

# Mentalization: Ontogeny, Assessment, and Application in the Treatment of Borderline Personality Disorder

Lois W. Choi-Kain, M.D., M.Ed.

John G. Gunderson, M.D.

This article aims to review the development of the concept of mentalization, its applications in the understanding and treatment of borderline personality disorder, and the issue of its assessment. While conceptually derivative of theory of mind, Fonagy's concept of mentalization concerns more affectively and interpersonally complex understandings of oneself and others, reflecting abilities that enable an individual not only to navigate the social world effectively but also to develop an enriched, stable sense of self. The components of mentalization can be organized around self-/other-oriented, implicit/explicit, and cognitive/affective dimensions. Concepts of mindfulness, psychological mindedness, empathy, and affect consciousness are shown to par-

tially overlap with mentalization within these three dimensions. Mentalization is assessed by the measure of reflective function, a scale to be used adjunctively on semistructured narrative interviews such as the Adult Attachment Interview. Its validity has not been fully tested, and its usage has been hampered by the time and expense it requires. Although the concept of mentalization is a useful heuristic that enables clinicians to adopt a coherent treatment approach, it may be too broad and multifaceted to be operationalized as a marker for specific forms of psychopathology such as borderline personality disorder. Research elucidating the relationship between reflective function, overlapping concepts, and features of borderline psychopathology is needed.

(*Am J Psychiatry* 2008; 165:1127–1135)

**T**he concept of mentalization, popularized in the last 15 years by Peter Fonagy and collaborators (1–3), describes the way humans make sense of their social world by imagining the *mental* states (e.g., beliefs, motives, emotions, desires, and needs) that underpin their own and others' behaviors in interpersonal interactions. Integrating parallel streams of thought from the fields of psychoanalysis, developmental psychology, and cognitive neuroscience, Fonagy has elaborated a theory of how the capacity to mentalize develops in early childhood and, alternatively, how deviations from this normal developmental path result in severe forms of adult psychopathology, most notably borderline personality disorder.

Using this theory of borderline personality disorder as a disorder of mentalization, Fonagy and colleague Anthony Bateman developed a psychodynamically oriented, manualized psychotherapy program for borderline personality disorder called mentalization-based treatment (3). Mentalization-based treatment became the second psychotherapeutic treatment for borderline personality disorder to be empirically validated by randomized, controlled trials as more effective than nonspecialized psychiatric treatment. Mentalization-based treatment reduced depressive symptom profiles, attempts at suicide and self-harm, and inpatient hospital stays while increasing social functioning in patients with borderline personality disorder,

both at the end of partial hospital treatment and with continued gains at 18 months after treatment (4, 5). Moreover, an 8-year follow-up study demonstrates that these outcomes in Bateman and Fonagy's randomized controlled trial continue to differentiate the mentalization-based treatment group from the group in treatment as usual (6). Transference-focused psychotherapy (7), the other empirically validated, manualized psychodynamic treatment for borderline personality disorder (8), appropriated the existing measure of mentalization, referred to as reflective function (unpublished Reflexive Function Manual: version 5.0 for application to the Adult Attachment Interview by Fonagy P, Steele M, Steele H, Target M), as an important outcome measure in their randomized, controlled trial (9).

Fonagy's mentalization concept was first introduced in the context of a theory and treatment for borderline personality disorder, but its usage has expanded into an impressive array of clinical domains, including the treatment of professionals in crisis (10), families (11), high-risk parent-infant dyads (12), eating disorders (13), and school-based communities to minimize violence (14). It has been proposed that mentalization is the essential mechanism by which all effective therapies work (15). Quickly adapted into psychiatric vernacular, the term *mentalization* has been used with variable meanings. The broad territory of

This article is the subject of a [CME](#) course (p. 1223).

the concept paradoxically contributes to its familiarity as well as to its ambiguity. Even proponents of the concept have referred to the concept as “ungainly” (16) and “all-encompassing...potentially beyond manageable bounds” (17).

This review will attempt to clarify and critically examine the concept, its application to borderline personality disorder, and its assessment. We will start by outlining Fonagy’s definition of mentalization; tracing the conceptual origins of the term; reviewing closely related concepts, notably mindfulness, psychological mindedness, empathy, and affect consciousness; and describing the existing methodology used for its assessment.

## Definition of the Concept

Bateman and Fonagy (3) define mentalization as “the mental process by which an individual *implicitly* and *explicitly* interprets the actions of *himself* or *herself* and *others* as meaningful on the basis of intentional mental states such as personal desires, needs, feelings, beliefs, and reasons. Within this definition, Bateman and Fonagy identify three dimensions of mentalization: the first related to two modes of functioning (i.e., implicit and explicit), the second related to two objects (i.e., self and other), and the third related to two aspects (i.e., cognitive and affective) of both the content and process of mentalizing.

*Implicit* mentalization refers to unconscious, automatic, or procedural operations of an individual’s ability to imagine his own and others’ mental states. Jon Allen (17) refers to conversational turn-taking as an example of *implicit* mentalization in operation. Without deliberate reflection, individuals naturally and instinctually hold the mind of their conversation partners in mind, anticipating when the other might want to respond in turn. In contrast, *explicit* mentalization involves deliberately exercised and conscious uses. The activities of psychotherapy provide an example of *explicit* mentalization. The therapist works to consciously and deliberately imagine the mental states of the patient and also encourages his patient to consciously and deliberately focus on her mental states. While the *implicit* and *explicit* modes define the two poles of the process of mentalizing, they are not mutually exclusive and completely discrete (17). Individuals can alternate between these two modes and use them simultaneously. For example, a therapist can be consciously reflecting on the patient’s mental states but can also be attuned to the patient in more unconscious, intuitive ways.

