Thinking about Assessment: Further Evidence of the Validity of the Movie for the Assessment of Social Cognition as a Measure of Mentalistic Abilities

Andrea Fossati [a], Serena Borroni [b], Isabel Dziobek [c], Peter Fonagy [d] Antonella Somma [a]

a: Department of Human Studies, LUMSA University, Rome, Italy, and Clinical Psychology and Psychotherapy Unit, San Raffaele Hospital, Milano Italy

b: Faculty of Psychology, Vita-Salute San Raffaele University and Clinical Psychology and Psychotherapy Unit, San Raffaele Hospital, Milano Italy

c: Department of Educational Science and Psychology, University of Berlin, Berlin, Germany

d: Research Department of Clinical, Educational and Health Psychology, UCL and Anna Freud Centre, London, UK

Correspondence concerning this article should be sent to Serena Borroni, Servizio di Psicologia Clinica e Psicoterapia, Ospedale San Raffaele Turro, via Stamira d'Ancona, 20, 20127 Milano, Italy. Telephone: +39 02 2643-3241; e-mail address: borroni.serena@hsr.it

Abstract

The present study aimed to evaluate the nomological network validity of the Movie for the Assessment of Social Cognition (MASC) in its Italian translation, addressing distinct research questions in three independent samples of Italian participants comprising adolescent nonclinical participants ($N = 393$), adult nonclinical participants ($N = 193$), and adult outpatients with a Personality Disorder (PD) diagnosis who sought psychotherapy treatment ($N = 59$). In all three samples the MASC proved to be a reliable measure of mentalizing ability, with Cronbach's α values ranging from .70 to .78. In both nonclinical adolescents and nonclinical adults, the MASC scores correlated significantly and meaningfully with the Reading the Mind in the Eyes Test scores. In nonclinical adults, the MASC scores showed significant (albeit modest) correlations with self-reported measures of attachment styles. Finally, in adult outpatients, the MASC "no ToM" scores, that are specific errors that indicating non mentalistic responses, correlated significantly with interview-based measures (Spearman $r = .41$, $p < .01$) and self-reported measures (Spearman $r = .37$, $p < .01$) of borderline personality disorder (BPD), as well as with measures of emotion dysregulation, (Spearman $r = .37$, $p < .01$). As a whole, these findings highlight the validity of the MASC as a measure of mentalization and are consistent with Fonagy and colleagues' (i.e., Bateman & Fonagy, 2004b; Fonagy, 1991) model of mentalization and its role in personality pathology.

Mentalization has been defined as the mental process by which an individual implicitly and explicitly interprets the actions of himself/herself and others as meaningful based on intentional mental states such as personal desires, needs, feelings, beliefs, and reasons (Bateman & Fonagy, 2004a; Fonagy, 1991). Mentalization is a broad concept that subsumes different social-cognitive functions, including emotion recognition, theory of mind, mindreading as well as reflective function. (Ha, Sharp, Ensink, Fonagy & Cirino, 2013). Although the construct of mentalization has been developed within the framework of object relations theory (e.g., Bion, 1967; Winnicott, 1971) and psychoanalytic approach to human attachment (e.g., Bowlby, 1988; Fonagy, 1991; Holmes, 1998), it bears close resemblance to the cognitive construct of "theory of mind" (e.g., Baron-Cohen, Tager-Flusberg, & Cohen, 2000; Dennett, 1987).

The concept of mentalizing (Fonagy, 1991) has been in use in psychoanalytic literature since the 1970s (Allen, 2003; Fonagy, 1991). During the 1980s and 1990s, it was picked up in the neurobiological literature (Morton & Frith, 1989), as well as in the developmental literature, where it has been used interchangeably with the more frequently used concept of "theory of mind." Premack and Woodruff (1978) coined the term theory of mind to refer to the capacity to interpret the behavior of others within a mentalistic framework. From this point of view, mentalizing, or theory of mind, is defined as the set of processes by which children and adults understand themselves and others in terms of how they think, feel, perceive, imagine, react, attribute, infer, and so on (Sharp, Fonagy, & Goodyer, 2008). Therefore, mentalization and theory of mind are related constructs, but mentalization goes one step further to refer to the reflection on others' minds as well as one's own mind (Amodio & Frith, 2006). This is distinct from theory of mind which usually refers to the reflection on the mind of others only. Both mentalizing and theory of mind are subsumed under the umbrella construct of social cognition, which refers to the mental processes involved in perceiving, attending to, remembering, thinking about, and making sense of the people in our social world (Moskowitz., 2005) or the ability to understand ourselves and others as individuals with beliefs, feelings and personality (Mitchell, Macrae, & Banaji, 2004).

The construct of mentalization prompted the development of a psychoanalytically-informed treatment for BPD and other severe personality disorders – e.g., mentalization-based treatment (MBT; Bateman & Fonagy, 2004) that proved to be highly effective in reducing BPD symptoms and improving overall functioning in a randomized clinical trial (Bateman & Fonagy, 1999, 2001). MBT efficacy appeared to be stable over time in an eight-year follow-up study (Bateman & Fonagy, 2008). Finally, Levy and colleagues (2009) showed that the efficacy of transference focused psychotherapy (TFP; Clarkin, Levy, Lenzenweger, & Kernberg, 2007) on BPD features was related to improved measures of mentalizing abilities.

Mentalization: Controversial Aspects and Measurement Issues

Notwithstanding these findings, mentalizing has not been without criticism. For instance, it has been suggested that mentalizing is not a clearly defined construct (Choi-Kain & Gunderson, 2008); concepts of mindfulness, psychological mindedness, empathy, and affect consciousness have been proposed to partially overlap with mentalization within three dimensions: namely, self-/other-oriented, implicit/explicit, and cognitive/affective dimensions (Choi-Kain & Gunderson, 2008). Recently, based on evidence coming from functional magnetic resonance studies, Luyten and colleagues (2012) proposed that mentalization may be underpinned by four functional polarities, a) automatic-controlled, b) internally focused-externally focused, c) self oriented-other oriented, and d) cognitive process-affective process. According to this perspective, mentalization represents a multifaceted construct; evaluation of individuals' mentalizing depends on their functioning with respect to each of the four polarities involved in mentalization. In other words, the "ideal" measure of mentalization should cover all functional polarities.

Issues of measurement and assessment have been a most problematic area of mentalizing theory (Choi-Kain & Gunderson, 2008). Fonagy and colleagues (1998) originally developed the Reflective Function Scale (RFS) as a measure designed to assess the construct of mentalization. Several studies supported the validity of the RF scale: for instance, it has been shown that mentalization is a mechanism for the transmission of attachment security from parents to their children (Fonagy, Steele, Moran, Steele, & Higgitt, 1993; Fonagy, Steele, Steele, Moran, & Higgitt, 1991; Slade, Grienenberger, Bernbach, Levy, & Locker, 2005):),that impairments in mentalizing represent a specific deficit in patients with BPD (Fonagy et al., 1997) that can be changed in psychotherapy (Levy et al., 2006); and that extremely low reflective functioning (RF) is linked to antisocial personality disorder and may serve as a framework to understand violent behavior (Levinson & Fonagy, 2004; Taubner, Wiswede, Nolte, & Roth, 2010). Recently, Taubner and colleagues (2013) convincingly showed that the RF scale possesses sound psychometric properties. Despite these positive findings, exclusive reliance on the RF scale to measure mentalization has been considered a major problem (Choi-Kain & Gunderson, 2008). The RF scale has been criticized for only generating a single, global score, which fails to encompass the complexity of the mentalizing process that it is meant to measure (Choi-Kain & Gunderson, 2008). Moreover, the RF scale has an elaborate coding procedure based on transcripts of Adult Attachment Interviews (AAIs; George, Kaplan, & Main, 1984; 1985; 1996); assessment of mentalization using the RF scale is likely to be time-consuming and costly in nature, thus limiting further research on this topic (Choi-Kain & Gunderson, 2008). Although the validity of the RF scale as a measure of mentalization has been supported by several studies (see Steele & Steele, 2008 and Katznelson, 2014 for a review), such psychological constructs represent abstract, latent variables that cannot be reduced to a single observable measure (Nunnally & Bernstein, 1994); a construct may be said to exist only when it can be measured using different instruments based on different methods (Nunnally & Bernstein, 1994).

Currently, a number of questionnaires, interviews/narrative coding systems and experimental/observational tasks are available to assess mentalization (Luyten, Fonagy, Lowyck, & Vermote, 2012). Notwithstanding their well-known limitations (e.g., rigid administration procedures, limited generalizability to real-life situations, etc.), experimental tasks remain useful tools in mentalization research because they are (1) highly standardized procedures, (2) based on methods that are markedly different from clinical assessment, (3) computer-administered (which strongly limits the effect of measurement error and rater bias), and (4) usually do not require extensive training.

According to Luyten and colleagues (2012)'s overview (2012), among the experimental/observational tasks assessing dimensions of mentalization, the Movie for the Assessment of Social Cognition (MASC; Dziobek et al., 2006) represents a promising measure of mentalizing. The MASC aims at operationalizing social cognition through video, by approximating social interactions the way they actually happen in everyday life (Dziobek et al., 2006). The task is presented as a 15-minute movie cut into 43 segments that represent the test items, each followed by a question regarding the three different mental state modalities. Testing starts with a slide that instructs the subjects that they are going to watch a 15-minute film and that they should try to understand what the characters are feeling and thinking. Four characters are introduced in the form of photographs and names. Participants are then instructed that the film shows these four people getting together for a Saturday evening and that the movie will be stopped at various points and questions will be asked. Subjects are told to try to imagine what the characters are thinking or feeling at the very moment the film is stopped. There are four characters (Sandra, Betty, Michael, and Cliff) and the movie shows the development of different dynamics between them. They have very different motives to participate to the meeting and each character displays stable characteristics (traits) that are different from one another (e.g. outgoing, timid, selfish). They experience different situations that elicit emotions and mental states such as anger, affection, gratefulness, jealousy, fear, ambition, embarrassment, or disgust (for a detailed description on the development of the MASC see Dziobek et al., 2006)

The MASC not only allows for the usual dichotomous (right/wrong) response format, that is reflected in its total score, but also includes a qualitative error analysis where wrong choices (distracters) correspond to one of three error categories: (1) "less theory of mind (ToM)" (undermentalizing), involving insufficient mental state reasoning, in which case a participant may choose an item referring to mental states in an impoverished way; (2) "no ToM" (no mentalizing);

in this case, a participant may fail to choose any item that refers to mental states in explaining behavior by making attributions of physical causality to social situations and mental states; and (3) “excessive ToM” (hypermentalizing), reflecting overinterpretative mental state reasoning (Dziobek et al., 2006; Montag et al., 2011) namely a mental state that is attributed when there is no mental explanation for the situations.

