship between arousal and mentalizing (Fonagy & Luyten, 2009; Fonagy & Luyten, in press). Research to date has provided considerable evidence that activation of the attachment system is closely linked to arousal and stress regulation (Heinrichs & Domes, 2008; Lieberman, 2007; Mayes, 2006). This might in fact reflect an adaptation strategy by which the individual attempts to prepare him/herself for future threat and adverse experiences (Tottenham & Sheridan, 2009). Individuals with a secure attachment history may show a relative relaxation of threat processing because of their repeated experiences of security. This enables the relaxation of interpersonal distrust and avoidance, which will foster the development of the capacity for controlled mentalizing, particularly in an environment that is conducive to the development of this capacity. In contrast, individuals with insecure attachment experiences seem to develop a hypersensitivity to threat in an attempt to deal with experiences of (perceived) insecurity and unpredictability of the availability and behavior of attachment figures. This may on the one hand lead to a pattern characteristically associated with attachment hyperactivation. As we have discussed above, hypersensitivity to threat, and the automatic processes it entails, is typically associated with an emphasis on externally focused mentalizing, to the neglect of more internally focused, controlled mentalizing. Although this is understandable as a "survival" strategy, the price these individuals pay is that they may increasingly hold biased and schematic assumptions about themselves and others, as well as being constantly hypervigilant toward others. We have recently linked this hypervigilance with problems with *epistemic trust*, that is, a lack of openness to others as a source of knowledge, which seriously impairs resilience and social learning more generally (Fonagy et al., in press). BPD might be a disorder that is characterized by this pattern. On the other hand, individuals with insecure attachment experiences may start to excessively deactivate the attachment system when confronted with stress because of the (perceived) unavailability of attachment figures. This strategy to adapt to circumstances characterized by repeated failures of attachment figures to coregulate stress seems to lead to an excessive emphasis on cognitive control, compulsive autonomy, and a general distrust of others. Others and relationships are simply not rewarding, and they are met with hypervigilance and distrust, which may also lead to hypermentalizing. Both insecure attachment strategies—although adaptive in the short term—are associated with high interpersonal and metabolic costs

because of the "wear and tear" of chronic hypervigilance and hyperactivity of the stress system (Fonagy & Luyten, 2009).

Strathearn and colleagues (Strathearn, Fonagy, Amico, & Montague, 2009; Strathearn et al., 2008), for instance, assessed the attachment security of 30 first-time mothers, using the Adult Attachment Interview, before the birth of their child. About 10 months after birth of their child, the same mothers viewed their own or other infants' smiling and crying faces while the mothers underwent fMRI scanning. Mothers with secure attachment showed greater activation of regions of the brain associated with reward, including the ventral striatum, and the oxytocin-associated hypothalamus/pituitary region. Peripheral oxytocin response during contact with their infant was also significantly higher in securely attached mothers, and the size of change from baseline oxytocin levels was positively correlated with brain activation to own infants in both brain regions. Importantly, securely attached mothers also showed greater activation in reward-processing regions when they viewed their own infants' sad faces, whereas insecure/dismissing mothers who predominantly used attachment deactivating strategies, in agreement with the findings described earlier, showed less activation of the reward system and greater insular activation in response to seeing their own infant's sad face. The insula may be a region associated with feelings of unfairness, pain, and disgust (see review by Montague & Lohrenz, 2007); we have discussed its role in the SR system, as a structure being involved in the automatic, immediate, embodied understanding (or misunderstanding) of others. Mothers with insecure/dismissing attachment histories thus 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. For securely attached mothers, infant cues, whether they were positive or negative in affect, seemed to act as an important affective signal of "incentive salience" (Berridge, 2007), reinforcing and motivating responsive maternal care. These mothers seemed to be "addicted" to their babies: viewing their babies was a rewarding experience. Insecure mothers, by contrast, particularly when viewing their infants' sad faces, showed a negative subjective reaction that would cause them to mirror their infant's sadness without being able to create a symbolic/mentalizing distance between their infant's and their own states of mind, thus illustrating the potential for an immediate misunderstanding of others by the SR system.

