



# The relations between self- and caregiver- focused reflective function and theory of mind in the context of borderline pathology in adolescence

Salome Vanwoerden<sup>a</sup>, Isabel Greiner<sup>a</sup>, Karin Ensink<sup>b</sup>, Carla Sharp<sup>a,\*</sup>

<sup>a</sup> Department of Psychology, University of Houston, Houston, TX, United States

<sup>b</sup> School of Psychology, Laval University, Quebec City, Quebec, Canada

## ARTICLE INFO

### Keywords:

Social cognition

Personality pathology

## ABSTRACT

Theory of mind (ToM) development is fostered by parent-child interactions characterized by accurate reflection on the child's mental states, or reflective function (RF), by the caregiver. Therefore, attachment-based RF is the foundation upon which children learn to reason about minds outside the attachment context (domain-general ToM). However, it is not known to what extent attachment-based RF of the self versus caregivers uniquely relates to domain-general ToM. Additionally, it is likely that for psychopathology associated with maladaptive parent-child dynamics (i.e., borderline pathology), domain-general ToM impairments are more strongly related to attachment-based RF disturbances. Therefore, the aim of the current study was to evaluate associations between domain-general ToM and attachment-based RF to determine whether RF of the self versus caregivers has unique relations to domain-general ToM. Second, we tested whether borderline pathology would moderate this relation. Among a sample of inpatient adolescents ( $N = 330$  adolescents;  $Mage = 15.40$ ,  $SD = 1.44$ ), findings suggest that RF of the self uniquely relates to domain-general ToM and that this relation is strongest among adolescents with high levels of borderline pathology. Therefore, evidence supports theory regarding the association between attachment-based RF and domain-general ToM. Additionally, interpersonal disturbance observed in borderline pathology, even in adolescence, is related to attachment-based social-cognition.

## 1. Introduction

Theory of mind (ToM) is a social-cognitive construct defined as our ability to attribute mental states to the self and others (Premack and Woodruff, 1978). ToM abilities have been shown to be important for building and maintaining interpersonal relationships (Bosacki, 2015; Caputi et al., 2012; Fink et al., 2015) and therefore has important implications for psychopathology characterized by interpersonal dysfunction. This is demonstrated by research showing ToM deficits among psychiatric disorders with a strong interpersonal component (Sharp and Venta, 2013). ToM is thought to develop in an attachment context where the child, with the help of the caregiver, learns to reflect on their own mind and thereby on others' minds. Specifically, as elaborated by simulation theory, caregivers use marked mirroring and the use of mental state terms to describe children's behavior and affective displays, which serves to enhance children's awareness and explicit knowledge of their own internal experiences, which later generalizes to internal experiences of others (Harris, 1992; Ruffman, 2014). This suggests that reflection upon minds in an attachment context likely precedes and is related to an individual's domain-general ToM ability.

However, traditional ToM tasks typically focus on inferring the mental states of non-self-referential others, usually based on vignettes, videos, images, or enactments of fictional characters and are therefore unable to tap into the attachment-based nature of social cognition (Zaki and Ochsner, 2009).

From the attachment tradition, a construct has been defined to operationalize the developmental process described above more closely: Reflective Function (RF). Similar to ToM, RF describes the complex process of interpreting others' and one's own behaviors by considering underlying mental states (Tessier et al., 2016). A key distinguishing factor is that RF is conceptualized and measured within attachment contexts. RF can be measured either in questionnaire form (e.g., the Reflective Functioning Questionnaire for youth and adults; Fonagy et al., 2016; Ha et al., 2013) or by coding the quality of attachment narratives (e.g., Adult and Child Reflective Function Scales; Ensink et al., 2015; Fonagy et al., 1998). While questionnaires focus on domain-general RF (i.e., across all interpersonal interactions and relationships), coding systems are derived from narratives regarding the parent-child attachment context and therefore are domain-specific (Ensink et al., 2016) as they access relationship-specific RF. Though the

\* Corresponding author at: 3695 Cullen Boulevard, Room 126, Houston, TX 77204-5022, United States.

E-mail address: [csharp2@central.uh.edu](mailto:csharp2@central.uh.edu) (C. Sharp).

<https://doi.org/10.1016/j.psychres.2019.01.042>

Received 28 June 2018; Received in revised form 15 November 2018; Accepted 11 January 2019

Available online 11 January 2019

0165-1781/ © 2019 Elsevier B.V. All rights reserved.

constructs of RF and ToM have significant overlap, RF has seldom been studied in relation to more traditional ToM methods. Therefore, more research is needed to understand how general ToM capacity specifically relates to ToM in the attachment context (RF).

Based on the conceptual overlap, it would be expected that there is a strong relation between RF and domain-general ToM capacity; however only two studies have evaluated the relations between attachment-based RF and domain-general ToM capacity among children (De Rosnay and Harris, 2002; Repacholi and Trapolini, 2004). These studies converged on the finding that attachment-based RF was related to domain-general ToM capacity; however, performance on each of those two tasks did not completely overlap, suggesting that these abilities are distinguishable. Specifically, RF about the self in an attachment separation task accounted for between 6–9% of unique variance in domain-general ToM performance when accounting for the effects of age and language (Milligan et al., 2007). However, both of these studies utilized relatively small sample sizes (around  $N = 50$ ) of young children (combined age range of 3–6 years old). It is likely that interrelations between these constructs differ based on age. Specifically, during early childhood, the family is the primary context in which social interactions take place, whereas during adolescence, the social context expands to include peers and romantic relationships (Steinberg et al., 2006) suggesting that there will be less overlap between domain-general ToM capacity and attachment-based RF by adolescence. Additionally, neural regions underpinning social cognition undergo dramatic growth during this age range, making this a unique developmental phase in which to study social cognition (Blakemore and Mills, 2014).

Additionally, prior work has not distinguished between RF of the self and caregivers when relating RF to domain-general ToM capacity. Although RF-self and RF-caregivers are related constructs, they are also distinct from each other (Ensink et al., 2014). Furthermore, fMRI studies have shown differential activation in neural regions when attributing mental states to the self versus others (Decety and Sommerville, 2003; Ochsner et al., 2004). It has been argued that self-understanding is a more developmentally complex achievement (Bogdan, 2000) that develops through understandings of others (Fonagy and Target, 2006). Specifically, Fonagy and Gergely (2007) describe how a caregiver facilitates the development of ToM through marked mirroring. Therefore, it is via attunement to their caregiver's display (other focus) that the infant learns to label and understand their own internal states (self focus and understanding). Eventually, this understanding becomes internalized and is the basis by which a child reflects on and regulates their own mental states. Particularly during adolescence, self-functioning undergoes dramatic changes as there is a sharp increase of self-consciousness and a stronger emphasis on the appraisals of others, which expands the adolescent's own self-reflection (Harter, 2012). Therefore, we expect that during adolescence, RF of the self would relate more strongly to domain-general ToM capacity compared to RF of caregivers.

