ANNUAL REVIEWS

Annual Review of Clinical Psychology

The Mentalizing Approach to Psychopathology: State of the Art and Future Directions

Patrick Luyten,[1,2] Chloe Campbell,[2] Elizabeth Allison,[2] and Peter Fonagy[2]

[1]Faculty of Psychology and Educational Sciences, KU Leuven, B-3000 Leuven, Belgium; email: patrick.luyten@kuleuven.be

[2]Research Department of Clinical, Educational and Health Psychology, University College London, London WC1E 7HB, United Kingdom

Annu. Rev. Clin. Psychol. 2020. 16:297–325

First published as a Review in Advance on February 5, 2020

The *Annual Review of Clinical Psychology* is online at clinpsy.annualreviews.org

https://doi.org/10.1146/annurev-clinpsy-071919-015355

Keywords

mentalizing, reflective functioning, attachment, mental health treatment, personality disorder, psychological treatments

Abstract

Mentalizing is the capacity to understand others and oneself in terms of internal mental states. It is assumed to be underpinned by four dimensions: automatic–controlled, internally–externally focused, self–other, and cognitive–affective. Research suggests that mental disorders are associated with different imbalances in these dimensions. Addressing the quality of mentalizing as part of psychosocial treatments may benefit individuals with various mental disorders. We suggest that mentalizing is a helpful transtheoretical and transdiagnostic concept to explain vulnerability to psychopathology and its treatment. This review summarizes the mentalizing approach to psychopathology from a developmental socioecological evolutionary perspective. We then focus on the application of the mentalizing approach to personality disorders, and we review studies that have extended this approach to other types of psychopathology, including depression, anxiety, and eating disorders. We summarize core principles of mentalization-based treatments and preventive interventions and the evidence for their effectiveness. We conclude with recommendations for future research.

Contents

INTRODUCTION

Research on mentalizing and its role in psychopathology is burgeoning. A search in the PsycINFO database, for instance, shows a huge increase in the number of studies on mentalizing, from only a handful of studies on the topic having been published by the end of the 1990s to more than 3,000 in 2019. Mentalizing (or reflective functioning) refers to the quintessential human capacity to understand the self and others in terms of intentional mental states, such as feelings, desires, wishes, attitudes, and goals. It is a fundamental capacity that enables people to navigate the complex social world they live in. Moreover, it is a species-specific capacity: It appears to be present only in humans and, in a rudimentary form, in our nearest primate relatives, and it is absent in most other animal species (Tomasello 2010, 2018). Without the capacity to mentalize, we would be lost in a world that demands ever-increasing flexibility to allow adaptation to ever more complex environments (Sng et al. 2018) determined by ever-changing interpersonal relationships that require a high degree of collaboration and cooperation based on mutual understanding (Fonagy et al. 2015, Tomasello & Vaish 2013).

In this article, we critically review the four main assumptions of the mentalizing approach to psychopathology:

1. Neuroscience findings strongly suggest that mentalizing is an evolutionarily prewired capacity: Normatively developing children typically show joint attention and shared intentionality—capacities that reflect mentalizing—from the beginning of life (Csibra & Gergely 2009, Tomasello & Vaish 2013).

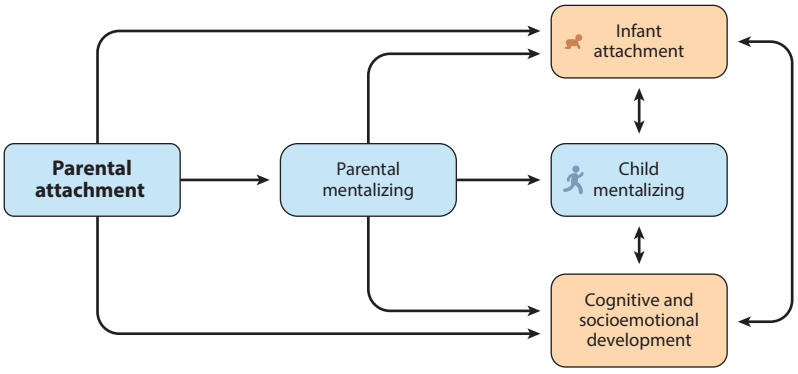

Figure 1

Simplified model of the initial mentalizing approach to psychological development. Parental attachment is assumed to influence infant attachment and socioemotional development more generally both directly and indirectly. Parents with high levels of attachment security provide a secure base to their children, which allows them to gradually explore and develop cognitive and socioemotional capacities that increasingly allow them to successfully navigate their interpersonal world. More indirectly, secure parental attachment is thought to foster parental mentalizing—that is, the capacity to see one's child as motivated by internal mental states. This latter capacity is thought to foster infant attachment and socioemotional development primarily through its influence on the development of child mentalizing, as high levels of parental mentalizing typically create a context for children in which they are able to learn to reflect on the inevitable challenges that are associated with psychological development, leading to resilience in the face of adversity. Adverse experiences, biological predisposition, or a combination of both may disrupt these virtuous cycles.

2. Developmental research suggests that considerable environmental input is needed to develop a fully balanced capacity for mentalizing. In this context, there has been a notable shift in our thinking in recent years (Fonagy et al. 2017a, 2017b). Whereas earlier formulations (Fonagy 2000, Fonagy et al. 1991b) focused on the unique role of dyadic attachment in fostering or hindering the development of mentalizing (see **Figure 1**), our current views have evolved to a more comprehensive set of considerations concerning the role of family, peers, and broader sociocultural factors in the development of mentalizing. In this context, we now stress the role of epistemic trust, the evolutionarily prewired capacity to trust others as sources of social information, both facilitated by and facilitating mentalizing, which in turn fosters resilience to adversity through a health-generating (salutogenic) process (see **Figure 2**) of social learning and deriving maximal benefit from the stream of relevant information accessible through the social environment (Fonagy et al. 2017a, 2017b).
3. Mentalizing is a transdiagnostic and transtheoretical concept that is (unsurprisingly, given its core role in species-specific adaptation) implicated in a wide range of psychological problems and disorders. Initially, research in this area focused on the investigation of imbalances in mentalizing in patients with severe psychopathology—specifically, borderline personality disorder (BPD). With the expanding scope of the mentalizing approach came the growing realization that if mentalizing is a central human capacity, most—if not all—forms of psychopathology can be expected to be characterized by temporary or chronic impairments and disruptions in this capacity.
4. Similarly, mentalizing may be commonly found as a factor associated with recovery in a range of psychotherapies, and interventions that may not explicitly focus on improving mentalizing nevertheless may be effective in reducing mental health problems as they may foster mentalizing and salutogenesis in particular through different routes.

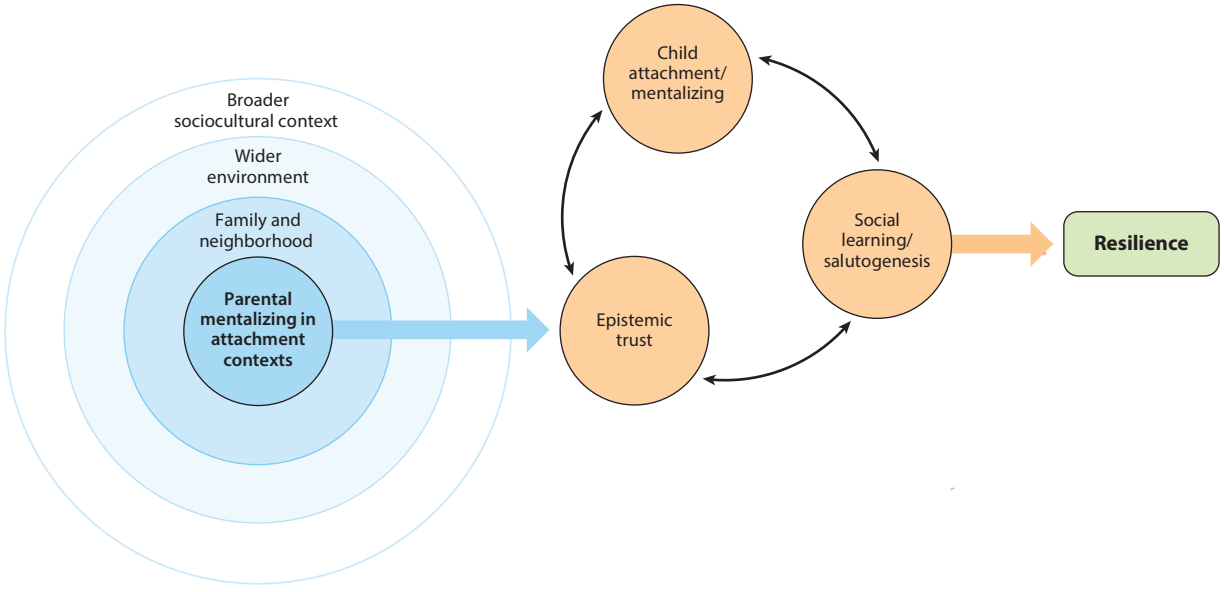

Figure 2

Social-evolutionary communicative model of the role of mentalizing in development. The capacity for parental mentalizing—that is, the capacity to understand one's child as motivated by internal mental states—is determined by and embedded within a broader set of factors influencing child development, including family and neighborhood, the wider environment, and the general sociocultural context. These factors influence the evolutionarily prewired capacity to identify knowledge conveyed by others as personally relevant and generalizable to other contexts (i.e., epistemic trust), which sets in motion a virtuous cycle characterized by social learning and salutogenesis (the capacity to benefit from positive influences in one's environment), as well as secure attachment and solid mentalizing. The use of secure attachment strategies when faced with adversity and the child's own capacity for mentalizing allow the child to recalibrate thoughts and feelings when faced with challenging adverse circumstances either him/herself and/or in interaction with others through co-regulation, thus fostering resilience.

We critically review evidence for these assumptions, identify important gaps in our knowledge, outline avenues for future research, and discuss implications for clinical practice. Wherever possible, we rely on meta-analytic findings and qualitative reviews.

NEUROBIOLOGY OF MENTALIZING

Neuroscientific studies have been of crucial importance in the realization that mentalizing provides a powerful lens through which to study human behavior. Findings from neuroscience have been particularly important in demonstrating that (*a*) a specific set of highly specialized, species-specific neural circuits is involved in mentalizing; (*b*) mentalizing is not a unitary but, rather, a multidimensional capacity, with possible imbalances and dissociations between the different dimensions underlying mentalizing being characteristic of different mental disorders; and (*c*) mentalizing is most helpfully considered as an umbrella concept that overlaps with a range of constructs and capacities, such as Theory of Mind (ToM), mindfulness, perspective taking, and empathy.

Mentalizing Is Evolutionarily Prewired in Humans

Neurobiological research is increasingly converging to suggest that mentalizing is an evolutionarily prewired, specifies-specific capacity of humans. Even in the first months of life, infants show

joint attention and joint intentionality—capacities that are present in only an elementary form in great apes and absent in most other species (Tomasello 2018, Tomasello & Vaish 2013). Moreover, from age 3 years onward, children develop the capacity for collective intentionality—that is, the ability to function in a group based on shared principles, norms, and conventions (Tomasello 2018). Hence, humans are essentially social creatures. The reasons why the capacity for mentalizing evolved are largely unknown, but the ability seems to have provided humans with a major adaptive advantage. The capacity of individuals to envision their own and others' goals, feelings, and wishes enabled complex collaboration and cooperation not seen elsewhere in the animal kingdom and permitted the transmission of shared goals, motives, morals, and knowledge across generations through social learning (Sng et al. 2018). Indeed, because of their capacity for shared intentionality, humans are equipped to acquire knowledge shared by a community (culture) very rapidly. The emergence of shared intentionality, underpinned by the capacity for mentalizing, is believed to facilitate meaningful social interactions within our species that exceed those of nonhuman primates. In addition, this propensity toward collaboration and cooperation seems to have mitigated competitive—and as a result often aggressive and violent—behavior and motives in humans, based on the principle of overriding or unlearning such impulses that would hinder collaboration and cooperation within a community (Fonagy & Luyten 2018). Yet, at the same time, the capacity to mentalize enables some humans to use other techniques to compete with others, including manipulation and deceit (Fonagy & Luyten 2018).

Mentalizing Is Multidimensional

Neuroscientific and behavioral studies suggest that mentalizing can be organized around four dimensions or polarities, with each pole subserved by relatively distinct underlying neural circuits. The four polarities are (*a*) automatic versus controlled mentalizing, (*b*) mentalizing with regard to the self and about others, (*c*) mentalizing based on external or internal features of the self and others, and (*d*) cognitive versus affective mentalizing. As discussed in the section titled Mentalizing and Personality Disorders, this view leads to the assumption that different types of psychopathology reflect different imbalances in the dimensions, resulting in different mentalizing profiles that are characteristic of each disorder.

