https://doi.org/10.1521/jscp.2018.37.9.725

Theory of Mind, Excessive Reassurance-Seeking, and Stress Generation in Depression: A Social-Cognitive-Interpersonal Integration

Chloe C. Hudson[a]

Amanda Shamblaw[a]

Gillian A. Wilson[b]

Meighen M. Roes[c]

Mark A. Sabbagh[a]

Kate L. Harkness[a*]

[a]Queen's University, Canada

[b]University of British Columbia, Canada

[c]Ryerson University, Canada

*Corresponding author. Department of Psychology, 62 Arch Street, Queen's University, Kingston, Ontario, Canada, K7L3N6. E-mail address: harkness@queensu.ca

The work described in this manuscript was performed at Queen's University.

Short title: An Integrated Model of Depression

Keywords: Depression; theory of mind; excessive reassurance seeking; stress; interpersonal model of depression.

Abstract

Objectives: The interpersonal model of depression states that individuals with depression engage in excessive reassurance-seeking (ERS), which leads to interpersonal rejection. We theorize that maladaptive ERS behaviors may be driven by difficulties with "theory of mind" – the ability to decode and reason about others' mental states.

Methods: Participants were 31 young adults in a current episode of unipolar depression, and 91 never-depressed participants. Theory of mind was assessed with standard laboratory tasks. Interpersonal stress was assessed with a gold-standard contextual interview.

Results: Consistent with hypotheses, in the depressed group only, lower theory of mind decoding accuracy was associated with greater ERS, which was significantly associated with exposure to greater interpersonal, but not non-interpersonal, stress. More surprising was that higher theory of mind reasoning accuracy was associated with greater ERS.

Conclusions: These results expand the interpersonal model of depression to include the foundational social cognitive processes that underlie effective social communication.

Theory of Mind, Excessive Reassurance-Seeking, and Stress Generation in Depression: A Social-Cognitive-Interpersonal Integration

Worldwide, major depressive disorder (MDD) affects 300 million people and is the leading cause of disability (World Health Organization, 2017). Individuals with MDD suffer dysfunction across social, occupational, and educational domains, but it is dysfunction in interpersonal relationships that is often reported by depression sufferers as most distressing and impairing (Barrett & Barber, 2007; Lee, Kleinman, & Kleinman, 2007). A number of theoretical models, including Coyne's (1976; Joiner, Alfano, & Metalsky, 1992) interpersonal model of MDD and Hammen's stress generation hypothesis (Hammen, 1991, 2006), seek to understand interpersonal dysfunction in depression as resulting, at least in part, from maladaptive behavior patterns that create conflict and rejection in interpersonal relationships. Excessive reassurance-seeking (ERS), defined as the tendency to repetitively seek assurances of one's worth and lovability from others, has been theorized as a central behavioral factor leading to interpersonal difficulties in depression. Specifically, Coyne (1976) proposes that, while others may first provide reassurance (e.g., "No, I am not mad at you." "Yes, I do love you."), the depressed individual doubts others' sincerity and seeks further reassurance, leading to frustration and annoyance in the depressed individuals' close others. The interpersonal conflict and worsening of depression symptoms that result paradoxically trigger more ERS behavior, thereby perpetuating the cycle and ultimately leading close others to reject the depressed individual.

Strong support has been provided for the interpersonal model. Prospective studies have reported that elevated ERS scores significantly predict conflict in romantic (Eberhart & Hammen, 2009; Shahar, Joiner, Zuroff, & Blatt, 2004) and other relationships (Potthoff, Holahan, & Joiner, 1995), as well as partner-initiated rejection (i.e., break-ups) in romantic relationships (Stewart & Harkness, 2015). The role of ERS as a specific risk factor in depression has also received

considerable empirical support. Depressed individuals report significantly higher levels of ERS compared to non-depressed individuals (the meta-analytic relation of ERS to depression is .32; Starr & Davila, 2008) and individuals with other disorders (e.g., anxiety and substance use disorders; Joiner & Metalsky, 2001). Elevated ERS scores also significantly prospectively predict increases in depression symptoms (Davila, 2001; Joiner & Metalsky, 2001). Further, the strength of the relation of ERS to interpersonal conflict and rejection is stronger in those with elevated levels of depression relative to non-depressed individuals (Starr & Davila, 2008; Stewart & Harkness, 2017).

Despite strong evidence for the interpersonal model of depression, the social cognitive processes that lie at the foundation of interpersonal dysfunction have received little theoretical or empirical attention. One of the reasons that ERS behaviors are perceived as annoying and frustrating by others may be that they constitute a breakdown in the pragmatic principles that guide social interactions (see e.g., Gibbs, 1994; Levinson, 1983). For instance, the pragmatic principle of "quality" describes how, in the normal flow of conversation, one is expected to make utterances that are truthful and accurate, and interpret others' utterances as truthful as well (e.g., Grice, 1975). Pragmatic principles allow conversation and interaction to flow smoothly, and contribute to a feeling of connectedness between partners (Pickering & Garrod, 2004). ERS behaviors may disrupt this feeling of connectedness because by continually asking questions that have already been answered, ERS behaviors signal implicit, and potentially insulting, doubt in the truthfulness of the responses ("I just told you I'm not angry, why do you keep asking?").

Current thinking about the origins of persistent and excessive reassurance-seeking behaviors is that they reflect an *unwillingness* to believe others' reassurance. This unwillingness to believe others may be due, for example, to the depressed individual's rigid cognitive schemas reflecting insecurity in relationships and fears of abandonment and rejection (see Evraire &

Dozois, 2011). Another possibility that has received less attention concerns the possibility that ERS behaviors reflect a more fundamental and basic *deficit* in "theory of mind" – the suite of cognitive abilities that allow one to decode and reason about others' mental states (e.g., beliefs, desires, emotions, intentions). Considering this latter possibility is important because identifying the range of basic mechanisms that drive maladaptive behavior and interpersonal dysfunction in depression expands the range of targets for intervention (e.g., schema work, remediating social cognitive deficits). Perhaps more intriguing, if theory of mind deficits are at play, this suggests a potential pre-existing vulnerability factor that could be targeted in depression prevention efforts. Finally, theory of mind was identified by the National Institute of Mental Health as a priority construct for research in the domain of social processes in psychopathology; therefore, an integrated understanding of its role in interpersonal function is likely to have transdiagnostic significance (Cuthbert & Kozak, 2009).

Engaging theory of mind has sometimes been called "mentalizing" or "mindreading" (e.g., Baron-Cohen, 2004), which captures the extent to which these judgments are probabilistic ones – we can never truly know the contents of others' mental states, but because we understand the critical role they play in causing others' behaviors, we are motivated to make guesses about them that are as accurate as possible. The cognitive processes that are recruited as individuals engage in theory of mind can be separated into at least two components (e.g., Sabbagh, 2004; Sabbagh & Bowman 2017). The first is an immediate, inductive "decoding" component that enables one to make probabilistic judgments about others' mental states based upon available observable information, such as facial expression or tone of voice. Mental state decoding is often measured with the "reading the mind in the eyes" task (RMET; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001), wherein participants are asked to identify the mental state that best matches a picture of the eye region of an actor's face. These immediate decoding inferences can

be accurate without any background knowledge about the person because the observable behavior is directly diagnostic of the mental state (e.g., wide open eyes are directly diagnostic of an emotional state of fear).

The second component is a theory-based conceptual "reasoning" process that relies on experience and knowledge to make sense of the causal connections between mental states and past or future behaviors (Sabbagh, 2004; Sabbagh & Bowman, 2017). This component is important for making accurate inferences about the contents of others' mental states that cannot be inferred from immediately available information alone. In the developmental and cognitive neuroscience fields, the basic ability to make theory-based judgments about others' mental states is commonly assessed with a range of experimental tasks, including the "false-belief task" (e.g., Wimmer & Perner, 1983). In false belief tasks, participants are asked to recognize situations in which story protagonists will make mistakes because their beliefs about the world do not comport with the way the world truly is. Performance on these tasks consistently and reliably activates the cortical regions that are critical for reasoning about mental states (e.g., Saxe, 2006) and, in the developmental context, individual differences in false belief performance both concurrently and longitudinally predict a variety of outcomes related to social competence (see Sabbagh & Bowman, 2017 for a recent review).

