

# Theory of mind impairment and its clinical correlates in patients with schizophrenia, major depressive disorder and bipolar disorder

Yan-yu Wang<sup>1,2,3</sup>, Yi Wang<sup>1\*</sup>, Ying-min Zou<sup>1,3</sup>, Ke Ni<sup>1,3,4</sup>, Xue Tian<sup>2</sup>, Hong-wei Sun<sup>2</sup>,  
Simon S. Y. Lui<sup>1,5</sup>, Eric F. C. Cheung<sup>5</sup>, John Suckling<sup>6</sup>, Raymond C. K. Chan<sup>1,3\*</sup>

1. Neuropsychology and Applied Cognitive Neuroscience Laboratory; CAS Key Laboratory of Mental Health, Institute of Psychology, Beijing, 100101, China

2. Department of Psychology, Weifang Medical University, Shandong, 261053, China

3. Department of Psychology, University of Chinese Academy of Sciences, Beijing, 100048, China

4. Mental Health Center of Qiqihar city, Heilongjiang, 161006, China

5. Castle Peak Hospital, Hong Kong Special Administrative Region, China

6. Department of Psychiatry, University of Cambridge, and Cambridgeshire and Peterborough NHS Trust, Cambridge, CB2 0SZ, United Kingdom

\*Correspondence should be addressed to:

Raymond Chan, Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Beijing 100101, China; Tel/Fax: +86(0)10-64836274; Email: [rckchan@psych.ac.cn](mailto:rckchan@psych.ac.cn)

Or

Yi Wang, Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Beijing 100101, China; Tel/Fax: +86(0)10-64877349; Email: [wangyi@psych.ac.cn](mailto:wangyi@psych.ac.cn)

## **Abstract**

**Background:** Although Theory of Mind (ToM) impairment has been observed in patients with a wide range of mental disorders, the similarity and uniqueness of these deficits across diagnostic groups has not been thoroughly investigated.

**Methods:** We recruited 35 participants with schizophrenia (SCZ), 35 with bipolar disorder (BD), 35 with major depressive disorder (MDD), and 35 healthy controls in this study. All participants were matched in age, gender proportion and IQ estimates. The Yoni task, capturing both the cognitive and affective components of ToM at the first- and second-order level was administered. Repeated-measure ANOVA and MANOVA were conducted to compare the group differences in ToM performance. A network was then constructed with ToM performances, psychotic and depressive symptoms, and executive function as nodes exploring the clinical correlates of ToM.

**Results:** Overall, ToM impairments were observed in all patient groups compared with healthy controls, with patients with SCZ performing worse than those with BD. In second-order conditions, patients with SCZ and MDD showed deficits in both cognitive and affective conditions, while patients with BD performed significantly poorer in cognitive conditions. Network analysis showed that second-order affective ToM performance was associated with psychotic and depressive symptoms as well as executive dysfunction, while second-order affective ToM performance and negative symptoms showed relatively high centrality in the network.

**Conclusions:** Patients with SCZ, MDD and BD show different variety and severity of impairments in ToM sub-components. Impairment in higher-order affective ToM appears to be closely related to clinical symptoms in both psychotic and affective disorders.

**Key words:** Theory of Mind, cross-diagnostic, schizophrenia, major depressive disorder, bipolar disorder, network analysis

**Word count:**

Abstract: 250

Text: 5027

Figure: 3

Table: 2

Supplementary: 1

## 1. Introduction

Theory of mind (ToM), one of the key aspects of social cognition, refers to the ability to infer others' aims, beliefs, intentions and emotional states (Premack and Woodruff, 1978; Schaafsma et al., 2015). ToM impairment has been reported in patients with mental disorders, such as schizophrenia (SCZ), bipolar disorder (BD) and major depressive disorder (MDD). In patients with SCZ, ToM impairments have been reported in both first-episode (Bora and Pantelis, 2013; Ho et al., 2015) and chronic phases (Bora et al., 2009a). Moreover, recent studies have also found these impairments in non-psychotic first-degree relatives of patients with SCZ and individuals with ultra-high risk for psychosis (Bora and Pantelis, 2013), suggesting that ToM impairments may be a trait marker of SCZ. Furthermore, meta-analytic studies have also found a large effect size for deficits in ToM performance in patients with BD, especially in acute episodes ( $d = 1.23$ ) (Bora et al., 2016a) and a moderate ToM deficit in patients with MDD ( $d = 0.51-0.58$ ) (Bora and Berk, 2016).