According to Luyten and colleagues’ (2012) overview of measures that assess mentalization, the MASC seems to cover the controlled (i.e. explicit) dimension of mentalization which reflects a serial and relatively slow process that is typically verbal and requires reflections, attention, intention, awareness, and effort (Fonagy & Luyten, 2009). On the contrary, the authors suggested that the automatic (i.e. implicit) dimension of mentalization is poorly covered. Moreover, the MASC assesses the internal-external polarity relying both on how a given character looks externally and is likely feeling inside. In addition, the MASC seems to evaluate only the other dimension (namely the ability to reflect about others’ mental states) of self-other mentalization polarity. Finally, the MASC represents a promising instrument to cover the cognitive –affective mentalization dimension assessing the ability to mark mental representations of others with affective information and subsequently integrate them with cognitive knowledge (Rochat & Striano, 1999). As a whole, the MASC can be considered a promising measure to assess different aspects of mentalization, although psychometric studies are needed to support this issue.

Extant Research Data on the MASC

Available data suggest that the total the MASC score possesses adequate internal consistency, as indicated by Cronbach’s alpha values greater than .80, at least in small-to-moderate adult samples with Asperger Syndrome (Dziobek et al., 2006) or BPD (Preissler et al., 2010), and nonclinical controls. The MASC has high test-retest reliability (r = 0.97; Dziobek et al., 2006). In terms of validity, adults with Asperger Syndrome score significantly and substantially lower than nonclinical controls on the MASC total score (Dziobek et al., 2006), and women with BPD score significantly lower than nonclinical controls on the total score as well as on the emotion, thought, and intention sub-scores (Preissler et al., 2010).

The clinical usefulness of the MASC has been extended in different clinical groups, e.g. paranoid schizophrenia (Montag et al., 2011), narcissistic personality disorder (Ritter et al., 2011), and major depression (Wolkenstein, Schönenberg, Schirm, & Hautzinger, 2011). Moreover, some studies focused on the relationship between the MASC and BPD. In a sample of 111 adolescent inpatients, Sharp and colleagues (2011) reported a relationship between BPD traits and "hypermentalizing" on the MASC, independent of age, gender, externalizing, internalizing and psychopathy symptoms. The relation between hypermentalizing and BPD traits was partially mediated by difficulties in emotion regulation, accounting for 43.5% of the hypermentalizing to BPD path (Sharp et al., 2011). Recently, Sharp and colleagues (2013) replicated the association between hypermentalizing on the MASC and BPD features in a sample of 164 adolescent inpatients; in the same study, hypermentalizing on the MASC, but not other forms of social-cognitive reasoning, was found to be malleable through a milieu-based inpatient treatment.

Open Questions

Thus, available evidence suggests that the MASC represents a reliable and valid measure of mentalizing ability (Luyten et al., 2012). These promising findings also indicate that the time has come to address some issues concerning the MASC that have not been answered yet. For instance, from a psychometric point of view, normative data on the distribution of correct answers and errors on the MASC in nonclinical samples are currently unavailable – i.e. we have no indications on the distribution of mentalistic abilities on the MASC in community dwelling subjects.

From a nomological network validity perspective, there is a dearth of studies reporting associations between the MASC and other measures of mentalistic abilities; up to now, only one study reported data showing significant associations between the MASC scores, and questionnaire-based and interview-based measures of reflective function in adolescent inpatients (Ha, Sharp, Ensink, Fonagy, & Cirino, 2013).

Moreover, most evidence is based on either the German or English version of the task, with the exception of a small study (22 participants with Asperger syndrome and 25 healthy controls) recently carried out in Spain (Lahera et al., 2014); to our knowledge, no data are currently published on the use of the MASC in other languages and cultures.

Finally, knowledge of the relationships between performance on the MASC and personality disorders (PDs) in clinical samples, particularly in adults, is still limited. Up to now, only one study (Ritter et al., 2011) tried to address the issue of mentalistic deficits in narcissistic personality disorder (NPD), reporting unaffected cognitive empathy (i.e. the MASC performance) in NPD and deficits in BPD compared to healthy controls. A significant association between BPD features and hypermentalizing was consistently observed in adolescent inpatients (Sharp et al., 2011, 2013), but this finding was not replicated in adult women suffering from BPD, although they showed a significantly lower the MASC total score than healthy comparison women (Preissler et al., 2010). As a whole, extant research suggests that further data on the specificity of the associations between the MASC scores and selected PD features, possibly assessed using both self-reported and observed rated measures of personality pathology in adult clinical samples may be useful.

This Study

Starting from these considerations, our study aimed to test the psychometric proprieties and the nomological network validity of the MASC, in its Italian translation, addressing distinct research questions in three independent samples of Italian participants composed of nonclinical adolescents, nonclinical adults, and adult outpatients with PD diagnosis who applied for psychotherapy treatment. In particular, we evaluated the internal consistency and unidimensionality of MASC items in both nonclinical adolescents and adults, respectively; in these samples, an item analysis of the distribution of error and correct answers was also carried out.

According to the nomological network approach, a proof of the extent to which a measure defines a construct would have to come from determining how well the measure fits lawfully into a network of expected relationship. From this point of view, we evaluated the convergent validity of the MASC towards a measure analogous to the construct assessed by the MASC (i.e. Reading the Mind in the Eyes Test; RET; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001) in a sample of nonclinical adolescents and nonclinical adults. The number of correct answers on the MASC was expected to show a positive, significant correlation with RET total score, whereas the MASC "exceeding ToM, "less ToM", and "no ToM" scale scores were expected to correlate negatively and significantly with RET total score.

Moreover, we considered the association between the MASC and other constructs (namely, BPD, attachment, and emotional dysregulation) that belong to a network based on Bateman and Fonagy's theory of mentalization (Bateman & Fonagy, 2004b; Fonagy, 1991). Accordingly, nonclinical adults received a measure of adult attachment style to evaluate if errors in mentalizing on the MASC were negatively correlated with measures of secure attachment, and positively correlated with measures of insecure attachment styles. In the adult outpatients, we tested whether the number of BPD traits, assessed using both a semi-structured interview and a self-report measure, was significantly associated with poor performance on the MASC. Based on Fonagy and colleagues' theory of BPD (Bateman & Fonagy, 2004b; Fonagy, 1991), a positive, significant correlation between BPD traits and "no ToM" scale scores should be expected, whereas Sharp and colleagues' (2011, 2013) data on adolescent inpatients would lead us to expect a significant association between BPD features and "exceeding ToM" scale scores. To evaluate the specificity of the associations between BPD traits and selected mentalization problems on the MASC, as well as to increase knowledge of the association between mentalization deficits and personality pathology, we evaluated the significance of the correlations between the MASC scale scores and interview-assessed and self-reported *DSM-IV* (American Psychiatric Association, 1994) PD features other than BPD.

Finally, in adult outpatients, we tested the hypothesis of a significant, positive association between poor performance on the MASC and self-reported measures of emotion dysregulation, that is considered a core feature of BPD and thought to be greatly influenced by mentalization deficits (Bateman & Fonagy, 2004b).

Method

Participants

Nonclinical adolescent participants. Participants were 373 adolescents attending a public high school in Northern Italy. The principal accepted to participate to a research study, which aimed at studying mentalization abilities in adolescents. The participation was voluntary for all the students. 238 participants (63.8%) were female, and 135 (36.2%) were male; mean age was 17.13 years, SD = 1.35 years. All students were unmarried.

Nonclinical adult participants. Participants were 193 Italian community dwelling adults who responded to advertisements requesting volunteers for psychological studies that were placed at the university campus and at San Raffaele Hospital outpatient waiting rooms (with the exclusion of neurology and psychiatry waiting rooms) during fall 2013. Of the non-clinical adult sample, 115 (59.3%) were female, and 78 (40.7%) were male; mean age was 32.77 years, SD = 11.38 years. One participant (0.5%) had a junior high school degree, 95 (49.2 %) had a high school degree, and 97 (50.3 %) had a University degree. 122 participants (63.2%) were unmarried, 60 (31.1%) were married, and 11 (5.7%) were divorced.

Clinical adult participants. This group consisted of 59 outpatients consecutively admitted to the private outpatient clinic of the Clinical Psychology and Psychotherapy Unit of the Scientific Institute San Raffaele, Milano, Italy, during 2012. All participants voluntarily sought psychotherapy treatment for personality pathology. Participants had to meet four criteria for inclusion: (a) IQ>75; (b) no schizophrenia, schizophreniform disorder, delusional disorder, dementia, or organic mental disorder diagnoses; (c) an education level higher than primary school; and d) a clinical diagnosis of

DSM-IV PD. Thirty-eight participants (64.4%) were female, and 21 (35.6%) were male; mean age was 37.02 years, *SD* = 10.42. Eleven participants (18.6 %) had a junior high school degree, 27 (45.8 %) had a high school degree, and 21 (35.6 %) had a University degree. 36 participants (61.0%) were unmarried, 17 (28.8%) were married, and 6 (10.2%) were divorced. 22 participants (33.9%) received at least one *DSM-IV* axis I diagnosis; mood disorders (*n* = 12, 20.3%) were the most frequently diagnosed axis I disorders. When administered the Structured Clinical Interview for DSM-IV Axis II Personality Disorders, Version 2.0 (First, Spitzer, Gibbon, Williams,& Benjamin, 1994), the presence of at least one PD diagnosis was confirmed in 47 participants (79.7%); not otherwise specified PD (*n* = 16, 27.1%), BPD (*n* = 13, 22.0%), and NPD (*n* = 12, 20.3%) were the most frequently diagnosed[1].

Procedure.

All participants gave written consent to participate in the study after it had been explained to them; when participants were of minor age, parents also had to give written consent to allow participation. None of the participants received an incentive, either directly or indirectly, for participating. All subjects volunteered to take part in the study and were treated in accordance with the Ethical Principles of Psychologists and Code of Conduct. All the measures were administered to the participants in random order.

Measures

Movie for the Assessment of Social Cognition. All participants were administered the Italian translation of the MASC in individual session at school (in the case of nonclinical adolescents) or at San Raffaele Hospital by trained M.Sc. Psychology students (in the case of

---

[1] Additional information on the comparisons among the three sample based on socio-demographic features are available upon request.

nonclinical adults and adult outpatients). Consistent with other studies (e.g., Ha et al., 2013; Sharp et al., 2011, 2013), we chose a multiple-choice format for the MASC (Dziobek et al., 2006).