Automatic mentalizing involves fast, parallel, and reflexive processes that require little effort. It is subserved by phylogenetically older neural circuits that rely primarily on sensory information. Controlled or explicit mentalizing, by contrast, is conscious, verbal, and reflective. Phylogenetically newer brain circuits that rely on linguistic/symbolic processing underlie this capacity, consistent with the notion that controlled mentalizing enabled the evolutionary leap of humans' capacity for mutuality, collaboration, and cooperation.

Automatic and controlled mentalizing play a key role in stress and emotion regulation. With increasing stress or arousal, there is a switch from neural systems associated with controlled mentalizing to those associated with automatic mentalizing (Lieberman 2007, Mayes 2006). The evolutionary advantage of this switch is clear: The fight/flight response that arises when faced with threat relies on fast, and thus automatic, processing of threatening information. However, in our complex interpersonal world, such automatic responses are often problematic; indeed, the many challenges that exist within the social world often require considerable "computational power" to develop models of the mind of others and the self (Fonagy et al. 2015). Overreliance on automatic mentalizing typically implies nonreflective—and therefore often overly simplistic and biased—assumptions about self and other. For instance, studies have amply demonstrated the rapid activation of biased views toward people of another ethnicity (Knutson et al. 2007). The complex demands for communication, collaboration, and competition that are typical of the interpersonal

world are particularly problematic for individuals with a low "switch point" from automatic to controlled mentalizing under stress or emotional arousal. Both biological (i.e., low levels of effortful control) and environmental (i.e., attachment history) factors, and their interactions, are known to contribute to interindividual differences in the capacity for controlled mentalizing (for a comprehensive review, see Long et al. 2020).

Reflecting on the self or others may be either internally focused (i.e., it may concern inferring mental states by perspective taking or contextual imputations) or focused on external cues, such as posture, facial expression, and tone of voice. Externally based mentalizing tends to recruit a lateral frontotemporoparietal network that involves less reflective processes. Internally based mentalizing relies more on a medial frontoparietal network that involves more active and controlled reflective processes (Lieberman 2007).

Consistent with the notion that mentalizing developed as an evolutionary adaptation to a growing need for mutuality and cooperation, a common neural circuitry underpins the capacity to reflect on the self and others. Moreover, two types of knowing the self and others—a basic and a more advanced way—seem to be distinguishable. Ripoll et al. (2013) distinguished between a shared representation (SR) system, in which empathic processing relies on shared representations of others' mental states, and a mental state attribution (MSA) system, which relies more on symbolic and abstract processing. The SR system involves a more implicit, visceral, bodily based, frontoparietal mirror neuron system without the need for high-level cognitive processing, based on a similarity of neural activation while experiencing, and observing others experiencing, states of mind. This system allows individuals to understand others through motor-simulation mechanisms. It is thought to be one of the key evolutionary mechanisms responsible for social empathy in humans and other mammals; it allows us to know how another feels from the inside. The SR system recruits the inferior frontal gyrus and inferior parietal lobule (both of which are rich in mirror neurons) and the anterior insula and anterior cingulate cortex (both of which are involved in observed and felt pain).

The MSA system is less bodily based; it processes information about the self and others in more abstract and symbolic ways. It is mainly shaped by interpersonal relationships. It has also been found in primates, and in humans it fully develops only in adolescence (Lackner et al. 2010). It involves a cortical midline system consisting of the ventromedial prefrontal cortex (VMPFC), dorsomedial prefrontal cortex, temporoparietal junction, and medial temporal pole (Lieberman 2007, Uddin et al. 2007). Studies suggest that the SR and MSA systems are mutually inhibitory and that the MSA system provides top-down regulation and correction of the SR system (Brass et al. 2009). Although mentalizing allows individuals to understand others, there is always the potential for conflating one's own mental state (based on immediate self-to-other mapping subserved by the SR system) and the mental states of others. Hence, while humans clearly have marked mentalizing capacities, the possibility of misunderstanding others seems to be built into the neuroarchitecture because of our tendency to assume that we understand others based on our own embodied simulation of others' experiences. This realization has become a key guiding principle of all mentalization-based treatment (MBT).

Finally, balanced mentalizing entails an integration of cognition and affect. Mentalizing clearly involves cognitive features, including the capacity for perspective taking (being able to see that others may have a different perspective) and belief–desire reasoning (the capacity to explain and predict another's behavior on the basis of understanding that person's desires and beliefs). But balanced mentalizing also includes embodied affective features that ground mentalizing in an affectively felt reality. Whereas cognitive aspects of mentalizing heavily rely on controlled mentalizing, affective mentalizing is, at least at the basic neural level, largely automatic and embodied (Sabbagh 2004). However, over the course of human development, these two features of

mentalizing became increasingly integrated. Cognitively oriented mentalizing recruits several areas in the prefrontal cortex, whereas the VMPFC appears to play a key role in affectively oriented mentalizing (Shamay-Tsoory & Aharon-Peretz 2007). These findings are also consistent with the notion that the capacity for empathy is underpinned by a more basic "emotional contagion" system and a more advanced cognitive perspective-taking system. Imaging studies suggest important behavioral and anatomical dissociations in cognitive empathy (which is associated with the VMPFC) and emotional empathy (which is associated with the inferior prefrontal gyrus) (Shamay-Tsoory et al. 2009).

Mentalizing Is an Umbrella Concept

Mentalizing encompasses and subsumes a wide range of related concepts that are focused on various aspects of social cognition, including empathy, mindfulness, ToM, psychological mindedness, alexithymia, and insightfulness (Choi-Kain & Gunderson 2008). Empathy and ToM focus on aspects of mentalizing others, while mindfulness and alexithymia concern core features of mentalizing the self (e.g., the capacity to be aware of and attend to one's own internal mental states). Concepts such as empathy and mindfulness concern affective components of mentalizing, while ToM focuses on cognitive features of mentalizing (e.g., belief–desire reasoning). Mentalizing is thus a broad concept that refers to processes involved in reflective functioning about self–other and cognition–affect based on internal and external features. Furthermore, mentalizing refers to dynamic state- and context-dependent processes rather than to a trait, as stress and arousal typically lead to a switch from controlled and reflective to fast and automatic mentalizing and also, as a result, often biased mentalizing. Effective mentalizing is therefore all about the balance between the different dimensions of mentalizing and the systems underlying them. Different types of psychological problems can be related to specific types of imbalances between the mentalizing dimensions; we discuss this issue in more detail in the section titled Mentalizing and Personality Disorders.

A DEVELOPMENTAL PSYCHOPATHOLOGY APPROACH TO THE EMERGENCE OF MENTALIZING

The Role of Attachment in the Emergence of Mentalizing

The notion that the capacity for mentalizing is first acquired in the context of attachment relationships has been a key feature of the mentalizing approach to both normal and disrupted development since its inception. In particular, the capacity for parental mentalizing, or parental reflective functioning (PRF)—that is, the caregiver's capacity to reflect upon his/her own internal mental experiences as well as those of the child—is assumed to play a key role in this context (Luyten et al. 2017b, Sharp & Fonagy 2008, Slade 2005). PRF represents a relationship-specific manifestation of the more general capacity for reflective functioning and is thought to foster the development of secure attachment in the child as well as the child's own capacity for reflective functioning and, consequently, emotion regulation and interpersonal functioning. Caregivers with high levels of PRF are assumed to be able to respond with contingent and marked affective displays of their own experience in response to the child's subjective experience, thus enabling the child to develop second-order representations of his/her own subjective experiences. Hence, from this perspective, a socializing context that focuses on mental states (rather than secure attachment, emotional availability, or parental sensitivity per se) is assumed to foster the development of secure attachment and reflective functioning in young children (see **Figure 1**), leading to virtuous cycles marked by adaptive emotion regulation and positive socioemotional development (Luyten et al. 2017b).

Table 1 Relationships among secondary attachment strategies, arousal, and mentalizing

	Threshold for switch from controlled to automatic mentalizing	Strength of activation of automatic mentalizing	Recovery of controlled mentalizing
Secure attachment	High	Moderate	Fast
Hyperactivating strategies	Low: hyperresponsivity to stress/arousal	Strong	Slow
Deactivating strategies	Relatively high: hyporesponsive, but downregulation fails under increasing stress	Weak, but moderate to strong under increasing stress, reflecting failure of the deactivation strategy	Relatively fast
Disorganized attachment	Incoherent: hyperresponsive but with often frantic attempts to downregulate	Strong	Slow

This view implies a loose coupling among attachment, emotional sensitivity or availability, and PRF. Parents' mentalizing capacities may fluctuate markedly even in securely attached individuals, although in general these capacities can be expected to be positively related (Sharp & Fonagy 2008). By contrast, it is very unlikely that insecurely attached caregivers will have high levels of PRF because disruptions in early attachment relationships typically impair individuals' capacity to mentalize, particularly in emotionally intense relationship contexts such as parent–child relationships.

Furthermore, mentalizing is viewed as fundamentally interactive in that the capacity to mentalize develops in the context of interactions with others, and as a result it is assumed to be continually influenced by the mentalizing capacity of those others. Mentalizing is thus, at least to an extent, relationship and context dependent.

Attachment-hyperactivating and deactivating strategies are assumed to play a key role in explaining the relationships among stress/arousal and mentalizing in different arousal/interpersonal contexts (see **Table 1**). They influence (*a*) the threshold at which the switch from controlled to automatic mentalizing occurs (as described in the section titled Mentalizing Is Multidimensional), (*b*) the strength of the relationship between the severity of stress/arousal and the activation of neural circuits involved in controlled versus automatic mentalizing, and (*c*) the time to recovery when controlled mentalizing is lost under stress/arousal (Luyten et al. 2019b).

Finally, the mentalizing approach proposes a heuristic to help recognize when individuals appear limited in their capacity for mentalizing; these modes of experiencing subjectivity reflect ineffective mentalizing that developmentally antedates the capacity for full mentalizing. The three modes are as follows:

1. The psychic equivalence mode. In this mode of functioning, thoughts and feelings become too real. The individual can consider no perspectives other than his/her own and believes that his/her own perspective is the only one possible; this mode reflects the domination of self over other, external over internal, and emotion over cognition.
2. The teleological mode. In this mode, only real, observable goal-directed behavior and objectively discernible events that may potentially constrain these goals are recognized; this mode reflects extreme exterior focus and momentary loss of controlled mentalizing.
3. The pretend mode. Here, thoughts and feelings are severed from reality (so-called hypermentalizing or pseudomentalizing), and the individual becomes entangled in endless cognitive or affectively overwhelming narratives that have no connection to reality and, in the extreme, lead to feelings of derealization and dissociation; this mode reflects domination of explicit mentalizing by implicit mentalizing, inadequate internal focus, poor belief–desire reasoning, and vulnerability to fusion with others.

A common factor of these prementalizing modes of experiencing subjectivity is that they are thought to lead to pressure to externalize unmentalized aspects of the self (so-called alien-self parts); these are expressed in attempts to dominate the mind of others, self-injury, or other types of behavior (e.g., substance abuse) that, in the teleological mode, are expected to relieve tension and arousal. A tendency toward revictimization represents a specific type of externalization of unmentalized experiences of neglect and/or abuse (Luyten & Fonagy 2019).

Empirical Evidence

Meta-analyses and qualitative reviews have provided support for the main tenets of the mentalizing approach depicted in **Figure 1**. Here, we briefly review evidence for the main assumptions of this approach and note important limitations of extant research that have led to reformulations of the theory, which we discuss in the section titled Toward a Broader Socioecological Evolutionary Perspective.

Stability of mentalizing. Mentalizing has been shown to have both trait and state features; it is to a large extent relationship specific, and controlled mentalizing tends to be inhibited with increasing arousal or stress (for a review, see Luyten et al. 2019b). Hence, considerable stability in mentalizing may coexist with substantial fluctuations across contexts. For instance, high relative and absolute stability in alexithymia, reflecting serious problems with internally based mentalizing, have been found in a large cohort ($n = 3{,}083$) of Finnish adults followed up for 10 years (Hiirola et al. 2017). However, considerable fluctuations in mentalizing abilities have been extensively documented in patients with BPD (Fonagy & Luyten 2016), in nonclinical samples when reflecting about in-group versus out-group members (Luyten et al. 2019b), and in community individuals in whom stress and arousal levels were experimentally manipulated (Nolte et al. 2013). More research using dynamic modeling approaches is needed to further study the factors involved in explaining stability and changes in mentalizing across different contexts and time spans.