Although many previous studies on ToM have been carried out in each of the diagnostic group mentioned above, it is still not clear whether patients with SCZ, BD and MDD share similar or distinct ToM impairments. Of the few studies directly comparing ToM performance across diagnostic groups, most were limited to a comparison between two groups. For example, Purcell et al. (2013) found comparable ToM impairments between remitted BD and MDD patients, while a number of other studies did not find significant ToM impairment in MDD patients (Bazin et al., 2009; Berecz et al., 2016; Mellick and Sharp, 2016; Wilbertz et al., 2010),

even a better ToM performance was found in depressed adolescent boys (Mellick and Sharp, 2016) and woman with a maternal history of depression (Harkness et al., 2011). Similarly, previous studies comparing ToM performance in SCZ with BD patients suggested that SCZ patients might be more impaired in ToM than BD patients (Caletti et al., 2013; Guastella et al., 2013; Lahera et al., 2015) while some studies reported similar extent of ToM impairment in these two groups of patients (Donohoe et al., 2012; Thaler et al., 2013) .

One possible explanation for the inconsistent findings may be the use of different tasks for ToM assessment. Recent studies have suggested that there are dissociable cognitive and affective components of ToM, with the cognitive component of ToM focusing on the inferences of others' beliefs, while the affective ToM refers to the inferences of others' feelings (Shamay-Tsoory and Aharon-Peretz, 2007; Shamay-Tsoory et al., 2007). Shamay-Tsoory *et al.* (2007) found that first-episode patients with SCZ performed significantly worse in affective ToM, but not cognitive ToM. Donohoe et al. (2012) found that BD patients presented comparable deficits to SCZ patients on the "Reading the Mind in the Eyes Task" (an affective ToM task), but only subtle deficits compared with SCZ patients on the "Hinting Task" (a cognitive ToM task). These findings suggest that patients with SCZ and BD may be similarly impaired in affective ToM while patients with BD may be less impaired in cognitive ToM.

ToM impairment was widely found in first-episode, chronic patients, even individuals at ultra-high risk for psychosis and first-degree relatives of SCZ and BD

(Bora, 2009; Bora and Pantelis, 2012, 2013; Bora et al., 2009a), thus considered to be trait related. However, there are still some other studies which found significant relationships between ToM impairment with clinical symptoms and general cognitive functions (Bora et al., 2016a; Bora et al., 2009b). Mitchell and Young (2015) think that the relationship between ToM impairments with positive psychotic symptoms may not be disease-specific, but symptom-specific. Marjoram et al. (2005) examined ToM abilities in patients with SCZ and affective disorders and found that hallucinations and delusions were associated with ToM impairment, regardless of diagnosis. Bora et al. (2009b) reviewed ToM studies in SCZ spectrum disorders, affective psychosis and BD and suggested that the ToM impairment may reflect executive dysfunction and residual symptom expression rather than a specific trait-marker.

It is well known that comorbid clinical symptoms exist in different mental disorders, and about 45% of people who meet diagnostic criteria of one mental disorder also receive another diagnosis (Kessler et al., 2005). It has also been reported that 18.6% of patients with MDD experience hallucinations and/or delusions (Ohayon and Schatzberg, 2002), while negative symptoms may also be stable trait like features of illness that persist into periods where BD patients are free from mood episodes (Strauss et al., 2016). In addition, the association between positive and negative psychotic symptoms, affective symptoms and cognitive symptoms is complex. According to the framework of Research Domain Criteria (RDoC) (Insel et al., 2010), trans-diagnostic study designs can improve the

identification of shared and unique biological mechanisms, thus promoting the development of improved prevention and treatment strategies. Although previous studies partly revealed the relationship between clinical symptoms and cognitive functions with ToM impairments in different psychoses using meta-analyses respectively (Bora et al., 2016a; Bora and Berk, 2016; Bora et al., 2009b), traditional correlation analysis is unable to take into account all the symptom dimensions in examining the correlations between ToM performance and symptomatology. Recently, researchers have proposed a network perspective in understanding the psychopathology of mental disorders (Borsboom, 2017; Cramer et al., 2010). The network analysis approach, which considers the disorders in question as a system of casually connected symptoms rather than latent variables, has recently been applied to studies in psychiatry and psychology to examine the relationships between symptoms (Borsboom and Cramer, 2013). Using this approach, a network based on partial correlations between symptoms and/or clinical manifestations is constructed to assess the independent associations controlling other variables.