Vrtička et al. (2008) similarly found that avoidant attachment was related to a relative downregulation of reward-related activity, linked to the dopaminergic system, in striatal circuits during socially reinforcing interactions. Hence, reward responses associated with the attachment system were blunted in avoidant individuals. Yet, avoidant attachment was positively related to activation in the MPFC and the ventral ACC, areas that have been implicated in controlled mentalizing as well as social rejection and emotion suppression. Anxious attachment, in contrast, was associated with increased activation in the left amygdala in response to negative social feedback; as we have discussed, this is a brain area that is typically associated with automatic processes involved in fear and arousal more generally. Finally, secure attachment was not associated with any distinct neural responses but mirrored the pattern found for avoidant and anxious attachment. Secure attachment thus 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. Hence, in line with our assumptions, securely attached individuals simultaneously showed greater activation of the reward system in response to positive social reinforcement, and lower activation of the amygdala—and thus fear and arousal—in response to negative social feedback; they seemed to be able to relax their vigilance to threat.

These findings are also congruent with studies showing that early adverse attachment experiences are associated with decreased oxytocin levels and increased cortisol response (Fries, Hesse, Hellhammer, & Hellhammer, 2005; Heim, Newport, Mletzko, Miller, & Nemeroff, 2008; Meinlschmidt & Heim, 2007). Attachment hyperactivating and deactivating styles have also been related to polymorphisms in the oxytocin receptor gene in patients with unipolar depression (Costa et al., 2009). Similarly, a study reported dysregulated peripheral oxytocin release in depressed women (Cyranowski et al., 2008), and Gotlib and colleagues found that adolescent girls at risk for depression exhibited decreased activation in the reward-processing system (and specifically in striatal areas), suggesting a markedly reduced sensitivity to reward (Gotlib et al., 2010). As we have discussed in more detail elsewhere (Fonagy and Luyten, in press), low endogenous levels of oxytocin, polymorphisms in oxytocin-related genes, and negative effects of oxytocin administration have also been documented in individuals with BPD (Bartz, Simeon, et al., 2011; Bertsch, Schmidinger, Neumann, & Herpertz, 2013; Cyranowski et al., 2008; Stanley & Siever, 2010).

Conclusions and Directions for Future Research

Over the past decades, our knowledge of the neurobiology underlying mentalizing has greatly increased. Perhaps somewhat paradoxically, this body of knowledge has also helped us to better understand the *psychological experiences* that are associated with mentalizing and that mentalizing gives rise to. The present review shows how, at the very least, neurobiology puts limits to psychological explanations, and rules out some views about the nature of mentalizing and impairments in this capacity as improbable, while rendering other assumptions more plausible.

However, more research is needed, and rapid advances in the neurosciences are likely to lead to considerable changes in our assumptions about mentalizing in both normal and disrupted development. With increasing technological innovations, more sophisticated views will emerge. Currently, studies in this field still suffer from many limitations, including small sample sizes, the use of relatively simple paradigms, and lack of consideration of individual differences (such as individual differences in attachment style or temperament) and contextual factors. For instance, attachment clearly is a dimensional construct, and thus individuals tend to rely on different attachment strategies to a greater or lesser extent, for example, depending on contextual factors. This necessarily complicates the interpretation of findings. As in psychological research, heterogeneity complicates neuroscience studies, particularly studies with small sample sizes. The translation of these findings to clinical samples remains to be determined. For instance, it is becoming clearer that there are qualitative differences between individuals with insecure but organized attachment strategies, and those with more disorganized attachment, as is often the case in BPD patients and in many patients with a history of attachment trauma (Main, 1991). Both functional and structural

differences have been identified between healthy controls and patients in many brain areas, including, as noted, with regard to the neural circuits involved in mentalizing and attachment (Vrticka & Vuilleumier, 2012).