In the translation process, the authors closely followed Denissen, Geenen, van Aken, Gosling, and Potter's (2008) indications. Four clinical psychologists dubbed the audio of the English MASC into Italian. Following Dziobek and colleagues' (2006) procedure, the Italian translation of the MASC was cut in the same 43 segments of the English version of the task that were embedded in a PowerPoint presentation (Dziobek et al., 2006).

As we stated above, participants were provided with four response options: (1) a hypermentalizing response, (2) an undermentalizing response, a (3) no mentalizing response and an (4) accurate mentalizing response. In order to derive a summary score of each of the subscales, points were added. Additionally, the MASC provided six control answers to evaluate individual's attention during the task.

Reading the Mind in the Eyes Test Revised Version. As an additional measure of mentalization, all nonclinical participants were administered the Italian version of the Reading the Mind in the Eyes Test (RET; Baron-Cohen et al., 2001). The RET presents 36 black-and-white photographs of the area of the face immediately surrounding and including the eyes. Participants were asked to choose one of four words (three distracter words and one correct word) that describe the mental state of the person in the photograph. Scores were calculated as total number of correct discriminations for all 36 items. Recently RET was validated in Italian sample and the data show adequate internal consistency and test retest stability (Vellante et al, 2012). In our study, the RET yielded moderately reliable scores in both Italian nonclinical adolescents, M = 25.53, SD = 3.79, Cronbach's α = .73, and nonclinical adults, M = 24.92, SD = 3.93, Cronbach's α = .70.

Attachment Style Questionnaire. The ASQ is a 40-item Likert-type self-administered questionnaire, designed to measure five dimensions of adult attachment: Confidence in Self and

Others (eight items), Discomfort with Closeness (10 items), Relationships as Secondary (seven items), Need for Approval (seven items), and Preoccupation with Relationships (eight items) (Feeney, Noller, & Hanrahan, 1994). Reliability and validity data were provided for both English (Feeney et al., 1994) and Italian (Fossati et al., 2003) versions of the ASQ. In our sample, Cronbach's α values for ASQ subscales ranged from .63 for Confidence to .79 for Relationships as Secondary.

Before being administered the MASC, adult outpatients were administered the Italian versions of the Structured Clinical Interview for DSM-IV Axis II Personality Disorders, Version 2.0 (SCID-II; First et al., 1994), Personality Diagnostic Questionnaire-4+ (PDQ-4+; Hyler, 1994), and Difficulties in Emotion Regulation Scale (DERS; Gratz & Roemer, 2004) as part of their routine clinical assessment.

Structured Clinical Interview for DSM-IV Axis II Personality Disorders, Version 2.0. The SCID-II (First et al., 1994) is a 140-item semistructured interview designed to provide both a categorical and dimensional (i.e. number of symptoms) assessment of *DSM-IV* PDs. SCID-II items are organized by diagnosis. In the present study, subjects with Axis I diagnoses were administered the SCID-II by expert rater only when the acute Axis I symptoms were judged by their clinicians to be in remission. Inter-rater reliability of the SCID-II diagnoses was assessed using a pairwise interview design. Interclass correlation coefficients (ICCs), based on a one-way random effects ANOVA, were computed to evaluate the inter-rater reliability of dimensionally assessed SCID-II personality disorders. As a whole, the inter-rater reliability of the dimensional SCID-II diagnoses was acceptable, median ICC value = .84, SD = .15; in particular, the ICC value for the dimensional BPD diagnosis was .88. The inter-rater reliability of the categorical SCID-II personality disorder diagnoses was assessed computing Cohen κ coefficient. The Cohen κ value for BPD diagnosis was .90; the inter-rater reliability of any PD diagnosis was adequate, κ = .88.

Personality Diagnostic Questionnaire–4+. The PDQ-4+ (Hyler, 1994) is a self-report questionnaire with 99 true/false items, designed to measure the 10 PDs included in DSM-IV Axis II and the 2 PDs (passive-aggressive and depressive) proposed for further research. The PDQ-4+ PD scales list one item for each DSM-IV PD criterion; the higher the PD scale total score, the higher the number of criteria for a given PD. The PDQ-4+ also yields an overall score that reflects the overall level of self-reported personality pathology. The Italian version of the PDQ-4+ showed adequate psychometric proprieties (Fossati et al., 1998). In our sample, the BPD PDQ-4+ diagnosis showed adequate internal consistency reliability, Cronbach's α = .72; all the other PDQ-4+ *DSM-IV* PDs yielded moderately reliable scores, with α values ranging from .69 (Obsessive-Compulsive PD), to .83 (Antisocial PD).

Difficulties in Emotion Regulation Scale. The DERS (Gratz & Roemer, 2004) is a 36-item measure that provides a comprehensive assessment of overall emotion dysregulation as well as six specific dimensions: nonacceptance of negative emotions (6 items), difficulties engaging in goal-directed behaviors when distressed (5 items), difficulties controlling impulsive behaviors when distressed (6 items), limited access to effective emotion regulation strategies (8 items), lack of emotional awareness (6 items), and lack of emotional clarity (5 items). In our study, we used the DERS total score as an overall score of emotional dysregulation. The DERS total score seemed to have adequate internal consistency reliability, M = 2.90, SD = 0.68, Cronbach α = 0.92.

Data Analyses

We performed dimensionality analyses of the MASC items based on differing criteria; in particular, the minimum average partial statistic (MAP; Zwick & Velicer, 1986), quasi-inferential parallel analysis (Buja & Eyuboglu, 1992), and Hull's method (Lorenzo-Seva, Timmerman, &

Kiers, 2011; Wilderjans, Ceulemans, & Meers, 2013) were used for determining the number of latent dimension underlying the tetrachoric correlation matrix of the MASC items.

Factor structure replicability across adolescent and adult participants was used as a further criterion for determining the correct number of dimensions underlying the tetrachoric correlations among the MASC items. Unweighted least square was used as the EFA algorithm; in the case of multi-factor solutions, Promax oblique algorithm ($k = 4$) was used to rotate the extracted factors. The replicability of the best fitting factor solution across sub-groups defined by gender, civil status, and educational level was evaluated by computing congruence coefficients (CC; Gorsuch, 1983). All dimensionality analyses were carried out using Version 8.1 of the FACTOR statistical software routine (Lorenzo-Seva & Ferrando, 2006). Cronbach's α coefficient based on item tetrachoric correlations was used to evaluate the MASC internal consistency reliability.

Within each nonclinical sample, Pearson r coefficient was used to evaluate the relationships between the MASC scores and participants' age and RET scores, respectively; t-tests were computed to evaluate the presence of a significant effect of gender on the MASC scores; for mean comparisons, Cohen d (Cohen, 1988) was used as an effect size measure.

Since the three MASC error scales were based on the same number of items, in both nonclinical samples repeated measure MANOVA, followed by Bonferroni pairwise contrasts, was used to evaluate if at least one of the error scales showed a significantly higher mean score when compared to the other two error scales. Pillai V was used as an effect size measure in MANOVA analyses.

The percentage of correct answers and the percentage of errors were used as "difficulty" indices for each item; Cochran Q tests, followed by pairwise McNemar tests with Bonferroni-corrected p-level, were used to identify the presence of significant differences on error distribution on each item (i.e. to detect if a specific category of errors was significantly more frequent than other categories of error on a given item). For descriptive purposes, the MASC total scores were divided by the possible maximum score (i.e. 45) to estimate the proportion of answers on the keyed direction; for ease of presentation, proportion is listed as corresponding percentages.

Student *t*-tests were computed to detect the presence of significant differences on the MASC mean scores between adolescents and adults; these analyses were followed by one-way ANCOVAs in which participants' age was entered as covariate, in order to assess if the differences in mean scores between the two samples became nonsignificant after controlling for age.

In the nonclinical adult sample, hierarchical regression analyses were carried out to evaluate if the ASQ scales significantly predicted the MASC scores while controlling for the confounding effect of age and gender. Each MASC scale was used in turn as dependent variable; participants' age and gender were entered in step 1 of the hierarchical models, while the ASQ scales were entered in step 2 of the hierarchical regression equation. Change in adjusted R^2 value and Cohen f^2 index (Cohen, 1988), were computed as effect size measures; the Variance Inflation Factor (VIF) was computed to assess collinearity (Hsieh, Lavori, Cohen, & Feussner, 2003).

The small sample size prevented from using a multivariate approach in testing the associations between the MASC scores and dimensionally-assessed (i.e. number of criteria) *DSM-IV* PDs and DERS scores, respectively. Considering that a relatively small number of adult outpatients participated in the present study, we relied on Spearman *r* coefficient to assess the significance of correlations between the MASC scores and the other key variable scores. In order to evaluate if categorical PD diagnoses were significantly associated with the MASC deficits, the Mann-Whitney U test was used; rank biserial correlation (r_b) was used as an effect size measure (Wendt, 1972).

In all statistical analyses, the *p*-level for statistical significance was $p < .05$.

Results

Dimensionality Analysis Results

In the nonclinical adolescent sample, the MAP values for the first five Unweighted least squares (ULS) for the MASC factors were .006, .007, .008, .01, and .02, respectively; thus, MAP statistic reached its minimum values when the first ULS factor was extracted, suggesting a unidimensional solution. Both quasi-inferential parallel analysis and Hull method consistently replicated the result of MAP analysis in the nonclinical adolescent sample.

When the tetrachoric correlations among the MASC items were computed in the nonclinical adult sample, dimensionality analyses yielded results that were less consistent than those observed in the nonclinical adolescent sample. For instance, the Hull method suggested a unidimensional solution, since the best balance between goodness-of-fit index values (CFI = .39) and number of estimated parameters (*df* = 90) was reached when the first ULS factor was extracted; however, MAP statistic and quasi-inferential parallel analysis suggested a two-factor solution.

Factor replicability results suggested that no more than one factor could be reliably extracted from the tetrachoric correlation matrix of the MASC items in our two nonclinical samples. When factors that were extracted in the nonclinical adolescent sample and in the nonclinical adult sample, respectively, were formally compared a CC value of .85 was observed for the one-factor model of the MASC items. Rather, the comparison of the two-factor solutions obtained in the adolescent sample and in the adult sample yielded CC values .74 and .58 for factor 1 and factor 2, respectively. Thus, according to current standards (Lorenzo-Seva & ten Berge, 2006) CC values suggested a fair replicability only in the case of the one-factor model of the MASC items.