Poorer performance on the RMET and tasks of theory of mind reasoning relative to healthy populations have been reported in clinical conditions in which marked interpersonal impairment is a key feature, including autism spectrum disorder (see Yirmiya, Erel, Shaked, & Solomonica-Levi, 1998), schizophrenia (see Brüne, 2005), and bipolar disorder (see Bora, Bartholomeusz, & Pantelis, 2016). Importantly, these differences are independent of broader neurocognitive deficits in these patient populations. Further, and of direct relevance to the current discussion, the severity of theory of mind impairment is significantly correlated with the severity

of interpersonal dysfunction in these populations, as assessed by social functioning questionnaire measures (e.g., Bora, Eryavuz, Kayahan, Sugu, & Veznedaroglu, 2006; Lerner, Hutchins, & Prelock, 2011; Schenkel, Chamberlain, & Towne, 2014) or laboratory behavioral assessment (e.g., Couture, Granholm, & Fish, 2011). Results from a recent meta-analysis also confirm significantly poorer performance on the RMET (d = .44) and various tasks of mental state reasoning (d = .50) in individuals with MDD relative to non-depressed individuals (Bora & Berk, 2016). Furthermore, lower scores on theory of mind tasks in depressed patients have been cross-sectionally associated with lower self-reported social functioning (Cusi, MacQueen, Spreng, & McKinnon, 2011; Cusi, Nazarov, MacQueen, & McKinnon, 2013; c.f. Szanto et al., 2012).

The current study is the first to examine whether individual differences in the ability to decode and reason about others' mental states are associated with the behaviors and real-world interpersonal outcomes theorized in the interpersonal model of MDD. The design of the current study is cross-sectional, with the aim of establishing the basic relation of theory of mind abilities and interpersonal processes to support future prospective, longitudinal work. Nevertheless, it includes a number of important methodological strengths. First, we include a well-characterized depressed group based on structured diagnostic interview. Second, theory of mind is assessed with well-validated laboratory tasks that map clearly onto the distinction between decoding and reasoning described above. Third, an important tenet of the interpersonal model is that ERS behaviors are particular toxic to interpersonal relationships and, thus, should be preferentially associated with stress in this domain. However, the corollary – that ERS should *not* predict elevations in the level of *non*-interpersonal life events (e.g., getting fired, developing a serious illness) has received very little empirical attention (see Hernandez, Trout, & Liu, 2016). In the current study, we assess stressful life events using a rigorous contextual life event interview and

independent rating system that allows for unbiased ratings of life event severity and contextually-based distinctions between interpersonal and non-interpersonal events.

We test the following hypotheses: 1. Lower accuracy on the RMET and a false belief (FB) task will be significantly associated with greater ERS in the depressed group. 2. Greater ERS will be significantly associated with higher levels of interpersonal, but not non-interpersonal, stress. 3. In the depressed group, ERS will cross-sectionally mediate the relation of theory of mind decoding and theory of mind reasoning accuracy to interpersonal stress.

Methods

Participants

The current sample included 122 participants (31 depressed, 91 non-depressed) recruited from an introductory psychology course and from advertisements placed across campus who were part of a larger study on individual differences in theory of mind and lifetime depression. The General Research Ethics Board at Queen's University approved the study protocol and all participants provided written informed consent. To be included in the depressed group of the current report, participants had to meet DSM-IV-TR criteria (American Psychiatrist Association, 2000) for a current episode of a unipolar depressive disorder based on a structured diagnostic interview (see below). Participants in the non-depressed group had to be free of any current or past psychiatric diagnosis. To increase the likelihood of recruiting participants with a depression history, individuals from the introductory psychology course were initially contacted for potential participation based on their responses to the Beck Depression Inventory (BDI-II; Beck, Steer, & Brown, 1996) given as part of a departmental pre-screening inventory. Individuals who scored less than 5 on the BDI-II were contacted for potential participation in the non-depressed group, and individuals with BDI-II scores over 15 were contacted for potential participation in the depressed group. Individuals who expressed interest in the study participated in an initial phone

screen to evaluate exclusion criteria (i.e., a lifetime diagnosis of a psychotic disorder, bipolar disorder, substance dependence, developmental disability, or medical disorder causing depression). Those who passed the telephone screen were invited to an in-person interview.

The initial sample included 213 participants. Of these, 58 met criteria for a past, but not current, episode of a depressive disorder, and 21 participants did not meet criteria for depression, but met criteria for another current psychiatric disorder. A further 12 participants were excluded because they were missing data on one or more of the primary measures due to attrition or technical difficulties. The excluded participants did not differ significantly from those included in the current report in terms of sex, age, ethnicity, BDI-II scores, accuracy on the theory of mind tasks, or level of interpersonal or non-interpersonal life events (all *p*s > .05).

Measures

Diagnostic interview. Participants were administered the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-IV-I/P; First, Spitzer, Gibbon, & Williams, 2002) by clinical psychology graduate student interviewers who were trained and supervised throughout the project by the senior author, a licensed clinical psychologist. Students were trained to "gold standard" status, which involved matching the diagnoses of a gold standard interviewer on six consecutive SCID-IV-I/P interviews (Grove, Andreasen, McDonald-Scott, Keller, & Shapiro, 1981). Based on the SCID-IV-I/P interview, 31 participants met DSM-IV criteria for a current episode of a unipolar depressive disorder (major depressive disorder, *n* = 27; adjustment disorder with depressed mood, *n* = 2; depressive disorder not otherwise specified, *n* = 2).[1] Eleven (35%) participants in the depressed group met criteria for a comorbid disorder: Social phobia (*n* = 7), post-traumatic stress disorder (*n* = 2), generalized anxiety disorder (*n* = 1), and panic disorder with agoraphobia (*n* = 1).

[1] The pattern of results did not change when only those with MDD were included in the model.

Depression severity. The BDI-II (Beck et al., 1996) is a 21-item self-report questionnaire that assesses the presence and severity of depression symptoms. Items are scored on a 4-point scale (0-3) with higher scores indicating greater severity of depressive symptoms. The standardized internal consistency estimate for the BDI-II in the current sample was .95.

Excessive reassurance seeking. The Depressive Interpersonal Relationships Inventory-Reassurance Seeking Subscale (DIRI-RS; Metalsky et al., 1991) is a 4-item measure of reassurance seeking and is the most widely-used measure of ERS in the literature. A sample item is "Do you frequently seek reassurance from the people you feel close to as to whether they really care about you?" Items are rated on a 7-point Likert-type scale ranging from 1-*No, Not at all* to 7-*Yes, very much*. Scores were obtained by summing across the four items. The standardized internal consistency estimate for the DIRI-RS in the current sample was .92.

Stressful life events. The Life Events and Difficulties Schedule (LEDS; Bifulco et al., 1989) is a semi-structured, contextual interview and rating system that assesses recent life events in ten domains: education, occupation, housing, finances, role changes, legal, health, romantic relationships, other relationships, and deaths. The current study focused on events occurring in the 6-month period prior to the interview to allow sufficient time for events to occur, but to also ensure the participants would remember adequate details surrounding the events.

Graduate student interviewers were trained and supervised by the senior author in the Bedford College method of obtaining objective contextual information regarding life events. Specifically, they were trained not to probe for details regarding the participants' subjective response to stressors and to focus their questions on the behavioral and contextual details of each event. The interviews were audio recorded and research assistants subsequently summarized each event in a vignette. Each event was then rated by a team of 2-4 independent judges who were unaware of participants' depression status or subjective responses to events using the LEDS

manual, which includes explicit rules and rating criteria, as well as over 5000 case examples to anchor the ratings. The contextual method minimizes depressive and other biases in subjective perceptions of event severity, resulting in superior reliability and validity of life event data than that provided by self-report stress checklists. As such, the LEDS is widely regarded as the gold-standard method for defining and rating life events (see Harkness & Monroe, 2016).