The question about relations between RF about self and caregivers with domain-general ToM capacity has conceptual as well as clinical implications. Recent research has demonstrated that personality pathology can be characterized by a core disruption in self and other processing (Hopwood et al., 2013). A paradigmatic disorder for self-other impairment is borderline personality disorder (BPD), which is characterized by severe deficits in identity development, interpersonal relationships, emotion regulation, and self-damaging behaviors (American Psychiatric Association, 2013). Research and theory has characterized self and other impairment in borderline personality pathology as consisting of dramatic shifts between extreme perspectives of the self and others that are not integrated (Bender and Skodol, 2007). While experimental work investigating deficits in other-functioning has been inconsistent, the pattern of results suggests a tendency for those with borderline personality pathology to hold negative views of others and to hypermentalize—to overattribute mental states to others (Sharp, 2014; Sharp and Vanwoerden, 2015). Because developmental models of

BPD identify maladaptive early attachment relationships as the source of these disturbances, it is likely that there is a close association between attachment-based RF and domain-general ToM capacity among individuals with high levels of borderline personality pathology. In fact, the few studies examining associations between social-cognitive function within an attachment context compared to non-attachment contexts has found that avoidant or dismissing attachment styles related to greater discordance between social-cognitive performance across relationship domains (Bączkowski and Cierpiąłowska, 2015; Humfress et al., 2002). These authors concluded that children with avoidant attachment utilize a deactivating strategy when faced with attachment-related stress. This may lead them to underperform when asked to mentalize attachment figures, but would not affect mentalizing other individuals. However, BPD has been associated with particularly with disorganized and preoccupied attachment styles (Bakermans-Kranenburg and IJzendoorn, 2009; Miljkovitch et al., 2018), which may relate to alternative self-regulatory strategies when the attachment system is activated. In fact, research has shown that individuals with BPD have a hyperreactive attachment system, which is likely related to social-cognitive impairments in this disorder (Fonagy et al., 2011; Nolte et al., 2013). We expect, therefore, that even when interacting with non-attachment figures, individuals with BPD features may demonstrate a lower threshold for activation of the attachment system and therefore may have difficulties with accurate perception and interpretation of others' mental states. In sum, testing the degree of overlap between domain-general ToM and attachment-based RF in the context of borderline personality pathology has the potential for elucidating the self-other impairment characteristic of this disorder.

To this end, the current study is the first to clarify the association between domain-general ToM ability and attachment-based RF among adolescents while evaluating whether there are unique effects of RF about self versus caregivers with domain-general ToM. While studying this in community samples is important, the current study uses a clinical sample, given that psychiatric samples represent a larger proportion of impaired social-cognitive functioning than the general population (Sharp, 2006). While initial research has demonstrated moderate relations between RF and ToM, more research is necessary to determine the differential relations between RF about self versus caregivers. Second, we examined the moderating role of borderline pathology in this relation to test the hypothesis that relations between attachment-based RF and domain-general ToM are stronger among individuals with higher levels of borderline pathology. These aims were carried out in a sample of adolescents given that previous literature has been limited to children, while adolescence represents a unique developmental phase in the interpersonal domain (Doyle and Cicchetti, 2017). The current study has implications for further elucidating attachment theories that root social cognition in reflective capacities within attachment relationships.

## 2. Methods

### 2.1. Subjects

Adolescents were recruited from a 16-bed inpatient psychiatric unit that usually serves individuals with severe behavioral and emotional disorders who have not responded to previous interventions. Length of stay ranged from 5–85 days ( $M = 36.77$ ;  $SD = 13.46$ ), during which adolescents participated in milieu-based treatment emphasizing improvement of social-cognitive capacity, emotion regulation, maladaptive behaviors, and family relations with a primary theoretical framework of interpersonal-psychodynamic. The inclusion criterion was sufficient proficiency in English to consent to research and complete the necessary assessments, and exclusion criteria were a diagnosis of schizophrenia or another psychotic disorder, an autism spectrum diagnosis, or an IQ of less than 70. Of  $N = 675$  adolescents and their parents who were approached for consent,  $n = 54$  declined and  $n = 78$  were

excluded based on aforementioned criteria. Additionally,  $n = 213$  participants were excluded due to missing one of the two main study variables ( $n = 165$  missing the CRFS due to incomplete coding and  $n = 61$  missing the MASC as a result of adolescent refusal or incomplete assessments due to abrupt discharges). Adolescents with and without missing data differed in age, with younger adolescents missing data on main study variables. Adolescents with and without missing data did not differ in gender, level of BPD features, or social-cognitive variables.

The final sample consisted of  $N = 330$  adolescents (64% female; ages 12–17,  $M = 15.40$ ,  $SD = 1.44$ ), with the following racial/ethnic breakdown: 76.7% White, 4.3% Hispanic, 3.9% Asian, 4.5% mixed or other, and 8.5% unspecified. Based on DSM-IV criteria, 44.9% were diagnosed at admission with a depressive disorder, 7.6% with a bipolar disorder, 8.8% with any eating disorder, 42.5% with any externalizing disorder, and 59.2% with any anxiety disorder. The study was approved by a human subjects review committee, and subjects participated after signing a written voluntary informed consent form. Adolescents were assessed by doctoral-level clinical psychology students and/or trained clinical research assistants. The assessments were conducted independently and in private within the first two weeks following admission.

## 2.2. Measures

### 2.2.1. Domain-general theory of mind capacity

The Movie for the Assessment of Social Cognition (MASC; Dziobek et al., 2006) is a computerized test for the assessment of mentalizing that approximates the demands of everyday life. Subjects were asked to watch a 15-minute film about four characters getting together for a dinner party. Themes of each segment covered friendship and dating issues. During administration of the task, the film is stopped at 45 points and multiple-choice questions referring to the characters' mental states (feelings, thoughts, and intentions) are asked (e.g., "What is Betty feeling?", "What is Cliff thinking?"). All items answered correctly are summed for a total score with higher scores indicating higher ToM capacity. The MASC is a reliable instrument that has proven sensitive in detecting subtle mindreading difficulties in adults with a normal IQ (Dziobek et al., 2006) and in adolescents (Sharp et al., 2011).