With respect to objects, i.e., the self and the other, in the mentalization framework, each has a set of *mental states*, including feelings, thoughts, motives, intentions, beliefs, desires, and needs, to name a few. Fundamentally, the two objects in this framework mentalize interactively. The process of imagining one’s own thoughts and feelings potentially determines one’s idea of what is in the other’s mind and vice versa. For example, a person may become aware

that he or she feels angry at a friend and then assumes the friend is also angry at him or her. Through talking and interacting, both parties can develop a more complex, enriched, and realistic idea of what is going on in their minds. In short, through mentalizing, the two friends can come to understand a misunderstanding. As illustrated by this example, the mental contents of both objects are dynamic: feelings, thoughts, and intentions constantly shift in response to changes in the interpersonal milieu.

A third dimension of the mentalization concept, in addition to the implicit/explicit dimension and the self/other dimension, relates to its cognitive and affective aspects. The *content* of mentalizing activity, that is, the “intentional mental states” in oneself and others, can be cognitively focused and affectively laden to varying degrees. Additionally, there are cognitive and affective aspects of the *process* of mentalizing. Mentalization requires a panoply of intact cognitive skills that enable individuals to imagine mental states with plausibility, flexibility, and complexity, but it optimally integrates this cognitive realm concerning reason and insight with emotion. The integration of cognitive and affective aspects of both the process and content of understanding mental states allows individuals to “feel clearly” and enhances “emotional knowing” (17).

The domain of mentalization appears “all-encompassing,” so we will use these three dimensions of the concept (i.e., implicit/explicit modes, self/other objects, and cognitive/affective aspects) to anchor this review of the mentalization concept and its relationship to its conceptual sources and conceptual overlaps.

## Ontogeny of the Concept

In a 1991 article titled “Thinking About Thinking,” Fonagy introduced his concept of mentalization, simply defined as “the capacity to conceive of conscious and unconscious mental states in oneself and others” (1). Fonagy’s new use of the term combined the psychoanalytic idea of symbolization with the scientific and philosophical concept of *theory of mind* (18). Frequently and without clear definition, the term *mentalization* had been intermittently scattered throughout psychoanalytic writings since the 1960s (19). In their review of this literature, Lecours and Bouchard (20) report that the concept of mentalization signifies the basic intrapsychic transformation of one’s own inchoate somatic experiences into increasingly organized images, ideas, and words that could be modified, linked, and communicated. Other terms used interchangeably with mentalization in this literature include “metabolization” and “representation” (21), “symbolization” or “symbol formation” (22), “secondary mental processes” (23), and “alpha-function” (24). An individual’s felt experience can be perceived in several forms ranging from physical (somatic and motor) to internally visualized (im-

ages and dreams) to interpersonally communicated, verbally articulated forms.

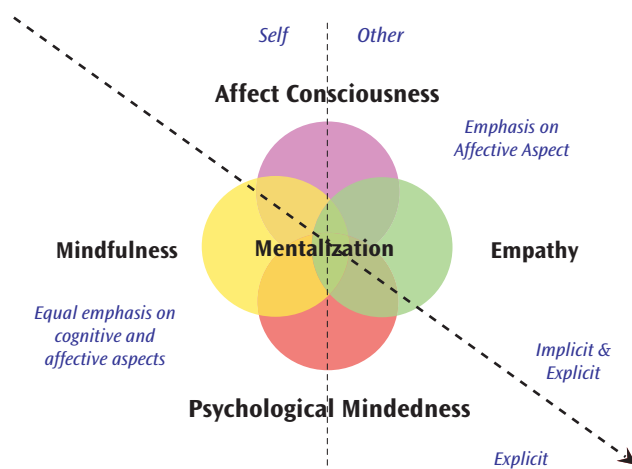
To illustrate the distinctions among these forms of mentalization just described, we can consider the experience of anxiety on these three levels of representation. Starting with somatic and motor representations, one can experience stomachache, sweaty palms, and pacing as markers of anxiety. At this level of representation, anxiety is experienced more viscerally than mentally. To move to an increasingly *conscious* level of mental representation, one can imagine anxiety-laden images or dreams, like showing up at school without one's homework. Finally, at the most mentalized or self-reflective level, one can conceive of a mental representation of anxiety in an idea or thought such as "I am anxious because I am getting close to my boyfriend, and I am afraid of intimacy." This verbally articulated representation of one's affect state is not only the most easily and unambiguously communicated form but also the easiest to potentially link to other representations of felt experience. Of importance, all of these representations of anxiety facilitate awareness of one's own internal state. Greater awareness of internal experience can be achieved through linking multiple representations of experience together.