Each event was rated for its contextual threat on a 5-point scale (1-*marked*, 2a-*high moderate*, 2b-*low moderate*, 3-*some*, 4-*little/none*). Raters also coded each event as interpersonal or non-interpersonal using the manual. Discrepancies among raters were resolved through discussion, and the consensus ratings were used in analyses. To create the variables for analyses, the threat ratings were reverse-coded so that higher scores represented greater threat, and then the ratings were summed across the number of events reported in the 6-month period. This is a commonly used convention for defining life event variables in tests of the stress generation hypothesis (e.g., Conway, Hammen, & Brennan, 2012; Harkness, Lumley, & Truss, 2008). Separate threat scores were calculated for interpersonal and non-interpersonal life events.

Theory of mind. Theory of mind decoding was assessed with the Reading the Mind in the Eyes task (RMET; Baron-Cohen et al., 2001). The RMET includes 36 black-and-white photographs of the eye region of different faces that are surrounded by four adjectives, one at each corner of the photograph (e.g., jealous, panicked, arrogant, hateful). Participants were told to choose the adjective that best describes the photograph as quickly as possible. The task was presented on a computer monitor and participants responded by pressing one of four letters on the keyboard (i.e., S, X, K, M) that were spatially analogous to the adjectives. There is a standardized correct response for each item. Accuracy was defined as percent correct and response time was recorded in milliseconds. The RMET task was designed as an “advanced theory of mind test” that is able to detect subtle differences in social intelligence among adults of normal intellectual

functioning. It is advantageous because it does not require free recall or visuospatial discrimination skills that may be biased or impaired in depressed groups (Harkness, Sabbagh, Jacobson, Chowdrey, & Chen, 2005). Further, it is challenging as it requires a high degree of sensitivity to the subtle features of eye expressions, and requires participants to "put themselves into the mind of others" (Baron-Cohen et al., 2001). Healthy adults perform at about 70% accuracy on this task, thereby minimizing ceiling effects.

To control for task demands and perceptual processing abilities not attributed to theory of mind in the RMET task, we administered the Animals task (Harkness et al., 2005). In this task participants were presented with 10 black-and-white photographs of animals surrounded by four adjectives in a format similar to the RMET (e.g., tiger: aloof, *ferocious*, timid and obedient; standardized correct answer italicized). Accuracy was defined as percent correct and response time was recorded in milliseconds. The RMET and Animals items were presented randomly in one block of 46 items.

Theory of mind reasoning was assessed with the False Belief task (FB task; Drover, 2012). The FB task consists of 13 animated scenes presented on a computer screen. In each scene, a cartoon protagonist watches two other cartoon characters enter two of three boxes. The protagonist then leaves the room. While the protagonist is gone, the two characters either remain in the same boxes (true belief [TB] trials) or move to different boxes (false belief [FB] trials). The protagonist then returns and the participants are instructed to indicate where the protagonist will look for each character. Participants responded using the numbers 1, 2, or 3 on the keyboard that corresponded to the boxes on the screen. Accuracy was defined as percent correct. Accuracy on TB items was used to control for motivation and cognitive abilities not attributable to theory of mind.

Procedure

Participants came into the laboratory for two sessions separated by approximately one week. During the first session, participants provided written informed consent and then completed the FB task, questionnaires (i.e., demographic, BDI-II, DIRI-RS), and the SCID-IV-I/P. During the second session, participants completed the RMET and the LEDS. Participants completed other measures at both time points not relevant to the current study. Participants were compensated $40 or course credit (i.e., 4% towards their introductory psychology final exam) for their participation in the study. All participants were provided with contact information for local mental health services upon completion of the study.

Data Analysis

SPSS statistical software version 23.0 was used for all analyses. All continuous variables were standardized. Moderated mediation analyses were conducted in PROCESS (Hayes, 2013), a regression-based macro for SPSS. We used 10,000 bootstrap estimates to obtain 95% bias-corrected confidence intervals (CIs) for the conditional indirect effects.

Four moderated mediation models were constructed to examine whether the relations of theory of mind accuracy (decoding and reasoning) to life event threat score (interpersonal and non-interpersonal) through ERS differed based on depression status (see Figure 1). Individual regression paths were tested to determine (1) the relation of theory of mind accuracy, depression status, and their interaction to ERS (paths a_1, a_2, a_3) and (2) the relation of ERS to life event threat score (path b). If the interaction term (path a_3) was significant, we examined the relation of theory of mind accuracy to ERS separately for depressed and non-depressed groups. The direct effect of theory of mind accuracy on life event threat score (path c) is not a necessary component of mediation models (MacKinnon & Fairchild, 2009), thus, we proceeded with moderated mediation models even when the direct path was not significant. The significance of the moderated mediation models was tested using Hayes' (2015) index of moderated mediation. The index of

moderated mediation is a numerical value that specifies whether the indirect effect (path a_3b) is significantly different than zero. If the CIs for the index of moderated mediation do not overlap with zero, the indirect effect is systematically larger or smaller for the depressed group relative to the non-depressed group (Hayes, 2015). To follow-up significant moderated mediation models, we examined the indirect effect for the depressed and non-depressed groups separately.

The interpersonal model of depression and the existing literature on theory of mind and depression provides the theoretical basis for the model we proposed in Figure 1. Testing alternative models is not appropriate because they would be of the same equivalent class, and thus, statistically indistinguishable. While alternative models may produce stronger or weaker effect sizes, these results are not indicative of which model is a better fit to the true model (see Lemmer & Gollwitzer, 2017; Thoemmes, 2015).

Results

Preliminary Analyses

Demographic and clinical characteristics of the sample, stratified by depression group, are presented in Table 1. As expected, the depressed group had significantly higher BDI-II scores than the non-depressed group. No other differences emerged as significant. Study variables, stratified by depression group, are presented in Table 2. The depressed group scored higher on the RMET, at a trend level, than the non-depressed group. As expected, the depressed group scored significantly higher than the non-depressed group on the DIRI-RS. Further, consistent with the stress generation hypothesis (Hammen, 2006), the depressed participants had significantly higher interpersonal (but not non-interpersonal) threat scores than the non-depressed participants.

Accuracy on the RMET was not significantly associated with sex or age ($ps > .13$). However, participants who identified as White performed significantly better on the RMET task

compared to participants who did not identify as White (*M*s = 74.01, 70.09, *SD*s = 9.74, 10.32; t[120] = 2.15, p = .03).[2] Within the depressed group, RMET accuracy was not associated with the number of episodes of depression, age of first episode, presence of a comorbid diagnosis, or antidepressant medication (ADM) use (*p*s > .34). Accuracy on the FB task was not significantly associated with sex, age, or ethnicity (*p* > .13). Further, within the depressed group, FB accuracy was not significantly associated with the number of episodes of depression, age of first episode, comorbid diagnosis, ADM use, or BDI-II scores (*p* > .34).

Women scored higher on the DIRI-RS than men, at a trend, (*M*s = 9.99, 7.52, *SD*s = 6.70, 4.76; *t*[120] = 1.85, *p* = .07) Further, higher ERS scores were significantly associated with younger age, *r*(120) = -.19, *p* = .04. ERS scores were not significantly associated with ethnicity (*p* = .69). Within the depressed group, ERS scores were not significantly associated with any of the clinical variables (*p*s > .12).

Finally, neither interpersonal nor non-interpersonal threat scores were significantly associated with sex (*p*s > .42). However, higher interpersonal, *r*(120) = -.21, *p* = .02, and non-interpersonal, *r*(120) = -.26, *p* = .004, threat scores were significantly associated with younger age, and White participants had a higher interpersonal life event threat score compared to non-White participants at a trend level (*M*s = 2.17, 1.39, *SD*s = 2.27, 2.03; *t*[120] = 1.98, *p* = .05). Within the depressed group, higher non-interpersonal life event threat scores were significantly associated with higher BDI-II scores, *r*(29) = .54, *p* = .002. Neither interpersonal nor non-interpersonal life event threat scores were significantly associated with any of the additional clinical variables (*p*s > .25).