Studies in larger samples using more ecologically valid and perhaps personalized paradigms, together with novel imaging methods such as brain connectivity studies, are likely to yield much more insight into the neurobiology of mentalizing. Furthermore, the field of neuroscience is plagued by a lack of a unifying theory, even in the field of social cognition/mentalizing, leading different authors to emphasize different aspects of similar neural circuits. It is clear that the field is not ready yet for such a unifying theory, which should humble anyone engaged in developing psychological theories about such a fundamentally human—and complex—capacity as mentalizing. The views expressed in this article can therefore be seen only as an approximation that will undergo major changes in the future.

References

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

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

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14,* 110–118. http://dx.doi.org/10.1016/j.tics.2009.12.006

Bartels, A., & Zeki, S. (2000). The neural basis of romantic love. *NeuroReport, 11,* 3829–3834. http://dx.doi.org/10.1097/00001756-200011270-00046

Bartels, A., & Zeki, S. (2004). The neural correlates of maternal and romantic love. *NeuroImage, 21,* 1155–1166. http://dx.doi.org/10.1016/j.neuroimage.2003.11.003

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

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

Beebe, B., Badalamenti, A., Jaffe, J., Feldstein, S., Marquette, L., Helbraun, E., . . . Ellman, L. (2008). Distressed mothers and their infants use a less efficient timing mechanism in creating expectancies of each other's looking patterns. *Journal of Psycholinguistic Research, 37,* 293–307. http://dx.doi.org/10.1007/s10936-008-9078-y

Beebe, B., Jaffe, J., Buck, K., Chen, H., Cohen, P., Blatt, S., . . . Andrews, H. (2007). Six-week postpartum maternal self-criticism and dependency and 4-month mother-infant self- and interactive contingencies. *Developmental Psychology, 43,* 1360–1376. http://dx.doi.org/10.1037/0012-1649.43.6.1360

Beeghly, M., & Cicchetti, D. (1994). Child maltreatment, attachment, and the self system: Emergence of an internal state lexicon in toddlers at high social risk. *Development and Psychopathology, 6,* 5–30. http://dx.doi.org/10.1017/S095457940000585X

Beer, J. S., John, O. P., Scabini, D., & Knight, R. T. (2006). Orbitofrontal cortex and social behavior: Integrating self-monitoring and emotion-cognition interactions. *Journal of Cognitive Neuroscience, 18,* 871–879. http://dx.doi.org/10.1162/jocn.2006.18.6.871

Bernhardt, B. C., & Singer, T. (2012). The neural basis of empathy. *Annual Review of Neuroscience, 35,* 1–23. http://dx.doi.org/10.1146/annurev-neuro-062111-150536

Berridge, K. C. (2007). The debate over dopamine's role in reward: The case for incentive salience. *Psychopharmacology, 191,* 391–431. http://dx.doi.org/10.1007/s00213-006-0578-x

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

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

Brass, M., & Haggard, P. (2008). The what, when, whether model of intentional action. *The Neuroscientist, 14,* 319–325. http://dx.doi.org/10.1177/1073858408317417

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

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

Bull, R., Phillips, L. H., & Conway, C. A. (2008). The role of control functions in mentalizing: Dual-task studies of theory of mind and executive function. *Cognition, 107,* 663–672. http://dx.doi.org/10.1016/j.cognition.2007.07.015

Carpendale, J., & Lewis, C. (2006). *How children develop social understanding*. Oxford, UK: Blackwell.