Theory of mind refers to mental faculties that allow an individual to first appreciate the existence of different mental states in *others* and then to accurately identify others' mental states (e.g., intentions, motives, beliefs, desires, and feelings) in order to interpret their behavior. The term *theory of mind* was introduced into the scientific literature by primatologists (18) who observed a chimpanzee's ability to understand the intentions of an actor in film clips, which enabled her to predict the actor's next move. Research on this particular social cognitive capacity expanded after developmental psychologists (25) introduced the "false-belief task," the first experimental paradigm for studying theory of mind. In this experimental paradigm, children are told a character named Maxi puts a candy bar in a cupboard in the kitchen and then leaves the kitchen, after which time his mother comes in and moves the candy bar to a drawer. The child is then asked where Maxi will look for the candy. A child who has developed theory of mind will understand that Maxi falsely believes the chocolate is in the cupboard. Using this and other versions of the false-belief task, researchers have shown that a majority of children master this task between 4 and 6 years of age. A number of other experimental paradigms targeting theory of mind functions have been developed for use in both children and adults. The operationalization of theory of mind into experimental protocols has enabled researchers to localize its function to specific regions in the brain (26) and also to differentiate individuals with high-functioning autism and Asperger's syndrome from normal comparison subjects (27). Baron-Cohen et al.'s conception of the interpersonal functioning difficulties in autism as theory of mind deficits (28) influenced Fonagy

to apply the theory of mind concept to borderline personality disorder. In his appropriation of the theoretically derived and clinically applied broad psychoanalytic term *mentalization*, Fonagy brought its internally or self-oriented and affectively rich dimensions to bear on the more empirically derived, externally or other-oriented, cognitively focused construct of theory of mind.

Fonagy outlined several salient propositions relevant to this new chimeric term. First and foremost, Fonagy emphasized a developmental model, contextualizing the formation of mentalization in the setting of secure early attachment relationships. More specifically, Fonagy argued that the primary caretaker's *marked* and *contingent* mirroring of a child's internal states within a secure attachment facilitates that child's development of a capacity to mentalize. To clarify through an example, consider an infant who is distressed and crying. His caretaker responds, not by crying and looking distressed herself but rather with an exaggerated frown and furrowed brow, expressing both concern and loving. This constitutes an expression that not only mirrors the imagined emotional state of the baby but modifies it as a "re-presentation" (29) so that it is clearly congruent with the infant's state but also differentiated as the caregiver's response to it (2). By mirroring the internal states of the child in a way that is both *marked* (i.e., the caretaker's metabolized "representation" of it) and *contingent* (i.e., accurate and responsive), the caretaker helps the child solidify an understanding of an internal experience that he is only dimly aware of initially (2, 30). The caretaker's mirroring helps the child convert a felt, physical, sensory experience into a contained mental, conscious awareness. This mirroring process facilitates the development of a child's capacity to mentalize his internal experience, which in turn enables him to regulate his affect and distress. As was earlier identified by Kohut (31), this mirroring process also theoretically contributes to a *sense of self*.

This developmental theory states that when the capacity to mentalize insufficiently develops, as is the case in borderline personality disorder, capacities for self-awareness and self-regulation remain impaired (2). Here, while Fonagy suggested that the borderline personality disorder subject's failures in mentalization represent a developmental deficit, born in the context of insecure attachment with insufficient mirroring, he also argued that subjects with borderline personality disorder who have backgrounds of trauma employ a defensive inhibition of mentalization, as a self-protective way to avoid considering the malicious intents of an abusive or neglectful figure. This original model of the mentalization deficits in borderline personality disorder combined theories of deficit and defense, which some authors have framed hard to reconcile (32), while others have claimed as quite compatible (33). In Fonagy's trauma model, the defensive inhibition of mentalization presumably arose in children who had an intact capacity to mentalize. But in his developmental

**FIGURE 1. Mentalization Map: Dimensional Features and Conceptual Overlaps<sup>a</sup>**

<sup>a</sup> This Venn diagram maps the conceptual overlaps between mentalization and four related concepts including mindfulness, psychological mindedness, empathy, and affect consciousness, which are represented by the four circles. The lines which bifurcate the diagram according to its three dimensions (i.e., self-/other-oriented, implicit/explicit, and cognitive/affective) are dashed to illustrate the permeable and nonabsolute nature of these divisions. In the self/other dimension, mindfulness focuses more on mental states within oneself, while empathy is primarily understood in terms of one's imagination of mental states within others. Both affect consciousness and psychological mindedness concern both sides of the self and other distinction. While mindfulness and psychological mindedness emphasize both cognitive and affective aspects of mental states and function explicitly, affect consciousness and empathy relate more primarily to affective mental contents and function both explicitly and implicitly. Mentalization lies at the intersection of these concepts but the boundaries between them are not distinctly drawn.

model, he suggests that in the absence of secure early attachments, the capacity to mentalize inadequately develops, leading to a deficit.

More recently, Fonagy and Bateman (34) proposed a more complex relationship between early attachment, trauma, and borderline personality disorder that incorporates three mechanisms by which mentalization becomes destabilized or impaired in borderline personality disorder: first as a deficit, second as a defense, and third as a derailment due to dysregulated affect. Their theory now notes that a variety of factors other than trauma, such as genetic contributions and temperament, may contribute to a suboptimal fit between infant and caregiver that interferes with the establishment of a secure attachment as well as the process of marked contingent mirroring. Incorporating research on family environment that suggests that factors such as neglect, lack of support, excessive control, and emotional maltreatment are predictive of borderline personality disorder, Fonagy and Bateman suggest that the family context may impair the development of mentalization both in combination with and apart from any incidents of trauma. This kind of family context, trauma, and even innate biological factors may cause a

dysfunction and hypersensitivity in the stress-response system, leading to a cascade of hyperarousal, affective dysregulation, and inhibition of the orbitofrontal cortex, a brain region associated with mentalizing activity (34).