Parental mentalizing and child attachment. The propensity of caregivers to treat their infant as a psychological agent is known to be conducive to the development of secure attachment in children. A recent meta-analysis (Zeegers et al. 2017) identified 20 effect sizes (total number of participants = 974) examining the impact of parental mentalizing on attachment security. A pooled correlation of $r = 0.30$ between parental mentalizing and infant attachment security was found. Sensitivity and mentalizing together explained 12% of the variance in attachment security. However, sensitivity measured behaviorally did not account for the association between mentalizing and attachment. Similarly, the relationship between sensitivity and attachment remained significant after parental mentalizing was controlled for ($r = 0.19$). Thus, sensitivity, observed behaviorally, and parental mentalizing, assessed primarily through verbal reports, appeared to be relatively independent influences on parent–infant attachment, although a small proportion of the impact of parental mentalizing seemed to be mediated by behavioral sensitivity ($r = 0.07$).

These findings are particularly impressive for two reasons. First, studies in this area have been based on three relatively independent research traditions. Research by Meins (2013) and colleagues on parental mind-mindedness has emphasized the distinction between appropriate mind-related comments (referring to the accurate interpretation of a child's behavior) and nonattuned mind-related comments (reflecting the caregiver's inaccurate understanding of the child's intentional state, referring to past events without current connection, or attempting to divert the child from a current activity against his/her preference, thus indicating a lack of awareness of the child's perspective or a superimposition of the parent's perspective). Oppenheim et al. (2001), in turn,

have advanced an overlapping construct, parental insightfulness, which assesses the caregiver's tendency to perceive intentions underpinning the child's behavior alongside a willingness to modify such beliefs in what they refer to as an "open to change" attitude (Oppenheim & Koren-Karie 2013). Finally, the concept of PRF emerged from studies that used the Reflective Functioning Scale (Fonagy et al. 1998) for the Adult Attachment Interview [AAI (George et al. 1985)] and Parent Development Interview [PDI (Fonagy et al. 1991a, Slade 2005)]. PRF focuses on the caregiver's capacity to reflect on the child's subjective experience, the caregiver's own attachment history, and how that history may influence the relationship with the child. These three approaches clearly overlap and address, from slightly different perspectives, the same underlying construct of the parent's awareness of—and capacity to step beyond—the opacity of the child's mental state, and highlight the profound impact of the capacity for reflective functioning on the quality of the parent–child relationship. These approaches also further stress the multidimensional nature of (parental) mentalizing highlighted throughout this review.

A second strength of this body of research is that parental mentalizing has been shown to prospectively predict the development of attachment in children. Fonagy et al. (1991b) were the first to find evidence for such a link, in a study using the AAI in a sample of 200 first-time mothers and fathers. Antenatal parental mentalizing predicted child attachment assessed using the Strange Situation Procedure [SSP (Ainsworth et al. 1978)] at 12 and 18 months after birth, even when verbal IQ was controlled for, and continued to be a predictor of young adult reflective functioning 17 years later (Steele et al. 2016). In this sample, security of attachment in infancy was also shown to be associated with better performance on a cognitive-emotion task when children were age 5.5 years (Steele & Steele 2005). As another example, Meins et al. (2001, 2002) found that mothers' maternal mind-mindedness (as an index of parental mentalizing) predicted attachment security in their infants as assessed with the SSP at 45- and 48-month follow-up, as well as their children's social-cognitive performance at 55-month follow-up (Meins et al. 2003) and effortful control at 18- and 26-month follow-up (Bernier et al. 2010). These findings are particularly impressive considering the limited stability of attachment in childhood (Pinquart et al. 2013) and general limitations in the ability to predict child development over time (Fearon et al. 2014). As we discuss in more detail in the section titled Toward a Broader Socioecological Evolutionary Perspective, causal processes involved in psychological development are multifactorial, and there is a need to reconsider simple deterministic models of child development.

Parental mentalizing and child mentalizing. Studies also converge to suggest that higher levels of parental mentalizing foster mentalizing in children (e.g., Meins et al. 2002) and adolescents (e.g., Rosso & Airaldi 2016, Rosso et al. 2015). Whereas associations between parental mentalizing and infant attachment typically represent small effect sizes (defined in terms of Cohen's $d = 0.20$), the association between parental and infant mentalizing is characteristically much stronger, representing medium to large effect sizes (Cohen's $d = 0.50$–0.80). Rosso & Airaldi (2016), for instance, found a particularly strong association between mothers' ability to mentalize negative and mixed-ambivalent mental states, but not positive mental states, and the corresponding ability in their adolescent children ($r \approx 0.40$–0.50). Findings such as these suggest that the capacity of caregivers to reflect on difficult and affect-charged mental states is particularly important in the context of the intergenerational transmission of mentalizing.

Studies on mentalizing and adversity have provided some of the strongest evidence for the potential role of caregivers' mentalizing capacities in child development. Early adversity and complex trauma (i.e., early negative life experiences involving neglect and/or abuse, typically within an attachment/caregiving context) in particular have been shown to have the potential to severely impair mentalizing, as indicated by strongly biased mentalizing, hypersensitivity to the mental states

of others, a defensive inhibition of mentalizing, or a combination of these features (for reviews, see Borelli et al. 2019, Luyten & Fonagy 2019). At the same time, increasing evidence suggests that high levels of caregivers' reflective functioning, and specifically reflective functioning with regard to their own traumatic experiences (trauma-RF) (Ensink et al. 2017), may be an important buffer in the relationship between early adversity and child outcomes (reviewed by Borelli et al. 2019). For instance, higher trauma-RF in parents with a history of sexual abuse and neglect has been shown to be related to a lower risk of infant attachment disorganization (Berthelot et al. 2015) and a substantially lower risk of exposure to childhood sexual abuse in their own infants (Borelli et al. 2019). These findings are particularly encouraging with regard to prevention and intervention (reviewed later in the section titled The Spectrum of Mentalization-Based Treatment Interventions) and further stress the need for more direct tests of the assumed role of parental mentalizing in the relationship between parent attachment history and child outcomes (Zeegers et al. 2017).

Children's attachment style and child mentalizing. There is good evidence from cross-sectional and longitudinal studies in children and adolescents to suggest that secure attachment in children is associated with higher levels of child mentalizing. Studies suggest that secure attachment in children fosters cognitive features of mentalizing, including joint attention, perspective taking, and ToM, as well as affective components, such as emotion processing, empathy, and the use of mental-state language (e.g., Becker Razuri et al. 2017, Claussen et al. 2002, Kobak et al. 2017, Kokkinos et al. 2016, McQuaid et al. 2008, Meins et al. 2008, Troyer & Greitemeyer 2018, Zaccagnino et al. 2015).

Child mentalizing and cognitive and socioemotional development. Impairments in mentalizing in childhood have been related to a wide array of cognitive and socioemotional problems—which range from attentional control, effortful control, and academic achievement to emotion regulation and interpersonal problems—and internalizing and externalizing problems (for reviews, see Fonagy & Luyten 2016, 2018; Luyten & Fonagy 2018). In the section titled Mentalizing and Personality Disorders, we focus specifically on associations between imbalances in mentalizing and vulnerability for different psychological disorders. With regard to the emergence of psychopathology in childhood and adolescence, more studies are needed to disentangle the relative roles of infant attachment and mentalizing in determining vulnerability for psychological problems, as well as their interaction with other child features (e.g., temperament), contextual factors (e.g., family, school, and sociocultural environments), and biological vulnerability and resilience.

The role of attachment and mentalizing in stress and arousal regulation. Developmental psychopathology and neuroscience studies have generally supported the hypothesized associations between attachment dimensions, mentalizing, and stress and arousal regulation depicted in **Table 1** (for reviews, see Feldman 2017, Long et al. 2020, Vrticka & Vuilleumier 2012). Secure attachment experiences seem to pave the way for the highly synchronous functioning of biobehavioral systems involved in mentalizing and in attachment in coordinating the stress response. Research in this area has provided strong support for the assumption that both the attachment and mentalizing systems are centrally implicated in stress and emotion regulation and show high levels of functional connectivity at both the behavioral and the brain level. The attachment system seems to be relatively distinct from the mentalizing system; it is underpinned by a mesocorticolimbic reward system in the brain consisting of mesolimbic pathways that originate from the ventral tegmental area and project to ventral striatal regions and to the nucleus accumbens in particular (as well as to the hippocampus and amygdala), and mesocortical pathways with projections to the prefrontal cortex and anterior cingulate cortex. Several key biological mediators are involved

in this system, including dopamine, oxytocin, opioids, and cannabinoids. Secure attachment experiences typically buffer the effects of stress in early development as a result of the activation of these biological mediators, resulting in so-called adaptive hypoactivity of the hypothalamic–pituitary–adrenal (HPA) axis, the main human stress system, in early development. Particularly during so-called sensitive periods, which in humans seem to extend into early adulthood, the described social regulation is essential: It lays the basis for more generalized patterns of stress and emotion regulation. Repeated experience of downregulation of stress and arousal in the context of caring, sensitive, and available attachment figures has been shown to lead to a pattern of biobehavioral synchrony between the reward and mentalizing systems, expressed in terms of behavioral, autonomic, hormonal, and brain-to-brain synchrony between parent and infant (Feldman 2017). For these infants, caregivers become not only a source of reward but also a constantly available source of opportunities to recalibrate the infants' own minds (i.e., to mentalize) when in need. There is now also evidence that in these infants, this pattern tends to generalize across development from parent–infant relationships to pair-bonds, peers, and others more generally. However, studies also suggest a general weakening of this pattern that moves from parent–infant and pair-bonds to peers and others, which clearly suggests that there are limitations to the human capacity for secure attachment and solid mentalizing in relation to fellow human beings.

For infants growing up in highly insecure attachment contexts, this pattern associated with the integrated functioning of the attachment and mentalizing systems may never fully develop. Insecure attachment experiences typically lead to increased vulnerability to stress, as expressed in impairments in the functioning of the HPA axis and the reward system and, more generally, in the synchronous functioning of the mentalizing and reward systems. For these individuals, interpersonal relationships are not rewarding and may even become highly aversive. This may be either because attachment figures were not available for these individuals, leading to a dismissive pattern of relating to others, or because attachment figures were available only intermittently, leading to the excessive use of hyperactivating attachment strategies to obtain love, care, and support but with the underlying belief that others will not be available. A pattern of either compulsive autonomy, marked by downplaying attachment relationships and hyporesponsivity to distress, or excessive seeking of love and care, combined with hyperresponsivity to distress associated particularly with rejection, ensues (Luyten & Fonagy 2015).

The reemergence of prementalizing modes and externalization of alien-self parts. Studies have extensively documented the associations between impairments in mentalizing and psychic equivalence functioning, as observed, for instance, in the rigid, highly simplistic, and often defensive narratives indicative of hypomentalizing in patients with BPD in studies using the Reflective Functioning Scale on the AAI, or in severely at-risk parents' accounts of their developmental history on the PDI (Katznelson 2014, Slade 2005, Taubner et al. 2013). Similarly, studies have amply demonstrated the tendency of BPD patients (who typically show severe imbalances in mentalizing) to use pretend mode functioning, as expressed in high levels of hypermentalizing or pseudomentalizing (Fonagy & Luyten 2016, Sharp et al. 2016). Similar tendencies have been identified in at-risk mothers (Luyten et al. 2017a, Slade 2005).

Likewise, impairments in mentalizing have been shown to be strongly associated with teleological mode functioning, as evidenced by relationships between such impairments and substance abuse, violence, eating disorders (EDs), and self-harm, which can all be seen as attempts to regulate, through specific goal-directed behavior, unbearable and unmentalized self-states (Fonagy et al. 2016, Jewell et al. 2016, Suchman et al. 2018, Taubner et al. 2016). These latter observations also provide support for the view that problems with mentalizing lead to a tendency to externalize unmentalized experiences; these behaviors can be understood as attempts to downregulate

stress and arousal associated with unmentalized self-states. Consistent with this view, in a recent meta-analysis of 64 studies representing a total of 275,183 participants, individuals engaging in self-harm reported greater levels of body dissatisfaction, body disownership, and deficits in the experience and evaluation of bodily sensations compared with control groups who did not engage in self-injurious behaviors (Hielscher et al. 2019). More experimental research is needed in this context to investigate the precise mechanisms involved in these associations.