[2] It has been extensively documented that individuals, even from infancy, are better at processing faces of their own race than faces of other races (the "cross-race recognition deficit"; Levin, 2000; Lee, Quinn, & Pascalis, 2017). Therefore, the ethnicity effect reported here on the RMET likely does not reflect ethnic differences in theory of mind ability, but instead reflects the fact that all of the eyes stimuli in the RMET are from White faces. Importantly, however, ethnicity did not emerge as a significant covariate in any of our primary models.

Preliminary model-building revealed that including the covariates of sex, age, and ethnicity in the primary models did not change the pattern of results. Therefore, for the sake of parsimony and ease of interpretability, the uncontrolled models are reported below.[3]

Zero-Order Relations Among Study Variables

The relations among all study variables are presented in Table 3. As expected, higher BDI-II scores were significantly associated with higher DIRI-RS scores, as well as with higher levels of interpersonal and non-interpersonal life event threat. In the full sample, BDI-II scores were not significantly associated with RMET or FB task accuracy; however, within the depressed group, consistent with the literature on depression, higher BDI-II scores were significantly related to *lower* RMET accuracy, $r(29) = -.41$, $p = .02$. As predicted by the interpersonal model, higher DIRI-RS scores were significantly associated with higher interpersonal, but not non-interpersonal, threat. As expected, RMET and Animals task accuracy were positively correlated at a trend level, and accuracy on the FB and TB items of the FB task were significantly positively correlated. Higher accuracy on the Animals task was significantly associated with higher accuracy on FB and TB trials. Finally, higher accuracy on the TB trials was significantly associated with lower non-interpersonal life event threat score.

Moderated Mediation Analyses

For models 1 and 2 including the RMET, we used the residuals of RMET accuracy regressing Animals task accuracy as the independent variable. For models 3 and 4 including the FB task, we used the residuals of the FB items accuracy regressing the TB items accuracy as the independent variable.

Model 1: RMET and interpersonal threat. As presented in Table 4, column 1, the relation of RMET to ERS was not significant. Depression status was significantly related to ERS,

[3] All models are available upon request.

and this was qualified by a significant RMET by depression status interaction. Following up this interaction revealed that, consistent with hypotheses, lower RMET accuracy was significantly associated with higher ERS scores in the depressed group, b = -.54, $t(29)$ = 3.38, p = .001, but not in the non-depressed group, b = -.04, $t(90)$ = 0.40, p = .69.

As presented in Table 4, column 2, the relation of RMET to interpersonal life event threat score was not significant. However, as predicted, higher ERS was significantly associated with higher interpersonal life event threat scores. Further, the overall index of moderated mediation was significant (b = -.10, SE = .06, 95% CI [-.27, -.004]). Follow-up analyses revealed that, consistent with hypotheses, the indirect effect of RMET accuracy on interpersonal life event threat score through ERS was significant for the depressed group (b = -.10, SE = .07, CI [-.28, -.01]) but not the non-depressed group (b = .01, SE = .01, CI [-.05, .02]).

Model 2: RMET and non-interpersonal threat. The results for paths a_1, a_2 and a_3 (Table 4, column 1) are identical to the model above. As presented in column 3 of Table 4, RMET and ERS were not significantly associated with non-interpersonal threat. Further, the overall index of moderated mediation in the model with non-interpersonal threat was not significant (b = -.07, SE = .07, 95% CI [-.27, .01]).

Model 3: FB and interpersonal threat. As presented in column 1 of Table 5, the relation of FB to ERS was not significant. Depression status was significantly related to ERS, and this was qualified by a significant FB by depression status interaction. Contrary to hypotheses, *higher* FB accuracy was associated with higher ERS scores in the depressed group, b = .39, $t(30)$ = 2.45, p = .02, but not the non-depressed group, b = -.05, $t(90)$ = 0.52, p = .61.

As displayed in column 2 of Table 5, the relation of FB to interpersonal threat score was not significant. However, as above, higher ERS was significantly associated with higher interpersonal threat. Further, the overall index of moderated mediation was significant (b = .08,

SE = .06, 95% CI [.004, .23]). Follow-up analyses revealed that the indirect effect of FB accuracy on interpersonal threat through ERS was significant for the depressed group (*b* = .08, SE = .08, 95% CI [.0001, .20]) but not for the non-depressed group (*b* = -.01, SE = .01, 95% CI [-.05, .01]).

Model 4: FB and non-interpersonal threat. The results for paths a_1, a_2 and a_3 (Table 5, column 1) are identical to the model above. As presented in Table 5, column 3, neither FB accuracy nor ERS was associated with non-interpersonal threat. Further, the overall index of moderated mediation was not significant (*b* = .06, SE = .05, 95% CI [-.01, .20]).

Discussion

The current results provide important support for the interpersonal model of depression, generally, and expand it in novel ways. First, this study is the first to integrate the foundational social cognitive skill of theory of mind into the interpersonal model of depression. Specifically, in the depressed group, individual differences in the ability to decode and reason about others' mental states were significantly associated with excessive reassurance-seeking, which in turn were associated with heightened interpersonal stress. These results suggest that maladaptive behaviors in depression may not only be driven by the content of depressed individuals' depressotypic schemas, but may also be driven by more fundamental differences in the basic social cognitive ability to decode and reason about others' mental states. Second, the above relations emerged in the depressed group only. Therefore, the preliminary results reported here provide further support for the specific toxicity of ERS in depression, and for the fundamental importance of theory of mind in scaffolding effective social communication and interpersonal functioning in depression. Third, we found that ERS was specifically associated with higher levels of stress in the interpersonal domain, and preferentially mediated the relation of theory of mind abilities to interpersonal stress. Previous studies of the interpersonal model in depression

have typically failed to test the primacy of interpersonal *over non*-interpersonal forms of stress, thus the current results provide important discriminant validity for the ERS construct.

The pattern of results revealed an intriguing dissociation in the relation of theory of mind to interpersonal functioning in depression depending on whether the construct under analysis was *decoding* others' mental states or *reasoning* about others' false beliefs. In terms of mental state decoding, as hypothesized, for the depressed group only, lower scores on the RMET were significantly associated with greater self-reported ERS, which was significantly associated with greater interpersonal stress. One of the intriguing potential interpretations suggested by these findings is that ERS in depression may be driven, at least in part, on difficulties in depressed individuals' abilities to decode others' mental states. People express a multitude of complex and subtle mental states during interpersonal interactions, and accurately decoding and labelling these mental states is crucial to responding to others in appropriate ways. For example, not being able to determine whether one's conversational partner is angry vs. confused, concerned vs. anxious, tired vs. bored is likely at the very least to result in confusion. If this confusion occurs in the context of depression, which is characterized by dependence and insecurity in relationships, it could provoke maladaptive attempts to seek reassurance about others' mental states. An important next step for research is to clarify the temporal sequence and external validity of these results by developing paradigms that enable the examination of mental state decoding accuracy in real-world interpersonal exchanges and its effect in provoking actual ERS behaviors in depression. Further, in the current study poorer accuracy on the RMET was associated with more severe depression. Therefore, future research is also required to determine whether depression severity further moderates the relation of theory of mind decoding accuracy and interpersonal outcomes within a depressed group (see further discussion of the dissociation of theory of mind accuracy across the depression spectrum below).

In direct contrast, *higher* scores on the false belief task were associated with higher ERS in the depressed group, which again were related to higher levels of interpersonal stress. The reasons for this unexpected pattern of results are not entirely clear and require further study. Further, it should be noted that the effects for the false belief task model were weaker than those for the RMET model and, thus, we interpret these results with caution. Nevertheless, we provide some speculations below to guide future inquiry. It is important to note that while theory of mind decoding and reasoning are related insofar as both are necessary for making accurate judgments about the contents of other's mental states, they are also dissociable skills. Indeed, performance on the RMET and FB tasks was not significantly correlated in the current study (see Table 3). Therefore, some insight into the issue might come from a consideration of the distinct cognitive architectures of these skills themselves.