In this study, we used the Yoni Task, an instrument specifically designed to capture the cognitive and affective components of ToM, to compare the ToM performance of well-matched patients with SCZ, BD and MDD. We then adopted network analysis to examine the independent associations between clinical symptoms, ToM performance and executive function in these patients. We hypothesized that patients with SCZ, BD and MDD would exhibit significant ToM impairments compared with healthy controls; and among the three patient groups, patients with

SCZ would show greater ToM impairment. We also hypothesized that ToM performance in these patients would be correlated with psychotic symptoms, depression and executive function regardless of diagnosis.

## **2. Method**

### **2.1 Participants**

According to effect sizes of previous studies about the comparisons in ToM between psychosis patients with health controls (Bora, 2009; Bora et al., 2016a; Bora and Berk, 2016), the sample size was calculated with G\*power 3.1 software (Faul, 2007) and a reasonable number of each group was 36 (effect size = 0.60 ). Based on this estimation, 35 participants in each group who were well matched in age ( $F=1.860$ ,  $p=0.139$ ), gender proportion ( $\chi^2=0.320$ ,  $p=0.956$ ) and IQ estimates ( $F=0.659$ ,  $p=0.578$ ) were recruited. Patients groups were recruited from either the Mental Health Center of Qiqihar city or the Mental Health Center of Weifang City. All of them fulfilled the diagnostic criteria of DSM-IV (APA, 1994) for SCZ, MDD, and BD respectively and in a relative stable phase under treatment. Diagnosis were ascertained by the Mini-International Neuropsychiatric Interview (MINI) version 6.0.0 (Sheehan et al., 2009). Exclusion criteria of each diagnostic group were: a co-morbid DSM-IV Axis I disorder; a history of head injury or neurological disorder; a lifetime history of alcohol and substance abuse; severe hearing or visual impairment; an IQ<80 measured by the Chinese version of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) (Gong, 1992). Healthy individuals were recruited from the



community as controls, who were screened by a qualified psychiatrist using the MINI to ascertain the absence of lifetime or family history of mental disorders. Exclusion criteria included a history of substance abuse in the past 12 months, a history of electroconvulsive therapy in the past six months, a history of neurological disorders, a history of head injury with loss of consciousness for more than 30 minutes, and learning difficulties.

This study was approved by the Ethics Committees of the Institute of Psychology, Chinese Academy of Sciences, the Mental Health Center of Qiqihar city, and the Mental Health Center of Weifang. All participants gave written informed consent.

## **2.2 Assessment**

### **2.2.1 Assessment of ToM**

The Chinese version of the computerized Yoni Task (Ho et al., 2015; Shamay-Tsoory et al., 2007) was used to assess affective and cognitive ToM. In total, there are ninety-eight trials, including 32 for first-order ToM and 66 for second-order ToM assessment in the task. During each trial, a cartoon face of “Yoni” (Chinese name Xiaoming) was shown in the middle and four objects belonging to a single semantic category or faces were shown in the four corners. Participants were required to infer which one Xiaoming was referring to according to its eye gaze and facial expression. According to the content of inferences, there are three conditions: cognitive, affective and physical condition. In the first-order cognitive and affective conditions, participants were required to infer which of the four subjects or faces

represented what Xiaoming was thinking or feeling. In the second-order cognitive and affective conditions, participants were required to infer which one Xiaoming was thinking or feeling of another character. Physical conditions served as a control condition. The mean accuracy of ToM performance of each condition (first order cognitive; second order cognitive; first order affective; second order affective; first order physical; second order physical) was calculated. An illustration of the Yoni task can be found as Supplementary Figure S1.

### **2.2.2 Executive function**

Executive function was assessed using the Chinese version of the Letter–Number Span Test (Chan et al., 2008), which requires participants to update and re-arrange the correct order of spoken words on the “Celestial Trunks and Number”. For example, the correct order of “3 Bing Jia 2” should be given as “2 3 Jia Bing.” The test was terminated when the participant failed all four trials at any length of string.

### **2.2.3 Clinical assessments**

The severity of clinical symptoms of patients with SCZ was measured using the Positive and Negative Syndrome Scale (PANSS) (Kay et al., 1987). The severity of clinical symptoms of patients with BD was measured with the PANSS, the 24-item Hamilton Rating Scale for Depression (HRSD-24) (Hamilton, 1960, 1967) and Young manic rating scale (YMRS) (Young et al., 1978) . The severity of depression of patients with major depressive disorder was also measured with the HRSD-24.