TOWARD A BROADER SOCIOECOLOGICAL EVOLUTIONARY PERSPECTIVE

Limitations of the Mentalizing Approach

In recent years, a number of findings have resulted in a major shift in our views concerning the role of attachment and mentalizing in normal and disrupted development (summarized in Fonagy et al. 2017a, 2017b). Although there is good evidence for the role of parental attachment and parental mentalizing in the intergenerational transmission of mentalizing, meta-analyses suggest that cross-generational associations are typically small (Zeegers et al. 2017). Moreover, other statistically significant contextual factors (controlling for verbal ability), such as socioeconomic status ($r = 0.18$) and number of siblings ($r = 0.14$), are implicated in the intergenerational transmission of mentalizing (Devine & Hughes 2018).

Developmental research increasingly points to the complex, multifactorial nature of human developmental processes, particularly in the field of attachment. It is becoming clear that evocative person–environment correlations (i.e., features of the child that evoke certain responses in the child's caregivers and the wider social environment) play a major role in these processes (Klahr & Burt 2014); this presents a major challenge for attachment theory, which has traditionally assumed that the attachment style of caregivers is relatively stable over time and plays a major role in determining child developmental trajectories (Fraley 2002). Yet, meta-analyses have consistently shown that contextual factors such as risk status lower the congruence between parent and child attachment classifications (Verhage et al. 2018).

Overall, developmental research provides five major challenges for contemporary attachment theory. First, the relationship between attachment in childhood and developmental outcomes is less strong than may be expected from some traditional assumptions within attachment theory (Fearon et al. 2010, Groh et al. 2012, Madigan et al. 2013). Second, meta-analyses suggest only moderate stability of attachment across development, which again contrasts with key assumptions of attachment approaches (Fraley 2002, Pinquart et al. 2013). Although the stability of attachment is somewhat greater in adolescence and adulthood than in childhood (Fraley & Roberts 2005, Jones et al. 2018), risk status (e.g., family conflict, parental separation, minority ethnic status, male sex) has typically been associated with lower stability in extant meta-analyses (e.g., Verhage et al. 2018). These findings make sense only if attachment is conceptualized as an interpersonal strategy to optimize adaptation to a particular environment. Hence, the stability (or lack of stability) of attachment seems to be largely a function of the stability of the environment, as also shown by simulation studies (Fraley & Roberts 2005). Third, historical, sociocultural, and environmental factors may determine the role and function of the attachment-behavioral system, which challenges Bowlby's original formulations of attachment as an innate, universal behavioral system (see Bowlby 1988). Fourth, a recent meta-analysis found that parental sensitivity, which is considered to play a key role in the intergenerational transmission of attachment, explained only a small proportion of the variance (effect sizes $r = 0.31$ for secure-autonomous transmission and $r = 0.21$ for unresolved transmission) in the association between parent and infant attachment (Verhage et al. 2016). As we have seen, parental mentalizing similarly accounts for only a small proportion

of the variance in explaining the intergenerational transmission of attachment and related features (Zeegers et al. 2017). Finally, there is increasing evidence for genetic factors in determining the course of attachment, suggesting that genes may play an important role in resetting developmental trajectories associated with attachment (Fearon et al. 2014). Again, such findings are difficult to accommodate within traditional attachment perspectives.

Broadening the Scope of the Mentalizing Approach

Recent evolutionary-based, developmental neurobiology accounts of human development appear to provide a more encompassing view of the role of mentalizing and attachment in psychological development (see **Figure 2**). In this context, we have been specifically influenced by Gergely and Csibra's natural pedagogy theory (Gergely 2013), Konner's description of childhood as a process of enculturation made possible by the evolution of a cultural acquisition device (Konner 2010), Tomasello's cultural intelligence hypothesis (Tomasello 2010), social baseline theory (Coan & Sbarra 2015), and Sperber's work on epistemic vigilance (Sperber et al. 2010). The body of research supporting these theories points to the same conclusion: that humans possess a species-specific capacity for the fast intergenerational transmission of cultural knowledge. Although the capacity for mentalizing is essential in this context, an even more fundamental role seems to be played by the capacity for epistemic trust: the capacity to identify knowledge conveyed by others as personally relevant and generalizable to other contexts. The acquisition of this capacity had major evolutionary advantages for humans and probably first emerged during the late Pleistocene era (Wilson & Wilson 2007). Instead of having to work out cultural knowledge oneself—a very time-consuming, difficult, and often impossible process—the recipient of information (e.g., a young child) can, through epistemic trust, rely on the authority and perceived trustworthiness of the person communicating that information (e.g., a caregiver or teacher). Epistemic trust thus enables a particular kind of species-specific learning: It involves encoding knowledge offered by others as significant, relevant to the recipient, and socially generalizable.

Ostensive cueing has been suggested to play a key role in this specific form of learning: Verbal and nonverbal ostensive cues are thought to trigger a pedagogic stance in the recipient, priming him/her that forthcoming communications are significant (Gergely 2013). Furthermore, ostensive cues typically lead the recipient to feel recognized as a subjective, agentive self (Gergely 2013). Developmental studies suggest that this feeling of being recognized opens up the channel for the fast transmission of knowledge and, in turn, the pathway for salutogenesis: the capacity to benefit from positive influences in one's environment. Hence, the virtuous cycle set in motion by epistemic trust may capture the essence of the human capacity for resilience (see **Figure 2**).

However, epistemic trust is not the default mode of functioning. Developmental studies suggest that epistemic vigilance—the ability to identify and filter out information conveyed by others that is perceived to be misleading, inaccurate, or deceitful—has to be overcome in the course of development. Studies indicate that infants show appropriate skepticism and distrust toward knowledge conveyed by others from very early on in development (Fonagy et al. 2017a). It is here that early attachment experiences have been shown to play a crucial role. It is primarily in the context of (early) attachment relationships that children learn to recognize who is trustworthy, authoritative, and knowledgeable (Corriveau et al. 2009). Mentalizing is an essential competence for discerning intention, to understand the opaque states of mind of others that would justify dependability on knowledge that they convey. Secure attachment experiences characterized by sensitive caregivers with genuine interest in the child's mind (i.e., high levels of mentalizing) probably provide the most consistent ostensive cueing and, as a result, the most fertile ground in which the child can develop epistemic trust and generalize it to new relationships and contexts. Yet, this process

goes beyond the attachment context. Other social contextual factors and learning processes also influence the development of epistemic trust. Peers, people in the community, and sociocultural influences more generally (e.g., those transmitted through social media) may further foster or inhibit the development of epistemic trust.

This broader social-communicative perspective allows us to accommodate the role of the wider social context and differing cultural norms and their influence on the development of epistemic trust and mentalizing. It has also resulted in an important shift in our views on attachment: We now view specific attachment styles as contexts for social communication that the familial setting is promoting about the most effective way to function in a prevailing environment. Normally, attachment or personality styles are attributed to the individual rather than to features of the types of interactions these individuals have with others and their social contexts more generally. From this perspective, insecure attachment and associated concepts such as disturbed personality functioning, personality disorders, and in fact aspects of most types of psychopathology can be conceptualized as manifestations of communicative strategies underpinning social learning to ensure adequate adaptation to changing social situations. Pervasive mistrust of social communication, for instance, may be considered a reasonable adaptation to severe childhood adversity.

A Transtheoretical Approach to Change in Psychological Interventions

This broader social-communicative approach to mentalizing has also resulted in substantial changes in our views concerning the role of mentalizing in psychological treatments (Fonagy et al. 2017b). Basically, we propose that changes resulting from psychological interventions are the outcome of particular forms of social learning from the patient's environment, and effective treatments are in essence a form of social relearning fostered by changes in what we have conceptualized as three communication systems:

1. Communication system 1 (lowering of epistemic vigilance) refers to the fact that all effective psychological treatments convey a particular model of mind to the patient that feels meaningful and self-relevant, often with the therapist using specific ostensive cues that, ideally, activate social learning in the patient. The channel for learning is opened to the extent that the patient recognizes benign intentions and feels recognized as an independent agent. Epistemic vigilance lessens, and the growth of epistemic trust creates the potential for learning and change. Mutual mentalizing plays a key role in this process because the therapist needs to tailor his/her intervention to the specific patient, demonstrating his/her ability to see the patient's problems from his/her perspective, and the patient needs to be able to recognize that the therapist is able to consider the patient's perspective (i.e., joint intentionality).
2. Communication system 2 (enabling mechanisms of social learning) is activated by the patient's increase in epistemic trust (achieved by communication system 1). The reactivation of the patient's mentalizing capacity is fostered by the background of trust and the social experience of the therapy; ideally, the patient models the mentalizing stance adopted by the therapist. The reemergence of mentalizing further facilitates epistemic trust. Hence, although we still believe that mentalizing is a common factor in most psychological interventions, we now argue that the aim of therapy is not to increase mentalizing as such; instead, the goal is for increased mentalizing to open up the patient's potential for learning and thus, with increased epistemic trust, help the patient benefit from the communications from the therapist, learn new skills, acquire self-knowledge, and restructure internal working models. The new learning enables a virtuous cycle marked by salutogenesis—the capacity of the patient to benefit from further positive social influences in the therapy and in the interpersonal world outside the treatment setting.

3. Communication system 3 (reengaging with the social world) reflects how being mentalized by another person frees the patient from his/her state of temporary or chronic social isolation and (re)activates the capacity to learn; this frees the person to grow in the context of relationships outside therapy. This view thus implies that it is not just the facts and techniques that are taught in treatment that are important but also, and perhaps primarily, that when the patient's capacity for social learning and social recalibration of the mind is activated, new experiences may be sought, and reconstruing existing relationships is likely to improve adaptation. The patient is enabled to "use" his/her environment in a different way. A further implication is, of course, that psychological interventions may need to also intervene at the level of the social environment when needed or appropriate.

Empirical Evidence

First, the basic tenets of this theoretical approach are well supported by developmental psychopathology, evolutionary, and comparative research. There is significant evidence for the species-specific nature of social learning based on joint intentionality and mentalizing in humans as well as evidence showing that the quality of the relationship of a child to a communicator determines in large part the extent to which the child will acquire and generalize information from that communicator (Lane & Harris 2015, Mascaro & Sperber 2009, Shafto et al. 2012). Second, the "hard-to-reach" character of individuals with a history of early adversity and deprivation that we see as prototypically leading to epistemic distrust has been well demonstrated (Fonagy et al. 2015). Findings concerning high dropout rates and problems with trust in these individuals can be interpreted along the same lines (e.g., Ekeblad et al. 2016). Third, findings concerning a general psychopathology factor (p factor)—which captures vulnerability to mental disorder, comorbidity among disorders, persistence of disorders over time, severity of symptoms, and therapeutic response—point to the relevance of a general factor in explaining numerous aspects of psychopathology (Caspi & Moffitt 2018). The association between the p factor and a family history of psychiatric problems, childhood adversity, and adult life impairment suggests that the p factor may be related to problems with epistemic trust and salutogenesis. Fourth, the equal effectiveness of common psychological interventions similarly points to a final common pathway involving social learning and salutogenesis. Finally, there is good evidence to suggest that the prevalence of mental health disorders is highly associated with levels of distrust in social institutions (Rozer & Volker 2016), reflecting the impact of epistemic trust on well-being not only at the level of the individual but also at the population level.

MENTALIZING AND PERSONALITY DISORDERS

Borderline Personality Disorder

The bulk of research in this area has focused on the role of attachment and mentalizing in BPD and associated conditions. Consistent with the theoretical assumptions outlined in this review, high levels of preoccupied attachment (reflecting attachment-hyperactivating strategies) and disorganized/unresolved patterns of attachment (reflecting the use of both attachment-hyperactivating and deactivating strategies) have been found in BPD patients (for a review, see Fonagy & Luyten 2016). Yet, because of the overlap in phenomenology between BPD and patterns of preoccupied and disorganized attachment, these studies are not particularly compelling. Prospective studies showing very high rates of (complex) trauma in patients with BPD provide more convincing evidence for hypothesized associations between disruptions in the development of the attachment-behavioral system and BPD. A recent review identified 39 prospective studies (covering 24 unique samples) and found that exposure to different types of trauma, including emotional abuse, neglect,

and physical and sexual abuse, was associated with increased risk of BPD (Stepp et al. 2016). The review by Stepp and colleagues (2016) also provides support for more recent formulations emphasizing the broader socioecological context in the etiology of BPD, as it found that BPD patients typically are exposed to a broader adverse context characterized by parental psychopathology, lower socioeconomic status, and/or violence.

