MENTALIZATION IN ADULT ATTACHMENT NARRATIVES: REFLECTIVE FUNCTIONING, MENTAL STATES, AND AFFECT ELABORATION COMPARED

Marc-André Bouchard, PhD, Université de Montréal
Mary Target, PhD University College London, and The Anna Freud Centre, London

Serge Lecours, PhD Université de Montréal
Peter Fonagy, PhD University College London, and The Anna Freud Centre, London

Louis-Martin Tremblay, BSc Université de Montréal
Abigail Schachter, PhD The Anna Freud Centre, London

Helen Stein, PhD Menninger Child and Family Center

Relationships between three measures of mentalization (reflective function, mental states, and verbal elaboration of affect), attachment status, and the severity of axis I and axis II pathology were examined. Seventy-three adults, both ex-psychiatric patients and nonclinical volunteers were administered the Adult Attachment Interview (AAI). Comparisons between the three measures indicate that they share some aspect of a core mentalization process and that each illuminates a specific component. Reflective function was the only predictor of attachment status. The number of axis I diagnoses is partly explained by attachment insecurity, but the capacity to generate high-level defensive mental
states as well as increments in verbal affect elaboration further contribute to the model. Finally, increments in affect elaboration, as well as augmentations in high-level defensive activity and reflective function are all associated with decreases in the number of axis II diagnoses, over and above the contribution of attachment status and axis I pathology.

**Keywords:** mentalization, attachment, reflective functioning, affect elaboration, mental states

Mentalization is a term used by both developmental psychologists and psychoanalysts to refer to a core process of human social functioning and self-regulation, involved in the establishment of robust links between personally meaningful early experiences and their representation. More elaborate mentalization, sometimes called “reflectiveness,” is linked with attachment, insofar as a caregiver’s mindfulness about a child’s mental states appears as a key mediator of the transmission of attachment (Koren-Karie, Oppenheim, Dolev, Sher, & Etzion-Carasso, 2002; Meins, Fernyhough, Wainwright, Clark-Carter, Das Gupta, Fradley, & Tuckey, 2003). Conversely, it is known that maltreatment impairs a child’s reflective capacities (e.g., Beeghly & Cicchetti, 1994; Schneider-Rosen & Cicchetti, 1984, 1991).

A secure representational model of adult attachment relationships appears as a key favorable developmental outcome, the enduring impact of early sensitive caregiving, particularly in the context of benign life experiences (Waters, Merrick, Treboux, Crowell, & Albersheim, 2000). Attachment has recently been reformulated in terms of regulatory processes (Polan & Hofer, 1999), and the pathway from infancy to adult adaptation is thought to be mediated by key self-regulation functions, including the regulation of the stress response, the regulation of attention and the mentalizing function (Fonagy & Target, 2002).

Thus, mentalization, particularly mentalized affectivity, is a potentially powerful conceptual tool that helps to develop bridges between attachment theory, developmental psychopathology, affect regulation, and psychoanalysis (Jurist, 2005, p. 428). However, mentalization has been diversely construed and measured. Three approaches have been offered. Fonagy and Target (1997, 2002) understand mentalizing as resulting from the development of representations of psychological states in the mind of the human infant. Their measures have operationalized this capacity as “reflective function” (RF). Second, from an ego-psychology and object relations theory perspective, focusing initially on therapist and then on patient mental states actualized in a psychotherapy situation (Racker, 1968), different qualities of mentalization are presumably dependent upon differences in ego attitudes toward emotional experiences (e.g., degree, maturity, and continuity of defensive function and of the ego’s capacity to observe it’s own activity). A third approach suggests that mentalization involves a transformation of affective experiences through increases in the quantity and complexity of representational networks resulting in better quality of elaboration, containment, and regulation. The present research was designed to carry forward the exploration of the interrelations and comparative predictive value of these three concepts of mentalization and their associated measures, with respect to attachment status, axis I and axis II psychiatric diagnosis.

RF is defined as the capacity to envision and think about mental states, in oneself and in others, in the service of building realistic models of why they behave, think, and feel as they do (see Fonagy & Target, 1997, p. 679). Secure attachment offers a child a chance
to explore the mind of the caregiver, when she ascribes mental states to the child and treats
the child as a mental agent. Thus, this is presumably perceived by the child who learns
about minds, initially his own mind as perceived and understood by his caregiver. It was
first demonstrated (Fonagy, Steele, & Steele, 1991) that reflective parenting contributed
significantly to the development of attachment security. For instance, ratings of the quality
of the RF of each parent before a child’s birth was associated with the infant’s security
with this parent at 12 and 18 months (Fonagy, Steele, Moran, Steele, & Higgitt, 1991).
Fonagy, Steele, Steele, Higgitt, and Target (1994) reported that all mothers with adverse
histories, who showed high RF, had secure children, whereas in similarly deprived
mothers with low RF, only 1 child out of 17 was rated as securely attached to his mother.
Importantly, however, for the nondeprived group, RF was a much less important predictor
of attachment security. Severely maltreated individuals are likely to meet Diagnostic and
Statistical Manual of Mental Disorders (DSM) criteria for borderline personality disorder
only if their RF is rated as low (Fonagy, Leigh, Steele, Steele, Kennedy, Mattoon, Target,
& Gerber, 1996). Recently, Ward, Ramsay, Turnbull, Steele, Steele, and Treasure (2001),
examining the attachment status of patients with severe anorexia nervosa and their
mothers, found their sample to show very low RF levels, and further, that the vast majority
of participants were rated insecure on the Adult Attachment Interview (AAI; George,
Kaplan & Main, 1985/1996). Finally, Bernbach (2002) found that RF scores of subjects
participating in a 30-session Brief Relational Psychotherapy treatment increased during
the midphase in good outcome patients, whereas the poor outcome group demonstrated no
change.

Mental States as a Measure of Structural Development

Psychoanalytic authors have identified a number of key structural elements contrib-
uting to a person’s personality functioning, long-term adaptation, and well being. Kern-
berg (1996) in particular, has underlined the maintenance of reality testing, the quality,
and maturity in defensive activity and in object relations, and the capacity for affect
tolerance and reflectiveness as key determinants of character structure. Measurement of
these structural components of the personality has been difficult to achieve although recent
progress has been encouraging (Westen, 1998). One fundamental premise holds that
psychic structure is activated and revealed in the analytic setting and its equivalents. The
study of countertransference has been one classic and clinically significant opportunity to
examine how “structure becomes process.” In particular, theoretical discussions of coun-
tertransference have revealed important, enduring distinctions between at least three
intentional analyst orientations: (1) productive and preconsciously and consciously re-
flective; (2) counterproductive, mostly unconsciously defensive and reactive; and (3)
comparatively neutral and consciously objective and rational (see Bouchard, Normandin,
& Séguin, 1995; Racker, 1968).