### 2.2.2. Attachment-based reflective function

The Child Reflective Function Scale (CRFS; Ensink et al., 2015, 2013) is an interview-based measure of reflective function coded from transcriptions from the Child Attachment Interview (CAI; Target et al., 1998). RF ratings were coded on an 11-point dimensional scale, ranging from  $-1$  to  $9$ , and anchored at six points in terms of ability to reflect on self and others in mental state terms such that a score between  $5$  and  $6$  reflects an overall average level of RF, with scores of  $7$  or higher indicating high RF, and scores of  $4$  or lower indicating low to impaired RF. The RF-Self scale was computed by averaging scores from three items on the CAI that elicit self-descriptions in addition to an item eliciting description of a time the subject felt hurt or upset. In the first three self-description questions, adolescents were asked to provide three words to describe themselves, and then prompted to provide examples. For instance, an adolescent may have described him/herself using the word "intelligent." An example of an average reflective function response would be "My teacher says I'm intelligent because I made an A on the math exam," which would be coded with a score of  $6$ . An example of a high reflective function score ( $9$ ) would be a response of "I feel intelligent when my big brother cannot complete a math problem and I help him figure it out. That makes me feel intelligent." An example of a response that would be coded with a ( $1$ ) would be one that is clearly attacking the interviewer, such as "I did not say that I'm intelligent. Why are you asking so many questions? Is this interview over yet?" The RF of caregivers consisted of an average of three items tapping into the child's relationship with their mother and three items regarding their relationship with their father, which were scored in a

similar manner to the self-scales. In addition, to two items asked about what happens when the mother and father get angry, respectively. Adolescent reflective function was coded by a team of trained coders who were directly trained on the coding system by the developer of the CRFS (the third author). The CRFS has been shown to be reliable and valid among healthy and clinic samples of children and adolescents, with ICCs ranging from  $0.60$  to  $1.0$  (median of  $0.93$ ) and high temporal stability for children over a three-month period (Ensink et al., 2015, 2013; Ha et al., 2013).

### 2.2.3. Borderline personality features

The Borderline Personality Features Scale for Children (BPFS-C; Crick et al., 2005) is a self-report questionnaire used to measure levels of borderline pathology among children ages  $9$ – $18$ . The assessment's  $24$  items were adapted from the Personality Assessment Inventory's BPD scale to be age-appropriate, and they reflect functioning in  $4$  key domains: negative relationships, self-harm, identity problems, and affect instability. Sample questions include: "I get into trouble because I do things without thinking" and "I feel there is something important missing about me, but I don't know what it is". Respondents respond on a  $5$ -point Likert scale ranging from  $1$  (*not at all true*) to  $5$  (*always true*), and scores for each of the  $24$  items yield a total score. A higher total score indicates greater levels of borderline pathology. Research demonstrates that the BPFS-C demonstrates good criterion and concurrent validity (Chang et al., 2011) and moderate stability across a  $12$ -month period (Crick et al., 2005).

### 2.2.4. Verbal comprehension

The Wechsler Intelligence Scale for Children-IV (WISC-IV; Wechsler, 2003) and Wechsler Adult Intelligence Scale-IV (WAIS-IV; Wechsler, 2008) are general tests of intelligence. For the purpose of this study, we used the verbal comprehension index, which is a measure of verbal comprehension. The Wechsler intelligence scales are widely used assessment tools that show excellent psychometric properties (Wechsler, 2008, 2003).

## 2.3. Data analytic strategy

Analyses were conducted using SPSS 24 (IBM Corp, 2016). Distributions for study variables approximated normality (skewness  $< |0.82|$ ) and kurtosis  $< |1.41|$ ), which is displayed in Table 1 (George and Mallery, 2003). First, bivariate relations were assessed using Pearson correlation coefficients between main study variables. To further unpack the relations between variables of interest while controlling for appropriate covariates, multiple linear regressions were run with domain-general ToM (as measured by the MASC) as the dependent variable, predicted by the independent variables of RF of the self and caregivers. Finally, the moderating role of borderline personality features was tested in the relation between ToM and RF using hierarchical regression analysis with domain-general ToM as the dependent variable and attachment-based RF and BPD as independent variables. Additionally, the interaction term between RF and BPD was also entered as an independent variable to test the moderation.

## 3. Results

### 3.1. Preliminary analyses

Bivariate correlations were run between main study variables. Results are displayed in Table 1, which demonstrate that age, gender, and verbal comprehension were positively related to RF of the self and caregivers such that older adolescents, females, and those with higher verbal comprehension demonstrated greater RF. Age and verbal comprehension were related to ToM as measured by the MASC; therefore, these variables were entered as covariates for subsequent analyses with ToM as the dependent variable. Additionally, as demonstrated in other

**Table 1**  
Descriptive statistics for and bivariate correlations among main study variables.

| Variable                | 1      | 2            | 3              | 4           | 5           | 6            | 7             |
|-------------------------|--------|--------------|----------------|-------------|-------------|--------------|---------------|
| 1. Gender               | 1      |              |                |             |             |              |               |
| 2. Age                  | 0.13*  | 1            |                |             |             |              |               |
| 3. Verbal comprehension | 0.20*  | 0.09         | 1              |             |             |              |               |
| 4. CRFS-self            | −0.19* | 0.14*        | 0.23*          | 1           |             |              |               |
| 5. CRFS-att. figures    | −0.22* | 0.22*        | 0.23*          | 0.76*       | 1           |              |               |
| 6. MASC-ToM             | −0.09  | 0.29*        | 0.35*          | 0.18*       | 0.13*       | 1            |               |
| 7. BPFS-C               | −0.21* | −0.11        | −0.12          | −0.04       | 0.00        | −0.14*       | 1             |
| Mean(SD)                |        | 15.40 (1.44) | 111.50 (15.48) | 2.87 (1.09) | 3.08 (1.12) | 32.31 (4.73) | 70.42 (16.12) |
| Range                   |        | 12–17        | 70–150         | −1–6.50     | −1–7.11     | 10–43        | 32–112        |
| Skew                    |        |              | −0.15          | −0.05       | −0.18       | −0.82        | −0.11         |
| Kurtosis                |        |              | −0.06          | 0.47        | 0.51        | 1.41         | −0.36         |