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

Choi-Kain, L. W., & Gunderson, J. G. (2008). Mentalization: Ontogeny, assessment, and application in the treatment of borderline personality disorder. *The American Journal of Psychiatry, 165,* 1127–1135. http://dx.doi.org/10.1176/appi.ajp.2008.07081360

Chung, Y. S., Barch, D., & Strube, M. (2014). A meta-analysis of mentalizing impairments in adults with schizophrenia and autism spectrum disorder. *Schizophrenia Bulletin, 40,* 602–616. http://dx.doi.org/10.1093/schbul/sbt048

Costa, B., Pini, S., Gabelloni, P., Abelli, M., Lari, L., Cardini, A., . . . Martini, C. (2009). Oxytocin receptor polymorphisms and adult attachment style in patients with depression. *Psychoneuroendocrinology, 34,* 1506–1514. http://dx.doi.org/10.1016/j.psyneuen.2009.05.006

Cusi, A. M., Nazarov, A., Holshausen, K., Macqueen, G. M., & McKinnon, M. C. (2012). Systematic review of the neural basis of social cognition in patients with mood disorders. *Journal of Psychiatry & Neuroscience, 37,* 154–169. http://dx.doi.org/10.1503/jpn.100179

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

Decety, J., & Michalska, K. J. (2010). Neurodevelopmental changes in the circuits underlying empathy and sympathy from childhood to adulthood. *Developmental Science, 13,* 886–899. http://dx.doi.org/10.1111/j.1467-7687.2009.00940.x

Dimaggio, G., Lysaker, P. H., Carcione, A., Nicolò, G., & Semerari, A. (2008). Know yourself and you shall know the other . . . to a certain extent: Multiple paths of influence of self-reflection on mindreading. *Consciousness and Cognition, 17,* 778–789. http://dx.doi.org/10.1016/j.concog.2008.02.005

Dunbar, R. I. M. (2008). Mind the gap: Or why humans aren't just great apes. *Proceedings of the British Academy, 154,* 403–423.

Dziobek, I., Preissler, S., Grozdanovic, Z., Heuser, I., Heekeren, H. R., & Roepke, S. (2011). Neuronal correlates of altered empathy and social cognition in borderline personality disorder. *NeuroImage, 57,* 539–548. http://dx.doi.org/10.1016/j.neuroimage.2011.05.005

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

Fertuck, E. A., Marsano-Jozefowicz, S., Stanley, B., Tryon, W. W., Oquendo, M., Mann, J. J., & Keilp, J. G. (2006). The impact of borderline personality disorder and anxiety on neuropsychological performance in major depression. *Journal of Personality Disorders, 20,* 55–70. http://dx.doi.org/10.1521/pedi.2006.20.1.55

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

Fonagy, P., & Luyten, P. (2009). A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Development and Psychopathology, 21,* 1355–1381. http://dx.doi.org/10.1017/S0954579409990198

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: Wiley.

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

Fonagy, P., Luyten, P., & Bateman, A. (2015). Translation: Mentalizing as treatment target in borderline personality disorder. *Personality Disorders: Theory, Research, and Treatment, 6,* 380–392. http://dx.doi.org/10.1037/per0000113

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* (pp. 37–87). Washington, DC: American Psychiatric Publishing.

Fonagy, P., Steele, M., Steele, H., Higgitt, A., & Target, M. (1994). The Emanuel Miller Memorial Lecture 1992. The theory and practice of resilience. *Child Psychology & Psychiatry & Allied Disciplines, 35,* 231–257. http://dx.doi.org/10.1111/j.1469-7610.1994.tb01160.x

Fonagy, P., & Target, M. (1997). Attachment and reflective function: Their role in self-organization. *Development and Psychopathology, 9,* 679–700. http://dx.doi.org/10.1017/S0954579497001399

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology. The broaden-and-build theory of positive emotions. *American Psychologist, 56,* 218–226. http://dx.doi.org/10.1037/0003-066X.56.3.218

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

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

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

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

Gotlib, I. H., Hamilton, J. P., Cooney, R. E., Singh, M. K., Henry, M. L., & Joormann, J. (2010). Neural processing of reward and loss in girls at risk for major depression. *Archives of General Psychiatry, 67,* 380–387. http://dx.doi.org/10.1001/archgenpsychiatry.2010.13