A careful review of this literature has inspired Normandin (1991) to develop a
therapist mental states rating system. In the reflective mental state, the subject demon-
strates that he or she is actively involved in a subjective self-perception and observation
of whatever experience or memory is presently activated or recalled. In a reactive mental
state the subject is unaware, and may variously mentally expel, impulsively discharge,
refuse, distort, or inhibit what is currently being activated, thus defending him or herself
against it. In the objective-rational mode, the focus is on objective facts and situations, and
emotional and affective elements are toned-down and being treated as if from a distance.
A first series of studies (Lecours, Bouchard, & Normandin, 1995; Normandin & Bouchard, 1993; Séguin & Bouchard, 1996) offered initial support for the notion of an antagonism or competition between the reflective (presumably more mentalized) and the reactive (diversely mentalized) mental states. They also pointed to the need in future study to distinguish between high level defensive (Hi-Def, where often a clear contrast is observed between a spontaneous first movement and expression, and its interruption and inhibition) and low level defensive (Lo-Def, where a more disinhibited subject importantly distorts a representation, expels an experience through some enactment, etc.) forms of regulation. Each corresponding to a distinct level of mentalization, thus understood. Finally, it was necessary to include concrete thinking (Co), the least mentally elaborate and least contactful mental state, which often creates the sense of an empty presence, and manifested in stereotyped expressions, the use of clichés, and an overuse of anecdotal forms and quotations.

Six mental states are currently distinguished with the Mental States Rating System (MSRS; Bouchard et al., 2001), and are positioned on a hypothetical continuum of increasing mental elaboration and implicit structural development as follows: Co, Lo-Def, reactive-impulsive (Rimp), Hi-Def, Obr, and Ref. We have found that these six categories of mental activity can be reliably identified from verbal productions of subjects, both therapists and patients, observed in a variety of contexts beyond the psychoanalytic and psychotherapy situation, including the AAI.

After a careful preliminary examination of some personality disordered patients’ mental states, some more mentalized than others, during intensive psychoanalytic psychotherapy (Audet, Bouchard, Wiethaeuper, & St-Amand, 2001; Bouchard, Audet, St-Amand, Picard, & Wiethaeuper, 1999), the relative predominance in mental states was hypothesized to act as a good indicator of structural personality development.

A sample of 42 participants at risk to develop alcoholism (Lepadatu, Milcent, Dongier, Bouchard, & De Bruille, 2002) was examined and divided into groups of predominantly reflective or nonreflective participants, based on the relative proportions in mental activity during the AAI. On the Wisconsin Card Sorting Method, the reflective group achieved more categories, took fewer trials to reach the first category, and demonstrated higher scores on the conceptual level index. This was interpreted as an indication of stronger dorso-lateral frontal cognitive activation in the presumably more “effortful” reflective group compared with the other group. Together these findings underscored once again the relevance of the reflective versus reactive distinction in mental states, seen to be associated in part with varying degrees of prefrontal cortical activation.

**Mentalization of Affect as Indicated by Levels of Affect Elaboration**

The psychoanalytic concept of mentalization also designates a particular form of cognitive-affective elaboration, achieved through a linking of words and images to unprocessed affective bodily activation. The multiplication and organization of these representations, seen in the verbal elaboration of affect (VEA) is presumed to contribute to the development of increasing affect tolerance. One VEA conceptual model distinguishes two dimensions: levels of affect tolerance and abstraction and modalities of representation (Lecours & Bouchard, 1997; Lecours, Bouchard, St-Amand, & Perry, 2000). Each dimension contains levels of increasing verbal elaboration, that is, levels of richer and more complex networks of word associations, reflecting increasing levels of abstraction, and control of affective experience. The GEVA (Grille de l’Élaboration
Verbale de l’Affect; Lecours, 1995), provides the operational description in levels of affect mentalization and specifies the forms of verbal affect expression in each dimension. The first dimension describes levels and forms of increasing affect tolerance and abstraction: disruptive impulsion, modulated impulsion, externalization, appropriation, meaning association. The second differentiates four modalities of representation: somatic and motor activity, imagery, and labeling verbalization (see Method section).

Prior work with the GEVA indicates that negative affects are more relevant to the description of clinical processes. Positively valenced affective experiences tend to show higher levels of verbal elaboration, compared with negatively valenced ones (Lecours et al., 2000). Further, the quality of negative affect elaboration has been associated to the quality of overall functioning and symptoms (GAS: \( r = .43, p < .05, n = 32 \) (Lecours & Perry, 2002), whereas a negative correlation was observed between a subject’s negative affect quality of elaboration and the number of DSM–IV borderline personality disorder criteria \( (r = -.39, p < .05, n = 37) \). Thus, negative affects will be retained in the present work.

**Objectives**

This study aimed to examine for the first time some convergent and discriminant properties of three distinct but closely related measures of reflectiveness and mentalization: reflective functioning (Fonagy & Target, 1997), mental states (MSRS; Bouchard, Audet, Picard, Carrier, & Milcent, 2001) and levels of affect elaboration (Lecours, 1995). It was expected that the three measures would overlap moderately, thus demonstrating both shared and specific characteristics resulting from the distinct conceptual domains underlying each instrument. Second, we wanted to compare the predictive value of each measure, in terms of the three important outcomes available to us in this study: attachment security, and number of axis I and axis II diagnoses. Based on attachment theory (Bowlby, 1973, 1988), and previous observations by Fonagy et al. (1996), it was hypothesized that RF would be most closely associated with attachment status, and, to a lesser extent, also with axis I and axis II disorders. Based mostly on Kernberg’s (1996) psychoanalytic developmental and structural model of the personality, it was expected that a subject’s mental state profile, considered as a likely valid indicator of character structure and associated personality dysfunction, as presently activated during an interview, would predict most closely to the number of axis II diagnoses, an indication of clinically meaningful difficulties in personality adjustment, whereas showing little or no relation to axis I. Poor quality in affect elaboration (lower levels of verbal elaboration of negative affects) was expected to be associated with an increased likelihood of the number of axis-I and/or axis-II disorders.

**Method**

**Participants**

This study examines 73 verbatim transcripts from as many participants’ AAIIs. Participants are from both clinical and nonclinical populations, randomly selected from six different samples. Each of the first four samples was studied in previous published work.