## **2.3 Data analysis**

Statistical Package for the Social Sciences (SPSS) version 17.0 was used for statistical analysis. Descriptive analyses were conducted for socio-demographic characteristics, including age, gender, years of education, duration of illness and performances on both the Yoni task and the various rating scales. Group differences in gender proportion were compared using the Chi-square test. ANOVA and post-hoc tests were conducted to compare the age, years of education, and estimated IQ between the four groups. Independent-samples *t* tests were used to compare the severity of clinical symptoms (the PANSS for SCZ vs BD; the HRSD for MDD vs BD).

For performance on the Yoni Task, we examined the main effect of group on the overall performance using repeated measures ANOVA, with group (SCZ, MDD, BD, HC) as the between-subject factor, order (first and second) and condition (cognitive, affective) of the Yoni Task as within-subject factors. The main effects of group and order/condition of Yoni task, together with the interactions between group and Yoni order/condition were also examined. Pairwise comparisons were conducted when a significant main effect of group was detected and simple effect tests were conducted when interaction effects were significant.

To further examine the differences between the four groups in each condition under different order of the Yoni Task, whilst covarying for ToM performance in the physical condition, MANOVA was conducted with group as the between-subject factor. Subsequently, post-hoc pairwise comparisons were undertaken when a significant main effect of group was detected to observe the direction of effects.

Significance level was set at  $p < 0.05$  and all of the above statistical significance in comparisons were corrected by Bonferroni method.

The Qgraph package (<https://CRAN.R-project.org/package=qgraph>) implemented in R statistical software (version 3.3.2, <https://www.r-project.org/>) was used for network construction, analysis and visualization.

To examine the independent relationships between clinical symptoms, executive function and ToM performance, we constructed networks for analysis by defining nodes and edges as follows: The nodes were variables including performance on the Yoni Task (first order cognitive/affective, second order cognitive/affective), clinical symptoms (PANSS and/or HRSD) and performance on executive function test; The edges were defined as the partial correlation coefficients between each pair of nodes after controlling for all the other variables. An undirected network was then constructed using a Gaussian Graphical Model (GGM) (Lauritzen and Wermuth, 1989). Graphical least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996) regularization with tuning parameter  $\lambda = 0.5$  (Beard et al., 2016; Foygel R, 2010; Isvoranu et al., 2017) was applied to shrink the regression coefficients and set small ones to zero to obtain a stable and interpretable network. The importance of each node in the networks was further investigated by examining nodal strength, betweenness and closeness.

To examine the associations between psychotic symptoms, depressive symptoms, executive function and ToM performance in a relatively large and heterogeneous

sample, we combined all patients into one group and conducted network construction and analyses. For this combined group, the network was constructed with the following variables as nodes: PANSS total and subscale scores, HRSD-24 total score, accuracy on first order cognitive/affective and second order cognitive/affective of the Yoni Task (ToM performance) and the total number of items passed on the Letter-Number Span Test (executive function). The betweenness, closeness and strength of each node were calculated based on the partial correlation coefficients to assess the centrality of each variable in the network. We repeated the network construction and analysis for each patient group, as well as the SCZ & BD combined group, the MDD & BD combined group. These results can be found in the Supplementary Materials.

### **3. Results**

#### ***3.1 Demographic and clinical variables***

Demographics and clinical characteristics of the participants are presented in Table 1. There were no significant differences in gender proportion, age, handedness and estimated IQ between the four groups ( $p > 0.05$ ). The SCZ group had a shorter duration of illness (post-hoc  $p = 0.006$ ), higher levels of positive symptoms ( $t = 2.456, p = 0.017$ ), negative symptoms ( $t = 5.512, p < 0.001$ ) and general psychopathological symptoms ( $t = 3.717, p < 0.001$ ) on the PANSS than the BD group. The MDD group had higher levels of depressive symptoms than the BD group ( $t =$

2.014,  $p = 0.048$ ). Significant differences in executive function performance were found between the four groups ( $F = 16.407$ ,  $p < 0.001$ ). Patients with SCZ and BD performed significantly worse than patients with MDD (post-hoc  $ps < 0.05$ ).