## Conceptual Overlaps

The broad nature of Fonagy's concept of mentalization contributes to its appeal as well as its potential to be confusing. The territory of the concept spans a number of other "conceptual cousins" (17), including mind blindness, emotional intelligence, insight, rationality, and imagination (17); theory of mind (18); and a number of psychoanalytic concepts including alpha function (24) and potential space (35, 36). Reviews of the overlap between mentalization and a number of these concepts have been published elsewhere (17, 20, 36). This review of conceptual overlaps is restricted to the related concepts of mindfulness, psychological mindedness, empathy, and affect consciousness, all of which have been operationalized into empirical measures and studied in relation to borderline personality disorder or integrated into psychotherapeutic treatments. In order to elucidate the domain of the mentalization concept, we will examine each of these four "conceptual cousins" and the ways in which they overlap with each other. A graphical depiction of the conceptual overlaps is summarized in Table 1 and Figure 1.

## Mindfulness

Mindfulness, defined as "keeping one's consciousness alive to the present reality" (37, 38), is a concept originally derived from Eastern meditation practice and later borrowed in a number of treatment modalities (39) including dialectical behavioral therapy (40), which signifies skills of observing and describing one's own experience while participating nonjudgmentally. Mindfulness has been conceptualized in a two-component model bifurcated into the domains of 1) attention regulation and 2) acceptance and openness to experience (41). Four skills that underpin mindfulness have been identified in factor analytic studies and include observing, describing, acting with awareness, and accepting without judgment (42). This clear and empirically developed deconstruction of the mindfulness concept has allowed it to be operationalized into research scales (38, 41) and several forms of psychotherapeutic treatment (39).

Mindfulness overlaps with mentalization within the observing and describing subscales. Both mindfulness and mentalization involve directing one's attention to one's own experience as a way to mitigate tendencies toward impulsivity and reactivity. Both also emphasize the integration of cognitive and affective aspects of mental states in encouraging simultaneous recognition and participation in internal experience. Mindfulness only overlaps with one of the two modes (explicit) and one of the two objects (self) within the mentalization concept (Table 1).



TABLE 1. Conceptual Overlaps of Mentalization, Mindfulness, Psychological Mindedness, Empathy, and Affect Consciousness

Components of Mentalization	Mindfulness	Psychological Mindedness	Empathy	Affect Consciousness
Implicit mode	No	No	Yes	No
Explicit mode	Yes	Yes	Yes	Yes
Self-oriented	Yes	Yes	Minor	Yes
Other oriented	No	Minor	Yes	Yes
Cognitive versus affective aspect	Cognitive=affective	Cognitive=affective	Affective>cognitive	Affective>cognitive

Three other distinctions exist between the two concepts. First, in mindfulness, one's own experience interacting with inanimate objects, and not just other people, is considered. Second, the practice of mindfulness is oriented to present experience, while the process of mentalization can concern the past, present, and future. Finally, mindfulness aims at acceptance of internal experience, whereas mentalization emphasizes the construction of representation and meaning related to these experiences.

### Psychological Mindedness

Appelbaum (43) defined psychological mindedness as "a person's ability to see relationships among thoughts, feelings, and actions, with the goal of learning the meanings and causes of his [own] experiences and behavior." Appelbaum identified four parts of the concept: 1) the skill to discern connections between meanings and causes of behaviors, which requires both intact cognition and "intuition and empathy"; 2) the goal of understanding the meaning of behaviors, which entails "an interest in the way minds work"; 3) "self-directed psychological thinking"; and 4) the ability to engage in psychological thinking in the context of treatment (43). Farber (44) proposed a similar definition of psychological mindedness, "a disposition to reflect upon the meaning and motivation of behaviors, thoughts, and feelings of oneself and others" that adds an interpersonal dimension to the term. Psychological mindedness was initially operationalized into a self-report measure to assess suitability for psychoanalysis (45). Factor analytic studies of the concept have been done (46) and its relationship to other psychological concepts has been studied (47, 48).

There are several areas of overlap between psychological mindedness and mentalization. The definitions of both terms overlap considerably, especially Farber's definition of psychological mindedness and Fonagy's definition of mentalization. Aspects of psychological mindedness as described by Appelbaum, such as "intuition and empathy" and "interest in the way minds work," appear to be relevant to mentalization. Although mentalization operates both implicitly and explicitly, psychological mindedness primarily concerns explicit or conscious consideration of mental states. Psychological mindedness also concerns the self and one's own mental states more so than it considers the others' mental states. Cognitive and affective aspects of internal experience are equally emphasized. Although interest in others' mental states is a factor within this concept, the actual ability to plausibly discern those mental states is not. The orientation of the

Psychological Mindedness Scale toward the conscious and deliberate aspects of thinking about one's own and other's internal experiences as well as interest and inclination to such thinking rather than an actual capability restricts its relevance to Fonagy's concept of mentalization (Table 1).

### Empathy

Human empathy has been the object of academic investigation in a number of disciplines. Decety and Jackson (49) published a comprehensive review of the concept of empathy, aimed at clarifying a model of the construct in order to facilitate further research. They note numerous definitions of the term but advance the definition that "empathy is a complex form of psychological inferences in which observation, memory, knowledge, and reasoning are combined to yield insights into the thoughts and feelings of others." What the many definitions and conceptions of the term *empathy* have in common is three parts: 1) an affective reaction that involves sharing of another person's emotional state, 2) a cognitive capacity to imagine other people's perspective ("perspective taking"), and 3) a stable ability to maintain a self-other distinction. A number of self-report measures as well as functional magnetic resonance imaging protocols developed to assess and elicit empathy have been applied in research (49).