---

1 Protocols are drawn from an available data bank of six different samples. The first is the London Twin Study group of normal mothers \( (N = 100) \). The second, also a group of normal mothers, was recruited as part of a study of child attachment \( (N = 28) \). The third is the Follow-Up Study of Child Psychoanalysis at the Anna Freud Center, London \( (N = 75) \). The fourth group was
or in doctoral theses (Fearon, 1998; Shmueli-Goetz, 2001; Target, 2002; Target, Shmueli-Goetz, & Fonagy, 2003). These included both ex-psychiatric patients and normal control participants recruited from an outpatient department of a university teaching hospital and a child psychotherapy clinic (many of the participants were mothers or caregivers recruited as part of studies of their children’s development). The present sample includes 30 men (41%) and 43 women (59%). The average age of the participants was 35.6 years. Forty-eight percent of the initial samples were of social class I or II (professional and managerial occupations) according to the United Kingdom Registrar General’s Classification of Occupations. The majority reported education beyond the high school level. Ethnic distribution was predominantly White but also included 4 participants of first generation Asian origin, 3 who were refugees from African states (for these 7 women English was not their first language but it was the language of their schooling), 3 who were born to mixed African American and White parents, 2 U.K.-born women who were of African Caribbean origin, 2 who came from other European countries (1 had English as her native language, the other did not), and 1 African American male.

**Procedures and Instruments**

Participants in the psychiatric samples were assessed by a research psychiatrist or psychologist who was not involved in the treatment of the patient, using standardized diagnostic interview schedules in which they had been trained (axes I and II modules of the Structured Clinical Interview for the *DSM–III–R* [Diagnostic and Statistical Manual of Mental Disorders, 3rd ed., revised; SCID-I and SCID-II, respectively; Spitzer, Williams, Gibbon, & First, 1990). The reported interrater reliability for DSM axis I and II diagnoses was high (Cohen’s $k = 0.85$). Normal participants were screened using the abbreviated General Health Questionnaire (GHQ-28; Goldberg, 1983). This questionnaire focuses on anxiety, depressive, and somatic symptoms. Potential participants were excluded if they were suffering from schizophrenia or any mental disorder of organic origin, had an estimated IQ below 75, or if their standard of English (either comprehension and expression) was estimated to be below that of an average 11-year-old native speaker.

Complete diagnostic information was not available for $n = 9$ participants who were recruited as part of a study of their child’s development; however, these participants were known not to be in current psychiatric treatment and are likely not to have been suffering from any moderate or severe psychiatric disorder. Fifty-two percent ($n = 33$ of 64) of the remaining sample met *DSM–III–R* criteria for at least one axis I or an axis II diagnosis. Forty-seven percent ($n = 30$; with $n = 18$ males) met criteria for an axis I. Eighteen participants met *DSM–III–R* criteria for a current affective disorder (major depressive disorder, $n = 14$; bipolar disorder, $n = 3$; dysthymia, $n = 3$; minor depression, $n = 3$; with some overlap). Other patients presented with an anxiety disorder ($n = 15$), substance abuse/dependence ($n = 16$), an obsessive–compulsive disorder ($n = 3$), PTSD ($n = 4$), morbid obesity and anorexia ($n = 2$), and impulsive disorder ($n = 1$). The mean number of axis I disorders was 0.72 for the 64 participants. Twenty-seven percent of these ($n =$ of female caregivers (mothers, grandmothers, or foster mothers) of children referred for psychiatric treatment in London ($N = 90$). The next group was from volunteer male subjects from the Montreal area, issued from a sample ($N = 14$) of mental health professionals interested in the AAI and patients presently undergoing psychotherapy in private practice. The last group of subjects came from a sample of subjects participating in a follow-up project at the Menninger Clinic Child and Family Program, in which children followed up decades after attendance at the Clinic Preschool were interviewed as adults.
17; with \( n = 13 \) males) met criteria for an axis II personality disorder. The average number of axis II diagnoses was 0.33 (range, 0 to 3). Personality disorders were distributed as follows: borderline \((n = 6)\), avoidant \((n = 5)\), narcissistic \((n = 5)\), obsessive–compulsive \((n = 5)\), antisocial \((n = 6)\), paranoid \((n = 4)\), histrionic \((n = 2)\), passive-aggressive \((n = 2)\), and unspecified \((n = 1)\), with some overlap.

The AAI interview. Forty-three of the AAI interviews were administered either in a university Psychology Department or in a family clinic, and the other 30 were administered in a private room in the participant’s home. The interviews were audiotaped, and later transcribed verbatim. Dysfluencies, grammatical errors, stuttering, interruptions, and pauses were retained and marked. Coding was conducted by coders trained, and certified as reliable in the use of the adult attachment classification system (Main & Goldwyn, 1998). The coding system for the AAI was applied a few years before this study was initiated, except for a series of 14 protocols, from two samples (3 in Montréal, 11 from the Menninger Clinic), which were recently rated by one of two trained and reliable coders. Reliability was not specifically assessed for these samples; it is based on the testing of 32 cases in the Main-Hesse reliability accreditation procedure, which was passed by both coders before coding began. AAI interviews are rated on a number of scales concerning experience (e.g., loving mother and neglecting father) and state of mind of the interviewee as reflected in the narrative (coherence, passivity of thought reflected in unfinished sentences, idealization or derogation of caregivers, etc.). The interrater reliabilities of these scales are all in excess of .80, as previously reported (Fonagy et al., 1996). Rater reliabilities on major attachment classification in previous work was 85% (100% agreement on Insecure vs. Secure classification), which is consistent with values reported in the literature (see Bakermans-Kranenburg & van Ijzendoorn, 1993).

All instruments to be described below were independently applied to assess the verbatim transcriptions from the AAI interviews. Most interviews \((n = 59)\) were conducted in England, 11 in the United States, 3 in Canada, all with English speaking participants. All RF ratings were completed by one reliable coder, MT, who is also the lead trainer for this measure, before this study began. The two other measures of mentalization were applied in Montréal, before all information was put together and analyzed. All raters worked solely from transcripts and were thus blind as to participants’ characteristics.

The RF measure. The RF scale (Fonagy, Steele, Steele, & Target, 1997) assesses the clarity and complexity of a person’s representation of mental states, whether of self or others. Raters are required to mark the presence or absence of a reflective stance in relation to self or other. They then use definitions and narrative descriptions of different types of reflective statements to score the participant’s AAI protocol on a scale from 1 to 9 (see Fonagy et al., 1996, for more detail). The reliability of the RF scale after training is typically high; in previous applications, pairs of raters arrived at correlations between .81 and .94. In the present study, all protocols were rated for RF by one of us (MT), who is the main trainer in the use of this method, and the standard against which others’ reliability is tested. The \(N = 22\) protocols were double coded for RF, by a second coder previously trained to reliability; Pearson correlation for these cases was \(r = .86\).