*Insert Table 1 About Here*

### **3.2 Overall performance on the Yoni Task**

Repeated measures ANOVA revealed significant main effects for Group ( $F = 5.83$ ,  $p = 0.001$ ,  $\eta^2 = 0.114$ ), Order ( $F = 76.56$ ,  $p < 0.001$ ,  $\eta^2 = 0.360$ ) and Condition ( $F = 5.31$ ,  $p = 0.023$ ,  $\eta^2 = 0.038$ ) for the Yoni Task. Post-hoc tests showed that all the patient groups showed significant impairment in their overall performance on the Yoni Task, with SCZ patients performing worse than patients with BD ( $p = 0.042$ ) (see Figure 1). The Order x Condition ( $F = 6.20$ ,  $p = 0.014$ ,  $\eta^2 = 0.044$ ) and Order x Group ( $F = 2.68$ ,  $p = 0.049$ ,  $\eta^2 = 0.056$ ) interactions were also significant. Further simple effect tests revealed that all three patient groups only showed significant impairments in the second-order performances of Yoni task compared to the healthy controls ( $p < 0.001$  for SCZ and MDD groups,  $p = 0.015$  for BD group; after Bonferroni correction), but not in the first-order tasks ( $p > 0.05$  for SCZ and MDD groups,  $p = 0.058$  for BD groups; after Bonferroni correction). The comparisons between four groups under different orders and conditions showed that patients with SCZ and MDD both showed significant impairments in second-order cognitive and affective ToM ( $ps < 0.01$ ), patients with BD showed significant impairment in second-order cognitive ToM ( $p = 0.011$ ) and marginal significance in affective ToM ( $p = 0.057$ ), while only SCZ patients

showed a marginal significant impairment in first-order cognitive ToM compared to healthy controls ( $p = 0.055$ ). (see Table 2)

*Insert Figure 1 About Here*

### **3.3 Group comparisons for each Yoni Task condition**

To further examine the differences between the four groups in each condition under different order of the Yoni Task, whilst covarying for ToM performance in the physical condition, we conducted MANOVA with Group as the between-subject factor.

For the first-order conditions, the results showed non-significant differences between the four groups in ToM performance in both the cognitive ( $F = 2.117, p = 0.101, \eta^2 = 0.045$ ) and the affective ( $F = 1.892, p = 0.134, \eta^2 = 0.040$ ) conditions.

For the second-order conditions, significant Group effects were found in the cognitive condition ( $F = 7.098, p < 0.001, \eta^2 = 0.136$ ) and the affective condition ( $F = 5.098, p = 0.002, \eta^2 = 0.102$ ). Pairwise comparisons with Bonferroni correction showed that patients with SCZ ( $p < 0.001$ ) and MDD ( $p = 0.005$ ) performed significantly worse than healthy controls in the second-order cognitive condition. In the second-order affective condition, only patients with SCZ performed significantly worse than healthy controls ( $p = 0.001$ ). There were no significant differences between the three patient groups in ToM performance. (see Table 2 & Figure 2)

*Insert Table 2 & Figure 2 About Here*

### ***3.4 Associations between ToM abilities, clinical symptoms and executive function***

In the combined sample of SCZ, MDD and BD patients, the network constructed using PANSS, Depression, ToM performance and EF with and without LASSO regularization is shown in Figure 3. With LASSO regularization, performance in first order affective ToM conditions (Y1A) was positively correlated with executive function; performance in first order cognitive ToM conditions (Y1C) was positively correlated with PANSS positive and negative symptoms; and performance in second order affective ToM conditions (Y2A) was negatively correlated with PANSS positive, negative and general symptoms, as well as depressive symptoms, and positively correlated with executive function (see Figure 3B). Regarding the centrality measures of the nodes, we found that second order affective ToM (Y2A) had the highest closeness, suggesting the shortest average distance between Y2A and other nodes in the network. First order cognitive ToM (Y1C) had the greatest strength of connections, while negative symptoms had the highest betweenness, which measures how often a node lies on the shortest path of the network (see Figure 3C).

*Insert Figure 3 About Here*

## **4. Discussion**

The present study examined ToM impairment and its correlation with clinical symptoms in patients with SCZ, MDD and BD. Our main findings showed that ToM impairment was found in patients with SCZ, BD and MDD compared with healthy



controls. Additionally, patients with SCZ performed significantly worse than BD patients. Subsequent analysis showed that patients with SCZ and MDD showed impairment in both cognitive and affective components, while patients with BD only showed impairment in the cognitive component of the second-order ToM. Network analyses showed that second-order affective ToM performance was associated with psychotic and depressive symptoms, as well as executive dysfunction. Performance on second-order affective ToM and first-order cognitive ToM and negative symptoms showed relatively high centrality in the network we constructed between clinical symptoms, ToM performance and executive function.