The unidimensional model of the MASC items showed adequate fit, at least in terms of root mean square difference between the observed correlation matrix and the reproduced correlation matrix; standardized root mean square error values were .05 and .08 in the nonclinical adolescent sample and nonclinical adult samples, respectively.

Internal Consistency Reliability, Descriptive Statistics, and Item Analysis of the MASC Scales

The MASC total score showed adequate internal consistency reliability in the nonclinical adolescent sample, as indicated by a Cronbach's *α* value of .76; Cronbach's α values were almost identical when items were coded according to specific errors. Similar findings were observed among nonclinical adults; in this sample, the Cronbach's *α* value for the MASC total score was .80.

The MASC descriptive statistics and gender comparisons in nonclinical adolescents and nonclinical adults, respectively, are listed in Table 1[2]. In the nonclinical adolescent sample, no significant correlation was observed between age and the MASC scores. In the nonclinical adult sample, age correlated significantly and negatively with the number of correct answers, $r = -.17$, $p <.05$, and showed positive, significant correlations with the number of overall errors, $r =.16$, $p <.05$, number of "less ToM" answers, $r = .14$, $p <.05$, and number of "no ToM" answers, $r = .16$, $p <.05$. Before performing ANCOVAs, the hypothesis of parallelism of regression lines of the MASC scores on age between the two nonclinical groups was formally tested; none of the age-by-group interaction terms was significant. After holding constant the effect of age, nonclinical adolescents did not differ significantly from nonclinical adults on number of correct answers, number of "no ToM" answers and number of "exceeding ToM" answers. A slight, albeit significant difference on the number of "less ToM" answers remained between the two groups even when the effect of age was held constant, $F(1, 563) = 8.87$, $p <.01$, $\eta^2 = .02$. In both nonclinical samples the MASC task elicited a higher number of correct answers than errors; repeated measure analyses showed that this difference was highly significant, Pillai $V =.48$, $p <.001$, although it was significantly more pronounced in the nonclinical adolescent sample than in the nonclinical adult sample, interaction effect Pillai $V = .02$, $p <.001$. Significant differences were observed among average numbers of specific errors, Pillai $V = .57$, $p <.001$, although they were sharper in the nonclinical adolescent sample than in the nonclinical adult sample (with the exception of the difference between "less ToM" and "no ToM"), interaction effect Pillai $V = .04$, $p <.001$. According to Bonferroni contrasts,

[2] Additional information about the association between the MASC and social demographic characteristics as well as gender differences in the three samples is available upon request.

both samples scored significantly higher on “exceeding ToM” than on “less ToM” (p <.001) and “no ToM” (p <.001); on average “no ToM” answers were significantly less common than “less ToM” answers (p <.001).

In the nonclinical adolescent sample, the MASC items differed significantly as to their “difficulty” level, Cochran $Q(44)$ = 1470,29, although on average they elicited correct answers, M = 62.5%, Mdn = 64.0%, SD = 22.1%. In particular, 14 (31.1%) items elicited more than 50% of error responses among nonclinical adolescents; items 6 (percentage of correct answers = 4.61%) and 4 (percentage of correct answers = 12.47%) were the most difficult. According to Cochran Q analyses and post-hoc Bonferroni-corrected McNemar tests, nine (20%) items (i.e. items 12, 15, 21, 25, 28, 34, 38, 42, and 45) did not show significant differences in the frequencies of “exceeding ToM”, “less ToM”, and “no ToM” answers. The remaining items yielded significant Cochran Q values, ranging from 6.30 (item 34) to 370.13 (item 8), all ps <.05, suggesting that the “exceeding ToM”, “less ToM”, and “no ToM” answers were not equally likely to occur. The majority of items (n = 28, 77.8%) that yielded significant Cochran Q values (i.e. 36 items) elicited “exceeding ToM” answers more frequently than either “less ToM” or “no ToM” answers.

Similar findings were observed in the nonclinical adult sample. On average, the MASC items elicited correct answers, M = 59.3%, Mdn = 62.8%, SD = 20.20%; however, they differed significantly as to “difficulty” level (i.e. number of errors), Cochran $Q(44)$ = 3448.27, p <.001. According to Cochran Q statistics, only eight (17.8%) items (i.e. items 6/7, 12, 21, 24, 25, 29, 34, and 45) elicited roughly the same number of “exceeding ToM”, “less ToM”, and “no ToM” answers, respectively, thus yielding nonsignificant Q values. All other items yielded significant Q values, ranging from 7.53 (item 2) to 210.52 (item 28), all ps <.05, suggesting that the “exceeding ToM”, “less ToM”, and “no ToM” answers were not equally likely to occur among nonclinical adults; Bonferroni-corrected McNemar tests showed that “exceeding ToM” answers were significantly more frequent than either “less ToM” or “no ToM” answers, respectively, on the majority of items (n = 22, 57.9%) that yielded significant Cochran Q values (i.e. 38 items).

Considering the adult outpatient sample, the MASC total score showed adequate internal consistency reliability, as indicated by a Cronbach's α value of .78 (M = 28.53, SD = 4.48). Among adult outpatients, age correlated significantly only with the overall number of errors on the MASC, Spearman r = .32, p <.05, and with the number of "less ToM" errors, Spearman r = .32, p <.05. The number of answers on the distractor scale of the MASC (M = 4.46, SD = 1.38) did not significantly differentiate adult outpatients from nonclinical adults (see Table 1); similar considerations held also for the number of "less ToM" errors (M = 5.89, SD = 3.41), and "no ToM" errors (M = 3.22, SD = 2.03), respectively. Among adult outpatients, age correlated significantly only with the overall number of errors on the MASC, Spearman r = .32, p <.05, and with the number of "less ToM" errors, Spearman r = .32, p <.05. The number of answers on the distractor scale of the MASC (M = 4.46, SD = 1.38) did not significantly differentiate adult outpatients from nonclinical adults (see Table 1); similar considerations held also for the number of "less ToM" errors (M = 5.89, SD = 3.41), and "no ToM" errors (M = 3.22, SD = 2.03), respectively.

Correlations between the MASC Scores and the RET Total Score

Two-way ANOVA results showed that nonclinical adolescents did not differ significantly from nonclinical adults on the RET total score; rather, in both groups males (M = 24.79, SD = 4.15) scored significantly lower than females (M = 25.64, SD = 3.61), $F(1, 562) = 6.08$, p <.05, η^2 = .01, with no significant group-by-gender interaction effect. Among nonclinical adolescents, the RET total score correlated positively and significantly with age, r = .27 p <.001, whereas among nonclinical adults the RET total score showed a nonsignificant, negative correlation with age.

In the nonclinical adolescent sample, the number of correct answers on the MASC showed a significant correlation with the RET total score, r = .30, p <.001; when this correlation was corrected for attenuation due to measurement error, the r value became .41. The RET total score showed significant, negative correlations with the overall number of errors on the MASC, r = -.32, p <.001, number of "exceeding ToM" answers, r = -.19, p <.001, number of "less ToM" answers, r

= -.17, p <.001, and number of “no ToM” answers, r = -.16, p <.005. The number of distractor items correctly answered on the MASC was positively correlated with the RET total score, r = .16, p <.005.

Among nonclinical adults, the RET showed a correlation with the MASC scores even stronger than those observed among nonclinical adolescents. The number of correct answers on the MASC showed a significant correlation with the RET total score, r = .45, p <.001; when this correlation was corrected for attenuation due to measurement error, the r value became .60. Correlations of the RET total scores with the overall number of errors on the MASC, number of “exceeding ToM” answers, number of “less ToM” answers, and number of “no ToM” answers were -.45, -.23, -.33, and -.28, all ps <.005, respectively. The number of distractor items correctly answered on the MASC was positively correlated with the RET total score, r = .30, p <.001.

The correlation between the RET total score and the number of correct answer on the MASC observed in the nonclinical adult sample was significantly larger than the corresponding correlation observed in the nonclinical adolescent sample, z = 1.96, 2-tailed p <.05; none of the other comparisons reached statistical significance (i.e. p <.05).

Correlations between the MASC Scores and ASQ Scale Scores in Nonclinical Adult Participants

ASQ scales descriptive statistics Cronbach’s α values and correlations with the MASC scores in nonclinical adults are listed in Table 2. One-way MANOVA did not evidence any significant effect of gender on ASQ scale scores. With the exception of the correlation between age and Discomfort with Closeness, r = .20, p <.01, none of the other correlation coefficients between age and the ASQ scales reached statistical significance (i.e. p <.05).

As can be observed in Table 2, the number of correct answers on the MASC correlated positively and significantly with ASQ Confidence scale, conversely, the overall number of errors on the MASC showed a significant, negative correlation with Confidence. The number of errors on the

MASC correlated positively and significantly with Discomfort with Closeness. The number of "Exceeding ToM" answers showed a negative, significant correlation with Confidence, and a positive and significant correlation with Relationships as Secondary. The number of "Less ToM" answers correlated positively and significantly with both Discomfort with Closeness and Need for Approval. Finally, the number of "no ToM" answers correlated positively and significantly with both Discomfort with Closeness and Preoccupation with Relationships.

Interestingly, the RET total score showed a significant, negative correlation only with the ASQ Need for Approval scale, $r = -.15$, $p < .05$; none of the remaining correlation coefficient between RET total score and the ASQ scales reached statistical significance.

Associations between the MASC Scores and Measures of Personality Pathology and Emotion Dysregulation in the Italian Adult Outpatient Sample.

Adult outpatients with at least one PD diagnosis ($M = 6.38$, $SD = 3.43$) showed a significantly higher number of "less ToM" answers than participants with no PD diagnosis ($M = 4.00$, $SD = 2.52$), $U = 167.0$, $z = 2.18$, $p < .05$, rank $r_b = .41$. None of the other MASC scale scores significantly discriminated participants with at least one PD diagnosis from participants with no PD diagnosis.

The presence of at least one axis I diagnosis was not significantly associated with a lower number of correct answers on the MASC; similarly, none of the MASC specific error scores significantly differentiated participants who received at least one axis I diagnosis from participants with no axis I diagnosis.