The mental states measure (MSRS). All AAI protocols were first segmented by an independent rater (LMT) for thematic contents, depending on lines of questioning, content answers, and potential changes in participant’s mental states. Units varied in length between 2 and 56 lines of text, with a median of 14, or 240 words as pronounced by the patient. Once identified, each thematic unit was assigned a value corresponding to which of the six mental states category would predominate at this point in time during the
interview, and within the unit. Two female and one male graduate student in clinical psychology scored the MSRS, after receiving approximately 45 hours of training. The MSRS scores are obtained by using standardized decision procedures defined in a detailed scoring manual (see Bouchard et al., 2001). Each protocol was independently rated by two of the three judges. Interrater agreement is assessed and then consensual scores are obtained. Reliability estimates from previous work (Audet et al., 2001; Bouchard et al., 1999; Lecours et al., 1995; Normandin & Bouchard, 1993) yielded satisfactory results. In the present work, the mean percentage of agreement for the identification of the thematic units was of 84%. Once determined, the mental states category values assigned separately by each judge to each section of the AAI were applied to the text under a different computer environment (SATO for Système d’Analyse de Texte par Ordinateur; Daoust, 1990). This enabled us to assess the total number of words produced by the participant, as well as the total number of words assigned by each judge under each of the mental states category, for the AAI protocol taken as a whole. Thus, percentage scores are used, with scores varying between 0 and 100. Pearson product–moment correlations for each mental state between the two judges were very satisfactory, except for the infrequent reactive-impulsive mental state, which will be deleted from further analyses: concrete thinking (Co, \(r = .870\)), lower defensive (Lo-Def, \(r = .792\)), reactive-impulsive (Rimp, \(r = .592\)), higher defensive (Hi-Def, \(r = .779\)), objective-rational (Obr, \(r = .788\)), and reflective (Ref, \(r = .807\)).

**Affect mentalization:** The GEVA. The GEVA (for Grille de l’Élaboration Verbale de l’Affect—verbal elaboration of affect scale; Lecours, 1995) defines two orthogonal dimensions to the quality of verbal expression of affects: Tolerance/abstraction, related to five levels of containment and modalities, specified through four modes of representation. The levels of tolerance and abstraction go as follows: (1) At a descriptive level, a disruptive impulsion is an uncontrolled manifestation of an emotional action-tendency (e.g., physical fighting as anger expression) or a disorganized or symptomatic form of affect expression (e.g., a panic attack instead of a normal fear response; psychomotor retardation instead of sadness). “Impulsion” is defined here as a nonreflexive verbal reference to an affect. It is not yet felt as a subjective, private experience, the person does not talk about the affect, but rather she or he expresses it through his actions, words, or imagery. The subject is either unaware of the significance of the affective experience (implicit representation) or she or he has no control over the expression of a known emotion (explicit representation). The disruptive aspect characterizes this lowest level of impulsion. (2) In a modulated impulsion, the integrity and adaptive function of the affect are preserved. The expression is still not reflected upon but is here more modulated. It corresponds to the manifestation of a state of action tendency or action readiness (Frijda, 1986), that is, the normal motivational “push or pull” of an affect. This level thus consists of spontaneous but adaptive affect expression such as some forms of laughing or crying. (3) Through externalization, an affect is explicitly recognized but is not yet completely tolerated as one’s own. It is either mentally externalized and perceived as if caused by some external event or agency, or disowned and generalized to a group of people. Common with the previous two levels, there is still some kind of disownment of the experience, but, contrary to the levels of impulsion, at least some explicit mental content is here acknowledged before being “externalized” (e.g., “He made me mad;” “One can only be saddened by this”). (4) With the appropriation level, the affect, now fully tolerated, is felt as an internal, private, subjective experience. Its outward manifestation is described in an abstract or symbolic way. It is owned and appropriated by the subject, but its meaning is not yet developed (e.g., “I feel angry”). (5) By means of further meaning
associations, the affect, which has been appropriated, has gained a fuller meaning through resorting to richer and more complex networks of representations. The person experiences and verbalizes a clear knowledge of why she or he lives through this particular affect, as when achieving insight.

The modalities of representation define a second dimension, and are somewhat related to the classic components of emotions, but the focus here is on their use as material for representation. With the somatic modality, the affect is expressed through the verbalization of an internal physiological sensation; the autonomic internal body is “talking.” Motor expression involves the description of behavior, that is, the voluntary external body. With the verbalization of mental contents through metaphors, affects are communicated through the imagery modality. Finally, through the labeling verbalization, the affect is expressed by the use of “objective” labels and of common, social, well-defined language.

The GEVA scoring procedure involves two steps. First affective units consisting of portions of the transcripts that involve an implicit or explicit discussion of an affect are identified on the transcript by a first judge-observer. Then a second judge proceeds to assign a score using the bidimensional system of GEVA categories. The level of affect mentalization achieved by a participant is given by a combination of the two dimensions, which specify 20 different forms (5 levels of tolerance × 4 modalities of expression). Thus, each verbal expression of an affect receives one of 20 GEVA individual scores. These 20 categories are combined to produce a supraordinate Weighted Geva Score of affect mentalization (WGS). The latter is a product of weights given to each modality (ver = 4, ima = 3, etc.) by the mean score for the levels of tolerance obtained by the channel, adjusted to the proportion of affective units scored by the channel; the result is a number ranging from 1 to 4, with higher scores indicating higher quality in the verbal elaboration of affect. Two weighted scores were used for the analyses: one for all negatively valenced affect experienced by the subject (WGSNS), and another for the negatively valenced affects described by a participant as belonging to another person (WGSNO), the latter being specifically included for an investigation of associations with a capacity to mentalize other people’s affective experiences.

Reliability estimates were based on five AAI protocols. Two male graduate students were trained until they obtained consistent target levels of agreement beyond 70% for the identification of affective units and Kappas over .65 for the GEVA categories. Results, based on a conservative method (see Stinson, Milbrath, Reidbord, & Bucci, 1994), yielded percentages of agreement ranging from 68.8% to 82.6% (M = 75.2%) for an average of 316.3 individual affective units per interview. For the GEVA categories, scores for each single AU were used for the computing of Kappas, each interview taken separately, which amounts to a rigorous test of agreement. The scoring proper of the GEVA categories yielded mean Kappas of .83 (range .67 to .94) for the modalities and of .71 (range .68 to .73) for the levels of tolerance/abstraction. Evaluation of valence produced a mean kappa of .89 (.84 to .92). This indicates very good to excellent agreement beyond chance (Fleiss, 1981).

Results

Preliminary Analyses

Sex was not related to the AAI classification with 20 female and 11 male participants classified as secure [$\chi^2 (1, N = 73) = 0.70, ns$]. However, sex was significantly related with RF and GEVA scores (WGSNS), females showing higher levels of mentalization.
than males (RF, Mfem = 5.14, SD = 1.96, Mmal = 3.53, SD = 1.62, t = 3.69, df = 71, p < .001; WGSNS, Mfem = 2.74, SD = .20, Mmal = 2.54, SD = .28, t = 3.46, df = 71, p < .01). Males were also more likely to present a personality disorder than females (χ²(1, 64) = 10.07, p < .01). Further as expected, attachment insecurity was related with the presence of an axis I (χ²(1.64) = 8.82, p < .01) and axis II diagnosis (χ²(1.64) = 7.25, p < .01). Because they are interrelated, sex, attachment status, and diagnosis will be included in the regression analyses.