#### ***4.1 ToM impairments across diagnostic groups***

Across the diagnostic groups, we also observed worse performance in the SCZ group than the BD group. This is consistent with most previous studies that directly compared the ToM ability of SCZ and BD patients, including those using different tasks, such as the Reading the Mind in the Eyes Task (Donohoe et al., 2012; Guastella et al., 2013), the Faux Pas Test (Donohoe et al., 2012), the Hinting Task (Lahera et al., 2015), and the Awareness of Social Inference Test (Rowland et al., 2012). In a recent meta-analysis, Bora and Pantelis (2016) examined social cognition deficits between patients with SCZ and BD, and found that ToM performance was worse in SCZ than BD patients ( $d = 0.57$ ).

As for the different conditions of the Yoni Task, we found a significant group effect on performance only in the second-order conditions. Shamay-Tsoory et al.

(2007) and Ho et al. (2015) examined 22 SCZ patients and 41 first-episode SCZ patients respectively, and in both studies significant ToM impairments were also only found in second-order conditions. In affective disorders, previous studies have also reported normal performance in first-order ToM, but impaired performance in second-order ToM compared with healthy controls (Cusi et al., 2013; Doody et al., 1998). Second-order ToM could be considered a more complex undertaking than first-order ToM, as participants are required to infer the beliefs/emotions of a third character. Studies in developmental psychology have suggested that first-order ToM ability first becomes evident at the age of four, while second-order ToM ability only emerges at the age of eight. More importantly, higher-order recursive thinking undergoes continuous development in adolescence and early adulthood and may play important roles in social interactions (Valle et al., 2015). Given that psychotic disorders and BD commonly emerge during adolescence and early adulthood (McGorry et al., 2011), coinciding with the development of higher-order ToM ability, it is not surprising that there is significant ToM impairment in patients with these disorders.

Regarding the cognitive and affective components of ToM, we observed impairments in both cognitive and affective ToM in SCZ patients after controlling for ToM performance in the physical conditions. In a previous study by Shamay-Tsoory et al. (2007), SCZ patients showed significant impairment only in affective conditions but not cognitive conditions, and the authors suggested a dissociation of cognitive and affective components of ToM in SCZ. However, in the present study, our findings

do not support this dissociation as SCZ patients showed impairment in both components, which is consistent with previous studies using the same task in first-episode and chronic SCZ patients (Ho et al., 2015; Zhang et al., 2016). Previous meta-analytic studies in MDD and BD patients have reported deficits in both cognitive ( $d = 0.49$  for MDD;  $0.68$  for BD) and affective ( $d = 0.52$  for MDD;  $0.46$  for BD) ToM (Bora et al., 2016a; Bora and Berk, 2016). In the present study, we found impairments in both cognitive and affective ToM in MDD and impairment only in cognitive ToM in BD patients. However, after controlling for accuracy of the physical conditions, impairment in cognitive ToM in the MDD group remained significant, whereas no significant differences were found in the BD group. Many studies using the “Read the Mind in the Eyes Test” have indicated that only severely depressed patients with MDD exhibit impaired ToM, whereas mild/moderately depressed MDD patients only exhibit ToM impairment approaching trend significance (Berecz et al., 2016; Lee et al., 2005). In the present study, the mean HRSD-24 score in patients with MDD was 26.20 ( $SD=11.91$ ), which indicates moderate depressive symptoms. This might explain the non-significant impairment in affective ToM in MDD patients in this study.

For BD patients, both positive and negative findings have been reported (Lahera et al., 2015; Lee et al., 2013; Purcell et al., 2013; Rowland et al., 2012). In a meta-analysis including 34 studies (Bora et al., 2016a), four of five studies including acute episode BD patients showed pronounced ToM impairment, six of eight studies with subsyndromal patients showed some level of ToM impairment, and about half

of 10 studies including remitted patients did not show any significant impairment in ToM. This suggests that the severity of BD symptoms may contribute to ToM impairment. In the present study, all patients were in relative stable phase under treatment. The relatively intact ToM ability of patients with BD in our study suggested that the social cognition profiles of BD patients may be state-related. Furthermore, the significant difference between BD patients and healthy controls disappeared after controlling the effect of physical conditions. One possible reason may be due to the impairments in general cognition such as attention in patients with BD, which led to the significant differences in physical condition and second-order cognitive ToM. The other may more possibly suggest an overlap between executive dysfunction and social cognition impairment in BD. According to a cross-diagnostic latent class analysis by Bora et al. (2016b), patients with BD were over-represented in the “neuropsychologically normal” cluster, which means that a larger proportion of high functioning patients with BD had normal cognitive function. As such, only a small proportion of BD patients with both executive dysfunction and impairment in ToM are characterized by severe deficits, which particularly impair premorbid social function in the context of neurodevelopmental abnormalities (Bora et al., 2016b).