In addition, although the prevalence of (complex) trauma in BPD is often up to 90%, and patients with BPD typically report higher levels of trauma than individuals with other personality disorders (Fonagy & Luyten 2016), genetic factors, including gene–environment correlations and interactions, may play an important role in the etiology of BPD. BPD has an estimated heritability of 40–50% (Distel et al. 2008). A study of over 5,000 twins and almost 1,300 siblings found that the unique environmental variance explaining BPD features increased linearly with the number of traumatic life events to which an individual had been exposed (from 54% with no events to 64% with six events) (Distel et al. 2011). In a nationally representative birth cohort of over 1,100 families with twins in the United Kingdom, maltreatment was highly associated with BPD, but only in those with a family history of psychopathology as an index of genetic vulnerability (Belsky et al. 2012). In families without a history of psychopathology, maltreatment was reported by only 7% of individuals with BPD compared with almost 50% of those with BPD and a family history of psychopathology. Hence, consistent with our more recent theoretical formulations, vulnerability to BPD is best considered within a multifactorial, socioecological framework (Luyten et al. 2020).

Insecure attachment has also been associated with impairments (often severe) in mentalizing that are typical of BPD patients. These impairments are typically expressed in terms of patients' overly simplistic or overanalytic/hyperactive accounts of their own mental states and those of others (Fonagy & Luyten 2009). Yet, research findings and clinical accounts have also reported apparently superior mentalizing capacities in BPD patients compared with normal controls—a phenomenon that has also been termed the empathy paradox (Carter & Rinsley 1977, Dinsdale & Crespi 2013, Krohn 1974). These seemingly conflicting findings concerning mentalizing in BPD can be understood when considering the typical imbalances between the four dimensions of mentalizing (Fonagy & Luyten 2016). The characteristic pattern of mentalizing in BPD is a rapid loss of controlled mentalizing and overreliance on fast, automatic mentalizing, followed by problems with cognitive mentalizing, particularly in complex interpersonal situations; there is also an overreliance on affectively dominated and highly externally based mentalizing at the expense of mentalizing that is directly focused on mental interiors, and a tendency to conflate mental states of the self and others (so-called identity diffusion), leading to increased susceptibility to emotional contagion. Hence, the presumed superiority of mentalizing of BPD patients in some circumstances appears to be largely based on a tendency toward hypermentalizing—an attempt to make sense of others' external cues (such as their facial expressions or posture) based on fast, automatic processing of such information. As a result, BPD patients often might "get it right," but the flip side is that they often jump to conclusions about others' internal mental states. This is also shown by findings concerning a negativity bias in BPD patients, which has been observed, for instance, in their interpretation of neutral faces (Herpertz & Bertsch 2015) or when they are presented with short silent video clips (Barnow et al. 2009). When presented with such material, BPD patients typically see characters as more negative and more aggressive. Finally, the tendency of BPD patients to conflate the mental states of the self and others can be seen as another consequence of fast, automatic, affect-driven mentalizing, which is also consistent with evidence of overactivation of neural circuits involved in the SR network and deficits in the MSA system in individuals with BPD (Fonagy & Luyten 2016, Ripoll et al. 2013).

Currently, there is only indirect evidence for the role of epistemic trust in BPD—for instance, from studies showing high levels of distrust in BPD patients. Individuals with BPD are

characterized by a bias in their perception of others as being hostile and untrustworthy; they tend to expect that others will reject, hurt, abandon, criticize, or neglect them or treat them dishonestly (for a recent review, see Fertuck et al. 2018). From a neurobiological perspective, the lack of trust in others typical of BPD patients appears to be mediated by the reward system (Herpertz & Bertsch 2015), which is also centrally implicated in attachment (see the section titled Neurobiology of Mentalizing). In double-blind placebo-controlled trials, activation of the reward system by administration of oxytocin leads BPD patients to become even less trusting of others (Herpertz & Bertsch 2015). Studies have shown that it can take decades for BPD patients to catch up with normative developmental trajectories in central life domains such as work and relationships, even after successful psychotherapy (Gunderson et al. 2018); this is also consistent with the broad, social-communicative approach to BPD outlined in this review, with its emphasis on impairments in the capacity for salutogenesis in BPD.

Other Personality Disorders

More recent work has extended the mentalizing approach to antisocial personality disorder (ASPD) (Bateman et al. 2019) and narcissistic and avoidant personality disorders (Simonsen & Euler 2019). As with BPD, these studies have begun to delineate specific mentalizing imbalances in each of these disorders. For instance, there is now increasing consensus that two major developmental pathways are involved in ASPD and its precursors in childhood and adolescence (Fonagy & Luyten 2018, Taubner et al. 2019). Both pathways involve different patterns of imbalances in mentalizing. One pathway leads to a cluster of individuals characterized by high levels of anxiety, hypervigilance to emotional states, and high levels of reactive aggression. Studies have consistently shown that these individuals are characterized by a fast switch to automatic, affect-dominated mentalizing (Fonagy & Luyten 2018). The other pathway appears to be typical of individuals with so-called callous–unemotional features: hyporeactivity to stress, severe deficits in affective mentalizing, and the use of instrumental (i.e., manipulative) aggression (Viding et al. 2014).

Consistent with the social-communicative approach outlined in this review, studies suggest that ASPD and its precursor in childhood and adolescence, conduct disorder, can best be conceptualized as adaptation strategies. This assumption is consistent with robust findings of problems with the unlearning of aggression in individuals with these disorders, particularly those with callous–unemotional traits. In these individuals, social learning based on social communication appears to be fundamentally disrupted because of biological vulnerability, an adverse social context, or a combination of both (Fonagy & Luyten 2018). In social environments marked by abuse, neglect, and violence, an automatic, fast, and affective focus on the mental states of others (as opposed to the self) would be an appropriate survival strategy, and extending epistemic trust to others may compromise chances of survival (Luyten et al. 2020).

MENTALIZING AND OTHER DISORDERS

Recent years have seen a steady increase in studies on common mental disorders and problems throughout the life span based on the mentalizing approach to psychopathology. With regard to depression and anxiety, studies have amply demonstrated the role of attachment disruptions and (early) adversity more generally in explaining vulnerability to these disorders, the negative impact of these problems on mentalizing, and the role of impairments in mentalizing with regard to both the self and others in pathways to depression and anxiety (Luyten & Fonagy 2018, Nolte et al. 2011). There is also evidence that mentalizing impairments are in part a consequence of, and are exacerbated by, mood disturbances and the duration of mood problems (Fischer-Kern & Tmej 2019, Li et al. 2015, Nolte et al. 2011). Mentalizing impairments have also been found in remitted

depressed patients and may increase the likelihood of relapse (Luyten & Fonagy 2018). Impairments in mentalizing thus present an important target in treatments for depression regardless of the theoretical orientation of specific treatments. Consistent with this assumption, both cognitive behavioral and psychodynamic approaches have shifted from treatments focusing on the content of psychological dynamics in depression (e.g., schemas or attachment representations) to treatments that also include a focus on the process of mentalizing or metacognition (e.g., in MBT and mindfulness-based approaches), particularly in patients with chronic depression (Luyten et al. 2013).

In anxiety disorders, the importance of patients' capacity to reflect on anxiety symptoms and their relation to interpersonal events in particular has been demonstrated. This panic-focused reflective functioning has been shown to be associated with the therapeutic alliance, an in-session focus on interpersonal relationships and emotional expression, and outcomes in both psychodynamic and cognitive behavioral therapy (Keefe et al. 2019, Solomonov et al. 2019). More studies such as these focusing on disorder-specific impairments in mentalizing are needed because they are consistent with research demonstrating the context-specific nature of mentalizing and may have direct implications for improving therapeutic interventions across theoretical orientations.

Studies in patients with EDs have similarly pointed to intrinsic associations between attachment disruptions and mentalizing impairments (for a recent comprehensive review, see Robinson et al. 2019). In more dysregulated patients with EDs, studies have typically found a pattern of severe impairments in mentalizing (in particular, embodied mentalizing) as well as affective hypermentalizing, particularly in those with comorbid BPD features. In higher-functioning, perfectionistic patients, by contrast, a combination of hypomentalizing and cognitive hypermentalizing has been reported. Particularly in high-functioning patients, hypermentalizing is often difficult to distinguish from genuine mentalizing (Fonagy et al. 2016).

Research on the mentalizing approach in somatoform disorders increasingly suggests a causal sequence from disruptions in attachment to impairments in mentalizing and stress dysregulation, leading to a pattern of hyperreactivity to stress (Luyten et al. 2019a). Because of the wear and tear of constant distress on the HPA axis system and the sympathetic nervous system as core structures of the stress system, this pattern may lead to a so-called biopsychosocial crash, resulting in a state of chronic dysregulation of the biological stress system and associated pain and immune regulation systems as well as depressed mood, anxiety, and fatigue. This state leads to further impairments in mentalizing (particularly embodied mentalizing) as the patient increasingly begins to experience his/her body as an alien object that threatens the self from within (Schattner et al. 2008).

The long-standing knowledge that individuals with autism spectrum disorder have problems with ToM sparked research that has documented (sometimes severe) deficits in a wide range of mentalizing capacities in these individuals. Given the marked heterogeneity within the autism spectrum, the precise role (if any) of these deficits in the origin of autism and in approaches to help people with autism navigate their social worlds needs to be further determined (Lombardo et al. 2019). In addition, although several trials have investigated the effect of intranasal oxytocin, a key biological mediator in attachment and mentalizing, its effects on social functioning in autism have been limited (DeMayo et al. 2017).

Studies have also begun to address the role of mentalizing in other disorders, including disorders dominated by teleological mode functioning, such as substance abuse disorder (Suchman et al. 2018), pathological gambling (Cosenza et al. 2019), attention-deficit/hyperactivity disorder (Perroud et al. 2018), and psychotic disorders (Debbané et al. 2016) characterized by seriously distorted (embodied) mentalizing.

Finally, in posttraumatic stress disorder (PTSD), a recent meta-analysis reported a consistent, large deficit in mentalizing in individuals with PTSD relative to trauma-exposed and healthy

controls (Stevens & Jovanovic 2019); this finding stresses the need for further research on mentalizing in PTSD, particularly as premorbid mentalizing deficits were found to increase the risk of PTSD.

THE SPECTRUM OF MENTALIZATION-BASED TREATMENT INTERVENTIONS

Psychological interventions that are rooted in the mentalizing approach have three features in common despite obvious differences related to their specific goals and target population. First, consistent with their theoretical roots, MBT interventions focus on improving mentalizing capacities through a focus on the patient's mental states as they are experienced moment by moment, and by emphasizing the therapeutic alliance with the active repair of ruptures in the patient–therapist relationship (Bateman & Fonagy 2016). The therapist adopts a stance that is simultaneously inquisitive and not-knowing: The therapist does not "know" what the patient is feeling or thinking but is curious to learn from the patient and, thus, needs the patient as a "teacher" in jointly developing a model of the patient's mind. Second, MBT interventions are structured, manualized interventions that focus on delivering treatments that are coherent, consistent, and continuous over time. Consistent with our recent theoretical shift toward a broader, social-communicative approach, we believe that adherence to these "three Cs" is a key common feature of all evidence-based psychotherapies, particularly in patients who lack a coherent self-structure, such as those with BPD (Fonagy et al. 2017c). Third, based on the socioecological model outlined in this review, MBT is increasingly emphasizing the fostering of the capacity for salutogenesis and thus resilience in patients.

A recent meta-analysis including 33 randomized controlled trials (RCTs) of specialized psychotherapies for BPD versus nonspecialized psychotherapies in adult patients diagnosed with BPD (Cristea et al. 2017) supports the efficacy of MBT in BPD patients. RCTs published more recently have similarly found that MBT for BPD in both adult and adolescent patients is typically associated with medium to large or very large effect sizes on a wide variety of outcome measures that range from core BPD features to educational attainment and interpersonal functioning (Volkert et al. 2019). In RCTs, BPD patients have shown maintenance of gains and further improvement in several life domains up to 3-year (Smits et al. 2020) and 8-year (Bateman & Fonagy 2008) follow-up. More than a dozen naturalistic studies have consistently shown similar effects of MBT in BPD patients (Volkert et al. 2019). Only one RCT has compared outpatient MBT and day-hospitalization-based MBT for BPD; it found no substantial differences in outcomes despite large differences in the intensity of the treatment programs (Smits et al. 2020). Hence, a less intensive treatment may be as effective as a high-intensity day-hospitalization treatment for BPD; this opens up interesting perspectives regarding the upscaling of MBT and other treatments for BPD. It also chimes with our conceptual approach of placing at least some of the effective components of psychotherapeutic interventions outside the therapeutic context and located in a changed approach that the patient takes to his/her social environment. As there is growing evidence for the cost-effectiveness of MBT for BPD (Blankers et al. 2019), issues related to treatment intensity and length may ultimately determine the optimal treatment for BPD.