Gweon, H., Dodell-Feder, D., Bedny, M., & Saxe, R. (2012). Theory of mind performance in children correlates with functional specialization of a brain region for thinking about thoughts. *Child Development, 83,* 1853–1868. http://dx.doi.org/10.1111/j.1467-8624.2012.01829.x

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. http://dx.doi.org/10.1016/j.biopsych.2012.03.027

Heim, C., Newport, D. J., Mletzko, T., Miller, A. H., & Nemeroff, C. B. (2008). The link between childhood trauma and depression: Insights from HPA axis studies in humans. *Psychoneuroendocrinology, 33,* 693–710. http://dx.doi.org/10.1016/j.psyneuen.2008.03.008

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

Herpertz, S. C., Jeung, H., Mancke, F., & Bertsch, K. (2014). Social dysfunctioning and brain in borderline personality disorder. *Psychopathology, 47,* 417–424. http://dx.doi.org/10.1159/000365106

Holtmann, J., Herbort, M. C., Wüstenberg, T., Soch, J., Richter, S., Walter, H., . . . Schott, B. H. (2013). Trait anxiety modulates fronto-limbic processing of emotional interference in borderline personality disorder. *Frontiers in Human Neuroscience, 7,* 54. http://dx.doi.org/10.3389/fnhum.2013.00054

Hurlemann, R., Hawellek, B., Maier, W., & Dolan, R. J. (2007). Enhanced emotion-induced amnesia in borderline personality disorder. *Psychological Medicine, 37,* 971–981. http://dx.doi.org/10.1017/S0033291706009792

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

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

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

Kahl, K. G., Bens, S., Ziegler, K., Rudolf, S., Dibbelt, L., Kordon, A., & Schweiger, U. (2006). Cortisol, the cortisol-dehydroepiandrosterone ratio, and pro-inflammatory cytokines in patients with current major depressive disorder comorbid with borderline personality disorder. *Biological Psychiatry, 59,* 667–671. http://dx.doi.org/10.1016/j.biopsych.2005.08.001

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., & Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science, 321,* 806–810. http://dx.doi.org/10.1126/science.1156902

Knutson, K. M., Mah, L., Manly, C. F., & Grafman, J. (2007). Neural correlates of automatic beliefs about gender and race. *Human Brain Mapping, 28,* 915–930. http://dx.doi.org/10.1002/hbm.20320

Koenigsberg, H. W., Fan, J., Ochsner, K. N., Liu, X., Guise, K. G., Pizzarello, S., . . . Siever, L. J. (2009). Neural correlates of the use of psychological distancing to regulate responses to negative social cues: A study of patients with borderline personality disorder. *Biological Psychiatry, 66,* 854–863. http://dx.doi.org/10.1016/j.biopsych.2009.06.010

Kovács, A. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science, 330,* 1830–1834. http://dx.doi.org/10.1126/science.1190792

Kraus, A., Valerius, G., Seifritz, E., Ruf, M., Bremner, J. D., Bohus, M., & Schmahl, C. (2010). Script-driven imagery of self-injurious behavior in patients with borderline personality disorder: A pilot FMRI study. *Acta Psychiatrica Scandinavica, 121,* 41–51. http://dx.doi.org/10.1111/j.1600-0447.2009.01417.x

Kuipers, G. S., & Bekker, M. (2012). Attachment, Mentalization and Eating Disorders: A review of studies using the Adult Attachment Interview. *Current Psychiatry Reviews, 8,* 326–336. http://dx.doi.org/10.2174/157340012803520478

Lackner, C. L., Bowman, L. C., & Sabbagh, M. A. (2010). Dopaminergic functioning and preschoolers' theory of mind. *Neuropsychologia, 48,* 1767–1774. http://dx.doi.org/10.1016/j.neuropsychologia.2010.02.027