Theory of mind decoding requires attention only to information available in the immediate environment at a discrete moment in time. Consistent with the very quick nature of theory of mind decoding, the neural time course of these judgments in a modified RMET is rapid; in event-related potential (ERP) studies, brain electrical activity uniquely associated with mental state judgments is detected at 270 ms post-stimulus over anterior temporal and inferior frontal regions that are similarly recruited when making emotion recognition judgements (e.g., Sabbagh, Moulson, & Harkness, 2004). In contrast, theory of mind reasoning, and in particular false belief reasoning, is a more effortful process that requires sustained focus on the relevant features of a scenario and integration of context and past knowledge (see e.g., Apperly & Butterfill, 2009). Consistent with the deliberative nature of theory of mind reasoning, false belief tasks recruit core frontal lobe sites and the neural time course of such judgments is more extended. ERP studies show that brain electrical activity associated with theory of mind reasoning begins to emerge at around 300-500 ms but is sustained for much longer and maximal around 700-900 ms post

stimulus (e.g., Liu, Sabbagh, Gehring, & Wellman, 2009). An important factor contributing to the capacity to *sustain* focus on others' mental states is motivation; that is, one must have a drive to understand others' mental states to justify the substantial cognitive effort that is necessary to make accurate judgments (see Andrews & Thompson, 2009; Harkness, Jacobson, Sinclair, Chan, & Sabbagh, 2012; Weary & Edwards, 1994). Therefore, individuals who are especially motivated to understand others' mental states might be expected to perform better on tasks of mental state reasoning than those who lack this interpersonal motivation. In depression, in particular, when this motivation may be driven, at least in part, by interpersonal insecurity and dependency, one would also expect those who are more highly motivated to understand others' mental states to engage in higher levels of ERS than those who are not so motivated. That is, it is possible that interpersonal motivation (or "interpersonal orientation"; e.g., neediness, connectedness; Little & Garber, 2000, 2004; Rude & Burnham, 1995) is a "driver" of the relation between enhanced ability to reason about others' mental states and negative interpersonal outcomes in depression.

The interaction between theory of mind decoding and reasoning may also provide important insights into the social cognitive mechanisms driving maladaptive functioning in depression. In particular, in depression, those who show enhanced motivation to understand others' minds (reflected in better theory of mind reasoning) coupled with poor abilities in decoding others' mental states, may be those who evidence the highest levels of interpersonal impairment. The size of the sample in the current study prevented us from examining this intriguing question, so it remains open for future research.

The depressed group in the current sample scored *higher* on the RMET, as a trend, than the non-depressed group. This finding is in contrast to meta-analytic results reporting lower RMET accuracy in depressed adult patient samples relative to healthy controls (Bora & Berk, 2016). However, while almost all of the literature on theory of mind decoding accuracy in major

depression has focused on adult outpatients with major depressive disorder, our young adult sample included subthreshold cases, and a wide range of severity (BDI-II scores = 9-47). In previous work, mild depression (i.e., dysphoria) has been associated with significantly *enhanced* accuracy on the RMET (Harkness et al., 2005; Harkness et al., 2012; Mellick & Sharp, 2016; Washburn et al., 2016). Importantly, in the current sample we found that, within our depressed group, higher depression severity scores were significantly associated with lower scores on the RMET. Therefore, our current results are also consistent with a dissociation across the depression spectrum, with enhanced abilities characterizing mild to moderate depression, and impaired abilities characteristic of more severe levels of depression (see Harkness et al., 2005; Lee, Harkness, Sabbagh, & Jacobson, 2005). When we constrained the current sample to only those with major depressive disorder in the severe range (BDI-II > 25), the depressed groups scored lower than the non-depressed group on the RMET and FB tasks, although not significantly so, perhaps due to reduced power. It is also important to note that for both RMET and FB tasks, the pattern of results for our moderation models were similar to the results reported above when we operationalized depression continuously using BDI-II scores (results available upon request). Nevertheless, future research with larger samples of individuals across the full spectrum of depression severity are required.

Further, we failed to find a significant difference between depressed and non-depressed groups on the FB task. Theory of mind reasoning in previous studies has been assessed with a very wide array of tasks, from recognition and interpretation of humor and social faux pas in simple stories (e.g., Stone & Baron-Cohen, 1998) to the assignment of mental states and intentions from videos of complex social interactions (e.g., Dziobek et al., 2006). False belief scenarios similar to the ones employed in the current study have only been investigated in three previous studies of depression. Two of these also failed to find significant differences between

depressed and healthy comparison groups (Corcoran et al., 2008; Cusi et al., 2013), and the third employed a sample with chronic depression (Mattern et al., 2015). The use of very diverse tasks limits conclusions that can be drawn about the relation of depression to theory of mind reasoning, and suggests that even this sub-category of theory of mind may be multi-dimensional. Nevertheless, it is intriguing that even in the absence of strong group differences on either the RMET or the FB tasks in current study, individual differences in theory of mind accuracy were still differentially related to interpersonal behaviors and stress in depressed versus non-depressed groups. Future research is now needed to examine whether the results reported here generalize to other aspects of mental state decoding and reasoning not captured by the RMET and FB tasks.

The results of the current study should be considered in light of the following limitations. First, this was a student sample and, thus, results may not generalize to patient or community depressed samples. Nevertheless, it is important to note that all individuals in the depressed group met formal DSM-IV-TR criteria for a unipolar depressive disorder. Second, the design was cross sectional. Longitudinal, prospective designs are now required to test the temporal predictive associations between individual differences in theory of mind, ERS, and the generation of interpersonal stress. Third, this study relied on self-reports of ERS behaviors. While the DIRI-RS is the standard in the field, and by far the most common approach to measuring ERS, it is not clear whether individuals' *beliefs* that they engage in ERS accurately reflect their observable behavior. Indeed, in one of the only studies to our knowledge to use objective codings of observable ERS behaviors, Stewart and Harkness (2017) found that such behavioral codings in the context of interactions with romantic partners were not significantly correlated with self-report scores on the DIRI-RS. Therefore, future development of objective, behaviorally anchored, assessment of ERS is needed. Finally, this study relied on retrospective reports of stressful life events, which can be subject to depressive and other biases. The LEDS includes a number of

important features to address the confounds introduced by such bias, such as the use of independent coders who are unaware of participants' depression status and subjective responses to events, and the use of an extensive manual, which serves to anchor and standardize the ratings. There is now overwhelming evidence that the LEDS and other contextual interview-based systems are the gold standard in the assessment of stress (see Harkness & Monroe, 2016).

Conclusions

In summary, individual differences in depressed individuals' abilities to decode and reason about others' mental states were significantly related to maladaptive interpersonal behavior, which was significantly associated with interpersonal stress. These novel, preliminary results suggest that depressed individuals who generate toxic interpersonal environments may do so not only as a product of their depressogenic schemas, but also as a result of their enhanced motivation to understand, and thus sustain attention on, others' beliefs, desires, and intentions, potentially coupled with a fundamental inability to decode those mental states. As such, the current results expand the interpersonal model of depression to include the foundational social cognitive processes that underlie effective social communication. Further, they suggest avenues for intervention that could supplement traditional cognitive and behavioral approaches. For example, patients with borderline personality disorder have been found to engage in a process of "over-interpretive mental state reasoning," or the over-attribution of mental states to others, consistent with a bias to over-attend to others' mental states in social interactions (e.g., Sharp & Vanwoerden, 2015). Mentalization treatments that include training in mindfulness, perspective-taking, and interpersonal skills have shown significant acute and long-term efficacy as adjuncts to dialectical behavior therapy for this population (Bateman & Fonagy, 2008). Similar strategies may be useful in preventing the distress and often disabling impairment in interpersonal

functioning experienced by individuals with depression, and in helping these individuals achieve healthy and rewarding relationships.