Consistent with previous reports (e.g., Fossati et al., 1998), in the present study the PDQ-4+ seemed to overdiagnose BPD, Wilcoxon $z = 5.95$, $p < .001$, as well as all the other dimensionally assessed (i.e. number of criteria) *DSM-IV* PDs; notwithstanding this finding, self-reported (i.e. PDQ-4+) and observer-rated (i.e. SCID-II) dimensional BPD diagnoses were substantially and significantly correlated, Spearman $r = .61$, $p < .001$. With the exception of Antisocial PD, Spearman

r = .52, p <.001, none of the other correlations between each PDQ-4+ PD scale scores and the corresponding dimensionally assessed SCID-II PD diagnosis reached statistical significance.

The correlations (Spearman r coefficients) between SCID-II and PDQ-4+ dimensionally assessed BPD diagnosis, and the MASC scale scores are listed in Table 3; in order to evaluate the specificity of these associations, the correlations (Spearman r coefficients) between the MASC scores and the remaining dimensionally assessed *DSM-IV* PD diagnoses are also reported. Dimensionally assessed BPD diagnosis based on both self-reports (i.e. PDQ-4+ BPD scale scores) and SCID-II interviews were significantly, positively correlated with the number of “no ToM” errors. None of the other dimensionally assessed *DSM-IV* PDs based on SCID-II interviews correlated significantly with any of the MASC scales. When categorical DSM-IV BPD diagnosis was considered, participants who met *DSM-IV* BPD diagnosis according to SCID-II interview showed a significantly higher mean number of “no ToM” errors than participants who did not meet *DSM-IV* criteria for BPD diagnosis based on SCID-II interview, U = 174.0, z = 2.34, p <.05, rank r_b = .42. Interestingly, the number of correct answers did not significantly differentiate participants who received a *DSM-IV* Not Otherwise Specified (NOS) PD (i.e. Mixed PD) diagnosis based on SCID-II interview (n = 16, M = 28.94, SD = 4.39) from participants who did not receive a NOS PD diagnosis (n = 43, M = 28.58, SD = 4.38); similarly, none of the other MASC scale scores was significantly associated with *DSM-IV* NOS PD diagnosis.

When PDQ-4+ PD self-reports were considered, the association between BPD and “no ToM” errors was less specific; indeed, the number of “no ToM” answers correlated positively and significantly also with PDQ-4+ Antisocial, Paranoid, Dependent, and Passive-Aggressive PD scale scores. Interestingly, the overall number of errors on the MASC showed positive correlations with PDQ-4+ Borderline, Paranoid, and Schizotypal PD scale scores.

No significant correlations were observed between the number of control answers on the MASC and PDQ-4+ BPD scale scores and a number of BPD criteria according to SCID-II interview.

Consistent with previous reports (Gratz & Roemer, 2004), the DERS total score correlated significantly with both dimensionally assessed SCID-II BPD diagnosis, Spearman $r = .52$, $p < .001$, and PDQ-4+ BPD scale scores, Spearman $r = .60$, $p < .001$. Considering dimensionally assessed SCID-II PD diagnoses, significant correlations were observed between the DERS total score and Dependent PD diagnosis, Spearman $r = .30$, $p < .05$; none of the remaining Spearman r coefficients reached statistical significance. When SCID-II categorical PD diagnoses were taken into consideration, participants who received a BPD diagnosis ($n = 13$, $M = 3.49$, $SD = 0.49$) scored significantly higher on the DERS total score than participants with no BPD diagnosis ($n = 46$, $M = 2.73$, $SD = 0.63$), $U = 89.5$, $z = 3.69$, $p < .001$, rank $r_b = .70$. No significant association was observed between the DERS total score and SCID-II NOS PD diagnosis.

Considering the correlations between the DERS total score and the individual PDQ-4+ PD scale scores, with the exception of PDQ-4+ Obsessive-Compulsive PD and Schizoid PD, all PDQ-4+ PD scales correlated significantly with the DERS total score, with Spearman r values ranging from .32 (Histrionic PD) to .63 (Passive-Aggressive PD), all ps $< .05$.

Considering the correlations between the MASC scale scores and the DERS total score, the DERS total score did not show any significant correlation with the overall number of correct answers on the MASC, as well as with the number of “exceeding ToM” errors, “less ToM” errors, and correct control answers; rather, a significant correlation was observed between the number of “no ToM” errors on the MASC and the DERS total score, Spearman $r = .37$, $p < .01$.

Partial correlation analyses based on Spearman r values indicated that the correlation between number of SCID-II BPD features and DERS total score remained significant when the effect of “no ToM” errors was held constant, Spearman partial $r = .43$, $p < .01$; similarly, the associations between number of SCID-II BPD criteria and “no ToM” features remained significant even after controlling for the effect of DERS total score, Spearman partial $r = .34$, $p < .05$. Rather, the correlation between DERS total score and “no ToM” errors dropped to non-significance when the effect of the number of SCID-II BPD features was held constant.

Discussion

Psychometric Characteristics and Normative Data of the MASC

The present study represents the first comprehensive attempt at testing how the MASC works as a measure of mentalization in two moderately large samples of nonclinical participants at different life stages – namely, adolescence and adulthood – as well as in adult outpatients, at least in its Italian translation.

Confirming and extending previous findings (e.g., Dziobek et al., 2006; Dziobek et al., 2011; Ha et al., 2013; Lahera et al., 2014; Preissler et al., 2010; Sharp et al., 2011; Sharp et al., 2013), the results of our study suggest that the MASC is a reliable measure of social cognition that taps into different aspects or dimensions of mentalization and can serve as proxy for the polarities underlying mentalization.

Dimensionality analyses and factor replicability criterion seemed to support the unidimensional structure of the MASC in both nonclinical samples. This finding seems to suggest that the MASC represent a construct implicating multiple aspects that are linked up to a unique dimension (namely the ability to read others' intentions). The MASC task elicited a substantially and significantly higher number of correct answers than errors in both nonclinical samples.

Consistent with previous reports (e.g., Dziobek et al., 2006; Preissler et al., 2010), nonclinical participants showed a majority of correct responses on the MASC; descriptively, both nonclinical adolescents and nonclinical adults correctly recognized the mental states of the characters in roughly 60% of scenes. This finding is consistent with Fonagy and colleagues' (e.g. Luyten et al., 2012) hypothesis that the ability to infer mental states fluctuates even in nonclinical subjects. Item analysis results of the MASC task provided further support for this point of view; they evidenced significant differences between scenes in eliciting difficulties in correctly recognizing the characters' mental states in both nonclinical adolescents and nonclinical adults.

As previously stated, in our study the Italian translation of the MASC showed reliability data akin to those previously reported in the literature (e.g., Dziobek et al., 2006; Dziobek et al., 2011;

Ha et al., 2013; Lahera et al., 2014; Preissler et al., 2010; Sharp et al., 2011; Sharp et al., 2013) and elicited a majority of correct answers in nonclinical participants who were expected to present preserved mentalistic abilities. However, the average MASC score observed in our nonclinical adult sample, as well as in its sub-groups based on gender, was markedly lower than the average MASC scores reported in the literature for healthy controls (e.g., Dziobek et al., 2006; Dziobek et al., 2011; Lahera et al., 2014; Preissler et al., 2010). This finding may be explained by linguistic differences and cultural differences in social interaction between Italy and other European countries; however, it is also possible that this finding reflects the use of community dwelling (i.e. nonclinical) participants rather than highly selected healthy controls. Since one of the study's aims was to obtain initial normative data for the MASC scores, we preferred to rely on community dwelling adolescents and adults, in order to obtain samples that could be more representative of the "average Italian" than healthy subjects explicitly assessed for the absence of any mental disorder and characterized by above-average IQ (e.g., Dziobek et al., 2006; Lahera et al., 2014; Preissler et al., 2010). Moreover, since in the adolescent sample we found significant differences between male and female scores for the number of correct answers (higher for female) and for the number of exceeding ToM answers (higher for males), the normative data should be split into male and female data.

Extending previous studies (Bleiberg, Rossouw, & Fonagy, 2012), our results in nonclinical adolescents and nonclinical adults suggest that mentalizing abilities, at least as they are measured by the MASC procedure, may undergo significant quantitative changes in the transition from adolescence to adulthood. Although comparisons between the MASC scores in the nonclinical adolescents and those in the nonclinical adults may have been influenced by sampling error, as well as by a number of sample characteristics (e.g., socio-economic status, IQ, school level, etc.), our findings did not support the hypothesis that adolescents experience a higher number of problems in mentalization when compared to adults.

Nonclinical adolescents performed slightly, albeit significantly, better than nonclinical adults in correctly inferring characters' mental states on the MASC. This is inconsistent with developmental findings concerning mentalizing using a physical perspective-taking task. Interestingly, the MASC control answers did not significantly discriminate nonclinical adolescents from nonclinical adults, suggesting that both groups showed roughly the same overall commitment to the task.

Our findings strongly stress the need for longitudinal studies designed to understand how mentalistic abilities develop and change across the life cycle. Although cross-sectional comparisons between adolescents and adults suggest a small decrease in hypermentalization and an increase of hypoactivation of mentalization in the transition from adolescence to adulthood, repeated measure analyses showed that the pattern of errors on the MASC was similar in both nonclinical adolescents and nonclinical adults. In both samples, "exceeding ToM" answers consistently appeared as the most frequent error, followed by "less ToM" answers and "no ToM", respectively. In other words, our data suggest that hypermentalization is the flaw in mentalistic abilities most frequently observed in nonclinical participants, whether adults or adolescents. Unfortunately, our findings do not allow us to tell whether the relatively high frequency of hypermentalization observed in both nonclinical samples represented an attempt to use an excessive number of cause-effect inferences in order to cope with a situation-specific deficit of mentalization, or if hypermentalization was the consequence of contest-specific difficulties in disentangling the subject's mental states from others' mental states. Extending Bateman and Fonagy's (2004b) thoughts on the relationships between certain motivated distortions of interpersonal attribution and mentalizing failures in BPD, our considerations suggest that future studies on the relationship between motivational structures such as defense mechanisms and mentalization may help increase the specificity of our understanding of failures of social cognition.

Convergent Validity of the MASC

Our findings support the convergent validity of the MASC as a measure of the ability to infer others' mental states in both nonclinical adolescents and nonclinical adults. Indeed, the MASC showed significant, positive correlations with the RET in both nonclinical samples. The moderate size of the correlation coefficients observed was consistent with the conceptualization of mentalization as a multifaceted construct (Luyten et al., 2012). According to this perspective, both the RET and the MASC seemed to assess the explicit and cognitive-affective dimensions of mentalization. The RET is based on external features of others, whereas the MASC also allowed the partial assessment of automatic (implicit) mentalization based on internal features of others (Luyten et al., 2012). Moreover, while the MASC is dependent on contextual cues, requiring inference of mental states from indicators that are not physically apparent, the RET calls on an individual's capacity to read the mental states of others from exclusively external cues (i.e. a static image of the eye region).