The concept of empathy overlaps with mentalization considerably. Both mentalization and empathy involve appreciation of mental states in others, yet empathy adds the dimensions of sharing in those mental states and having empathic concern for others. Empathy is more other-oriented while mentalization is equally self- and other-oriented. Empathy can function in both implicit and explicit modes but is generally regarded in its more implicit mode. Although the process of empathy involves cognitive skill and experience of affect, its content is primarily affectively focused (Table 1).

### Affect Consciousness

Affect consciousness refers to the relationship between the "activation of basic affects and the individual's capacity to consciously perceive, reflect on and express these affect experiences" in terms of nine basic affects (50). Affect consciousness has been operationalized into empirical measures and a psychotherapy treatment model (50). The absence of affect consciousness may be understood in part as *alexithymia*, which literally means "no words for moods" (51). Of importance, alexithymia has been operationalized into a measure that has findings relevant to borderline personality disorder (52, 53).

The overlap between affect consciousness and mentalization (54) is partial but significant. The awareness, representation, and interpersonal communication of affectively laden mental states are at the heart of the mentalization concept as it is theoretically at the basis of affect regulation and conversely the locus of dysfunction in borderline personality disorder. *Mentalized affectivity* (2) involves the same components of processing affect as affect consciousness: identification of affects, processing affects, and expressing affects. Mentalized affectivity involves a re-appraisal of affective experience that contributes to affect regulation. Affect regulation enables individuals to mentalize and, conversely, mentalization enables individuals to regulate affect. However, distinctions between affect consciousness and mentalization exist. Affect consciousness focuses more on its explicit function, in the sense that it focuses on conscious awareness and expression of affect states. The affect consciousness concept is relevant to the affect states of both the self and the other (Table 1). However, the scope of mental states involved in affect consciousness is restricted to affectively oriented mental contents, whereas the scope of mental states in mentalization is much broader.

### Summary of Conceptual Overlap

Mentalization involves unconscious, automatic, and conscious deliberate application of one's capacity to understand both cognitive and affective aspects of one's own and others' mental states. As reviewed, mentalization overlaps significantly with other codified psychological constructs, such as mindfulness, psychological mindedness, empathy, and affect consciousness. Table 1 and Figure 1 summarize the conceptual overlaps of these other concepts with the essential components of the mentalization concept. These conceptual cousins have been deconstructed into constituent parts and empirically investigated (41–53). A critical review of these synonymous and overlapping concepts facilitates the mapping of the boundaries around and territory within the mentalization construct. By using a map of the overlaps between mentalization and its conceptual cousins, it is possible to use validated measures of mindfulness, psychological mindedness, empathy, and affect consciousness to conduct further research on the role of mentalization in borderline personality disorder and psychotherapy.

### Conceptual Application: Formulation of Borderline Personality Disorder

Fonagy and collaborators have formulated borderline personality disorder as a syndrome organized around an *unstable* capacity for mentalization. The capacity for mentalization normally develops in the context of secure early attachments; in later attachment contexts, that capacity is vulnerable (2, 3). When an individual is unable to mentalize, three “prementalistic” or “nonmentalistic” modes of

thinking become evident, which Fonagy refers to as *psychic equivalence mode*, *pretend mode*, and *teleological stance*. In *psychic equivalence*, individuals equate what is in their mind with what is in the world, such that no alternative perspectives on reality exist and the contents of the mind become unbearably real (e.g., posttraumatic stress disorder flashbacks, transient paranoid ideation). Pretend mode, the converse of psychic equivalence, involves “decoupling” (2) or “cutting loose” (17) what is in the mind from reality. For example, an individual with borderline personality disorder might find himself or herself lost in a self-critical pessimistic train of thought that is unanchored to and unmodulated by reality. Developmentally, children acquire the ability to mentalize when they integrate these two modes and understand internal experience and external reality as “linked, but separate” (3). Finally, in the teleological stance, an individual only accepts the existence of mental states when they are observed in concrete and physically evident ways. In the teleological stance, communication of internal mental states only occurs through action; ideas and even words mean very little. For example, cutting provides an index of internal pain and sexual intercourse or physical affection serves as an index for love or caring.

In the mentalization-based treatment model, the borderline personality disorder symptom profile is organized around these primitive nonmentalizing modes of thinking. The mentalization-based treatment model instructs clinicians to identify moments in which mentalizing fails and these prementalistic modes of functioning become manifest. Then the clinician and patient can work together to revive the patient's capacity to mentalize instead of reverting to psychic equivalence mode, pretend mode, or teleological stance. Fully functioning mentalization capacities are established when patients can integrate psychic equivalence and pretend modes of functioning, thereby linking what is in their mind with what is observable in reality in a way that is separate but related.