Although there have not been any direct comparative trials, studies suggest that there are no substantial differences in efficacy between MBT and other specialized treatments for BPD, or between specialized and bona fide nonspecialized treatments (Cristea et al. 2017, Volkert et al. 2019); this suggestion is consistent with our emphasis on the importance of coherence, continuity, and consistency in treatments for this patient group and our emphasis on various generic ways

in which epistemic trust and social learning may be recovered and reengaged in patients with personality disorder. Only one nonrandomized study has compared MBT with dialectical behavior therapy, and that study found no differences in any outcomes (Barnicot & Crawford 2019).

Possible strengths of MBT are that it has low dropout rates (Barnicot & Crawford 2019) and might be more effective than nonspecialized interventions in patients whose symptoms are more severe (Bateman & Fonagy 2013, Kvarstein et al. 2019), but more research is needed. In any case, research on the implementation of MBT in routine care settings shows that the quality of implementation may have a large impact on the effects of MBT; effect sizes of poorly implemented MBT are up to three times smaller than those of well-implemented MBT (Bales et al. 2017).

The evidence base for MBT in other psychological disorders is rapidly growing, although much more research is needed before any firm conclusions can be drawn. Recent RCTs have provided preliminary support for the efficacy of MBT in ASPD (Bateman et al. 2016), EDs comorbid with BPD (Robinson et al. 2016), substance abuse disorder (Philips et al. 2018, Suchman et al. 2018), and depression (Fonagy et al. 2019). Several RCTs concerning the efficacy of MBT in a variety of psychological disorders are currently ongoing.

There is also growing evidence for MBT interventions that focus on the family and broader social context, consistent with the social-ecological approach outlined in this review. Several naturalistic studies and RCTs have supported the effectiveness of MBT in substance-abusing mothers and their infants (Suchman et al. 2018); fostered and adopted children (Redfern et al. 2018); mothers living in underserved, poor, urban communities with children at high risk of maltreatment (Byrne et al. 2019, Slade et al. 2020); individuals supporting a family member with BPD (Bateman & Fonagy 2019a); and school-based preventive intervention programs (Fonagy et al. 2009). These types of system-level interventions might be most effective at addressing the problems that nonmentalizing social environments (e.g., neighborhoods beset by crime and violence, schools with a culture of bullying) tend to generate, by creating a mentalizing climate as a counterweight against competitive, hostile, and aggressive wishes and tendencies. An important part of this view is that it acknowledges the need to provide a supportive mentalizing system around mental health professionals in light of the many internal and external pressures and anxieties that are generated by this type of work.

Although studies have relatively consistently shown across different therapeutic approaches and conditions that changes in patients' mentalizing are associated with changes in their symptoms (e.g., Cologon et al. 2017, De Meulemeester et al. 2018, but see Levy et al. 2006 for a negative finding), it is important to remember that the improvement in mentalizing itself may not be of direct benefit but, rather, that it may lead to the recovery of normal social functions that enable the interpersonal transmission of knowledge.

FUTURE DIRECTIONS

Although the mentalizing approach to normal and disrupted psychological development has become increasingly popular over the past decades, a number of important limitations hamper the empirical study of its validity. Here, we discuss what are, in our opinion, the five most pressing issues that should be addressed in future research.

First, there is a clear need for more research on the validity of measures of mentalizing and related constructs. Given the multidimensional nature of mentalizing and the influence of contextual factors on mentalizing, measures with greater ecological validity that can be used in both behavioral and neurobiological studies are needed. Similarly, there is an urgent need for valid measures of epistemic trust and salutogenesis to be developed.

Second, there is a need for better understanding of the neural systems involved in mentalizing, particularly in real-time interactions between individuals. So far, we have studied the brain mostly outside its natural context—that is, in interaction with other brains. As social neuroscience is increasingly transitioning from a first-person to a second-person approach focusing on biobehavioral synchrony and interbrain coherence (Long et al. 2020), these changes will undoubtedly revolutionize research on the approach outlined in this review.

Third, throughout this review, we have identified the need for sufficiently powered longitudinal research documenting the development of mentalizing in relation to other psychosocial and biological factors, providing a more comprehensive test of the theoretical models outlined in **Figures 1** and **2**.

Fourth, there is a need for large-scale, adequately powered studies of the efficacy and effectiveness of MBT, particularly given the risk of bias and publication bias identified in a recent review of psychological therapies for BPD (Cristea et al. 2017).

Finally, more research on the mechanisms of change in MBT and other treatments is needed. If mentalizing and epistemic trust are embedded within the human behavioral repertoire as key learning mechanisms, they should also be centrally involved in explaining the effects of psychosocial interventions. Sophisticated multilevel studies of the relationship between process and outcome in psychosocial interventions are needed to investigate these assumptions in frameworks that include the patient's social environment and explore the role of psychological therapy for the interaction of changes in social adaptation driven by and driving further changes.

CONCLUSION

This review shows that the mentalizing approach to normal and disrupted psychological development is coming of age, as evidenced by a steady increase in basic research and intervention studies. This approach provides a unified, transtheoretical, and transdiagnostic perspective on psychological development that is firmly rooted in evolutionary science and developmental psychopathology.

Earlier formulations of the mentalizing approach focused on dyadic attachment-related processes and their impact on normal and disrupted development. More recent formulations have centrally emphasized mentalizing and the associated capacities for epistemic trust and salutogenesis as key mechanisms enabling the fast intergenerational transmission of knowledge. This framing has led to a broad, socioecological perspective on psychological disorders as reflecting impairments in social communication rather than a consideration of vulnerability to psychopathology as residing solely or largely within the individual. This new perspective, while retaining the roots of the theory in psychoanalytic thinking, also bridges psychological and socioecological models of psychopathology and opens up new avenues for research as well as prevention and intervention strategies. We hope that the present review will spark further interest in the mentalizing approach and its application to understanding psychological problems.

SUMMARY POINTS

1. Mentalizing (or reflective functioning) refers to the highly developed, evolutionarily prewired human capacity to understand the self and others in terms of intentional mental states, such as feelings, desires, wishes, attitudes, and goals.
2. Mentalizing impairments are transdiagnostic and transtheoretical vulnerability factors for psychopathology; temporary or chronic impairments in mentalizing are implicated in a wide range of psychological problems and disorders.

3. Recent formulations have shifted from an emphasis on the role of dyadic attachment in the development of mentalizing in earlier formulations to a broader, sociocommunicative approach that emphasizes the role of family, peers, and broader sociocultural factors in the development of mentalizing and the capacity for epistemic trust, the evolutionarily prewired capacity to trust others as sources of social information.
4. A growing body of research supports the (cost-)effectiveness of mentalization-based treatment (MBT)—that is, treatments that focus on the recovery of the capacity for mentalizing and epistemic trust.
5. Although the evidence base for MBT is growing, there is a need for large-scale trials to further investigate the (cost-)effectiveness of MBT, its purported mechanisms of change, and its potential to be implemented in routine clinical care.
6. Similarly, there is a need for more research on the assessment of the various dimensions of mentalizing, which will also enable research concerning the neurobiological bases of mentalizing and associated psychological processes.

DISCLOSURE STATEMENT

P.L. and P.F. are involved in the development, training, and dissemination of mentalization-based treatments. While P.F. derives no personal gain from these activities, he is chief executive of the Anna Freud Centre, a charity that financially benefits from trainings in this modality. Apart from these, the authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Ainsworth MDS, Blehar MC, Waters E, Wall S. 1978. *Patterns of Attachment: A Psychological Study of the Strange Situation*. Hillsdale, NJ: Lawrence Erlbaum

Bales DL, Verheul R, Hutsebaut J. 2017. Barriers and facilitators to the implementation of mentalization-based treatment (MBT) for borderline personality disorder. *Personal. Ment. Health* 11:118–31

Barnicot K, Crawford M. 2019. Dialectical behaviour therapy *v.* mentalisation-based therapy for borderline personality disorder. *Psychol. Med.* 49:2060–68

Barnow S, Stopsack M, Grabe HJ, Meinke C, Spitzer C, et al. 2009. Interpersonal evaluation bias in borderline personality disorder. *Behav. Res. Ther.* 47:359–65

Bateman A, Fonagy P. 2008. 8-year follow-up of patients treated for borderline personality disorder: mentalization-based treatment versus treatment as usual. *Am. J. Psychiatry* 165:631–38

Bateman A, Fonagy P. 2013. Impact of clinical severity on outcomes of mentalisation-based treatment for borderline personality disorder. *Br. J. Psychiatry* 203:221–27

Bateman A, Fonagy P. 2016. *Mentalization-Based Treatment for Personality Disorders: A Practical Guide*. Oxford, UK: Oxford Univ. Press

Bateman A, Fonagy P. 2019a. A randomized controlled trial of a mentalization-based intervention (MBT-FACTS) for families of people with borderline personality disorder. *Personal. Disord.* 10:70–79

Bateman A, Fonagy P, eds. 2019b. *Handbook of Mentalizing in Mental Health Practice*. Washington, DC: Am. Psychiatr. Publ. 2nd ed.

Bateman A, Motz A, Yakeley J. 2019. Antisocial personality disorder in community and prison settings. See Bateman & Fonagy 2019b, pp. 335–49

Bateman A, O'Connell J, Lorenzini N, Gardner T, Fonagy P. 2016. A randomised controlled trial of mentalization-based treatment versus structured clinical management for patients with comorbid borderline personality disorder and antisocial personality disorder. *BMC Psychiatry* 16(1):304

Becker Razuri E, Hiles Howard AR, Purvis KB, Cross DR. 2017. Mental state language development: the longitudinal roles of attachment and maternal language. *Infant Ment. Health J.* 38:329–42

Belsky DW, Caspi A, Arseneault L, Bleidorn W, Fonagy P, et al. 2012. Etiological features of borderline personality related characteristics in a birth cohort of 12-year-old children. *Dev. Psychopathol.* 24:251–65

Bernier A, Carlson SM, Whipple N. 2010. From external regulation to self-regulation: early parenting precursors of young children's executive functioning. *Child Dev.* 81:326–39

Berthelot N, Ensink K, Bernazzani O, Normandin L, Luyten P, Fonagy P. 2015. Intergenerational transmission of attachment in abused and neglected mothers: the role of trauma-specific reflective functioning. *Infant Ment. Health J.* 36:200–12

Blankers M, Koppers D, Laurenssen EMP, Peen J, Smits ML, et al. 2019. Mentalization-based treatment versus specialist treatment as usual for borderline personality disorder: economic evaluation alongside a randomized controlled trial with 36-month follow-up. *J. Personal. Disord.* In press. **https://doi.org/10.1521/pedi_2019_33_454**

Borelli JL, Cohen C, Pettit C, Normandin L, Target M, et al. 2019. Maternal and child sexual abuse history: an intergenerational exploration of children's adjustment and maternal trauma-reflective functioning. *Front. Psychol.* 10:1062

Bowlby J. 1988. *A Secure Base: Parent-Child Attachment and Healthy Human Development*. New York: Basic Books

Brass M, Ruby P, Spengler S. 2009. Inhibition of imitative behaviour and social cognition. *Philos. Trans. R. Soc. B* 364:2359–67

Byrne G, Sleed M, Midgley N, Fearon P, Mein C, et al. 2019. Lighthouse Parenting Programme: description and pilot evaluation of mentalization-based treatment to address child maltreatment. *Clin. Child Psychol. Psychiatry* 24:680–93

Carter L, Rinsley DB. 1977. Vicissitudes of 'empathy' in a borderline adolescent. *Int. Rev. Psychoanal.* 4:317–26

Caspi A, Moffitt TE. 2018. All for one and one for all: mental disorders in one dimension. *Am. J. Psychiatry* 175:831–44

Excellent review of the origins of the mentalizing approach to borderline personality disorder.

Choi-Kain LW, Gunderson JG. 2008. Mentalization: ontogeny, assessment, and application in the treatment of borderline personality disorder. *Am. J. Psychiatry* 165:1127–35

Claussen AH, Mundy PC, Mallik SA, Willoughby JC. 2002. Joint attention and disorganized attachment status in infants at risk. *Dev. Psychopathol.* 14:279–91

Coan JA, Sbarra DA. 2015. Social baseline theory: the social regulation of risk and effort. *Curr. Opin. Psychol.* 1:87–91

Cologon J, Schweitzer RD, King R, Nolte T. 2017. Therapist reflective functioning, therapist attachment style and therapist effectiveness. *Adm. Policy Ment. Health* 44:614–25

Corriveau KH, Harris PL, Meins E, Fernyhough C, Arnott B, et al. 2009. Young children's trust in their mother's claims: longitudinal links with attachment security in infancy. *Child Dev.* 80:750–61

Cosenza M, Ciccarelli M, Nigro G. 2019. The steamy mirror of adolescent gamblers: mentalization, impulsivity, and time horizon. *Addict. Behav.* 89:156–62

Cristea IA, Gentili C, Cotet CD, Palomba D, Barbui C, Cuijpers P. 2017. Efficacy of psychotherapies for borderline personality disorder: a systematic review and meta-analysis. *JAMA Psychiatry* 74:319–28

Comprehensive review of research on importance of ostensive cues in human learning.