Note: \* $p \leq 0.05$ ; CRFS = Child Reflective Function Scale; MASC = Movie for the Assessment of Social Cognition; BPFS-C = Borderline Personality Features Scale for Children; Verbal Comprehension measured with Wechsler Intelligence Scale for Children;  $N = 330$  for all variables except Verbal Comprehension with  $n = 149$ ).

clinical samples (Sansone and Sansone, 2011), females displayed higher borderline personality features. Only a subset of the sample completed IQ testing as this measure was only administered at the hospital in the case of educational concerns. Therefore, only  $n = 149$  adolescents had data for this variable. Independent samples  $t$ -tests were run to determine whether adolescents with data for verbal comprehension differed from the rest of the sample on demographic variables (i.e., gender and age) and the main study variables (i.e., CRFS, MASC, BPFS-C). Results demonstrated that adolescents with and without data for verbal comprehension only differed in regard to RF about self ( $t(328) = 3.39$ ,  $p < 0.01$ ). Adolescents with missing data demonstrated overall higher scores on RF about self. Additionally, we conducted all bivariate analyses in only the subsample of 149 adolescents, given the relatively higher correlations between verbal comprehension and measures of social cognition. Differences in magnitude of these correlations were compared statistically, and we found that none of the correlations were significantly different when computed in the subsample. The biggest differences (in terms of  $Z$ -scores) were found for the correlation between RF about self and ToM ( $r_{\text{subsample}} = 0.29$  compared to  $r_{\text{full sample}} = 0.18$ ) and for the correlation between ToM and gender ( $r_{\text{subsample}} = 0.05$  compared to  $r_{\text{full sample}} = -0.09$ ).

### 3.2. Evaluating relations between reflective function and ToM

Correlational analyses revealed that ToM was positively related to both RF about self ( $r = 0.18$ ) and RF about caregivers ( $r = 0.13$ ), although the latter was related to a lesser extent. We tested the magnitude of difference between these correlations using a Fisher  $r$  to  $Z$  transformation and found this difference was not significant ( $Z = 1.48$ ,  $p = 0.07$ ). Finally, we observed a strong positive relation ( $r = 0.76$ ) between the two domains of RF, suggesting some degree of overlap (Table 1). Due to this overlap, we evaluated the relation between variables using a linear regression with ToM as the dependent variable and the two domains of RF as independent variables. Additionally, VIF was estimated to provide a measure of multicollinearity given the strong bivariate correlation between these two domains. Because only a subset of the sample had completed IQ testing, we conducted two separate linear regressions. For the first regression, we used the full sample and controlled for the effects of age on ToM. In a second regression, we utilized a subset of the sample who had completed IQ testing and controlled for both age and verbal IQ. Table 2 displays the results for the full sample, which demonstrate that while all predictors were significant and positively related to ToM, effect size was smaller for RF of caregivers compared to RF of self. Among the subset of the sample displayed in Table 3, we further controlled for verbal IQ and found that RF about caregivers was no longer a significant predictor of ToM. In the full sample, which does not account for the effects of verbal IQ, RF about self demonstrated a small effect size in the prediction of ToM ( $\eta^2 = 0.024$ ) whereas RF of the self demonstrated a moderate

**Table 2**  
Linear regressions predicting ToM from reflective function ( $N = 330$ ).

| Variable                | B     | SE B | $\beta$ | $t$    | $\eta^2$ | VIF  |
|-------------------------|-------|------|---------|--------|----------|------|
| Age                     | 0.92  | 0.18 | 0.28    | 5.20*  | 0.077    | 1.05 |
| CRFS-Self               | 0.98  | 0.35 | 0.23    | 2.81*  | 0.024    | 2.36 |
| CRFS-Att. Figures       | −0.44 | 0.34 | −0.11   | −1.29  | 0.005    | 2.44 |
| $R^2$                   |       |      |         | 0.108  |          |      |
| $F$ for change in $R^2$ |       |      |         | 13.18* |          |      |

Note: \* $p \leq 0.001$ .

**Table 3**  
Linear regressions predicting ToM from reflective function ( $n = 149$ ).

| Variable                | B     | SE B | $\beta$ | $t$    | $\eta^2$ | VIF  |
|-------------------------|-------|------|---------|--------|----------|------|
| Age                     | 0.79  | 0.26 | 0.23    | 3.06*  | 0.061    | 1.04 |
| Verbal Comprehension    | 0.10  | 0.02 | 0.31    | 4.05*  | 0.102    | 1.07 |
| CRFS-Self               | 1.63  | 0.50 | 0.34    | 3.23*  | 0.067    | 2.09 |
| CRFS-Att. Figures       | −0.98 | 0.48 | −0.22   | −2.05  | 0.028    | 2.10 |
| $R^2$                   |       |      |         | 0.237  |          |      |
| $F$ for change in $R^2$ |       |      |         | 11.19* |          |      |

Note: \* $p \leq 0.05$ .

effect in the subsample while controlling for verbal IQ ( $\eta^2 = 0.067$ ). In both analyses, VIF estimates suggested that multicollinearity did not bias estimates; rather, given the VIF estimate of 2.44 for the effect of RF of caregivers within the full sample, standard errors of this estimate were only 1.6 times larger than if this variable were uncorrelated with all other predictors (Dormann et al., 2013). Given the satisfactory multicollinearity estimate, we conducted a Wald test to determine whether the unique effects of the two domains of RF were significantly different from one another in the prediction of ToM. Results suggest that the difference of unstandardized effects were indeed significant ( $B_{\text{difference}} = 1.42$ ,  $SE = 0.65$ ,  $p = 0.04$ ). Therefore, after accounting for the overlap between the two domains of RF, RF about self had a significantly greater association with ToM than did RF about caregivers.