### Conceptual Measurement: Reflective Function

In revising the mentalization concept, Fonagy's ambition was to build a conceptual framework based on psychoanalytic theory, validated by scientific evidence, and applied effectively in clinical programs. Of importance, while he was developing the concept of mentalization, Fonagy was also involved in developing its measure, originally referred to as reflective self-function (55) and subsequently as reflective function. Initially designed as an adjunct to the Adult Attachment Interview (unpublished coding manual by Kaplan, Goldwyn, and Main, 1985), reflective function is a coding scheme that can be applied to a variety of clinical interviews, including the Parent Development Interview (unpublished protocol by Aber et al., 1985). The original scale expanded a subscale of the Adult

Attachment Interview called the Metacognitive Monitoring Scale, which regarded a subject's ability to reflect on his own thought process. Reflective function incorporated the core features of the Metacognitive Monitoring Scale, extending its scope to rate a subject's ability to reflect on his own and others' mental states.

The Reflective Function Scale's validity is primarily derived from research demonstrating the ability of mothers' prenatal reflective functioning to predict infant attachment security (55), a finding that has been replicated (29). Reflective function is also shown to differentiate "difficult to treat" personality disordered inpatients at Cassel Hospital from matched normal comparison subjects recruited from an ambulatory general medical clinic (56). While Fonagy did demonstrate statistically significant differences in Reflective Function Scale scores in subjects with borderline personality disorder compared to subjects without personality disorder, these findings supporting the scale's validation were completed before the manualization of the Reflective Function Scale in 1998 and have not been replicated since then.

The Reflective Function Scale has been applied as an outcome measure in the Borderline Personality Disorder Research Foundation/Personality Disorder Institute randomized, controlled trial comparing dialectical behavioral therapy, transference-focused psychotherapy, and supportive psychotherapy (9). However, its validity and stability as an outcome measure are uncertain because adequate reports of its test-retest reliability and convergent and divergent validity are still needed. While interrater reliability is well established in a number of studies (intra-class correlation coefficient ranging from 0.70 to 0.91) (9, 29, 55, 56), test-retest reliability of the scale is not established. Scores for reflective function may vary within the same therapist regarding different patients and can be below average for a therapist regarding their patients after 1 year of treatment with transference-focused psychotherapy (57). In short, the reflective function of an individual can vary depending on who he or she is mentalizing about, and even therapists can have below-average scores for reflective functioning.

Aside from the need for further empirical elaboration of the validity and reliability of the Reflective Function Scale, there are limitations in interpreting the meaning of a given reflective functioning score. The capacity that is assessed by the Reflective Functioning Scale is multidimensional, with factors such as plausibility, consistency, complexity, and originality. However, the grading is done using a unidimensional score that cannot be submitted to factor analysis. In two different transcripts coded for reflective functioning with a score of 3, one transcript may reflect a consistently superficial, clichéd, and general understanding of mental states, while the other transcript reflects a highly variable capacity to understand mental states with some moments of antireflectiveness and other moments of marked reflectiveness.

Adequate evidence demonstrating that deficits in reflective functioning distinguish individuals with borderline personality disorder from individuals with other forms of psychopathology is lacking. At present, reflective functioning has been shown to be predictive of borderline personality disorder diagnosis only in subjects with histories of abuse (56). Additionally, Fonagy and colleagues readily admit that mentalization capacities of individuals with borderline personality disorder are variable. The reflective functioning of an individual with borderline personality disorder is expected to vary depending on the relationship context in question, the level of distress, or the intensity of affect. Reflective functioning scores in individuals with borderline personality disorder might be sensitive to the context or timing of the assessment. Therefore, the assessment of mentalization as a marker of borderline personality disorder remains highly problematic.

Further research using the Reflective Functioning Scale is limited by the time-consuming and costly nature of the instrument. Adjacent concepts such as theory of mind, mindfulness, psychological mindedness, empathy, and affect consciousness have been operationalized into less cumbersome, albeit partial self-report measures. An exploration of findings related to these concepts may indicate simpler ways to evaluate the different dimensions of the mentalization concept in relation to borderline personality disorder psychopathology. As mentioned in the introduction, enhancing reflective function is thought to be a common feature of all psychotherapeutic approaches. Fonagy and Bateman propose that without changes in mentalizing, no improvements in functioning will be seen as a result of such treatments. Using self-report measures of overlapping constructs that isolate different dimensions of mentalizing relevant to different psychotherapeutic modalities would provide empirical evidence of their particular mechanisms.

## Conclusion

A review of the concept of mentalization reveals the breadth of its territory, spanning from the psychoanalytic domain of understanding and transforming internal experience into mentally contained forms to the social cognitive and developmental research domain of imagining the mental states of others in interpersonal interactions. An analysis of the overlap between mentalization and other related concepts, such as mindfulness, psychological mindedness, empathy, and affect consciousness, helps to clarify similarities and distinctions that are outlined in Table 1 and Figure 1. Research investigating the correlation of measures that operationalize these concepts may elucidate these relationships and also potentially lead to localization of mentalization deficits, in particular its implicit, explicit, self-oriented, other-oriented, cognitive, and affective components.

Ultimately, Fonagy's introduction of the concept of mentalization has catalyzed the development of novel and effective treatments as well as promising avenues of inquiry in psychopathology, psychotherapy, developmental psychology, and neuroscientific realms of research. However, one obstacle to its empirical utility is that it remains difficult to assess. While the assessment tool of reflective functioning importantly measures cognitive processes related to mentalization in an attachment context, the validity of this measure is underdeveloped, and it remains difficult to employ in large-scale research and routinely in clinical settings. Research to further develop the validity of reflective functioning, its relationship with related scales, and its relationship to borderline personality disorder and other diagnostic groups is needed to elucidate the boundaries of the concept and its usefulness in understanding and treating borderline personality disorder.