Before analysis, all mentalization variables were examined to establish the nature of the observed distributions for each. It turned out that only 22% of the protocols contained a concrete mental state rating, 30% of participants received a low-defensive rating, and only 28 participants got more than 10% of their words rated under the objective-rational category. Thus, it seemed best to reconsider these three variables as dichotomous, based on an absence or presence rating. The Hi-Def and Ref, GEVA (WGSNS, WGSNO), and RF scores showed acceptable distributions and were used as continuous scores in further analyses.

**Correlations Among Mentalization Measures**

Pearson correlations between each mentalization measure (see Table 1) show several observations consistent with expectations. First, the MSRS scores indicate patterns of intercorrelations that are consistent with previous observations and with theoretical predictions: (1) a positive association, at the lower end of the continuum, between Conc and Lo-Def; (2) a negative correlation between these two and Ref mental states; (3) no association between Obr and all other states. However, as indicated above, because Conc, Lo-Def, and Obr all show dichotomous distributions, this raised potential questions as to their metric properties. To test this further, a principal component factor analysis among the mental states variables produced a two-factor solution explaining 74% of the variance, with Hi-Def (with Conc and Lo-Def) and Ref (with Obr) showing the highest loadings. Thus, it was decided that only the Hi-Def and Ref mental states should be retained in further analyses of predictive power. Second, as expected, the two WGS scores of self and other’s affects show a significant positive correlation between them (r = .56, p < .01).

Third, the extremes of the mental states continuum show significant relations with the RF score (r = -.49, p < .01 with the concrete, r = -.30, p < .05 with the lower defensive,

Table 1

**Relationships Between the Major Mentalization Scores**

<table>
<thead>
<tr>
<th></th>
<th>Lo-Def</th>
<th>Obr</th>
<th>Hi-Def</th>
<th>Ref</th>
<th>RF</th>
<th>WGSNS</th>
<th>WGSNO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conc</td>
<td>.66**</td>
<td>-.18</td>
<td>-.61**</td>
<td>-.34**</td>
<td>-.49**</td>
<td>-.23</td>
<td>-.35**</td>
</tr>
<tr>
<td>Lo-Def</td>
<td>-.20</td>
<td>.02</td>
<td>-.68**</td>
<td>-.23</td>
<td>-.30**</td>
<td>.13</td>
<td>.14</td>
</tr>
<tr>
<td>Obr</td>
<td></td>
<td></td>
<td></td>
<td>.15</td>
<td>.08</td>
<td></td>
<td>.20</td>
</tr>
<tr>
<td>Hi-Def</td>
<td>-.28**</td>
<td></td>
<td></td>
<td>.08</td>
<td>.10</td>
<td>.06</td>
<td>.12</td>
</tr>
<tr>
<td>Ref</td>
<td>.42**</td>
<td></td>
<td></td>
<td>.06</td>
<td>.20</td>
<td>.44**</td>
<td>.56**</td>
</tr>
<tr>
<td>RF</td>
<td></td>
<td></td>
<td></td>
<td>.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WGSNS</td>
<td></td>
<td></td>
<td></td>
<td>.44**</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. N = 73.
Pearson rs: * = p < .05. ** = p < .01.
Conc = concrete state; Lo-Def = low level defensive state; Obr = objective rational state; Hi-Def = higher level defensive mental state; Ref = reflective mental state; RF = Reflective Functioning Score; WGSNS = Weighed Geva Score for negatively valenced affects (Self).
and $r = .42, p < .01$ with the reflective mental state). The intermediate states of the continuum are not related to reflective capacity. This observation underlines a partial convergence between the two constructs. Next, an interesting convergence is seen between WGSNO (the degree of mental elaboration in the subject’s verbal expression of affects attributed to others) and RF ($r = .44, p < .01$), as well as both the Conc and Lo-Def mental states ($r = -.35, p < .01; r = -.42, p < .01$, respectively). Taken together these observations suggest the need to differentiate in a subject’s discourse whether expressed affects are attributed to the self or to another person. It seems that the RF score and its closely (but negatively) associated concrete mental state are more powerfully indicative of an adult’s full representation and/or preoccupation of the other’s intentional stance, rather than one’s own, which is consistent with Fonagy and Target’s initial theory of mind impetus. Next WGSNS, a subject’s degree of affect (of self) elaboration relates to the objective-rational mental state index, but not with both extremes of either concrete or reflective mental state productions. This suggests the presence of a potential mutual positive interaction between the degree of affect transformation achieved and the establishment of mental distance or “cooling-off.”

To simplify our investigation of their links with clinical measures, only nonredundant mentalization variables were kept for further analysis. Thus, in addition to the single RF index, the WGSNS score will be chosen over the WGSNO, as it is thought to be a better and more unique indicator of a subject’s level of affect elaboration, also avoiding unnecessary overlap with RF.

**Mentalization and Attachment, Axis I and Axis II**

As expected, RF is the only mentalization measure to be associated with attachment ($r = -.40, p < .01$; higher scores are negatively related with attachment insecurity, see Table 2). RF, Hi-Def, and WGSNS are all significantly and very similarly associated with the number of axis I and axis II diagnoses, with lower levels of mentalization associated with the severity of both axis I and axis II pathology, perhaps even more strongly with axis II. Correlations show that the Ref mental state is not related with attachment status and number of axis I and II diagnoses. Ref scores will thus not be included in the following regression analyses.

The specific contribution of the representative mentalization variables will now be assessed with regression analyses. These analyses will also attempt to indicate the importance of the intervention of mentalization phenomena in the prediction of diverse forms of psychopathology. To clarify further interpretation, hierarchical logistic and

---

**Table 2**

<table>
<thead>
<tr>
<th></th>
<th>Att</th>
<th>Nb axis I</th>
<th>Nb axis II</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hi-Def</td>
<td>-.04</td>
<td>-.41**</td>
<td>-.42**</td>
<td>40.72</td>
<td>27.11</td>
</tr>
<tr>
<td>Ref</td>
<td>-.18</td>
<td>.04</td>
<td>-.04</td>
<td>25.10</td>
<td>21.67</td>
</tr>
<tr>
<td>RF</td>
<td>-.40**</td>
<td>-.29*</td>
<td>-.46”</td>
<td>4.48</td>
<td>1.99</td>
</tr>
<tr>
<td>WGSNS</td>
<td>-.04</td>
<td>-.32”</td>
<td>-.36”</td>
<td>2.66</td>
<td>.25</td>
</tr>
</tbody>
</table>

*Note.* N = 73 subjects for attachment and N = 65 for Axes I and II. Pearson $r$: * = $p < .05$, ** = $p < .01$. Higher scores on attachment mean insecurity. Hi-Def = higher level defensive state; Ref = reflective mental state; RF = Reflective Functioning Score; WGSNS = Weighed Geva Score for negatively valenced affects (Self).
multiple regressions will be computed to control for the contribution of sex and axis I and/or II comorbidity.