### 3.3. Evaluating the moderating effect of borderline personality features on the relation between attachment-based RF and domain-general ToM

Finally, we evaluated the moderating effect of BPD features on the relation between RF of the self and ToM given the more central role that RF of the played in the prediction of ToM, as compared to RF of caregivers. In a hierarchical regression analysis predicting ToM, covariates of age, verbal IQ, and RF of caregivers were entered in the first step. RF of the self and BPD features were entered in the second step, mean-centered, to estimate main effects. In the third and final step, the interaction term of RF of the self and BPD features was entered. Post hoc simple slope analyses were conducted for a significant interaction using

**Table 4**  
Main and interactive effects of RF of the self and borderline personality features on ToM.

|                      | $\beta$ | $t$   | $p$    |
|----------------------|---------|-------|--------|
| <b>Step 1</b>        |         |       |        |
| Age                  | 0.21    | 2.94  | 0.004  |
| Verbal comprehension | 0.29    | 3.96  | <0.001 |
| CRFS-Att. Figures    | −0.18   | −1.67 | 0.094  |
| <b>Step 2</b>        |         |       |        |
| BPFS-C               | −0.11   | −1.40 | 0.163  |
| CRFS-self            | 0.32    | 3.11  | 0.002  |
| <b>Step 3</b>        |         |       |        |
| BPFS-C*CRFS-self     | 0.16    | 2.08  | 0.039  |

values of  $+/-1$  SD (high/low) from the mean of the moderator variable (i.e., BPD features). Results of the moderation analysis are displayed in Table 4, and simple slopes are plotted in Fig. 1. Results demonstrated that BPD features did exert a moderating effect such that only at high levels of BPD features was there a significant relation between RF of the self and ToM.

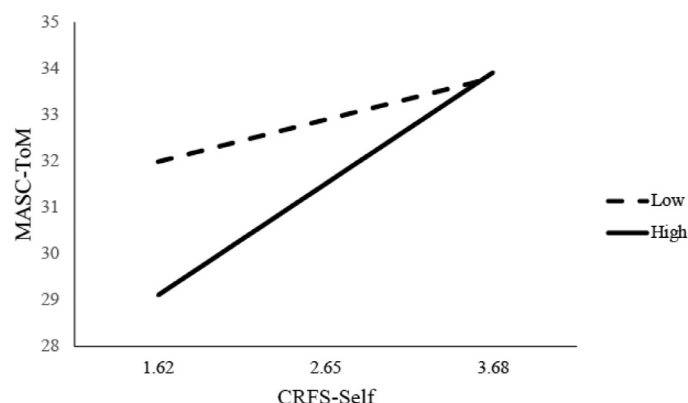
#### 4. Discussion

Our ability to reflect on our minds and others' in the context of relationships is a central component of healthy relationship functioning and therefore has significant implications for psychopathology. Research in this area has been impeded by a lack of clarity and conceptual overlap between related constructs (Happé et al., 2017). This is problematic in the field of social cognition, which sometimes differentiates and sometimes confounds the constructs of RF and ToM. Therefore, the aim of this study was to evaluate the interrelations between these constructs, while differentiating between RF of the self and caregivers, respectively. Second, we examined the role of borderline pathology in this relation. Findings indicated that RF about self, but not RF about caregivers, was related to general ToM capacity and this relation only held at high levels of borderline features. Theories of social cognition rooted in attachment relationships have suggested that the understanding of mental states of the self and caregivers builds the basis from which children learn to generalize and attribute mental states to generalized others (domain-general ToM). The current study extends these theories by demonstrating that RF about self is uniquely related to general ToM abilities as assessed outside of the attachment context, with specific implications for BPD.

Our findings demonstrated that, as expected and in line with previous research, successful reflection on the minds of the self and caregivers in an attachment interview was positively related to correctly attributing mental states to characters from a movie that had no personal connection to subjects. These findings fit with theories of attachment which posit that mental state understanding between infants

and caregivers form the basis of social-cognitive development (Fonagy and Target, 1997). This study's findings support the possibility that social-cognitive functions in the attachment context generalize to others with no personal relationship to the subject. This highlights the importance of secure attachment relationships in childhood, which foster mental state understanding within the attachment relationship, which have implications beyond childhood and into stages of life marked by greater independence from caregivers (e.g., adolescence).

At a bivariate level, the constructs of RF and ToM were only related to a small degree, which suggests that there was some overlap between these constructs. In fact, de Rosnay and Harris (2002) found that the discrepancy between mental state understanding between mothers and a neutral adult female was accounted for by indices of insecure attachment. Specifically, they found that among children with an avoidant attachment style, false belief understanding of mothers was relatively impoverished based on what would be expected from mental state understanding of a neutral adult female. Conversely, among securely attached children, there were no differences between false belief understanding of mothers compared to the neutral adult female. This finding provides some context for understanding the small association between attachment-based RF and domain-general ToM in our sample. Our sample demonstrated a smaller proportion of securely attached individuals than found in the general population (approximately 75% of the current sample were rated as having an insecure attachment on the CAI with both mother and father). However, despite the fact that attachment security may account for the discrepancy between attachment-based RF and ToM, it is still the case that insecurely attached individuals can achieve adequate mental state understanding with others outside of the attachment context. It may be that closer associations between these processes are particularly apparent in younger children as their social environment largely consists of parent-child interactions. As children reach adolescence, their social network grows and includes more interpersonal experiences with individuals outside their family network, which may act to further enrich social-cognitive abilities (Scharf and Mayseless, 2007). Therefore, we believe that part of the reason we found such low correlations between ToM and the domains of RF is due to the age of our sample, compared to previous studies conducted in young children. Studies have shown statistically significant increases in ToM development between the ages of 7–12 (Peterson and Wellman, 2018). This age range corresponds to a developmental phase in which children's social networks begin to expand through participation in extra-curricular activities and greater school engagement (Gifford-Smith and Brownell, 2003). Therefore, it is possible that with increased diversity of interactions, domain-specificity of social cognition is amplified such that there are weaker associations between social-cognitive ability as displayed within different contexts (i.e., attachment versus non-attachment relationships). Finally, there is also the fact that methodological differences partially account for the low associations between ToM and domains of RF. Previous studies



**Fig. 1.** Plots of simple slopes for interactive effects of borderline personality features and RF of the self.

utilized adapted versions of the same task to represent attachment versus non-attachment figures when evaluating domain-specificity of ToM. Therefore, while our tasks both tap into similar domains of social-cognition, task-based demands were quite different.

During adolescence, increased cognitive abilities contribute to ToM development such that understanding of others can become more complex and differentiated (Blakemore, 2012). It is notable, in fact that the relation between verbal comprehension and ToM—as well as RF across both domains—was higher than the relation between ToM and RF. However, the magnitude of this relation is in line with previous studies finding at least moderate associations between verbal ability and social cognition (Milligan et al., 2007). This can partly be ascribed to the fact that both of our social-cognitive measures are verbal tasks that rely on both expressive and receptive language skills. Furthermore, there is a large body of research demonstrating that language plays a causal role in the development of ToM (e.g., Astington and Baird, 2005) due to the fact that language is typically the tool used to represent and communicate the mental states of others. Additionally, more complex language abilities such as complementation syntax has been argued to provide a necessary format to represent false beliefs (e.g., complex sentences comprised of a tensed subordinate clause embedded under a mental verb: Sally thinks that the marble is in the basket), which is an early developing component of ToM (de Villiers and Pyers, 2002).