Presented in part at the 160th annual meeting of the American Psychiatric Association, San Diego, May 19–24, 2007. Received Aug. 28, 2007; revision received Feb. 24, 2008; accepted May 1, 2008 (doi: 10.1176/appi.ajp.2008.07081360). From the Department of Psychiatry, Harvard Medical School, Belmont, Mass. Address correspondence to Dr. Choi-Kain, Department of Psychiatry, McLean Hospital, 115 Mill St., Belmont, MA 02478; lchoikain@partners.org.

Supported by grants from the Program for Minority Research Training in Psychiatry and the McLean Hospital Psychosocial Fellowship (LWC).

The authors report no competing interests.

The authors thank Glen Gabbard, M.D., for his critical review and revisions on this manuscript.

## References

- Fonagy P: Thinking about thinking: some clinical and theoretical considerations in the treatment of a borderline patient. *Int J Psychoanal* 1991; 72:639–656
- Fonagy P, Gergeley G, Jurist E, Target M: Affect Regulation, Mentalization, and the Development of the Self. New York, Other Press, 2002
- Bateman A, Fonagy P: Psychotherapy for Borderline Personality Disorder: Mentalization-Based Treatment. Oxford, UK, Oxford University Press, 2004
- Bateman A, Fonagy P: Effectiveness of partial hospitalization in the treatment of borderline personality disorder: a randomized controlled trial. *Am J Psychiatry* 1999; 156:1563–1569
- Bateman A, Fonagy P: Treatment of borderline personality disorder with psychoanalytically oriented partial hospitalization: an 18-month follow-up. *Am J Psychiatry* 2001; 158:36–42
- Bateman A, Fonagy P: Eight-year follow-up of patients treated for borderline personality disorder: mentalization-based treatment versus treatment as usual. *Am J Psychiatry* 2008; 165: 631–638
- Clarkin JF, Yeomans F, Kernberg OF: Psychotherapy of Borderline Personality: Focusing on Object Relations. Arlington, Va, American Psychiatric Publishing, 1999
- Clarkin JF, Levy KN, Lenzenweger MF, Kernberg OF: Evaluating three treatments for borderline personality disorder: a multi-wave study. *Am J Psychiatry* 2007; 164:922–928
- Levy KN, Meehan KB, Kelly KM, Reynoso JS, Weber M, Clarkin J, Kernberg OF: Change in attachment patterns and reflective function in a randomized control trial of transference-focused psychotherapy for borderline personality disorder. *J Consult Clin Psychol* 2006; 24:1027–1040
- Bleiberg E: Treating professionals in crisis: a framework focused on promoting mentalizing. *Bull Menninger Clin* 2003; 67:212–226
- Fearon P, Target M, Sargent J, Williams LL, McGregor J, Bleiberg E, Fonagy P: Short-term mentalization and relational therapy (SMART): an integrative family therapy for children and adolescents, in *Practice in Handbook of Mentalization-Based Treatment*. Edited by Allen JG, Fonagy P. West Sussex, UK, John Wiley & Sons, 2006, pp 201–222
- Sadler L, Slade A, Mayes LC: Minding the baby: a mentalization-based parenting program, in *Practice in Handbook of Mentalization-Based Treatment*. Edited by Allen JG, Fonagy P. West Sussex, UK, John Wiley & Sons, 2006, pp 271–288
- Skarderud F: Eating one's words, part 3: mentalization based psychotherapy for anorexia nervosa, an outline for treatment and a training manual. *Eur Eat Disord Rev* 2007; 15:323–339
- Twemlow SW, Fonagy P: Transforming violent social systems into nonviolent mentalizing systems: an experiment in schools, in *Practice in Handbook of Mentalization-Based Treatment*. Edited by Allen JG, Fonagy P. West Sussex, UK, John Wiley & Sons, 2006, pp 289–306
- Allen JG, Fonagy P, Bateman AW: Mentalizing in Clinical Practice. Arlington, VA, American Psychiatric Publishing, 2008
- Holmes J: Mentalizing in psychoanalytic perspective, in *Practice in Handbook of Mentalization-Based Treatment*. Edited by Allen JG, Fonagy P. West Sussex, UK, John Wiley & Sons, 2006, pp 31–52
- Allen JG: Mentalizing, in *Practice in Handbook of Mentalization-Based Treatment*. Edited by Allen JG, Fonagy P. West Sussex, UK, John Wiley & Sons, 2006, pp 3–30
- Premack D, Woodruff G: Does the chimpanzee have a theory of mind? *Behav Brain Sci* 1978; 1:243–259
- Zetzel ER: The so-called good hysteric. *Int J Psychoanal* 1968; 49:256–260
- Lecours S, Bouchard M: Dimensions of mentalisation: outlining levels of psychic transformation. *Int J Psychoanal* 1997; 78: 855–875
- Aulagnier P: La Violence de L'Interprétation. Paris, Presses Universitaires de France, 1975
- Segal H: Notes on symbol formation. *Int J Psychoanal* 1957; 38: 391–397
- Freud S: The Interpretation of Dreams: The Standard Edition of the Complete Psychological Works of Sigmund Freud, vol IV: The Interpretation of Dreams. 1900 (first part), 4:ix–627
- Bion WR: Learning From Experience. New York, Basic Books, 1962
- Wimmer H, Perner J: Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 1983; 13:103–128
- Frith U, Frith CD: Development and neurophysiology of mentalizing, in *The Neuroscience of Social Interaction: Decoding, Imitating, and Influencing the Actions of Others*. Edited by Frith CD, Wolpert DM. Oxford, UK, Oxford University Press, 2003, pp 45–76
- Baron-Cohen S, Jolliffe T, Mortimer C, Robertson M: Another advanced test of theory of mind: evidence from very high-functioning adults with autism or Asperger syndrome. *J Child Psychol Psychiatry* 1997; 38: 813–822
- Baron-Cohen S, Leslie AM, Frith U: Does the autistic child have a "theory of mind?" *Cognition* 1985; 21:37–46
- Slade A, Grienenberger J, Bernbach E, Levy D, Locker A: Maternal reflective functioning, attachment, and the transmission gap: a preliminary study. *Attach Hum Dev* 2005; 7:283–298
- Stern DH: The Interpersonal World of the Infant: A View From Psychoanalysis and Developmental Psychology. New York, Basic Books, 1985