**Attachment status.** A logistic regression was performed to predict secure versus insecure attachment. The predictors underwent a z transformation before analysis to produce comparable odds ratios. Seventy-three cases were available for this first analysis. Results show that mentalization variables contribute significantly to the prediction of attachment status, over and above sex and the presence of an axis I or axis II diagnosis (ΔNagelkerke $R^2 = .158$, see Table 3). The model including the whole set of variables shows that only the presence of an axis I or II diagnosis and RF are significantly related with attachment security. The absence of a diagnosis increases a subject’s chance of having a secure attachment by 6.56. In addition, a one standard deviation increase in RF increases a subject’s chance of having a secure attachment by 2.74.

**Axis I and II diagnosis.** Hierarchical multiple regression analyses were computed for the prediction of axis I and II disorders. Sixty-four participants were available for these analyses. The number of diagnoses was preferred to the presence/absence of a diagnosis as a criterion because it provides important additional information on within-axis comorbidity and results in a more substantial index of severity of axis I or II pathology. Logistic regressions using presence/absence of axis I or II diagnosis lent very similar results with somewhat weaker effect sizes. Again, to provide an estimate of the unique role of mentalization variables to pathological phenomena, sex and the presence of comorbid clinical conditions were controlled for. Results show that all three mentalization facets contribute significantly to the prediction of the number of axis I disorders, over and above sex, attachment status, and the number of axis II disorders (Δ $R^2 = .160$, see Table 4). The final model indicates that attachment insecurity is positively associated with axis I pathology (part $r = .22$), negatively with Hi-Def scores (part $r = -.33$) and also negatively with the quality of negative affect elaboration (self) (part $r = -.24$).

Results for axis II again indicate that all three mentalization variables significantly contribute to the prediction of the number of diagnoses, over and above sex, attachment status, and the number of axis I diagnoses (Δ $R^2 = .198$, see Table 5). At least one element from each of the three mentalization facets is significantly associated with axis II pathology (Hi-Def part $r = -.32$, RF part $r = -.24$, WGSNS part $r = -.21$).

Table 3
**Summary of Hierarchical Logistic Regression Analysis Predicting Attachment Status From Patient Characteristics**

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>95% CI</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nagelkerke $R^2$</td>
<td>.239</td>
<td>$\chi^2(2) = 12.60, p &lt; .01$</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>0.45</td>
<td>0.14–1.43</td>
<td>.176</td>
</tr>
<tr>
<td>Axis I or II diagnosis</td>
<td>4.93</td>
<td>1.55–15.64</td>
<td>.007</td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td>$\Delta R^2 = .158, \Delta \chi^2(3) = 9.89, p &lt; .05$</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>0.70</td>
<td>0.18–2.75</td>
<td>.604</td>
</tr>
<tr>
<td>Axis I or II diagnosis</td>
<td>6.56</td>
<td>1.45–29.77</td>
<td>.015</td>
</tr>
<tr>
<td>RF</td>
<td>2.74</td>
<td>1.29–5.84</td>
<td>.009</td>
</tr>
<tr>
<td>Hi-Def</td>
<td>0.87</td>
<td>0.42–1.80</td>
<td>.712</td>
</tr>
<tr>
<td>WGSNS</td>
<td>0.71</td>
<td>0.35–1.46</td>
<td>.356</td>
</tr>
</tbody>
</table>

*Note.* $N = 73$. For a clearer interpretation of OR, dichotomous attachment and diagnostic scores are positively keyed: a positive score is associated with security and with the absence of an axis I or II diagnosis. OR = Odds Ratio; CI = Confidence Interval; RF = Reflective Functioning Score; Hi-Def = higher level defensive mental state; WGSNS = Weighed Geva Score for negatively valenced affects (Self).
The present study aimed at furthering the exploration of the important concept of mentalization, first by attempting to evaluate the relationships between various views of mentalization by comparing findings obtained from the use of three different measures and coding systems on verbatim transcriptions of AAI protocols. We then applied these measures to the prediction of three clinical conditions, namely attachment status, axis I and axis II pathology.

The three mentalization measures and their associated coding systems emphasize theoretically different facets of the process. RF stresses the protective function of the acquisition of an as-if or pretend mode of thinking centered on psychological explanations.

Table 4
Summary of Hierarchical Multiple Regression Analysis Predicting Number of Axis I Diagnoses From Patient Characteristics (n = 64)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Part r</th>
<th>B</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>.07</td>
<td>.08</td>
<td>.531</td>
</tr>
<tr>
<td>Attachment</td>
<td>.24</td>
<td>.25</td>
<td>.047</td>
</tr>
<tr>
<td>Nb axis II</td>
<td>.21</td>
<td>.23</td>
<td>.080</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>.00</td>
<td>.01</td>
<td>.967</td>
</tr>
<tr>
<td>Attachment</td>
<td>.22</td>
<td>.25</td>
<td>.041</td>
</tr>
<tr>
<td>Nb axis II</td>
<td>-.03</td>
<td>-.04</td>
<td>.786</td>
</tr>
<tr>
<td>RF</td>
<td>-.09</td>
<td>-.11</td>
<td>.421</td>
</tr>
<tr>
<td>Hi-Def</td>
<td>-.33</td>
<td>-.37</td>
<td>.003</td>
</tr>
<tr>
<td>WGSNS</td>
<td>-.24</td>
<td>-.27</td>
<td>.030</td>
</tr>
</tbody>
</table>

Note. Attachment = secure (1) versus insecure (2); Nb = total number of diagnoses; RF = Reflective Functioning Score; Hi-Def = higher level defensive mental state; WGSNS = Weighed Geva Score for negatively valenced affects (Self).

Discussion

The present study aimed at furthering the exploration of the important concept of mentalization, first by attempting to evaluate the relationships between various views of mentalization by comparing findings obtained from the use of three different measures and coding systems on verbatim transcriptions of AAI protocols. We then applied these measures to the prediction of three clinical conditions, namely attachment status, axis I and axis II pathology.

The three mentalization measures and their associated coding systems emphasize theoretically different facets of the process. RF stresses the protective function of the acquisition of an as-if or pretend mode of thinking centered on psychological explanations.