Also in adolescence, there is an expansion of social networks corresponding to increases in social awareness that contributes to ToM development as well as self-reflection (Harter, 2012). Therefore, we expected that RF regarding self would demonstrate greater relations with general ToM capacity in our sample. In line with this hypothesis, we found that when accounting for the effects of verbal IQ and age, the association between RF regarding self was uniquely related to ToM above and beyond RF about relationships with caregivers. In fact, RF about caregivers was no longer a significant predictor of domain-general ToM capacity. Therefore, it is possible that by adolescence, when the self-concept is more differentiated, RF about caregivers is less relevant to social understanding of others outside of the attachment context. However, this conclusion should be tempered by the fact that the two domains of RF evidenced a strong statistical overlap with one another (more than 50%). While an estimate of multicollinearity suggested that this overlap should not have strongly affected resulting parameters, findings must be replicated using alternative measures. In fact, high overlap is likely due to the use of the Child Attachment Interview in the current study. While the two domains of RF were obtained from separate prompts on the interview, in practice, discussion within these prompts sometimes goes beyond the original subject of the prompt. This is particularly true of the prompts that explore relationships with caregivers because adolescents are asked to consider their own experience within the examples provided.

Finally, we found that when examining the role of borderline pathology in the relation between RF about self and domain-general ToM capacity, it was only among adolescents with high levels of borderline features that there was a positive and statistically significant association between RF about self and ToM. This was not a surprising finding as self-other disturbance in BPD has been suggested to be a result of maladaptive parent-child dynamics during childhood (Fonagy and Bateman, 2008). Further, disruptions in ToM are more likely to be found in the context of close, attachment relationships among individuals with BPD (Sharp and Vanwoerden, 2015). As such, it is likely that maladaptive parent-child dynamics characterized by impaired RF is carried over into future interactions that a child may have. Specifically, among interactions with generalized others, individuals with BPD have a lower threshold for activation of their attachment system, which leads to emotion dysregulation and activation of representations of the self and other. Therefore, particularly among those with BPD, ToM impairments will be highly related to attachment-based disruptions in RF.

These findings must be interpreted with some caution in context of

the following limitations—although the theory is based on a developmental argument, our study design was cross-sectional and therefore directional interpretations cannot be made. Additionally, although examining social-cognitive functioning in an inpatient sample was beneficial, as these individuals have stronger impairments in this domain, it is unclear whether these findings would generalize to healthy functioning adolescents. In fact, given that the sample likely had a larger proportion of insecurely attached individuals than what would be found in the general population, this could account for the stronger relation between RF and ToM. Lastly, given that the method for assessing RF was likely the cause of high statistical overlap ( $r = 0.79$ ) in our scales for RF of the self and RF of attachment figures (i.e., within the same prompts, often, subjects are asked to reflect on mental states of both the self and attachment figures), future research should replicate these findings either using multiple methods to differentiate RF across these domains or with a measure that can more clearly distinguish between the objects of reflection. Despite these limitations, the current study has important implications for understanding the overlap between different forms of social cognition. Specifically, it would be inaccurate for future studies to equate ToM, as measured by instruments like the MASC, and RF to one another given that these constructs were not found to overlap completely. In fact, it is likely that discrepancies in how various social-cognitive constructs are operationalized and labeled, more generally, can partially account for heterogeneity in findings from studies attempting to understand the interpersonal difficulties characteristic of BPD (Lazarus et al., 2014). While the field would benefit from greater research to delineate the structure of social cognition (e.g., Happé et al., 2017), future studies should aim to maintain consistency with established operationalizations of the construct under investigation and to establish artifacts that lead to discrepant findings within the same construct. Additionally, in intervention contexts, addressing the attachment relationship, and specifically reflection upon minds in this relationship may have benefits for individuals' broader social network. Particularly, focusing on reflection on one's own mind in a coherent, integrated way may be an important target to improving social-cognitive abilities.

## Funding

This work was supported by the McNair Family Foundation [no grant number available].

## References

- American Psychiatric Association, 2013. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. American Psychiatric Publishing, Arlington, VA.
- Astington, J.W., Baird, J.A., 2005. *Why Language Matters for Theory of Mind*. Oxford University Press, New York, NY, US.
- Bączkowski, B.M., Cierpiakowska, L., 2015. Mentalization within close relationships: the role of specific attachment style. *Pol. Psychol. Bull.* 46, 285–299. <https://doi.org/10.1515/ppb-2015-0035>.
- Bakermans-Kranenburg, M.J., IJzendoorn, M.H.van, 2009. The first 10,000 adult attachment interviews: distributions of adult attachment representations in clinical and non-clinical groups. *Attach. Hum. Dev.* 11, 223–263. <https://doi.org/10.1080/14616730902814762>.
- Bender, D.S., Skodol, A.E., 2007. Borderline personality as a self-other representational disturbance. *J. Pers. Disord.* 21 (5), 500–517. <https://doi.org/10.1521/pedi.2007.21.5.500>.
- Blakemore, S.-J., 2012. Development of the social brain in adolescence. *J. R. Soc. Med.* 105 (3), 111–116. <https://doi.org/10.1258/jrsm.2011.110221>.
- Blakemore, S.-J., Mills, K.L., 2014. Is adolescence a sensitive period for sociocultural processing? *Ann. Rev. Psychol.* 65 (1), 187–207. <https://doi.org/10.1146/annurev-psych-010213-115202>.
- Bogdan, R.J., 2000. *Minding Minds: Evolving a Reflexive Mind by Interpreting Others*. The MIT Press, Cambridge, MA. [https://books.google.com/books?hl=en&lr=&id=HBYQjmrJ5v8C&oi=fnd&pg=PR11&dq=interpreting+minds+bogdan+2003&ots=OEGb55mTkt&sig=sZ0d\\_MqpKQViAbVzWu8IIfoNfc](https://books.google.com/books?hl=en&lr=&id=HBYQjmrJ5v8C&oi=fnd&pg=PR11&dq=interpreting+minds+bogdan+2003&ots=OEGb55mTkt&sig=sZ0d_MqpKQViAbVzWu8IIfoNfc).
- Bosacki, S.L., 2015. Children's theory of mind, self-perceptions, and peer relations: a longitudinal study. *Infant Child Dev.* 24 (2), 175–188. <https://doi.org/10.1002/icd.1878>.
- Caputi, M., Lecce, S., et al., 2012. Longitudinal effects of theory of mind on later peer relations: the role of prosocial behavior. *Dev. Psychol.* 48 (1), 257–270. <https://doi.org/10.1037/a0027000>.