31. Kohut H: *The Analysis of the Self*. New York, International Universities Press, 1971
32. Storolow RD, Lachman FM: *Psychoanalysis of Development Arrests: Theory and Treatment*. New York, International University Press, 1980
33. Eagle MN: *Recent Developments in Psychoanalysis: A Critical Evaluation*. New York, McGraw-Hill, 1984
34. Fonagy P, Bateman AW: Mentalizing and borderline personality disorder. *J Ment Health* 2007; 16:83–101
35. Winnicott DW: *Playing With Reality*. New York, Routledge, 1971
36. Bram AP, Gabbard GO: Potential space and reflective functioning: Towards conceptual clarification and preliminary clinical implications. *Int J Psychoanal* 2001; 82: 685–699
37. Hahn TN: *Miracle of Mindfulness*. Boston, Beacon Press, 1976
38. Brown KW, Ryan RM: The benefits of being present: mindfulness and its role in psychological well-being. *J Pers Soc Psychol* 2003; 84:822–848
39. Baer RA: Mindfulness training as a clinical intervention: a conceptual and empirical review. *Clin Psychol: Science and Practice* 2003; 10:125–143
40. Linehan MM: *Skills Training Manual for Treating Borderline Personality Disorder*. New York, Guilford Press, 1993
41. Bishop SR, Lau M, Shapiro S, Carlson L, Anderson ND, Carmody J, Segal ZV, Abbey S, Specca M, Velting D, Devins G: Mindfulness: a proposed operational definition. *Clin Psychol: Science and Practice* 2004; 11:230–241
42. Baer RA, Smith GT, Allen KB: Assessment of mindfulness by self-report: the Kentucky Inventory of Mindfulness Skills. *Assessment* 2004; 11:191–206
43. Appelbaum SA: Psychological-Mindedness: Word, Concept and Essence. *Int J Psychoanal* 1973; 54:35–46
44. Farber BA: The genesis, development and implications of psychological-mindedness in psychotherapists. *Psych Theory Res Prac* 1985; 22:170–177
45. Conte HR, Plutchik R, Jung BB, Picard S, Karasu TB, Lotterman A: Psychological mindedness as a predictor of psychotherapy outcome: a preliminary report. *Compr Psychiatry* 1990; 31: 426–431
46. Shill MA, Lumley MA: The Psychological Mindedness Scale: factor structure, convergent validity, and gender in a nonpsychiatric sample. *Psychol Psychother* 2002; 75(part 2):131–150
47. Beitel M, Cicero JJ: Predicting psychological mindedness from personality style and attachment security. *J Clin Psychol* 2003; 59:163–172
48. Beitel M, Cicero JJ: Psychological mindedness and awareness of self and others. *J Clin Psychol* 2005; 61:739–750
49. Decety J, Jackson PL: The functional architecture of human empathy. *Behav Cogn Neurosci Rev* 2004; 3:71–100
50. Monsen JT, Monsen K: A psychotherapy model integrating Silvan Tomkin's affect- and script-theory within the framework of self psychology. *Progress in Self Psychol* 1999; 15:287–306
51. Lesser IM: A review of the alexithymia concept. *Psychosom Med* 1981; 43:531–540
52. Bagby RM, Parker JDA, Taylor GJ: The twenty-item Toronto Alexithymia Scale I: item selection and cross-validation of the factor structure. *J Psychosom Res* 1994; 38:23–32
53. Guttman H, Laporte L: Alexithymia, empathy, and psychological symptoms in a family context. *Compr Psychiatry* 2002; 43: 448–455
54. Mohaupt H, Holgersen H, Binder P, Nielsen GH: Affect consciousness or mentalization? a comparison of two concepts with regard to affect development and affect regulation. *Scand J Psychol* 2006; 47:237–244
55. Fonagy P, Steele M, Steele H, Moran GS, Higgitt AC: The capacity for understanding mental states: the reflective self in parent and child and its significance for security of attachment. *Infant Ment Health J* 1991; 12:201–218
56. Fonagy P, Leigh T, Steele M, Steele H, Kennedy R, Mattoon G, Target M, Gerber A: The relation of attachment status, psychiatric classification, and response to psychotherapy. *J Consult Clin Psychol* 1996; 64:22–31
57. Diamond D, Stovall-McClough C, Clarkin JA, Levy KN: Patient-therapist attachment in the treatment of borderline personality disorder. *Bull Menninger Clin* 2003; 67:224–257