Acknowledgements

We gratefully acknowledge the assistance of Sarah Saperia, Mateya Dimnik, Dustin Washburn, Nicholas Tollefson and Heather Brooks with recruitment, data collection and data management.

Financial Disclosure Statement

The authors declare no conflict of interest.

Funding Source

This work was supported by a grant from the Social Sciences and Humanities Research Council [435-2012-1536, 2012].

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Andrews, P. W., & Thomson, J. A. (2009). The bright side of being blue: depression as an adaptation for analyzing complex problems. *Psychological Review*, *116*, 620–54. http://doi.org/10.1037/a0016242

Apperly, I. A. & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*, 953-970. doi: 10.1037/a0016923.

Baron-Cohen S. (2004). The cognitive neuroscience of autism. *Journal of Neurology and Neurosurgery in Psychiatry, 75*, 945-948.

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.

Barrett, M. S. & Barber, J. P. (2007). Interpersonal profiles in major depressive disorder. *Journal of Clinical Psychology, 63*(3), 247-266. doi: 10.1002/jclp.20346

Bateman, A., & 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*(May), 631–638.

Beck, A.T., Steer, R.A., & Brown, G.K. (1996). *Manual for the Beck Depression Inventory-II.* San Antonio, TX: Psychological Corporation.

Bifulco, A., Brown, G. W., Edwards, A., Harris, T., Neilson, E., Richards, C., et al. (1989). *Life Events and Difficulties Schedule (LEDS-2): Vol. 1. Life events manual.* London, England: Royal Holloway and Bedford New College, University of London.

Bora, E., Bartholomeusz, C., & Pantelis, C. (2016). Meta-analysis of Theory of Mind (ToM) impairment in bipolar disorder. *Psychological Medicine, 46*(2), 253-264. doi: 10.1017/S0033291715001993

Bora, E. & Berk, M. (2016). Theory of mind in major depression disorder: A meta-analysis. *Journal of Affective Disorders, 191*, 49-55. doi: 10.1016/j.jad.2015.11.023

Bora, E., Eryavuz, Kayahan, B., Sugu, G., & Veznedaroglu, B. (2006). Social functioning, theory of mind and neurocognition in outpatients with schizophrenia; mental state decoding may be a better predictor of social functioning than mental state reasoning. *Psychiatry Research, 145*(2-3), 95-103. doi: 0.1016/j.psychres.2005.11.003

Brüne, M. (2005). "Theory of mind" in schizophrenia: A review of the literature. *Schizophrenia Bulletin, 31*(1), 21-42. doi: 10.1093/schbul/sbi002

Conway, C. C., Hammen, C., & Brennan, P. A. (2012). Expanding stress generation theory: test of a transdiagnostic model. *Journal of Abnormal Psychology*, *121*(3), 754–66. http://doi.org/10.1037/a0027457

Corcoran R., Rowse, G., Moore, R., Blackwood, N., Kinderman, P., Howard, R.,… Bentall, R. P. (2008). A transdiagnostic investigation of 'theory of mind' and 'jumping to conclusions' in patients with persecutory delusions. *Psychological Medicine, 38*, 1577-1583.

Couture, S. M., Granholm, E. L., & Fish, S. C. (2011). A path model investigation of neurocognition, theory of mind, social competence, negative symptoms and real-world functioning in schizophrenia. *Schizophrenia Research, 125*, 152-160. doi: 10.1016/j.schres.2010.09.020

Coyne, J. C. (1976). Toward an interactional description of depression. *Psychiatry, 39*, 28-40.

Cusi, A. M., MacQueen, G. M., Spreng, R. N., & McKinnon, M. C. (2011). Altered empathic responding in major depressive disorder: Relation to symptom severity, illness burden,

and psychosocial outcome. *Psychiatry Research, 188*, 231-236. doi: 10.1016/j.psychres.2011.04.013

Cusi, A. M., Nazarov, A., MacQueen, G. M., & McKinnon, M. C. (2013). Theory of mind deficits in patients with mild symptoms of major depressive disorder. *Psychiatry Research, 210*, 672-674. doi: 10.1016/j.psychres.2013.06.018

Cuthbert, B. N., & Kozak, M. J. (2009). Constructing constructs for psychopathology: the NIMH research domain criteria. *Journal of Abnormal Psychology, 122*, 928-937. doi: 10.1037/a0034028.

Davila, J. (2001). Refining the association between excessive reassurance seeking and depressive symptoms: The role of related interpersonal constructs. *Journal of Social and Clinical Psychology, 20*(4), 538-559.

Drover, S. M. (2012). *Event-related potentials associated with theory of mind updating in adults* (Unpublished undergraduate honours thesis). Queen's University, Ontario.

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

Eberhart, N. K., & Hammen, C. L. (2009). Interpersonal predictors of stress generation. *Personality and Social Psychology Bulletin, 35*, 544-556. doi: 10.1177/0146167208329857

Evraire, L. E. & Dozois, D. J. A. (2011). An integrative model of excessive reassurance seeking and negative feedback seeking in the development and maintenance of depression. *Clinical Psychology Review, 31*, 1291–1303. doi: 10.1016/j.cpr.2011.07.014

First, M.B., Spitzer, R.L., Gibbon, M., & Williams, J.B. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition (SCID-I/P)*. New York, NY: Biometrics Research.

Gibbs, R. (1994). *The poetics of mind: Figurative thought, language, and understanding.* New York: Cambridge University Press.

Grice, H. P. (1975). *Logic and Conversation.* London: Elsevier.

Grove, W., Andreasen, N., McDonald-Scott, P., Keller, M., & Shapiro, R. (1981). Reliability studies of psychiatric diagnosis. *Archives of General Psychiatry, 38*, 408–413.

Hammen, C. (1991). Generation of stress in the course of unipolar depression. *Journal of Abnormal Psychology*, *100*(4), 555–561. http://doi.org/10.1037//0021-843X.100.4.555

Hammen, C. (2006). Stress generation in depression: Reflections on origins, research, and future directions. *Journal of Clinical Psychology*, *62*, 1065–1082.

Harkness, K. L., Jacobson, J. A., Sinclair, B., Chan, E., & Sabbagh, M. A. (2012). For love or money? What motivates people to know the minds of others? *Cognition and emotion, 26*(3), 541-549. doi: 10.1080/02699931.2011.588689

Harkness, K. L., Lumley, M. N., & Truss, A. E. (2008). Stress generation in adolescent depression: The moderating role of child abuse and neglect. *Journal of Abnormal Child Psychology*, *36*, 421–432. http://doi.org/10.1007/s10802-007-9188-2

Harkness, K. L. & Monroe, S. M. (2016). The assessment and measurement of adult life stress: Basic premises, operational principles, and design requirements. J*ournal of Abnormal Psychology, 125*(5), 727-745. doi: 10.1037/abn0000178

Harkness, K., Sabbagh, M., Jacobson, J., Chowdrey, N., & Chen, T. (2005). Enhanced accuracy of mental state decoding in dysphoric college students. *Cognition & Emotion*, *19*(7), 999–1025. http://doi.org/10.1080/02699930541000110

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York: The Guilford Press.

Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research, 50*, 1-22. doi: 10.1080/00273171.2014.962683

Hernandez, E. M., Trout, Z. M., & Liu, R. T. (2016). Child Abuse & Neglect Vulnerability-specific stress generation : Childhood emotional abuse and the mediating role of depressogenic interpersonal processes. *Child Abuse & Neglect*, *62*, 132–141. http://doi.org/10.1016/j.chiabu.2016.10.019

Joiner, T. E., Alfano, M. S., & Metalsky, G. I. (1992). When depression breeds contempt: Reassurance seeking, self-esteem, and rejection of depressed college students by their roommates. *Journal of Abnormal Psychology, 101*, 165-173.

Joiner, T. E., & Metalsky, G. I. (2001). Excessive reassurance seeking: Delineating a risk factor involved in the development of depressive symptoms. *Psychological Science, 12*, 371-378.