- org/10.1037/a0025402.
- Chang, B., Sharp, C., et al., 2011. The criterion validity of the borderline personality features scale for children in an adolescent inpatient setting. *J. Pers. Disord.* 25 (4), 492–503. <https://doi.org/10.1521/pedi.2011.25.4.492>.
- Crick, N.R., Murray-Close, et al., 2005. Borderline personality features in childhood: a short-term longitudinal study. *Dev. Psychopathol.* 17 (4), 1051–1070. <https://doi.org/10.1017/S0954579405050492>.
- De Rosnay, M., Harris, P.L., 2002. Individual differences in children's understanding of emotion: the roles of attachment and language. *Attach. Human Dev.* 4 (1), 39–54. <https://doi.org/10.1080/14616730210123139>.
- de Villiers, J.G., Pyers, J.E., 2002. Complements to cognition: a longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cogn. Dev.* 17, 1037–1060. [https://doi.org/10.1016/S0885-2014\(02\)00073-4](https://doi.org/10.1016/S0885-2014(02)00073-4).
- Decety, J., Sommerville, J.A., 2003. Shared representations between self and other: a social cognitive neuroscience view. *Trends Cog. Sci.* 7 (12), 527–533. <https://doi.org/10.1016/j.tics.2003.10.004>.
- Dormann, C.F., Elith, J., et al., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36 (1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>.
- Doyle, C., Cicchetti, D., 2017. From the cradle to the grave: the effect of adverse caregiving environments on attachment and relationships throughout the lifespan. *Clin. Psychol.* 24 (2), 203–217. <https://doi.org/10.1111/cpsp.12192>.
- Dziobek, I., Fleck, S., et al., 2006. Introducing MASC: a movie for the assessment of social cognition. *J. Autism Dev. Disord.* 36 (5), 623–636. <https://doi.org/10.1007/s10803-006-0107-0>.
- Ensink, K., Bégin, M., et al., 2016. Maternal and child reflective functioning in the context of child sexual abuse: pathways to depression and externalising difficulties. *Euro. J. Psychotraumatol.* 27 (7). <http://dx.doi.org/10.3402/ejpt.v7.30611>.
- Ensink, K., Normandin, L., et al., 2015. Friendlessness and theory of mind: a prospective longitudinal study. *Br. J. Dev. Psychol.* 33 (2), 203–217. <https://doi.org/10.1111/bjdp.12074>.
- Ensink, K., Target, M., et al., 2013. Child Reflective Functioning Scale Scoring Manual: For Application to the Child Attachment Interview. Anna Freud Center, University College London, London, UK Unpublished Manuscript.
- Ensink, K., Berthelot, N., et al., 2014. Another step closer to measuring the ghosts in the nursery: preliminary validation of the trauma reflective functioning scale. *Front. Psychol.* 5 (1471), 1–12. <https://doi.org/10.3389/fpsyg.2014.01471>.
- Fink, E., Begeer, S., et al., 2015. Friendlessness and theory of mind: a prospective longitudinal study. *Br. J. Dev. Psychol.* 33 (1), 1–17. <https://doi.org/10.1111/bjdp.12060>.
- Fonagy, P., Bateman, A., 2008. The development of borderline personality disorder—a mentalizing model. *J. Pers. Disord.* 22 (1), 4–21. <https://doi.org/10.1521/pedi.2008.22.1.4>.
- Fonagy, P., Gergely, G., et al., 2007. The parent-infant dyad and the construction of the subjective self. *J. Child Psychol. Psychiatry* 48 (3–4), 288–328. <https://doi.org/10.1111/j.1469-7610.2007.01727.x>.
- Fonagy, P., Luyten, P., et al., 2016. Development and validation of a self-report measure of mentalizing: the reflective functioning questionnaire. *PLoS One* 11 (7). <https://doi.org/10.1371/journal.pone.0158678>.
- Fonagy, P., Luyten, P., Strathearn, L., 2011. Borderline personality disorder, mentalization, and the neurobiology of attachment. *Infant Ment. Health J.* 32, 47–69. <https://doi.org/10.1002/imhj.20283>.
- Fonagy, P., Target, M., 2006. The mentalization-focused approach to self pathology. *J. Pers. Disord.* 20 (6), 544–576.
- Fonagy, P., Target, M., 1997. Attachment and reflective function: their role in self-organization. *Dev. Psychopathol.* 9 (4), 679–700.
- Fonagy, P., Target, M., et al., 1998. Reflective-Functioning Manual, Version 5.0, For Application to Adult Attachment Interviews. University College London, London, UK. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.466.3207&rep=rep1&type=pdf>.
- George, D., Mallery, P., 2003. SPSS For Windows Step by Step: A Simple Guide and Reference, 11.0 Update, Fourth ed. Allyn and Bacon, Boston, MA.
- Gifford-Smith, M.E., Brownell, C.A., 2003. Childhood peer relationships: social acceptance, friendships, and peer networks. *J. Sch. Psychol.* 41, 235–284. [https://doi.org/10.1016/S0022-4405\(03\)00048-7](https://doi.org/10.1016/S0022-4405(03)00048-7).
- Ha, C., Sharp, C., et al., 2013. The measurement of reflective function in adolescents with and without borderline traits. *J. Adol.* 36 (6), 1215–1223. <https://doi.org/10.1016/j.adolence.2013.09.008>.
- Happé, F., Cook, J.L., et al., 2017. The structure of social cognition: in(ter)dependence of sociocognitive processes. *Annu. Rev. Psychol.* 68 (1), 243–267. <https://doi.org/10.1146/annurev-psych-010416-044046>.
- Harris, P.L., 1992. From simulation to folk psychology: the case for development. *Mind Lang* 7 (1–2), 120–144. <https://doi.org/10.1111/j.1468-0017.1992.tb00201.x>.
- Harter, S., 2012. Emerging self-processes during childhood and adolescence. In: Leary, M.R., Tangney, J.P. (Eds.), *Handbook of Self and Identity*, Second ed. Guilford Press, New York, pp. 680–715.