Csibra G, Gergely G. 2009. Natural pedagogy. *Trends Cogn. Sci.* 13:148–53

De Meulemeester C, Vansteelandt K, Luyten P, Lowyck B. 2018. Mentalizing as a mechanism of change in the treatment of patients with borderline personality disorder: a parallel process growth modeling approach. *Personal. Disord.* 9:22–29

Debbané M, Salaminios G, Luyten P, Badoud D, Armando M, et al. 2016. Attachment, neurobiology, and mentalizing along the psychosis continuum. *Front. Hum. Neurosci.* 10:406

DeMayo MM, Song YJC, Hickie IB, Guastella AJ. 2017. A review of the safety, efficacy and mechanisms of delivery of nasal oxytocin in children: therapeutic potential for autism and Prader-Willi syndrome, and recommendations for future research. *Paediatr. Drugs* 19:391–410

Devine RT, Hughes C. 2018. Family correlates of false belief understanding in early childhood: a meta-analysis. *Child Dev.* 89:971–87

Dinsdale N, Crespi BJ. 2013. The borderline empathy paradox: evidence and conceptual models for empathic enhancements in borderline personality disorder. *J. Personal. Disord.* 27:172–95

Distel MA, Middeldorp CM, Trull TJ, Derom CA, Willemsen G, Boomsma DI. 2011. Life events and borderline personality features: the influence of gene–environment interaction and gene–environment correlation. *Psychol. Med.* 41:849–60

Distel MA, Trull TJ, Derom CA, Thiery EW, Grimmer MA, et al. 2008. Heritability of borderline personality disorder features is similar across three countries. *Psychol. Med.* 38:1219–29

Ekeblad A, Falkenstrom F, Holmqvist R. 2016. Reflective functioning as predictor of working alliance and outcome in the treatment of depression. *J. Consult. Clin. Psychol.* 84:67–78

Ensink K, Begin M, Normandin L, Fonagy P. 2017. Parental reflective functioning as a moderator of child internalizing difficulties in the context of child sexual abuse. *Psychiatry Res.* 257:361–66

Fearon P, Shmueli-Goetz Y, Viding E, Fonagy P, Plomin R. 2014. Genetic and environmental influences on adolescent attachment. *J. Child Psychol. Psychiatry* 55:1033–41

Fearon RP, Bakermans-Kranenburg MJ, van Ijzendoorn MH, Lapsley AM, Roisman GI. 2010. The significance of insecure attachment and disorganization in the development of children's externalizing behavior: a meta-analytic study. *Child Dev.* 81:435–56

Feldman R. 2017. The neurobiology of human attachments. *Trends Cogn. Sci.* 21:80–99

Authoritative review of the nature and functions of human attachment and its roots in neurobiology.

Fertuck EA, Fischer S, Beeney J. 2018. Social cognition and borderline personality disorder: splitting and trust impairment findings. *Psychiatr. Clin. North Am.* 41:613–32

Fischer-Kern M, Tmej A. 2019. Mentalization and depression: theoretical concepts, treatment approaches and empirical studies—an overview. *Z. Psychosom. Med. Psychother.* 65:162–77

Fonagy P. 2000. Attachment and borderline personality disorder. *J. Am. Psychoanal. Assoc.* 48:1129–46

Fonagy P, Lemma A, Target M, O'Keeffe S, Constantinou MP, et al. 2019. Dynamic interpersonal therapy for moderate to severe depression: a pilot randomized controlled and feasibility trial. *Psychol. Med.* In press. **https://doi.org/10.1017/S0033291719000928**

Fonagy P, Luyten P. 2009. A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Dev. Psychopathol.* 21:1355–81

Fonagy P, Luyten P. 2016. A multilevel perspective on the development of borderline personality disorder. In *Developmental Psychopathology*, Vol. 3: *Maladaptation and Psychopathology*, ed. D Cicchetti, pp. 726–92. New York: John Wiley & Sons. 3rd ed.

Comprehensive review of the mentalizing approach to borderline personality disorder and associated conditions.

Fonagy P, Luyten P. 2018. Conduct problems in youth and the RDoC approach: a developmental, evolutionary-based view. *Clin. Psychol. Rev.* 64:57–76

Fonagy P, Luyten P, Allison E. 2015. Epistemic petrification and the restoration of epistemic trust: a new conceptualization of borderline personality disorder and its psychosocial treatment. *J. Personal. Disord.* 29:575–609

Fonagy P, Luyten P, Allison E, Campbell C. 2017a. What we have changed our minds about: part 1. Borderline personality disorder as a limitation of resilience. *Borderline Personal. Disord. Emot. Dysregul.* 4:11

This paper discusses in detail the shift in the mentalizing approach to a social-communicative approach.

Fonagy P, Luyten P, Allison E, Campbell C. 2017b. What we have changed our minds about: part 2. Borderline personality disorder, epistemic trust and the developmental significance of social communication. *Borderline Personal. Disord. Emot. Dysregul.* 4:9

Fonagy P, Luyten P, Bateman A. 2017c. Treating borderline personality disorder with psychotherapy: Where do we go from here? *JAMA Psychiatry* 74:316–17

Fonagy P, Luyten P, Moulton-Perkins A, Lee YW, Warren F, et al. 2016. Development and validation of a self-report measure of mentalizing: the Reflective Functioning Questionnaire. *PLOS ONE* 11:e0158678

Fonagy P, Steele H, Steele M. 1991a. Maternal representations of attachment during pregnancy predict the organization of infant-mother attachment at one year of age. *Child Dev.* 62:891–905

Fonagy P, Steele M, Steele H, Moran GS, Higgitt AC. 1991b. The capacity for understanding mental states: the reflective self in parent and child and its significance for security of attachment. *Infant Ment. Health J.* 12:201–18

Fonagy P, Target M, Steele H, Steele M. 1998. *Reflective Functioning Scale Manual*. London, UK: Univ. Coll. London

Fonagy P, Twemlow SW, Vernberg EM, Nelson JM, Dill EJ, et al. 2009. A cluster randomized controlled trial of child-focused psychiatric consultation and a school systems-focused intervention to reduce aggression. *J. Child Psychol. Psychiatry* 50:607–16

Fraley RC. 2002. Attachment stability from infancy to adulthood: meta-analysis and dynamic modeling of developmental mechanisms. *Personal. Soc. Psychol. Rev.* 6:123–51

Pioneering paper demonstrating the importance of environment in explaining stability and change in attachment.

Fraley RC, Roberts BW. 2005. Patterns of continuity: a dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course. *Psychol. Rev.* 112:60–74

George C, Kaplan N, Main M. 1985. *The Adult Attachment Interview*. Train. Man., Dep. Psychol., Univ. Calif., Berkeley

Gergely G. 2013. Ostensive communication and cultural learning: the natural pedagogy hypothesis. In *Agency and Joint Attention*, ed. J Metcalfe, HS Terrace, pp. 139–51. Oxford, UK: Oxford Univ. Press

Groh AM, Roisman GI, van Ijzendoorn MH, Bakermans-Kranenburg MJ, Fearon RP. 2012. The significance of insecure and disorganized attachment for children's internalizing symptoms: a meta-analytic study. *Child Dev*. 83:591–610

Gunderson JG, Herpertz SC, Skodol AE, Torgersen S, Zanarini MC. 2018. Borderline personality disorder. *Nat. Rev. Dis. Primers* 4:18029

Herpertz SC, Bertsch K. 2015. A new perspective on the pathophysiology of borderline personality disorder: a model of the role of oxytocin. *Am. J. Psychiatry* 172:840–51

Hielscher E, Whitford TJ, Scott JG, Zopf R. 2019. When the body is the target—representations of one's own body and bodily sensations in self-harm: a systematic review. *Neurosci. Biobehav. Rev.* 101:85–112

Hiirola A, Pirkola S, Karukivi M, Markkula N, Bagby RM, et al. 2017. An evaluation of the absolute and relative stability of alexithymia over 11 years in a Finnish general population. *J. Psychosom. Res.* 95:81–87

Jewell T, Collyer H, Gardner T, Tchanturia K, Simic M, et al. 2016. Attachment and mentalization and their association with child and adolescent eating pathology: a systematic review. *Int. J. Eat. Disord.* 49:354–73

Jones JD, Fraley RC, Ehrlich KB, Stern JA, Lejuez CW, et al. 2018. Stability of attachment style in adolescence: an empirical test of alternative developmental processes. *Child Dev*. 89:871–80

Katznelson H. 2014. Reflective functioning: a review. *Clin. Psychol. Rev.* 34:107–17

Keefe JR, Huque ZM, DeRubeis RJ, Barber JP, Milrod BL, Chambless DL. 2019. In-session emotional expression predicts symptomatic and panic-specific reflective functioning improvements in panic-focused psychodynamic psychotherapy. *Psychotherapy* 56:514–25

Klahr AM, Burt SA. 2014. Elucidating the etiology of individual differences in parenting: a meta-analysis of behavioral genetic research. *Psychol. Bull.* 140:544–86

Knutson KM, Mah L, Manly CF, Grafman J. 2007. Neural correlates of automatic beliefs about gender and race. *Hum. Brain Mapp.* 28:915–30

Kobak R, Zajac K, Abbott C, Zisk A, Bounoua N. 2017. Atypical dimensions of caregiver–adolescent interaction in an economically disadvantaged sample. *Dev. Psychopathol.* 29:405–16

Kokkinos CM, Kakarani S, Kolovou D. 2016. Relationships among shyness, social competence, peer relations, and theory of mind among pre-adolescents. *Soc. Psychol. Educ.* 19:117–33

Konner M. 2010. *The Evolution of Childhood*. Cambridge, MA: Belknap

Krohn A. 1974. Borderline "empathy" and differentiation of object representations: a contribution to the psychology of object relations. *Int. J. Psychoanal. Psychother.* 3:142–65

Kvarstein EH, Pedersen G, Folmo E, Urnes O, Johansen MS, et al. 2019. Mentalization-based treatment or psychodynamic treatment programmes for patients with borderline personality disorder—the impact of clinical severity. *Psychol. Psychother.* 92:91–111

Lackner CL, Bowman LC, Sabbagh MA. 2010. Dopaminergic functioning and preschoolers' theory of mind. *Neuropsychologia* 48:1767–74

Lane JD, Harris PL. 2015. The roles of intuition and informants' expertise in children's epistemic trust. *Child Dev*. 86:919–26

Levy KN, Meehan KB, Kelly KM, Reynoso JS, Weber M, et al. 2006. Change in attachment patterns and reflective function in a randomized control trial of transference-focused psychotherapy for borderline personality disorder. *J. Consult. Clin. Psychol.* 74:1027–40

Li S, Zhang B, Guo Y, Zhang J. 2015. The association between alexithymia as assessed by the 20-item Toronto Alexithymia Scale and depression: a meta-analysis. *Psychiatry Res.* 227(1):1–9

Lieberman MD. 2007. Social cognitive neuroscience: a review of core processes. *Annu. Rev. Psychol.* 58:259–89

Lombardo MV, Lai MC, Baron-Cohen S. 2019. Big data approaches to decomposing heterogeneity across the autism spectrum. *Mol. Psychiatry* 24:1435–50

Long M, Verbeke W, Ein-Dor T, Vrticka P. 2020. A functional neuro-anatomical model of human attachment (*NAMA*): insights from first- and second-person social neuroscience. *Cortex* 126:281–321

Luyten P, Blatt SJ, Fonagy P. 2013. Impairments in self structures in depression and suicide in psychodynamic and cognitive behavioral approaches: implications for clinical practice and research. *Int. J. Cogn. Ther.* 6:265–79

Luyten P, Campbell C, Fonagy P. 2020. Borderline personality disorder, complex trauma, and problems with self and identity: a social-communicative approach. *J. Personal.* 88:88–105

Luyten P, De Meulemeester C, Fonagy P. 2019a. Psychodynamic therapy in patients with somatic symptom disorder. In *Contemporary Psychodynamic Psychotherapy: Evolving Clinical Practice*, ed. D Kealy, JS Ogrodniczuk, pp. 191–206. Philadelphia: Academic