Table 5
Summary of Hierarchical Multiple Regression Analysis Predicting Number of Axis II Diagnoses From Patient Characteristics (n = 64)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Part r</th>
<th>B</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>.28</td>
<td>.29</td>
<td>.016</td>
</tr>
<tr>
<td>Attachment</td>
<td>.15</td>
<td>.16</td>
<td>.191</td>
</tr>
<tr>
<td>Nb axis I</td>
<td>.20</td>
<td>.22</td>
<td>.080</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>.11</td>
<td>.12</td>
<td>.290</td>
</tr>
<tr>
<td>Attachment</td>
<td>.10</td>
<td>.11</td>
<td>.330</td>
</tr>
<tr>
<td>Nb axis I</td>
<td>-.03</td>
<td>-.03</td>
<td>.786</td>
</tr>
<tr>
<td>RF</td>
<td>-.24</td>
<td>-.28</td>
<td>.021</td>
</tr>
<tr>
<td>Hi-Def</td>
<td>-.32</td>
<td>-.36</td>
<td>.002</td>
</tr>
<tr>
<td>WGSNS</td>
<td>-.21</td>
<td>-.24</td>
<td>.036</td>
</tr>
</tbody>
</table>

Note. Attachment = secure versus insecure; Nb = total number of diagnoses; RF = Reflective Functioning Score; Hi-Def = higher-level defensive mental state; WGSNS = Weighed Geva Score for negatively valenced affects (Self).
of relationships in terms of intentional mental models. This developmental achievement presumably facilitates the regulation particularly of disruptive affects by providing a symbolic buffer between the behavior of significant others and their impact on the developing mind. It also permits a reflective stance toward self by facilitating the narrative coherence and meaningfulness provided by mentalized or intentional constructions about internalized interactions. The mental states approach used here refers to implicit intentional attitudes of a subject toward her or his mental contents. These are reduced to sensorimotor transactions, often virtually devoid of affect in the Conc mode, diffracted through the lens of primitive defenses (splitting, omnipotence, etc.) in the Lo-Def mode, treated with a comparatively more abstract but objectifying attitude in the Obr state, managed with more mentally elaborate defenses in the Hi-Def mode, and felt as subjective events accessible to exploration in the Ref state. Verbal elaboration of affect highlights the quality of verbal representation associated with affect and its associated power to transform experience from impulsively expressed somatized forms of affectivity to the range of increasingly tolerated emotional events, eventually highly saturated with abstract (verbal) meaning.

Bivariate correlations between these measures show a coherent pattern of overlap and divergence (see Table 1): the three measures share some aspect of a core mentalization process while each illuminates a specific facet. High levels of reflective functioning or elaboration of affect and a reflective state all contribute to and are presumably indicative of a central component of mentalization, seen as a core capacity to treat some events as mental internal events, part of an internal state of awareness. This capacity to recognize and elaborate the subjective quality of experience is not a biological given. It seems rather a key relational acquisition, connected with the ability to use symbolic processing, with its resulting increased meaning and tolerance.

Our findings also indicate that mentalization is not a homogeneous process. Three groupings are apparent from the intercorrelations, which are confirmed by an exploratory principal components factor analysis (a three-factor solution explaining 72% of the variance): Factor A: Hi-Def (negatively), Lo-Def and Conc; Factor B: WGSNS, WGSNO, and Obr; Factor C: Ref and RF. This suggests the existence of three potential components of mentalization: predominant hi-level defensive functioning over low-level defensive activity; abstract verbal articulation and objectivation, and an attitude of focusing on mental processes.

Each measure is more clearly delineated by the comparison made with the other two and with the clinical conditions. The RF continuum shows many similarities with the MSRS model, particularly with the two poles of the mental states continuum. This indicates that RF, with its focus on representations of intentionality in self or others, also encompasses a subject’s implicit and actualized stance toward her or his mental contents as they are evoked in the here-and-now of the AAI interview, a characteristic of the MSRS that scores presently active mental states. This is particularly true when what emerges for the subject is defended against via a concrete (mentally empty) or low-level (primitive and often disinhibited) mode or, at the other pole, if it is reflected upon (in fully mental and aware status). Differences between models are also apparent, however, between measures since RF does not appear to consider a subject’s use of more mature defenses (an important predictor of pathology) or of a more rational attitude.

Interestingly, RF seems to be more sensitive to a subject’s quality of affect elaboration of others rather than of self. This is consistent with the theory of mind construct (Fonagy & Target, 1997) that underlines the importance of “figuring out” a caregiver’s intentional stance, as a key coacquisition toward safe attachments. This specific focus on the
internalization of others’ mental models might explain why RF is the only mentalization index significantly related to attachment status (which is only a partial explanation because WGSN0s relationship with attachment is of a lesser amplitude: \( r = -.20 \)). Nevertheless, a unique contribution of the RF approach to mentalization appears to be its emphasis on mental models, particularly those concerned with early interpersonal scenarios.

The MSRS covers a wide range of often discrete phenomena, both primitive and more mature, defensive-reactive and reflective. Among them, the use of primitive defenses and a concrete mental state are highly related and appear to represent two facets of a single domain of primitive mental functioning. These two mental states are negatively associated with RF and affect elaboration. This implies that both concrete and low-level defensive states partly and implicitly address RF and affect elaboration in their lower level ratings. Interestingly, these primitive states are more antithetical to more mature defenses than to a reflective stance. This is true despite the fact that the mature defensive and reflective states are negatively correlated. Yet, the most significant finding with the MSRS highlights the important positive contribution of high-level defenses to the process of mentalization. This unique series of defensive mental states is not taken into consideration by the other measures whereas being the most strongly associated with the clinical conditions studied here.

Findings with the Ref mental state are more paradoxical and intriguing, however. Within the MSRS, it is posited to be the most mentalized mental state, which is supported by its moderate association with RF (see Table 1). Nevertheless, Ref also shows no association with attachment security, or either the presence/absence or intensity of pathology. It taps a component of reflectiveness that is rather independent of a health-pathology feature, which would not have been predicted from theory. A participant’s predominant use of the Ref mental state during an AAI interview does not seem to be an unambiguous and valid indicator of structural progress, which may serve as a cautionary note for its use as an outcome indicator of “structural change.” A better candidate for which would seem to be a predominance, within the AAI, of the high-level defenses. Rather we speculate that Ref is a presently consented “mental and conscious effort,” presumably the result of frontal activation (Lepadatu et al., 2002), indicating an attempt to resolve a personal affectively charged problem. Its function needs to be further investigated. To summarize, the specific contribution of the MSRS to the study of mentalization appears to be the delineation of a high-level defensive facet, whereas the study of primitive states appears promising but needs further work.

The GEVAs stress on representation makes it the most “cognitive” of the mentalization instruments, as its relationship with the Obr mental state tends to confirm. The emphasis is on the quality (form) of representation, of present or past origin, with no consideration of whether or not it is presently playing an active, dynamic role. Further, scoring is limited to segments of verbalized affect, in contrast to the more encompassing grasp on mental events provided by the RF or MSRS systems. Nevertheless, the quality of affect representation is significantly associated with axes I and II pathologies. Thus, the GEVA uniquely underscores the role of key representational forms and transformations within the larger mentalization process in achieving the necessary symbolic-affective distance in the face of painful, sometimes overwhelming experience.