Lee, K., Quinn, P. C., & Pascalis, O. (2017). Face Race Processing and Racial Bias in Early Development : A Perceptual-Social Linkage. *Current Directions in Psychological Science*, *26*, 256–262. http://doi.org/10.1177/0963721417690276

Lee, L., Harkness, K. L., Sabbagh, M. A., & Jacobson, J. A. (2005). Mental state decoding abilities in clinical depression. *Journal of Affective Disorders, 86*, 247-258. doi: 10.1016/j.jad.2005.02.007

Lee, D. T., Kleinman, J., & Kleinman, A. (2007). Rethinking depression: An ethnographic study of the experiences of depression among Chinese. *Harvard Review of Psychiatry, 15*(1), 1-18. doi: 10.1080/10673220601183915

Lemmer, G. & Gollwitzer, M. (2017). The “true” indirect effect won't (always) stand up: When and why reverse mediation testing fails. *Journal of Experimental Social Psychology, 69*, 144-149. doi: 10.1016/j.jesp.2016.05.002

Lerner, M. D., Hutchins, T. L., & Prelock, P. A. (2011). Brief report: Preliminary evaluation of the Theory of Mind Inventory and its relationship to measures of social skills. *Journal of Autism and Developmental Disorders, 41*, 512-517. doi: 10.1007/s10803-010-1066-z

Levin, D. T. (2000). Race as a Visual Feature : Using Visual Search and Perceptual Discrimination Tasks to Understand Face Categories and the Cross-Race Recognition Deficit. *Journal of Experimental Psychology: General*, *129*(4), 559–574.

Levinson, S. C. (1983). *Pragmatics.* Cambridge: Cambridge University Press.

Little, S. A., & Garber, J. (2000). Interpersonal and achievement orientations and specific stressors predicting depressive and aggressive symptoms in children. *Cognitive Therapy and Research, 24,* 651–671.

Little, S. A., & Garber, J. (2004). Interpersonal and achievement orientations and specific stressors predict depressive and aggressive symptoms. *Journal of Adolescent Research, 19,* 63–84.

Liu, D., Meltzoff, A. N., & Wellman, H. M. (2009). Neural correlates of belief- and desire-reasoning. *Child Development*, *80*(4), 1163–71. http://doi.org/10.1111/j.1467-8624.2009.01323.x

MacKinnon, D. P., & Fairchild, A. J. (2009). Current Directions in Mediation Analysis. *Current Directions in Psychological Science, 18,* 16-20.

Mattern, M., Walter, H., Hentze, C., Schramm, E., Drost, S., Schoepf, D.,…, Schnell K. (2015). Behavioral evidence for an impairment of affective theory of mind capabilities in chronic depression. *Psychopathology, 48,* 240-250. doi: 10.1159/000430450.

Mellick, W. & Sharp, C. (2016). Mental state decoding in adolescent boys with major depressive disorder versus sex-matched healthy control. *Psychopathology, 49*, 53-59. doi: 10.1159/000443989

Metalsky, G. I., Joiner, T. E., Jr., Potthoff, J., Pacha, L., Alfano, M. S., & Hardin, T. (1991). *The depressive interpersonal relationships inventory.* Unpublished manuscript.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavior and Brain Science, 27,* 169-190.

Potthoff, J. G., Holahan, C. J., & Joiner, T. E. (1995). Reassurance seeking, stress generation, and depressive symptoms: An integrative model. *Journal of Personality and Social Psychology, 68*, 664-670.

Rude, S. S., & Burnham, B. L. (1995). Connectedness and Neediness: Factors of the DEQ and SAS dependency scales. *Cognitive Therapy and Research, 19,* 323–340.

Sabbagh, M.A., (2004). Understanding orbitofrontal contributions to theory-of-mind reasoning: implications for autism. *Brain and Cognition. 55*, 209–219. doi: 10.1016/j.bandc.2003.04.002

Sabbagh, M. A., & Bowman, L. C. (2017). Theory of Mind. In S. Ghetti (Ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, 4th Ed.* New York: Wiley.

Sabbagh, M. A., Moulson, M. C., & Harkness, K. L. (2004). Neural Correlates of Mental State Decoding in Human Adults : An Event-related Potential Study. *Journal of Cognitive Neuroscience*, 415–426.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology,16*, 235-239

Schenkel, L. S., Chamberlain, T. F., & Towne, T. L. (2014). Impaired Theory of Mind and psychosocial functioning among pediatric patients with Type I versus Type II bipolar disorder. *Psychiatry Research, 215*(3), 740-746. doi: 10.1016/j.psychres.2013.10.025

Shahar, G., Joiner, T. E., Zuroff, D. C., & Blatt, S. J. (2004). Personality, interpersonal behavior, and depression: Co-existence of stress-specific moderating and mediating effects. *Personality and Individual Differences, 36*, 1583-1596. doi: 10.1016/j.paid.2003.06.006

Sharp, C., & Vanwoerden, S. (2015). Hypermentalizing in Borderline Personality Disorder : A Model and Data Hypermentalizing in Borderline Personality Disorder : A Model and Data. *Journal of Infant, Child, and Adolescent Psychotherapy*, *14*(1), 33–45. http://doi.org/10.1080/15289168.2015.1004890

Starr, L. R. & Davila, J. (2008). Excessive ressurance seeking, depression, and interpersonal rejection: A meta-analytic review. *Journal of Abnormal Psychology, 117*(4), 762-775. doi: 10.1037/a0013866

Stewart, J. G., & Harkness, K. L. (2015). The interpersonal toxicity of excessive reassurance-seeking: Evidence from a longitudinal study of romantic relationships. *Journal of Social and Clinical Psychology, 34*(5), 392-410. doi: 10.1521/jscp.2015.34.5.392

Stewart, J. G., & Harkness, K. L. (2017). Testing a revised interpersonal theory of depression with a laboratory measure of excessive reassurance seeking. *Journal of Clinical Psychology, 73,* 331-348.

Stone, V.E., Baron-Cohen, S. & Knight, R.T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10,* 640-656.

Szanto, K., Dombrovsky, A. Y., Sahakian, B. J., Mulsant, B. H., Houck, P. R., Reynolds, C. F., & Clark, L. (2012). Social emotion recognition, social functioning, and attempted suicide in

late-life depression. *American Journal of Geriatric Psychiatry, 20*(3), 257-265. doi: 10.1097/JGP.0b013e31820eea0c.

Thoemmes, F. (2015). Reversing arrows in mediation models does not distinguish plausible models. *Basic and Applied Social Psychology, 37*, 226-234. doi: 10.1080/01973533.2015.1049351

Washburn, D., Wilson, G., Roes, M., Rnic, K., & Harkness, K. L. (2016). Theory of mind in social anxiety disorder, depression, and comorbid conditions. *Journal of Anxiety Disorders, 37*, 71-77. doi: 10.1016/j.janxdis.2015.11.004

Weary, G., & Edwards, J. A. (1994). Social cognition and clinical psychology: Anxiety, depression, and the processing of social information. In Wyer, R. S. & Srull, T. K., (Eds.), *Handbook of social cognition* (pp. 289-338). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13,* 103-128.

World Health Organization (2017, February). *Depression: Fact Sheet.* Retrieved from http://www.who.int/mediacentre/factsheets/fs369/en/

Yirmiya, N., Erel, O., Shaked, M., & Solomonica-Levi, D. (1998). Meta-analyses comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals. *Psychology Bulletin, 124*(3), 283-307.

Table 1

Demographic and Clinical Characteristics by Depression Status.