Ladegaard, N., Larsen, E. R., Videbech, P., & Lysaker, P. H. (2014). Higher-order social cognition in first-episode major depression. *Psychiatry Research, 216,* 37–43. http://dx.doi.org/10.1016/j.psychres.2013.12.010

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

Lewis, P. A., Rezaie, R., Brown, R., Roberts, N., & Dunbar, R. I. M. (2011). Ventromedial prefrontal volume predicts understanding of others and social network size. *NeuroImage, 57,* 1624–1629. http://dx.doi.org/10.1016/j.neuroimage.2011.05.030

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

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

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

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

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

Main, M. (1991). Metacognitive knowledge, metacognitive monitoring, and singular (coherent) vs. multiple (incoherent) model of attachment: Findings and directions for future research. In C. M. Parkes, J. Stevenson-Hinde, & P. Marris (Eds.), *Attachment across the life cycle* (pp. 127–159). London, UK: Tavistock/Routledge.

Mayes, L. C. (2000). A developmental perspective on the regulation of arousal states. *Seminars in Perinatology, 24,* 267–279. http://dx.doi.org/10.1053/sper.2000.9121

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

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

Mier, D., Lis, S., Esslinger, C., Sauer, C., Hagenhoff, M., Ulferts, J., . . . Kirsch, P. (2013). Neuronal correlates of social cognition in borderline personality disorder. *Social Cognitive and Affective Neuroscience, 8,* 531–537. http://dx.doi.org/10.1093/scan/nss028

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

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: Neuroimaging, 155,* 231–243. http://dx.doi.org/10.1016/j.pscychresns.2007.03.006

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

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

Nater, U. M., Bohus, M., Abbruzzese, E., Ditzen, B., Gaab, J., Kleindienst, N., . . . Ehlert, U. (2010). Increased psychological and attenuated cortisol and alpha-amylase responses to acute psychosocial stress in female patients with borderline personality disorder. *Psychoneuroendocrinology, 35,* 1565–1572. http://dx.doi.org/10.1016/j.psyneuen.2010.06.002

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

New, A. S., Hazlett, E. A., Newmark, R. E., Zhang, J., Triebwasser, J., Meyerson, D., . . . Buchsbaum, M. S. (2009). Laboratory induced aggression: A positron emission tomography study of aggressive individuals with borderline personality disorder. *Biological Psychiatry, 66,* 1107–1114. http://dx.doi.org/10.1016/j.biopsych.2009.07.015

Nolte, T., Bolling, D. Z., Hudac, C. M., Fonagy, P., Mayes, L., & Pelphrey, K. A. (2013). Brain mechanisms underlying the impact of attachment-related stress on social cognition. *Frontiers in Human Neuroscience, 7,* 816. http://dx.doi.org/10.3389/fnhum.2013.00816

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

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

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

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

Ruocco, A. C., Medaglia, J. D., Tinker, J. R., Ayaz, H., Forman, E. M., Newman, C. F., . . . Chute, D. L. (2010). Medial prefrontal cortex hyperactivation during social exclusion in borderline personality disorder. *Psychiatry Research: Neuroimaging, 181,* 233–236. http://dx.doi.org/10.1016/j.pscychresns.2009.12.001

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

Salavert, J., Gasol, M., Vieta, E., Cervantes, A., Trampal, C., & Gispert, J. D. (2011). Fronto-limbic dysfunction in borderline personality disorder: A 18F-FDG positron emission tomography study. *Journal of Affective Disorders, 131,* 260–267. http://dx.doi.org/10.1016/j.jad.2011.01.001

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

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews, 42,* 9–34. http://dx.doi.org/10.1016/j.neubiorev.2014.01.009

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

Sebastian, C. L., Fontaine, N. M. G., Bird, G., Blakemore, S-J., Brito, S. A., McCrory, E. J. P., & Viding, E. (2012). Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Social Cognitive and Affective Neuroscience, 7,* 53–63. http://dx.doi.org/10.1093/scan/nsr023