Adult Attachment and Mentalization in the Nonclinical Adult Sample

Despite the modest size of the coefficients, all correlations of the MASC scores with self-reported measures of adult attachment styles (i.e., the ASQ scales) in nonclinical adults were in the predicted direction and were consistent with the hypothesis that mentalization is influenced at least to some extent by attachment style (Bateman & Fonagy, 2004b; Fonagy, 1991). The number of correct answers showed a positive and significant, albeit modest correlation with a measure of secure attachment – namely, the ASQ Confidence in Self and Others scale – and a negative, significant correlation with ASQ Discomfort with Closeness scale which is central to Hazan and Shaver's (1987) concept of avoidant attachment. Conversely, the overall number of errors on the MASC was negatively related to ASQ Confidence scale scores, and positively associated with ASQ Discomfort with Closeness scale scores. As a whole, these findings were consistent with recent fMRI studies implying an inhibition of mentalizing networks associated with the intense activation of neural systems linked to attachment (Zeki, 2007), at least in BPD subjects. Recent neuroimaging evidence of the deactivation of regions associated with mentalizing specifically linked to attachment related stress was reported using a modified version of the RET measure (Nolte et al. 2013).

The number of “exceeding ToM” errors showed a significant and negative correlation with the ASQ Confidence scale, and a positive and significant correlation with ASQ Relationships as secondary to achievement scale, a measure of dismissing attachment (Bartholomew, 1990) which describes an attachment style characterized by protection against disappointment through avoiding close relationships and maintaining a sense of independence and invulnerability. Avoidant attachment, at least as operationalized by the ASQ Discomfort with Closeness scale, was significantly and positively associated with “less ToM” and “no ToM” answers; interestingly, the number of “less ToM” answers and “no ToM” answers was also associated with measures of anxious attachment styles in nonclinical adults. In particular, the frequency of “less ToM” answers was significantly correlated with respondents’ need for acceptance and confirmation from others in attachment relationships in adulthood, at least as they were assessed by the ASQ Need for Approval scale, which was designed to measure an adult attachment style that is central to Bartholomew’s (1990) fearful and preoccupied groups. Finally, the frequency of blindness to characters’ mental states (i.e. “no ToM” answers) correlated significantly with Preoccupation with Relationships, that involves an anxious and dependent approach to relationships and represents a core feature of Hazan and Shaver’s (1987) original conceptualization of anxious/ambivalent attachment.

According to our findings, in nonclinical adults each adult attachment style assessed by the ASQ had a definite, albeit modest relationship with an error category on the MASC. Moreover, whereas the RET task showed a significant and positive relationship only with the ASQ Need for Approval scale and failed to detect other significant associations, the more ecologically valid the MASC clearly identified significant correlation in the predicted direction between mentalizing, at least as operationalized by the MASC, and adult attachment styles, as assessed by the ASQ. This finding suggests the relevance of assessing attachment in adults in order to understand difficulties in mentalization and is coherent with the hypothesis that attachment dynamics may be associated with fluctuations in mentalistic functioning (Fonagy, Bateman, & Luyten, 2012). Moreover, the association was weak, suggesting that a number of participants with insecure attachment may have not developed difficulties in mentalization. This is highly consistent with Bateman and Fonagy's (2004b) considerations that attachment disturbances per se are not likely to produce mentalization deficits; rather, these may be the results of specific caregiver-infant interactions based on non-contingent or non-marked mirroring of the infant's inner states by the caregiver in the context of insecure attachment. Finally, in a nomological network framework, a self-report measure of adult attachment covers only the self polarity of mentalization, specifically the explicit and cognitive affective internal states of the self.

As a whole, the associations between the MASC and the RET and the ASQ supported the hypothesis that the MASC is a valid instrument able to capture the explicit, cognitive, external and other dimensions of mentalization (RET) and the explicit, affective/cognitive, internal and self dimensions (ASQ), respectively. Further data are needed to evaluate its ability to cover remaining (e.g. implicit) dimension of mentalization.

Mentalization, BPD Features, and Emotion Dysregulation in the Outpatient Clinical Sample

Notwithstanding the limited sample size, in adult outpatients the MASC showed meaningful relationships with key variables, namely BPD features and difficulties in emotion regulation, which represent relevant aspects of the nomological network validity of mentalization measures (Fonagy et al., 2012; Luyten et al., 2012). Although adult outpatients did not significantly differ in mentalistic abilities from nonclinical adults, this it was not unexpected. These data are consistent with Karlsson and Kermott's study (2006), suggesting that mentalizing abilities (i.e. high RF scores) and good outcome are associated with patients who are committed to treatment. Indeed, our adult outpatient sample comprised participants who voluntarily sought psychotherapy treatment for personality pathology in a private outpatient clinic. In other words, our adult outpatients were likely to represent a sub-group with better social adjustment and higher mentalization skills compared to the wider population of adult outpatients with personality pathology. Moreover, the lack of differences in mentalistic abilities between adult clinical and nonclinical sample could be related to some confounding factors including education, IQ, and social economic status.

In our sample insufficient mental state reasoning resulting in incorrect, “reduced” mental state attribution (i.e., “less ToM” answers) was significantly associated with the presence of at least one PD diagnosis. Moreover, both dimensional and categorical BPD diagnoses were associated with a specific deficit in mentalizing abilities. More specifically, differently to Sharp et al. (2013) who found an association between BPD features and hypermentalizing in inpatient adolescents, our data showed a selective correlation between specific “no ToM” errors and BPD traits. We could hypothesize that the physiological relational instability, which characterizes adolescence, could represent a trigger to a hypermentalization deficit in adolescents with BPD features. On the contrary, BPD adults who are in a more stable life stage could manifest a more general mentalization deficit (i.e., “no ToM” answers). Other studies are needed to clarify how specific mentalistic deficit are related to BPD features in different life stages.

Consistent with Bateman and Fonagy’s (2004b) model of BPD and extending previous results (Preissler et al., 2010), dimensionally assessed BPD features in adult outpatients significantly correlated with failures to use mental state terms in explaining behavior (i.e. “no ToM” answers), using both self-reports and interview-based measures of *DSM-IV* BPD criteria. The association between the number of “no ToM” answers and BPD held also when categorical BPD diagnosis was taken into account. The relationship between mentalization deficit and dimensionally assessed BPD diagnosis was highly specific when BPD criteria were assessed using SCID-II interview, whereas it was less specific when BPD features were assessed using PDQ-4+ self-reports.

Confirming previous reports (e.g., Gratz & Roemer, 2004), in our adult outpatient sample emotion dysregulation, as operationalized by the DERS total score, was significantly and substantially associated with both self-reported and interview-based number of *DSM-IV* BPD criteria; the effect size for the association between BPD and DERS total score was large also when SCID-II BPD categorical diagnosis was taken into account. Thus, our data supports the hypothesis that difficulties in emotion regulation represent a core feature of BPD (e.g., Bateman & Fonagy, 2004b; Gratz & Roemer, 2004; Crowell, Beauchaine, & Linehan, 2009).

In our adult outpatient sample, difficulties in emotion regulation, as operationalized by the DERS total score, did not show an association with the overall number of errors on the MASC; rather, they showed a single, specific, significant correlation with the number of "no ToM" errors. Interestingly, "no ToM" answers were the only mentalization deficit that was significantly associated with BPD features. Partial correlation analyses showed that mentalization deficits only partially explained the association between observer-rated, dimensionally assessed BPD diagnosis and emotion dysregulation, at least as it was operationalized by the DERS total score; just as emotion dysregulation only partially explained the relationship between number of SCID-II BPD features and "no ToM" answers. Rather, when the number of *DSM-IV* BPD criteria was held constant in partial correlation analyses, the association between "no ToM" answers and emotion dysregulation became nonsignificant.

These findings suggest that the relationships between BPD, emotion dysregulation, and mentalization deficits that were observed were largely consistent with Bateman and Fonagy's (2004b) model of BPD development, although their mutual relationships are likely to be highly complex and not fully understood. For instance, our partial correlation analysis results suggest that much observed emotion dysregulation in BPD is unrelated to mentalization deficits, and a substantial amount of mentalizing impairment in BPD does not seem to be explained by emotion dysregulation. Rather, the relationship between mentalizing impairment and emotion dysregulation was explained by the effect of number of observer-rated BPD traits. As a whole, these findings suggest that both mentalization deficits and emotion dysregulation represent partially overlapping core features of BPD, which may have partially overlapping developmental trajectories.

Limitations

Our findings should be considered in the light of several limitations. Nonclinical sample sizes were fairly large and were not composed by random selection; the adult outpatient sample was of limited size and comprised reasonably adapted psychotherapy patients. Furthermore, we used community dwelling participants rather than healthy subjects explicitly selected for the absence of any mental disorder and characterized by above-average IQ. Our choice aimed at obtaining samples that could be more representative of the “average Italian”, although this aspect could limit the comparison between clinical and nonclinical sample. Moreover, the present study is cross-sectional in nature and, therefore, does not allow us to draw conclusions about causation. Although we used the RET as a convergent validity measure, we did not rely on the Reflective Function scale to evaluate the convergent validity of the MASC; even though the MASC allows for the assessment of more mentalizing polarities (Luyten et al., 2012) than the RET, neither the MASC nor the RET allowed for assessment of mentalization with regard to the self polarity (Luyten et al., 2012). Moreover, since we did not test its convergence with an implicit measure of mentalization we are not able to support the ability of the MASC to cover the automatic polarity of mentalization. We did not rely on multitrait-multimethod approach (Campbell and Fiske, 1959) but on nomological network to test the MASC construct validity; future studies could replicate our findings using a multitrait-multimethod approach. Finally, in our study, we did not consider the role of traumatic experiences and childhood adversity in the associations among BPD, emotion dysregulation, and attachment style. These variables are relevant and could have an important impact on our findings.

Keeping these limitations in mind, we nevertheless feel that our results represent a useful contribution to assessing mentalization, as well as to understanding the links between mentalization, and developmentally (e.g., attachment styles) and clinically (e.g., BPD, emotion dysregulation) relevant constructs in nonclinical adolescents, nonclinical adults, and adult outpatients.