Luyten P, Fonagy P. 2015. The neurobiology of mentalizing. *Personal. Disord.* 6:366–79

Luyten P, Fonagy P. 2018. The stress–reward–mentalizing model of depression: an integrative developmental cascade approach to child and adolescent depressive disorder based on the Research Domain Criteria (RDoC) approach. *Clin. Psychol. Rev.* 64:87–98

Luyten P, Fonagy P. 2019. Mentalizing and trauma. See Bateman & Fonagy 2019b, pp. 79–99

Luyten P, Malcorps S, Fonagy P, Ensink K. 2019b. Assessment of mentalizing. See Bateman & Fonagy 2019b, pp. 37–62

Luyten P, Mayes LC, Nijssens L, Fonagy P. 2017a. The parental reflective functioning questionnaire: development and preliminary validation. *PLOS ONE* 12:e0176218

Luyten P, Nijssens L, Fonagy P, Mayes LC. 2017b. Parental reflective functioning: theory, research, and clinical applications. *Psychoanal. Study Child* 70:174–99

Madigan S, Atkinson L, Laurin K, Benoit D. 2013. Attachment and internalizing behavior in early childhood: a meta-analysis. *Dev. Psychol.* 49:672–89

Mascaro O, Sperber D. 2009. The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition* 112:367–80

Mayes LC. 2006. Arousal regulation, emotional flexibility, medial amygdala function, and the impact of early experience: comments on the paper of Lewis et al. *Ann. N.Y. Acad. Sci.* 1094:178–92

McQuaid N, Bigelow AE, McLaughlin J, MacLean K. 2008. Maternal mental state language and preschool children's attachment security: relation to children's mental state language and expressions of emotional understanding. *Soc. Dev.* 17:61–83

Meins E. 2013. Sensitive attunement to infants' internal states: operationalizing the construct of mind-mindedness. *Attach. Hum. Dev.* 15:524–44

Meins E, Fernyhough C, Fradley E, Tuckey M. 2001. Rethinking maternal sensitivity: mothers' comments on infants' mental processes predict security of attachment at 12 months. *J. Child Psychol. Psychiatry* 42:637–48

Meins E, Fernyhough C, Wainwright R, Clark-Carter D, Das Gupta M, et al. 2003. Pathways to understanding mind: construct validity and predictive validity of maternal mind-mindedness. *Child Dev. Perspect.* 74:1194–211

Meins E, Fernyhough C, Wainwright R, Das Gupta M, Fradley E, Tuckey M. 2002. Maternal mind-mindedness and attachment security as predictors of theory of mind understanding. *Child Dev.* 73:1715–26

Meins E, Harris-Waller J, Lloyd A. 2008. Understanding alexithymia: associations with peer attachment style and mind-mindedness. *Personal. Individ. Differ.* 45:146–52

Nolte T, Bolling DZ, Hudac CM, Fonagy P, Mayes L, Pelphrey KA. 2013. Brain mechanisms underlying the impact of attachment-related stress on social cognition. *Front. Hum. Neurosci.* 7:816

Nolte T, Guiney J, Fonagy P, Mayes LC, Luyten P. 2011. Interpersonal stress regulation and the development of anxiety disorders: an attachment-based developmental framework. *Front. Behav. Neurosci.* 5:55

One of the first studies demonstrating longitudinal associations between parental mentalizing and child mentalizing.

Oppenheim D, Koren-Karie N. 2013. The insightfulness assessment: measuring the internal processes underlying maternal sensitivity. *Attach. Hum. Dev.* 15:545–61

Oppenheim D, Koren-Karie N, Sagi A. 2001. Mothers' empathic understanding of their preschoolers' internal experience: relations with early attachment. *Int. J. Behav. Dev.* 25:16–26

Perroud N, Badoud D, Weibel S, Nicastro R, Hasler R, et al. 2018. Mentalization in adults with attention deficit hyperactivity disorder: comparison with controls and patients with borderline personality disorder. *Psychiatry Res.* 256:334–41

Philips B, Wennberg P, Konradsson P, Franck J. 2018. Mentalization-based treatment for concurrent borderline personality disorder and substance use disorder: a randomized controlled feasibility study. *Eur. Addict. Res.* 24(1):1–8

Pinquart M, Feussner C, Ahnert L. 2013. Meta-analytic evidence for stability in attachments from infancy to early adulthood. *Attach. Hum. Dev.* 15:189–218

Redfern S, Wood S, Lassri D, Cirasola A, West G, et al. 2018. The Reflective Fostering Programme: background and development of a new approach. *Adopt. Foster.* 42:234–48

Ripoll LH, Snyder R, Steele H, Siever LJ. 2013. The neurobiology of empathy in borderline personality disorder. *Curr. Psychiatry Rep.* 15:344

Robinson P, Hellier J, Barrett B, Barzdaitiene D, Bateman A, et al. 2016. The NOURISHED randomised controlled trial comparing mentalisation-based treatment for eating disorders (MBT-ED) with specialist supportive clinical management (SSCM-ED) for patients with eating disorders and symptoms of borderline personality disorder. *Trials* 17(1):549

Robinson P, Skårderud F, Sommerfeldt B. 2019. Eating disorders and mentalizing. In *Hunger: Mentalization-Based Treatments for Eating Disorders*, pp. 35–49. Cham, Switz.: Springer Int.

Rosso AM, Airaldi C. 2016. Intergenerational transmission of reflective functioning. *Front. Psychol.* 7:1903

Rosso AM, Viterbori P, Scopesi AM. 2015. Are maternal reflective functioning and attachment security associated with preadolescent mentalization? *Front. Psychol.* 6:1134

Rozer JJ, Volker B. 2016. Does income inequality have lasting effects on health and trust? *Soc. Sci. Med.* 149:37–45

Sabbagh MA. 2004. Understanding orbitofrontal contributions to theory-of-mind reasoning: implications for autism. *Brain Cogn.* 55:209–19

Schattner E, Shahar G, Abu-Shakra M. 2008. "I used to dream of lupus as some sort of creature": chronic illness as an internal object. *Am. J. Orthopsychiatry* 78:466–72

Shafto P, Eaves B, Navarro DJ, Perfors A. 2012. Epistemic trust: modeling children's reasoning about others' knowledge and intent. *Dev. Sci.* 15:436–47

Shamay-Tsoory SG, Aharon-Peretz J. 2007. Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia* 45:3054–67

Shamay-Tsoory SG, Aharon-Peretz J, Perry D. 2009. Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain* 132:617–27

Sharp C, Fonagy P. 2008. The parent's capacity to treat the child as a psychological agent: constructs, measures and implications for developmental psychopathology. *Soc. Dev.* 17:737–54

Sharp C, Venta A, Vanwoerden S, Schramm A, Ha C, et al. 2016. First empirical evaluation of the link between attachment, social cognition and borderline features in adolescents. *Compr. Psychiatry* 64:4–11

Simonsen S, Euler S. 2019. Avoidant and narcissistic personality disorders. See Bateman & Fonagy 2019b, pp. 351–67

Slade A. 2005. Parental reflective functioning: an introduction. *Attach. Hum. Dev.* 7:269–81

Slade A, Holland ML, Ordway MR, Carlson EA, Jeon S, et al. 2020. *Minding the Baby®*: enhancing parental reflective functioning and infant attachment in an attachment-based, interdisciplinary home visiting program. *Dev. Psychopathol.* 32:123–37

Smits ML, Feenstra DJ, Eeren HV, Bales DL, Laurenssen EMP, et al. 2020. Day hospital versus intensive out-patient mentalisation-based treatment for borderline personality disorder: multicentre randomised clinical trial. *Br. J. Psychiatry* 216(2):79–84

Sng O, Neuberg SL, Varnum MEW, Kenrick DT. 2018. The behavioral ecology of cultural psychological variation. *Psychol. Rev.* 125:714–43

Solomonov N, Falkenström F, Gorman BS, McCarthy KS, Milrod B, et al. 2019. Differential effects of alliance and techniques on Panic-Specific Reflective Function and misinterpretation of bodily sensations in two treatments for panic. *Psychother. Res.* 30(1):97–111

Sperber D, Clement F, Heintz C, Mascaro O, Mercier H, et al. 2010. Epistemic vigilance. *Mind Lang.* 25:359–93

Steele H, Perez A, Segal F, Steele M. 2016. Maternal Adult Attachment Interview (AAI) collected during pregnancy predicts reflective functioning in AAIs from their first-born children 17 years later. *Int. J. Dev. Sci.* 10:117–24

Steele H, Steele M. 2005. Understanding and resolving emotional conflict: the London Parent-Child Project. In *Attachment from Infancy to Adulthood: The Major Longitudinal Studies*, ed. KE Grossmann, K Grossmann, E Waters, pp. 137–64. New York: Guilford

Stepp SD, Lazarus SA, Byrd AL. 2016. A systematic review of risk factors prospectively associated with borderline personality disorder: taking stock and moving forward. *Personal. Disord.* 7:316–23

Stevens JS, Jovanovic T. 2019. Role of social cognition in post-traumatic stress disorder: a review and meta-analysis. *Genes Brain Behav.* 18:e12518

Suchman NE, DeCoste C, Borelli JL, McMahon TJ. 2018. Does improvement in maternal attachment representations predict greater maternal sensitivity, child attachment security and lower rates of relapse to substance use? A second test of Mothering from the Inside Out treatment mechanisms. *J. Subst. Abuse Treat.* 85:21–30

Taubner S, Gablonski T-C, Fonagy P. 2019. Conduct disorder. See Bateman & Fonagy 2019b, pp. 301–21

Taubner S, Horz S, Fischer-Kern M, Doering S, Buchheim A, Zimmermann J. 2013. Internal structure of the Reflective Functioning Scale. *Psychol. Assess.* 25:127–35

Taubner S, Zimmermann L, Ramberg A, Schroder P. 2016. Mentalization mediates the relationship between early maltreatment and potential for violence in adolescence. *Psychopathology* 49:236–46

Tomasello M. 2010. *Origins of Human Communication*. Cambridge, MA: MIT Press

Tomasello M. 2018. Great apes and human development: a personal history. *Child Dev. Perspect.* 12:189–93

Tomasello M, Vaish A. 2013. Origins of human cooperation and morality. *Annu. Rev. Psychol.* 64:231–55

Troyer D, Greitemeyer T. 2018. The impact of attachment orientations on empathy in adults: considering the mediating role of emotion regulation strategies and negative affectivity. *Personal. Individ. Differ.* 122:198–205

Uddin LQ, Iacoboni M, Lange C, Keenan JP. 2007. The self and social cognition: the role of cortical midline structures and mirror neurons. *Trends Cogn. Sci.* 11:153–57

Verhage ML, Fearon RMP, Schuengel C, van IJzendoorn MH, Bakermans-Kranenburg MJ, et al. 2018. Examining ecological constraints on the intergenerational transmission of attachment via individual participant data meta-analysis. *Child Dev.* 89:2023–37

Verhage ML, Schuengel C, Madigan S, Fearon RM, Oosterman M, et al. 2016. Narrowing the transmission gap: a synthesis of three decades of research on intergenerational transmission of attachment. *Psychol. Bull.* 142:337–66

Viding E, McCrory E, Seara-Cardoso A. 2014. Psychopathy. *Curr. Biol.* 24:R871–74

Volkert J, Hauschild S, Taubner S. 2019. Mentalization-based treatment for personality disorders: efficacy, effectiveness, and new developments. *Curr. Psychiatry Rep.* 21(4):25

Vrticka P, Vuilleumier P. 2012. Neuroscience of human social interactions and adult attachment style. *Front. Hum. Neurosci.* 6:212

Wilson DS, Wilson EO. 2007. Rethinking the theoretical foundation of sociobiology. *Q. Rev. Biol.* 82:327–48

Zaccagnino M, Cussino M, Preziosa A, Veglia F, Carassa A. 2015. Attachment representation in institutionalized children: a preliminary study using the Child Attachment Interview. *Clin. Psychol. Psychother.* 22:165–75

Zeegers MAJ, Colonnesi C, Stams GJM, Meins E. 2017. Mind matters: a meta-analysis on parental mentalization and sensitivity as predictors of infant–parent attachment. *Psychol. Bull.* 143:1245–72

Authoritative review of prospective studies of the association between early adversity and borderline personality disorder.

Comprehensive review of comparative research on the genesis and role of shared intentionality in humans.

Groundbreaking meta-analysis of the association between parental mentalizing, sensitivity, and infant attachment.