- Hopwood, C.J., Wright, A.G., et al., 2013. The interpersonal core of personality pathology. *J. Pers. Disord.* 27 (3), 270–295. <https://doi.org/10.1521/pedi.2013.27.3.270>.
- Humfress, H., O'Connor, T.G., Slaughter, J., Target, M., Fonagy, P., 2002. General and relationship-specific models of social cognition: explaining the overlap and discrepancies. *J. Child Psychol. Psychiatry* 43, 873–883. <https://doi.org/10.1111/1469-7610.00137>.
- IBM Corp., 2016. IBM SPSS Statistics for Windows, Version 24.0. IBM Corp., Armonk, NY.
- Miljkovitch, R., Deborde, A.-S., Bernier, A., Corcos, M., Speranza, M., Pham-Scott, A., 2018. Borderline personality disorder in adolescence as a generalization of disorganized attachment. *Front. Psychol.* 9. <https://doi.org/10.3389/fpsyg.2018.01962>.
- Milligan, K., Astington, J.W., Dack, L.A., 2007. Language and theory of mind: meta-analysis of the relation between language ability and false-belief understanding. *Child Dev.* 78, 622–646. <https://doi.org/10.1111/j.1467-8624.2007.01018.x>.
- Nolte, T., Bolling, D.Z., Hudac, C., Fonagy, P., Mayes, L.C., Pelphrey, K.A., 2013. Brain mechanisms underlying the impact of attachment-related stress on social cognition. *Front. Hum. Neurosci.* 7. <https://doi.org/10.3389/fnhum.2013.00816>.
- Ochsner, K.N., Knierim, K., et al., 2004. Reflecting upon feelings: an fMRI study of neural systems supporting the attribution of emotion to self and other. *J. Cog. Neuro.* 16 (10), 1746–1772. <https://doi.org/10.1162/0898929042947829>.
- Peterson, C.C., Wellman, H.M., 2018. Longitudinal theory of mind (ToM) development from preschool to adolescence with and without ToM delay. *Child Dev.* <https://doi.org/10.1111/cdev.13064>.
- Premack, D., Woodruff, G., 1978. Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1 (04), 515–526. <https://doi.org/10.1017/S0140525X00076512>.
- Repacholi, B., Trapolini, T., 2004. Attachment and preschool children's understanding of maternal versus non-maternal psychological states. *Br. J. Dev. Psychol.* 22 (3), 395–415. <https://doi.org/10.1348/0261510041552693>.
- Ruffman, T., 2014. To belief or not belief: children's theory of mind. *Dev. Rev.* 34 (3), 265–293. <https://doi.org/10.1016/j.dr.2014.04.001>.
- Sansone, R.A., Sansone, L.A., 2011. Gender patterns in borderline personality disorder. *Innov. Clin. Neurosci.* 8 (5), 16–20.
- Scharf, M., Mayseless, O., 2007. Putting eggs in more than one basket: a new look at developmental processes of attachment in adolescence. *New Dir. Child Adolesc. Dev.* (117), 1–22. <https://doi.org/10.1002/cd.191>.
- Sharp, C., 2006. Mentalizing problems in childhood disorders. In: Allen, J.G., Fonagy, P. (Eds.), *Handbook of Mentalization-Based Treatment*. John Wiley & Sons, Ltd., pp. 101–121. <http://onlinelibrary.wiley.com/doi/10.1002/9780470712986.ch4/summary>.
- Sharp, C., 2014. The social-cognitive basis of BPD: a theory of hypermentalizing. In: Sharp, C., Tackett, J.L. (Eds.), *Handbook of Borderline Personality Disorder in Children and Adolescents*. Springer, New York, pp. 211–225. [http://link.springer.com/chapter/10.1007/978-1-4939-0591-1\\_15](http://link.springer.com/chapter/10.1007/978-1-4939-0591-1_15).
- Sharp, C., Pane, H., et al., 2011. Theory of mind and emotion regulation difficulties in adolescents with borderline traits. *J. Am. Acad. Child Adolesc. Psychiatry* 50 (6), 563–573. <https://doi.org/10.1016/j.jaac.2011.01.017>.
- Sharp, C., Vanwoerden, S., 2015. Hypermentalizing in borderline personality disorder: a model and data. *J. Infant, Child Adolesc. Psychother.* 14 (1), 33–45. <https://doi.org/10.1080/15289168.2015.1004890>.
- Sharp, C., Venta, 2013. Mentalizing problems in children and adolescents. In: Midgley, N., Vrouva, I. (Eds.), *Minding the Child: Mentalization-Based Interventions with Children, Young People and Their Families*. Routledge, New York, NY.
- Steinberg, L., Dahl, R., et al., 2006. The study of developmental psychopathology in adolescence: integrating affective neuroscience with the study of context. In: Cicchetti, D., Cohen, D.J. (Eds.), *Developmental Psychopathology, Vol 2: Developmental Neuroscience*, second ed. John Wiley & Sons Inc., Hoboken, NJ, pp. 710–741.
- Target, M., Fonagy, P., et al., 1998. The Child Attachment Interview (CAI) Protocol, Revised Edition VI. University College London, London, UK Unpublished Manuscript.
- Tessier, V.P., Normandin, L., et al., 2016. Fact or fiction? A longitudinal study of play and the development of reflective functioning. *Bull. Menninger Clin.* 80 (1), 60–79. <https://doi.org/10.1521/bumc.2016.80.1.60>.
- Wechsler, D., 2003. Wechsler Intelligence Scale for Children, fourth ed. (WISC-IV). The Psychological Corporation, San Antonio, TX. <https://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8979-044&Mode=resource>.
- Wechsler, D., 2008. Wechsler Adult Intelligence Scale, Fourth ed. (WAIS-IV). NCS Pearson, San Antonio, TX. <http://www.statisticsolutions.com/academic-solutions/resources/directory-of-survey-instruments/wechsler-adult-intelligence-scale-fourth-edition-wais-iv/>.
- Zaki, J., Ochsner, K., 2009. The need for a cognitive neuroscience of naturalistic social cognition. *Ann. NY Acad. Sci.* 1167, 16–30. <https://doi.org/10.1111/j.1749-6632.2009.04601.x>.