Taken together, our findings suggest that all the patient groups exhibited significantly poorer ToM performances, especially in high order ToM ability, which indicates that ToM performance has the sensitivity to differentiate patients from controls. At the same time, patients with SCZ may present stable deficits on both cognitive and affective ToM deficits, while ToM performance in MDD and BD may

change according to their illness state and severity of clinical symptoms. Thus, considering the severity and persistence of ToM impairment offers specificity for both differential diagnosis and as a state marker of illness.

#### ***4.2 Clinical correlates and network analysis***

Previous studies have found that psychotic symptoms of patients with BD may be related to ToM impairment, suggesting that they may not be specific to SCZ, but may represent a general deficit related to psychotic symptoms (Bora et al., 2016a). Similarly, the severity of depressive symptoms has also been found to be correlated with ToM impairments (Bora and Berk, 2016).

Instead of assuming a latent variable or “disorder” causing the observable manifestations, the relationships between symptoms irrespective of diagnostic entities may be an alternative approach in explaining the progress of a disorder (Borsboom and Cramer, 2013). Exploring the associations between symptoms may shed light on the aetiology of a disorder as well as treatment strategies targeting specific connections between symptoms. In the present study, we analyzed the direct associations between ToM performance and depression symptoms, positive and negative psychotic symptoms, and executive functions using a network approach. We found that impairments in second-order affective ToM was associated with higher levels of depressive and psychotic symptoms, and were correlated with poorer executive function. Performance in second-order affective ToM conditions also showed the highest closeness, which indicates the short distance with other

nodes in the network. These results were consistent with previous meta-analysis results conducted by Bora et al. in patients with SCZ, MDD and BD (Bora et al., 2016a; Bora and Berk, 2016; Bora et al., 2009a) and our hypothesis that ToM performance was related to both depressive and psychotic symptoms. Furthermore, according to the idea of Bora et al. (2009a), in acute phase of the illness, the impact of cognitive deficits on ToM may be masked by the relationship between symptoms and mentalising ability, while the ToM impairment in symptom free patients could be explained by working memory. The opinion has not been tested by empirical study partly because of the limitation of traditional correlation analysis method. Our study confirmed this opinion with network analysis which revealed the relationship between ToM performance with executive function independently excluding the effect of symptoms on ToM. In addition, the higher centrality and closeness of second-order affective ToM with symptoms also suggests the sensibility of higher-order affective ToM condition which might be a possible candidate for the assessment of early deficits and treatment effects in social cognition in patients with psychosis.

In this study, negative symptoms showed the greatest “betweenness” within the network we constructed. A high betweenness indicates that a node frequently lies on the shortest paths across the networks suggesting a mediating effect of negative symptoms in the constellation of symptoms experienced by the patient group, an interpretation reinforced by the observation that negative symptoms were correlated with positive symptoms, general symptoms, and depressive symptoms.

Lincoln et al. (2011) reported an association between ToM impairments and negative symptoms and suggested a potential perspective for the intervention of negative symptoms in SCZ patients. For example, the Social Cognition and Interaction Training that was specifically designed to target social cognition has been used to intervene in negative symptoms in SCZ patients (Roberts et al., 2014). Based on the results of our study, more focus intervention strategies such as programs aiming to improve the ability of higher-order ToM may be designed and put into effect on the rehabilitation of social function in psychosis.

### ***4.3 Limitations***

Our study has several limitations. Firstly, we acknowledge the time varying nature of experiencing symptoms, especially in BD patients, and the results and conclusions based on the samples in our study were derived from a cross-sectionally designed experiment. Longitudinal studies would be helpful to illustrate the ToM changes that might occur as symptoms change. Secondly, not all the clinical symptom assessments were measured in all patient groups, and some clinical assessment such as YMRS was only administered to patients with BD. As a result, the associations between symptoms and ToM performance in the merged group was mainly driven by a subset of patients. Thirdly, although the duration of illness among our participants was relatively short, we were unable to avoid the effects of medication and hospitalization on social cognitive function. Lastly, the small sample size in this study may be insufficient to fully capitalize on the opportunities of network analysis. However, it should be noted that the present study only took the advantage of

network analysis to explore the associations of ToM performances with clinical symptoms. Future studies should consider recruiting a large sample of trans-diagnostic patients to validate and extend the present results if we would really like to adopt the network analysis (Guloksuz et al., 2017) to address this kind of trans-diagnostic question.