Furthermore, as mentalization represents a relevant construct in the psychoanalytic perspective concerning both the psychopathological development and treatment, reliable and valid instruments are needed to assess this construct. According to this point of view, the MASC is an experimental task which seems to be a promising measure tapping into different mentalization features.

References

Allen, J. G. (2003). Mentalizing. *Bulletin of the Menninger Clinic, 67* (2: Special Issue), 91-112.

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: Author.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7*(4), 268-277.

Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. J. (eds.) (2000). *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience.* Oxford: Oxford University Press.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42*, 241-251.

Bartholomew, K. (1990). Avoidance of intimacy: An attachment perspective. *Journal of Social and Personal Relationships, 7*, 147–178.

Bateman, A. W., & Fonagy, P. (1999). The effectiveness of partial hospitalization in the treatment of borderline personality disorder: A randomized controlled trial. *American Journal of Psychiatry, 156*, 1563-1569.

Bateman, A. W., & Fonagy, P. (2001). Treatment of borderline personality disorder with psychoanalytically oriented partial hospitalization: An 18-month follow-up. *American Journal of Psychiatry, 158*, 36-42.

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

Bateman, A. W., & Fonagy, P. (2004b). *Psychotherapy for borderline personality disorder: Mentalization based treatment.* Oxford, UK: Oxford University Press.

Bateman, A. W., & Fonagy, P. (2008). 8-year follow-up of patients treated for borderline personality disorder: Mentalization-based treatment versus treatment as usual. *American Journal of Psychiatry, 165,* 631-638.

Bion, W. R. (1967). *Second Thoughts: Selected Papers on Psychoanalysis*. London, UK: Heinemann.

Bleiberg, E., Rossouw, T., & Fonagy, P. (2012). Adolescent breakdown and emerging borderline personality disorder. In A. W. Bateman & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice* (pp. 463-511). Washington, DC: American Psychiatric Publishing.

Bowlby, J. (1988). *A Secure Base: Clinical Applications of Attachment Theory.* London, UK: Routledge.

Buja, A. & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research,* 27, 509-540.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, *56*(2), 81.

Choi-Kain, L.W., & Gunderson, J.G. (2008). Mentalization: Ontogeny, assessment, and application in the treatment of borderline personality disorder. *American Journal of Psychiatry, 165*, 1127-1135.

Clarkin, J. F., Levy, K. N., Lenzenweger, M. F., Kernberg, O.F. (2007). Evaluating three treatments for borderline personality disorder: a multiwave study. *American Journal of Psychiatry, 164*, 922–928.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences* (2nd ed.). New York, NY: Academic Press.

Crowell, S. E., Beauchaine, T. P., & Linehan, M. M. (2009). A biosocial developmental model of borderline personality: Elaborating and extending Linehan's theory. *Psychological Bulletin, 135*(3), 495.

Denissen, J. J., Geenen, R., van Aken, M. A., Gosling, S. D., & Potter, J. (2008). Development and validation of a Dutch translation of the Big Five Inventory (BFI). *Journal of Personality Assessment*, *90*, 152-157.

Dennett, D. (1987). *The Intentional Stance.* Cambridge, Mass: MIT Press.

Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science, 13*(2), 331-338.

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., Kessler, J., Woike, J. K., Wolf, O. T., Convit, A. (2006). Introducing MASC: A movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders, 36,* 623-636.

Dziobek, I., Preissler, S., Grozdanovic, Z., Heuser, I., Heekeren, H.R., and Roepke, S. (2011). Neuronal correlates of altered empathy and social cognition in borderline personality disorder. *Neuroimage, 15*, 539–548.

First, M.B., & Spitzer, R.L., Gibbon, M., Williams, J.B.W., & Benjamin, L. (1994). Structured Clinical Interview for DSM-IV Axis II personality disorders (SCID-II).

Feeney, J. A., Noller, P., Hanrahan, M. (1994) Assessing adult attachment. In: Sperling MB, Berman WH, editors. *Attachment in adults: clinical and developmental perspectives*. New York, NY: Guilford Publications. p. 128–52.

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

Fonagy, P. (1997). Can we use observations of infant-caregiver interactions as the basis for a model of the representational world? Commentary on paper by Beebe, Lachmann, and Jaffe. *Psychoanalytic Dialogues: The International Journal of Relational Perspectives*, *7*(2), 207-213.

Fonagy, P., Bateman, A.W., & Luyten, P. (2012). Introduction and Overview. In A.W. Bateman & P. Fonagy (Eds.). *Handbook of Mentalizing in Mental Health Practice* (pp. 3–42). Arlington, VA: American Psychiatric Publishing.

Fonagy, P., & Luyten, P. (2009). A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Development and psychopathology*, *21*(04), 1355-1381.

Fonagy, P., Steele, M., Moran, G., Steele, H., & Higgitt, A. (1993). Measuring the ghost in the nursery: An empirical study of the relation between parents' mental representations of childhood experiences and their infants' security of attachment. *Journal of the American Psychoanalytic Association*, *41*(4), 957-989.

Fonagy, P., Steele, M., Steele, H., Moran, G. S., & Higgitt, A. C. (1991). The capacity for understanding mental states: The reflective self in parent and child and its significance for security of attachment. *Infant Mental Health Journal*, *12*(3), 201-218.

Fonagy, P., Target, M., Steele, H., & Steele, M. (1998). *Reflective-Functioning Manual, Version 5.0, For Application To Adult Attachment Interviews.* London, UK: University College London.

Fossati, A., Maffei, C., Bagnato, M., Donati, D., Donini, M., Fiorilli, M., ... & Ansoldi, M. (1998). Brief communication: Criterion validity of the Personality Diagnostic Questionnaire-4+ (PDQ-4+) in a mixed psychiatric sample. *Journal of Personality Disorders*, *12*(2), 172-178.

Fossati, A., Feeney, J. A., Donati, D., Donini, M., Novella, L., Bagnato, M., ... & Maffei, C. (2003). On the dimensionality of the Attachment Style Questionnaire in Italian clinical and nonclinical participants. *Journal of Social and Personal Relationships*, *20*(1), 55-79.

George, C., Kaplan, N., & Main, M. (1984/1985/1996). *The Berkeley Adult Attachment Interview.* Unpublished manuscript, Berkeley, CA.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment*, *26*(1), 41-54.

Ha, C., Sharp, C., Ensink, K., Fonagy, P., & Cirino, P. (2013). The measurement of reflective function in adolescents with and without borderline traits. *Journal of Adolescence, 36*(6), 1215-1223.

Hazan, C., & Shaver, P. R. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology, 52*, 511–524.

Hyler, S. E. (1994). *Personality questionnaire, PDQ-4+*. New York, NY (US): New York State Psychiatric Institute.

Holmes, J. (1996). *Attachment, intimacy, autonomy: Using attachment theory in adult psychotherapy*. Rowman & Littlefield.

Holmes, J. (1998). Defensive and creative uses of narrative in psychotherapy: An attachment perspective. Narrative *in psychotherapy and psychiatry*, 49-68.

Hsieh, F. Y., Lavori, P. W., Cohen, H. J., & Feussner, J. R. (2003). An overview of variance inflation factors for sample-size calculation. *Evaluation & the Health Professions*, *26*(3), 239-257.

Karlsson, R, & Kermott, A. (2006). Reflective-functioning during the process in brief psychotherapies. *Psychotherapy: Theory, Research, Practice, Training, 43*, 65-84.

Katznelson, H. (2014). Reflective functioning: A review. *Clinical psychology review, 34*(2), 107-117.

Lahera, G., Boada, L., Pousa, E., Mirapeix, I., Morón-Nozaleda, G., Marinas, L., ... & Parellada, M. (2014). Movie for the Assessment of Social Cognition (MASC): Spanish Validation. *Journal of Autism and Developmental Disorders*, 1-11.

Levinson, A., & Fonagy, P. (2004). Offending and Attachment: The Relationship between Interpersonal Awareness and Offending in a Prison Population with Psychiatric Disorder. *Canadian Journal of Psychoanalysis/Revue Canadienne de Psychanalyse,12*(2), 225-251.

Levy, K. N., Clarkin, J. F., Yeomans, F. E., Scott, L. N., Wasserman, R. H., & Kernberg, O. F. (2006). The mechanisms of change in the treatment of borderline personality disorder with transference focused psychotherapy. *Journal of Clinical Psychology*, *62*(4), 481-501.

Levy, R. A., & Ablon, S. (2009). *Handbook of evidence-based psychodynamic psychotherapy.* New York, NY: Springer.

Lorenzo-Seva, U., & Ferrando, P.J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavioral Research Methods, Instruments and Computers, 38*(1), 88-91.

Lorenzo-Seva, U., Timmerman, M.E., & Kiers, H.A.L. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research, 46*, 340-364.

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

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2004). Encoding-specific effects of social cognition on the neural correlates of subsequent memory. *The Journal of Neuroscience*, *24*(21), 4912-4917.

Montag, C., Dziobek, I., Richter, I. S., Neuhaus, K., Lehmann, A., Sylla, R., ... & Gallinat, J. (2011). Different aspects of theory of mind in paranoid schizophrenia: Evidence from a video-based assessment. *Psychiatry research*, *186*(2), 203-209.

Morton, J., & Frith, U. (1989). The origins of autism. *New Scientist, 1694*, 44-47.

Moskowitz, G. B. (2005). *Social cognition: Understanding self and others*. New York, NY: Guilford Press.

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

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw Hill.

Preissler, S., Dziobek, I., Ritter, K., Heekeren, H.R., & Roepke, S. (2010). Social cognition in borderline personality disorder: Evidence for disturbed recognition of emotions, thoughts, and intentions of others. *Frontiers in Behavioral Neuroscience, 4*, 1-8.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a 'theory of mind'? *Behavior and Brain Sciences, 4,* 515-526.

Ritter, K., Dziobek, I., Preissler, S., Rüter, A., Vater, A., Fydrich, T., ... & Roepke, S. (2011). Lack of empathy in patients with narcissistic personality disorder. *Psychiatry Research, 187*(1), 241-247.

Rochat, P., & Striano, T. (1999). Emerging self-exploration by 2-month-old infants. Developmental Science, 2(2), 206-218.

Research subjects: Participants' perspectives. *Journal of Medical Ethics*, *26*, 126-130.

Sharp, C., Fonagy, P., & Goodyer, I. M. (2008a). Introduction. In C. Sharp, P. Fonagy, & I. M. Goodyer (Eds.), *Social cognition and developmental psycho–pathology (pp. 1-28).* Oxford: Oxford University Press.