Seyfarth, R. M., & Cheney, D. L. (2013). Affiliation, empathy, and the origins of theory of mind. *Proceedings of the National Academy of Sciences of the United States of America, 110*(Suppl 2), 10349–10356. http://dx.doi.org/10.1073/pnas.1301223110

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

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

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

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

Sharp, C., & Fonagy, P. (2008). The parent's capacity to treat the child as a psychological agent: Constructs, measures and implications for developmental psychopathology. *Social Development, 17,* 737–754. http://dx.doi.org/10.1111/j.1467-9507.2007.00457.x

Sharp, C., Pane, H., Ha, C., Venta, A., Patel, A. B., Sturek, J., & Fonagy, P. (2011). Theory of mind and emotion regulation difficulties in adolescents with borderline traits. *Journal of the American Academy of Child & Adolescent Psychiatry, 50,* 563–573.e1. http://dx.doi.org/10.1016/j.jaac.2011.01.017

Silbersweig, D., Clarkin, J. F., Goldstein, M., Kernberg, O. F., Tuescher, O., Levy, K. N., . . . Stern, E. (2007). Failure of frontolimbic inhibitory function in the context of negative emotion in borderline personality disorder. *The American Journal of Psychiatry, 164,* 1832–1841. http://dx.doi.org/10.1176/appi.ajp.2007.06010126

Simmons, W. K., Avery, J. A., Barcalow, J. C., Bodurka, J., Drevets, W. C., & Bellgowan, P. (2013). Keeping the body in mind: Insula functional organization and functional connectivity integrate interoceptive, exteroceptive, and emotional awareness. *Human Brain Mapping, 34,* 2944–2958. http://dx.doi.org/10.1002/hbm.22113

Skårderud, F. (2007). Eating one's words: Part III. Mentalisation-based psychotherapy for anorexia nervosa—An outline for a treatment and training manual. *European Eating Disorders Review, 15,* 323–339. http://dx.doi.org/10.1002/erv.817

Slade, A. (2005). Parental reflective functioning: An introduction. *Attachment & Human Development, 7,* 269–281. http://dx.doi.org/10.1080/14616730500245906

Sleed, M., & Fonagy, P. (2010). Understanding disruptions in the parent-infant relationship: Do actions speak louder than words? In T. Baradon (Ed.), *Relational trauma in infancy*. London, UK: Routledge.

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

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

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

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

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

van Baaren, R. B., Holland, R. W., Kawakami, K., & van Knippenberg, A. (2004). Mimicry and prosocial behavior. *Psychological Science, 15,* 71–74. http://dx.doi.org/10.1111/j.0963-7214.2004.01501012.x

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

Van Overwalle, F. (2011). A dissociation between social mentalizing and general reasoning. *NeuroImage, 54,* 1589–1599. http://dx.doi.org/10.1016/j.neuroimage.2010.09.043

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

Vrtička, P., Andersson, F., Grandjean, D., Sander, D., & Vuilleumier, P. (2008). Individual attachment style modulates human amygdala and striatum activation during social appraisal. *PLoS ONE, 3,* e2868. http://dx.doi.org/10.1371/journal.pone.0002868

Vrtička, P., & Vuilleumier, P. (2012). Neuroscience of human social interactions and adult attachment style. *Frontiers in Human Neuroscience, 6,* 212. http://dx.doi.org/10.3389/fnhum.2012.00212

Walter, M., Bureau, J. F., Holmes, B. M., Bertha, E. A., Hollander, M., Wheelis, J., . . . Lyons-Ruth, K. (2008). Cortisol response to interpersonal stress in young adults with borderline personality disorder: A pilot study. *European Psychiatry, 23,* 201–204. http://dx.doi.org/10.1016/j.eurpsy.2007.12.003

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

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