A combination of at least some elements of each of the three models reviewed here significantly contributes to the formation and/or maintenance of attachment insecurity and to the presence and severity of axis I and II pathology. The moderate size obtained by the mentalization model used in the regression analyses underlines the important role that the
quality of mentalization plays in an individual’s mental functioning and state of psychological health. Findings indicate that each measure has both a shared and a specific effect on the outcomes here examined (mean zero-order $r = -.34$ and mean part $r = -.22$ for axis I—a difference of $.12$; mean zero-order $r = -.41$ and mean part $r = -.26$ for axis II—a difference of $.15$). The specific facets of the instruments used, taken together, have a larger impact on the prediction of the presence and/or severity of a mental disorder. This suggests the need in future work to consider mentalization as a complex phenomenon, likely involving at least three dimensions: (1) the capacity to use high-level affective regulatory (defensive) procedures, over low-level defensive ones; (2) the capacity to elaborate, transform, and objectify affects into verbally articulate abstract form; (3) the capacity to focus on affective and interactive exchanges as mental processes.

Understanding of the clinical conditions investigated in this study is also improved by an examination of their relationships with the mentalization measures. The prediction of attachment status clearly demonstrated the unique relation of RF scores, compared with the other mentalization measures. This is in line with several previous observations and provides additional evidence of the privileged pathway of association between the mentalizing function as defined by Fonagy and Target (2002) and attachment security. No other indicator of mentalization was retained as part of that prediction model, which underscores the differences between constructs derived partly from a developmental tradition (RF), and those issued from the clinical-psychodynamic tradition (mental states, including the high- vs. low-defensive activity dimension, and affect elaboration).

The number of axis I diagnoses is partly explained by the presence of an insecure attachment (part $r = .22$, see Table 4), which is consistent with the hypothesis that attachment security is a protection factor against an axis I diagnosis (Riggs & Jacobvitz, 2002). However, a decrease in high-level defensive mental state is predominant (part $r = -.33$) and a poorer quality of affective verbal elaboration is present as well (part $r = -.24$). The specific contribution of RF does not add to the prediction of axis I pathology despite its significant bivariate relationship ($r = -.29$, $p < .05$), likely because of the removal of the shared variance with attachment status. Thus, compared to attachment insecurity, the presence, and severity of axis I pathology is likely associated with a more complex deficit in mentalization because more facets are involved in its prediction. High-level defensive activity is characterized by a capacity to withhold, inhibit, repress, and mentally contain within the self a given emotionally meaningful, often threatening, or unacceptable affective experience. Whatever the causal chain of factors, and taking a restricted synchronic view, we may speculate that the presence of a significant problem (on axis I) in itself creates a pressure for more mentalization, as it simultaneously reflects intrapsychic conflict. Based on the present observations, a clinically meaningful deficit in mentalizing, involving some structural pathological character development, would thus likely imply several aspects combined: (1) In the face of emotionally charged conflict, high-defensive states cannot be generated (presumably being substituted instead by either Ref, Conc, or Lo-Def); (2) The available psychic material (prior transformation and quality of representation) being processed is itself of a comparatively lower quality, as described in the VEA model; (3) The teleological stance predominates, and actions are understood in terms of their concrete, physical contexts and outcomes (Bateman & Fonagy, 2004). Such deficits in mentalizing capacity leave the subject ill equipped to deal with current stressors.

Examination of the axis II predictions indicates even more clearly that all three mentalization measures are significantly involved, each uniquely contributing to the model. Increases in high-level defensive activity appears to be associated with a reduction
of axis II severity (part $r = -0.32$, see Table 5), which supports Kernberg (1996) in predicting that a higher level of defensive activity, as one key structural-dynamic aspect, would act as a crucial determinant of less severe personality disorders. The quality of affect elaboration and level of RF also substantially contribute to the reduction of the number of personality disorder diagnoses. This is in line with findings relating low RF with borderline pathology (Fonagy et al., 1996) and with views that underscore the major role played by stable patterns of affect dysregulation in the development of personality disorders. The size of the relationships between the number of axis II diagnoses ($\Delta R^2$ of the mentalization model = 0.198, see Table 5), as well as the fact that all three mentalization measures contribute strongly to the prediction with their own specific focus (with a mean part $r$ of $-0.26$), provide evidence to suggest that personality disorders, compared to attachment security and axis I disorders, are associated with more severe and complex deficits in mentalization. Again one can surmise that a subject’s acquisitions in the realm of mentalization, as demonstrated by specifically three of its interactive facets (the capacity to generate Hi-Def states, higher degrees of affect transformation and levels of representation, intentionality as a developmental-intersubjective acquisition), importantly contribute not only to dealing with past and present stressors, including from axis I disorders, but actually defines a key structural component of normal personality development.

The present study suffers from a number of limitations. The AAI itself shows some limitations when used to assess the other mentalization facets. Other interview formats could focus on specific areas of functioning, and help generate emotional regulation tasks that cover a wider range of affective-emotional categories. A case in point would be a recent use of RF to code Parent Development Interview narratives to examine the link between mental representations and maternal behavior within the intergenerational transmission of attachment (Grienenberger, Kelly, & Slade, 2005). Findings indicated that maternal behavior plays a mediating role, and may be a primary vehicle through which a mother’s reflective functioning is transformed into an appropriately communicated actual presence to her child. Second, our diagnostic measures were used in global fashion, lumping together heterogeneous pathological conditions, leading to obliterate important differences between them, which may interact with any or all of the three mentalization measures here examined. Psychotic conditions and the schizophrenias in particular have not been examined, although differences in mentalizing capacity may well be associated with other factors contributing to relapse prevention, severity of psychotic pathology, and so forth Limitations in statistical power also prevented us from examining in detail the differences in mentalization between the various negatively valenced affects, which would be indicated in future work. To illustrate, a participant may indicate good mentalizing of his or her anxiety and poor mentalizing of his or her sadness or anger.

References


Grienenberg, J., Kelly, K., & Slade, A. (2005). Maternal reflective functioning, mother-infant communication, and infant attachment: Exploring the link between mental states and observed
Mentalization Measures in the AAI


Towards a multidimensional model of the psychoanalytic process: The Montréal transference and countertransference measure. [CD-ROM]. In H. Kächele, E. Mergenthaler, & R. Krause (Eds.), *Psychoanalytic process research strategies II, 20th Ulm workshop of empirical research in psychoanalysis*. University of Ulm, Ulm, Germany.


---

**E-Mail Notification of Your Latest Issue Online!**

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at http://notify.apa.org/ and you will be notified by e-mail when issues of interest to you become available!