Sharp, C., Ha, C., Carbone, C., Kim, S., Perry, K., Williams, L., & Fonagy, P. (2013). Hypermentalizing in adolescent inpatients: Treatment effects and association with borderline traits. *Journal of personality disorders*, *27*(1), 3-18.

Sharp, C., Pane, H., Ha, C., Venta, A., Patel, A. B., Sturek, J., & Fonagy, P. (2011). Theory of mind and emotion regulation difficulties in adolescents with borderline traits. *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*(6), 563-573.

Slade, A., Grienenberger, J., Bernbach, E., Levy, D., & Locker, A. (2005). Maternal reflective functioning, attachment, and the transmission gap: A preliminary study. *Attachment & Human Development*, *7*(3), 283-298.

Steele, H., & Steele, M. (Eds.). (2008). *Clinical applications of the adult attachment interview*. Guilford Press.

Taubner, S., Hörz, S., Fischer-Kern, M., Doering, S., Buchheim, A., & Zimmermann, J. (2013). Internal structure of the Reflective Functioning Scale. *Psychological Assessment*, *25*(1), 127-135.

Taubner, S., Wiswede, D., Nolte, T., & Roth, G. (2010). Mentalisierung und externalisierende Verhaltensstörungen in der Adoleszenz. *Psychotherapeut*, *55*(4), 312-320.

Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D.R., Masala, C., Preti, A. (2012). The "Reading the Mind in the Eyes" test: Systematic review of psychometric properties and a validation study in Italy. *Cognitive Neuropsychiatry, 18*, 326-354.

Wendt, H. W. (1972). Dealing with a common problem in social science: A simplified rank-biserial coefficient of correlation based on the U statistic. *European Journal of Social Psychology*, *2,* 463–465.

Wilderjans, T.F., Ceulemans, E., & Meers, K. (2013). CHull: A generic convex-hull- based model selection method. *Behavior Research Methods, 45*, 1-15.

Winnicott, D. W. (1971). Transitional objects and transitional phenomena. In: *Playing and Reality* (ed. D. W. Winnicott), pp. 1–25. London, UK: Tavistock.

Wolkenstein, L., Schönenberg, M., Schirm, E., & Hautzinger, M. (2011). I can see what you feel, but I can't deal with it: Impaired theory of mind in depression. *Journal of Affective Disorders*, *132*(1), 104-111.

Zeki, S. (2007). The neurobiology of love. *FEBS Letters*, *581*(14), 2575–2579.

Zwick, W. R., & Velicer, W. F. (1986). Factors influencing five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432.

Table 1.
Descriptive Statistics and Gender Comparison of the Movie for the Assessment of Social Cognition Scores in Italian Nonclinical Adolescents (N = 373) and Nonclinical Adults (N = 193).

	Nonclinical Adolescents (*N* = 373)						Nonclinical Adults (*N* = 193)					
	Whole Sample (*N* = 373)		Female Subjects (*n* = 238)		Male Subjects (*n* = 135)		Whole Sample (*N* = 193)		Female Subjects (*n* = 115)		Male Subjects (*n* = 78)	
MASC Scores	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Correct answers	28.08	4.47	28.68	4.42	27.04	4.39	26.59	5.33	27.07	5.29	25.88	5.35
Error answers	16.98	4.44	16.38	4.42	18.00	4.31	18.39	5.31	17.96	5.24	19.04	5.39
"Exceeding ToM" answers	10.01	3.19	9.61	3.02	10.68	3.38	9.27	3.29	8.97	3.39	9.70	3.10
"Less ToM" answers	3.94	2.26	3.86	2.14	4.08	2.46	5.40	2.84	5.33	2.78	5.50	2.93
"No ToM" answers	3.02	2.27	2.90	2.20	3.24	2.38	3.76	2.67	3.65	2.53	3.91	2.88
Distractor answers	4.64	1.16	4.71	1.08	4.51	1.28	4.42	1.37	4.45	1.38	4.38	1.36

Note. MASC: Movie for the Assessment of Social Cognition
Percentile of the MASC correct score distribution in the nonclinical adolescent sample (*N* = 373): 5^{th} percentile = 20.00; 25^{th} percentile = 25.00; 50^{th} percentile = 29.00; 75^{th} percentile = 31.00; 95^{th} percentile = 35.00. Percentile of the MASC error score distribution in the nonclinical adolescent sample (*N* = 373): 5^{th} percentile = 10.00; 25^{th} percentile = 14.00; 50^{th} percentile = 16.50; 75^{th} percentile = 20.00; 95^{th} percentile = 25.00. "Exceeding ToM" answers score distribution: 5^{th} percentile = 5.00; 25^{th} percentile = 8.00; 50^{th} percentile = 10.00; 75^{th} percentile = 12.00; 95^{th} percentile = 16.00. "Less ToM" answers score distribution: 5^{th} percentile = 1.00; 25^{th} percentile = 2.00; 50^{th} percentile = 4.00; 75^{th} percentile = 5.00; 95^{th} percentile = 8.00. "No ToM" answers score distribution: 5^{th} percentile = 0.00; 25^{th} percentile = 1.00; 50^{th} percentile = 3.00; 75^{th} percentile = 4.00; 95^{th} percentile = 7.00.
Percentile of the MASC correct score distribution in the nonclinical adult sample (*N* = 193): 5^{th} percentile = 17.00; 25^{th} percentile = 23.25; 50^{th} percentile = 27.00; 75^{th} percentile = 30.00; 95^{th} percentile = 34.00. Percentile of the MASC error score distribution in the nonclinical adult sample (*N* = 193): 5^{th} percentile = 11.00; 25^{th} percentile = 15.00; 50^{th} percentile = 18.00; 75^{th} percentile = 21.00; 95^{th} percentile = 28.00. "Exceeding ToM" answers score distribution: 5^{th} percentile = 4.00; 25^{th} percentile = 7.00; 50^{th} percentile = 9.00; 75^{th} percentile = 12.00; 95^{th} percentile = 14.00. "Less ToM" answers score distribution: 5^{th} percentile = 1.00; 25^{th} percentile = 3.00; 50^{th} percentile = 5.00; 75^{th} percentile = 7.00; 95^{th} percentile = 11.00. "No ToM" answers score distribution: 5^{th} percentile = 0.65; 25^{th} percentile = 2.00; 50^{th} percentile = 3.00; 75^{th} percentile = 5.00; 95^{th} percentile = 10.00.

Table 2.
Attachment Style Questionnaire Scales: Descriptive Statistics, Internal Consistency Reliability, and Correlations with the Movie for the Assessment of Social Cognition Scores in the Italian Nonclinical Adult Sample (N = 193).

				Correlations (Pearson *r* Coefficients) with MASC Scales				
Attachment Style Questionnaire Scales	*M*	*SD*	*α*	Correct Answers	Error Answers	Exceeding ToM	Less ToM	No ToM
Confidence	31.54	4.24	.63	.16*	-.17*	-.16*	-.08	-.05
Discomfort with Closeness	33.46	5.82	.73	-.18*	.18*	-.03	.15*	.23**
Relationships as Secondary (to Achievement)	16.38	5.74	.79	-.08	.09	.15*	-.04	.03
Need for Approval	20.20	5.40	.70	-.11	.11	-.05	.17*	.10
Preoccupation with Relationships	27.69	6.42	.71	-.11	.11	-.05	.11	.16*

Note. MASC: Movie for the Assessment of Social Cognition
* p <.05
** p <.01

Table 3.
Spearman Correlations of Dimensionally Assessed DSM-IV Borderline Personality Disorder, as well as of the Other Dimensionally Assessed DSM-IV Personality Disorders, with Movie for the Assessment of Social Cognition Scale Scores in the Italian Adult Outpatient Sample (N = 59) based on Structured Clinical Interview for DSM-IV Axis II Personality Disorders, Version 2.0 (SCID-II) and Personality Diagnostic Questionnaire-4+ (PDQ-4+) Administration, Respectively.

			Correlations (Spearman *r* Coefficients) with MASC Scales				
SCID-II Personality Disorders	*M*	*SD*	Correct Answers	Error Answers	Exceeding ToM	Less ToM	No ToM
Borderline	1.88	2.59	-.21	.21	-.06	-.01	.41**
Paranoid	0.44	0.82	-.02	.01	.15	-.07	-.01
Schizoid	0.12	0.53	-.10	.10	.03	.15	.05
Schizotypal	0.15	0.71	-.11	.11	.11	.11	.04
Antisocial	0.08	0.65	.13	-.13	.05	-.14	-.08
Histrionic	0.95	1.50	.01	-.01	-.17	.02	.12
Narcissistic	2.00	2.00	.09	-.09	.21	-.21	-.13
Avoidant	0.36	0.87	.11	-.11	-.05	.05	-.18
Dependent	0.53	0.77	.14	-.14	-.27	.01	.13
Obsessive-Compulsive	0.37	0.79	-.13	.13	.16	.01	.03
Passive-Aggressive	1.00	1.26	.13	-.13	-.10	.03	-.11
Depressive	0.86	1.12	.10	-.10	-.21	.16	-.13
PDQ-4+ Personality Disorders							
Borderline	4.58	2.20	-.29*	.29*	.10	.05	.37**
Paranoid	3.19	1.89	-.30*	.30*	.11	.07	.35**
Schizoid	2.07	1.57	-.10	.10	.17	-.04	.13
Schizotypal	2.65	1.97	-.37**	.37**	.21	.17	.24
Antisocial	1.12	1.51	-.23	.23	.05	-.01	.42**
Histrionic	2.74	1.75	.02	-.02	.13	-.08	.07
Narcissistic	2.77	1.79	-.06	.06	.04	-.04	.23
Avoidant	3.54	1.97	-.11	.10	.10	.02	.17
Dependent	2.42	2.13	-.02	.02	.00	-.04	.30*
Obsessive-Compulsive	3.53	1.51	-.04	.04	.16	-.12	.08
Passive-Aggressive	3.25	1.79	-.18	.18	.10	-.05	.30*
Depressive	4.42	1.99	.09	-.09	.20	-.13	-.15
M			28.53	16.57	7.34	5.90	3.22
SD			4.48	4.50	2.83	3.41	2.03

Note. SCID-II: Structured Clinical Interview for DSM-IV Axis II Personality Disorders, Version 2.0; PDQ-4+: Personality Diagnostic Questionnaire-4+; MASC: Movie for the Assessment of Social Cognition.
* $p < .05$
** $p < .01$