	Depressed (n = 31)	Non-Depressed (n = 91)	t/χ^2	p
Sex (female), *n* (%)	25 (80.6)	68 (74.7)	0.45	.50
Age, *M* (*SD*)	20.13 (2.38)	20.41 (2.22)	0.59	.56
Ethnicity			1.83	.18
White, *n* (%)	21 (67.7)	47 (52.8)		
Asian, *n* (%)	6 (19.4)	29 (32.6)		
Black, *n* (%)	2 (6.5)	7 (7.9)		
Other, *n* (%)	2 (6.5)	6 (6.7)		
BDI-II, *M* (*SD*)	27.06 (10.20)	7.13 (6.23)	12.84	< .001
Age at first onset, *M* (*SD*)	15.67 (4.45)			
Number of episodes, *M* (*SD*)	2.13 (1.36)			
Comorbid diagnosis (yes), *n* (%)	11 (35.5)			
AD treatment (yes), n (%)	13 (41.9)			

Note: AD = Antidepressant

Table 2.

Means and Standard Deviations of Study Variables by Depression Status.

	Depressed (n = 31)	Non-Depressed (n = 91)	t	p
RMET accuracy	75.18 (9.35)	71.37 (10.26)	1.83	.07
Animals task accuracy	74.19 (14.65)	71.70 (15.11)	0.80	.43
FB accuracy	68.07 (37.28)	75.93 (36.82)	1.03	.31
TB accuracy	92.44 (17.23)	94.65 (13.45)	0.74	.46
ERS	14.23 (8.61)	7.76 (4.35)	5.44	< .001
Interpersonal threat	2.74 (2.34)	1.53 (2.07)	2.73	.001
Non-interpersonal threat	2.22 (1.78)	1.58 (1.63)	1.86	.07

Note: ERS = Excessive Reassurance-Seeking; RMET = Reading the Mind in the Eyes Task; FB = False Belief; TB = True Belief

Table 3.

Zero-Order Relations Among Study Variables.

	BDI-II	ERS	RMET	Animals Task	FB trials	TB trials	Interpersonal Threat
ERS	.52**	—	—	—	—	—	—
RMET	-.03	-.07	—	—	—	—	—
Animals Task	.06	.10	.16[t]	—	—	—	—
FB trials	-.13	.04	.12	.21*	—	—	—
TB trials	-.08	.02	.03	.18*	.31**	—	—
Interpersonal Threat	.25**	.19*	-.003	.06	-.07	-.03	—
Non-interpersonal Threat	.27**	.14	-.09	.04	-.07	-.23*	.12

Note: [t]$p < .01$, *$p < .05$, **$p < .01$; BDI-II = Beck Depression Inventory-II; ERS = Excessive Reassurance-Seeking; RMET = Reading the Mind in the Eyes Task; FB = False Belief; TB = True Belief.

Table 4. *Models for RMET Accuracy*.

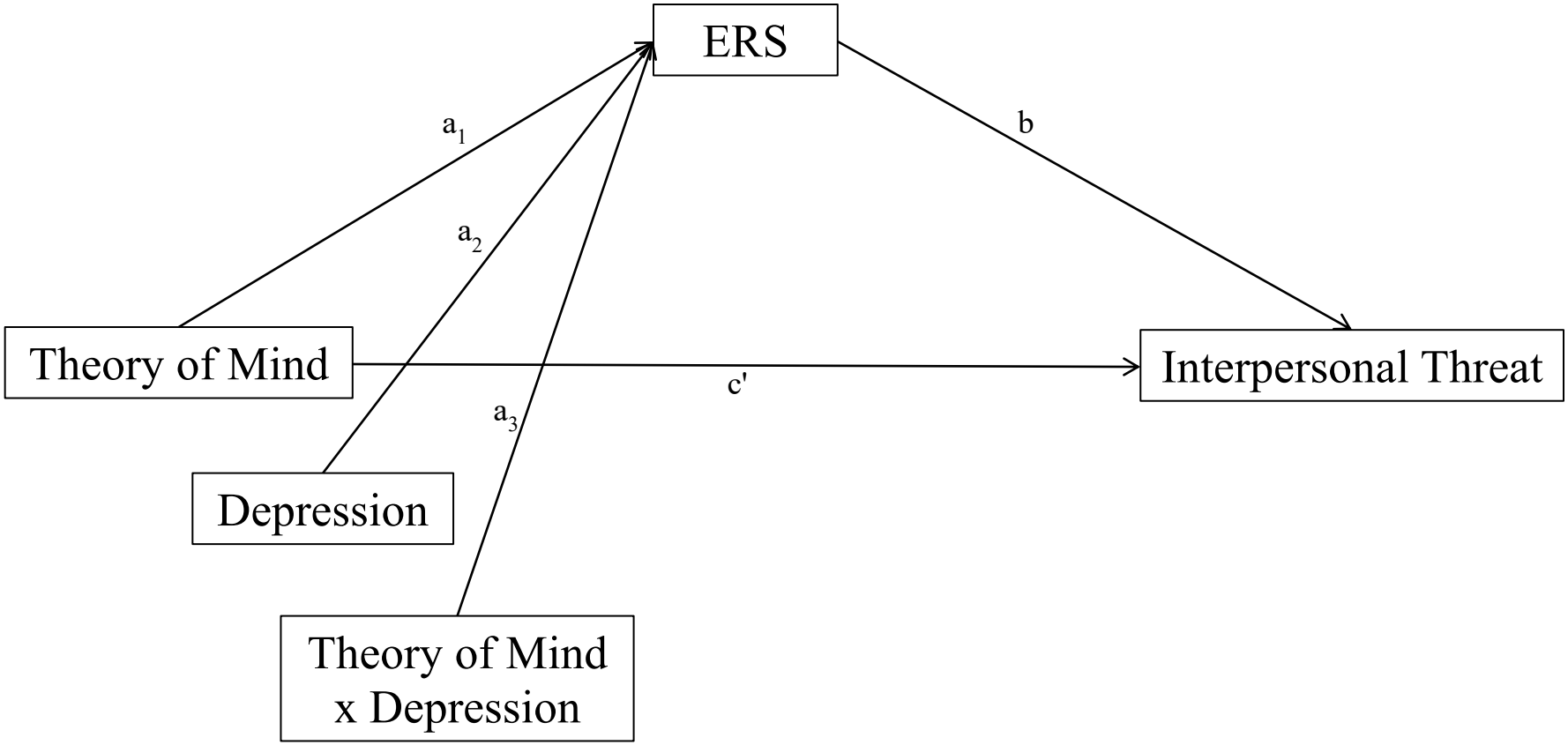
	Paths a_1, a_2, a_3 (Outcome: ERS)			Paths b & c' (Outcome: Interpersonal Threat)			Paths b & c' (Outcome: Non-interpersonal Threat)		
Predictor	*b*	SE	*p*	*b*	SE	*p*	*b*	SE	*p*
RMET	-.04	.09	.69	.004	.09	.97	-.09	.09	.34
Depression	1.16	.19	< .001	—	—	—	—	—	—
RMET x Depression	-.50	.18	.008	—	—	—	—	—	—
ERS	—	—	—	.19	.09	.04	.13	.09	.15
Model R^2	.27	—	< .001	.04	—	.11	.03	—	.20

Note. Paths are depicted on Figure 1. RMET = Reading the Mind in the Eyes Task; ERS = Excessive Reassurance-Seeking

Table 5. *Models for False Belief Accuracy*.

	Paths a_1, a_2, a_3 (Outcome: ERS)			Paths b & c' (Outcome: Interpersonal threat)			Paths b & c' (Outcome: Non-interpersonal threat)		
Predictor	*b*	SE	*p*	*b*	SE	*p*	*b*	SE	*p*
False Belief (FB)	-.05	.09	.61	-.07	.09	.46	.00	.09	.99
Depression	1.06	.19	< .001	—	—	—	—	—	—
FB x Depression	.44	.18	.02	—	—	—	—	—	—
ERS	—	—	—	.19	.09	.03	.14	.09	.13
Model R^2	.24	—	< .001	.04	—	.08	.02	—	.31

Note. Paths are depicted on Figure 1; ERS = Excessive Reassurance-Seeking.



Note: ERS = excessive reassurance seeking

Figure 1. *Schematic Model of the Relation of Theory of Mind, Excessive Reassurance-Seeking, and Interpersonal Stress, as Moderated by Depression.*