#### ***4.4 Conclusions***

In this study, significant ToM impairments were found in patients with SCZ, BD and MDD, especially in the higher-order ToM conditions. The different ToM performance across diagnostic groups and the association between ToM performance and clinical symptoms suggests that severity and persistence of the ToM impairments might offer information in addition to differential diagnosis.



## References

- American Psychiatric Association (APA), 1994. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV. American Psychiatric Association.
- Bazin, N., Brunet-Gouet, E., Bourdet, C., Kayser, N., Falissard, B., Hardy-Bayle, M.-C., Passerieux, C., 2009. Quantitative assessment of attribution of intentions to others in schizophrenia using an ecological video-based task: A comparison with manic and depressed patients. *Psychiatry Res* 167(1-2), 28-35.
- Beard, C., Millner, A.J., Forgeard, M.J.C., Fried, E.I., Hsu, K.J., Treadway, M.T., Leonard, C.V., Kertz, S.J., Bjoegvinsson, T., 2016. Network analysis of depression and anxiety symptom relationships in a psychiatric sample. *Psychol Med* 46(16), 3359-3369.
- Berecz, H., Tenyi, T., Herold, R., 2016. Theory of Mind in Depressive Disorders: A Review of the Literature. *Psychopathology* 49(3), 125-134.
- Bora, E., 2009. Theory of Mind in Schizophrenia Spectrum Disorders. *Turk Psikiyatri Dergisi* 20(3), 269-281.
- Bora, E., Bartholomeusz, C., Pantelis, C., 2016a. Meta-analysis of Theory of Mind (ToM) impairment in bipolar disorder. *Psychol Med* 46(2), 253-264.
- Bora, E., Berk, M., 2016. Theory of mind in major depressive disorder: A meta-analysis. *J Affect Disord* 191, 49-55.
- Bora, E., Pantelis, C., 2012. Theory of mind impairment at risk conditions to psychosis and in first-degree relatives of schizophrenia: systematic review and meta-analysis. *Schizophr Res* 136, S142-S142.
- Bora, E., Pantelis, C., 2013. Theory of mind impairments in first-episode psychosis, individuals at ultra-high risk for psychosis and in first-degree relatives of schizophrenia: Systematic review and meta-analysis. *Schizophr Res* 144(1-3), 31-36.
- Bora, E., Pantelis, C., 2016. Social cognition in schizophrenia in comparison to bipolar disorder: A meta-analysis. *Schizophr Res* 175(1-3), 72-78.
- Bora, E., Veznedaroglu, B., Vahip, S., 2016b. Theory of mind and executive functions in schizophrenia and bipolar disorder: A cross-diagnostic latent class analysis for identification of neuropsychological subtypes. *Schizophr Res* 176(2-3), 500-505.
- Bora, E., Yucel, M., Pantelis, C., 2009a. Theory of mind impairment in schizophrenia: Meta-analysis. *Schizophr Res* 109(1-3), 1-9.
- Bora, E., Yucel, M., Pantelis, C., 2009b. Theory of mind impairment: a distinct trait-marker for schizophrenia spectrum disorders and bipolar disorder? *Acta Psychiatr Scand* 120(4), 253-264.
- Borsboom, D., 2017. A network theory of mental disorders. *World Psychiatry* 16(1), 5-13.
- Borsboom, D., Cramer, A.O.J., 2013. Network Analysis: An Integrative Approach to the Structure of Psychopathology, in: NolenHoeksema, S. (Ed.), *Annual Review of Clinical Psychology*, Vol 9, pp. 91-121.
- Caletti, E., Paoli, R.A., Fiorentini, A., Cigliobianco, M., Zugno, E., Serati, M., Orsenigo, G., Grillo, P., Consonni, D., Zago, S., 2013. Neuropsychology, social cognition and global functioning among bipolar, schizophrenic patients and healthy controls: preliminary data. *Front. Hum. Neurosci.* 7, 661.
- Chan, R.C.K., Wang, Y., Deng, Y., Zhang, Y., Yiao, X., Zhang, C., 2008. The development of a Chinese equivalence version of Letter-Number Span Test. *Clin. Neuropsychol.* 22(1), 112-121.
- Cramer, A.O.J., Waldorp, L.J., van der Maas, H.L.J., Borsboom, D., 2010. Comorbidity: A network

- perspective. *Behav. Brain Sci.* 33(2-3), 137-+.
- 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 Res* 210(2), 672-674.
- Donohoe, G., Duignan, A., Hargreaves, A., Morris, D.W., Rose, E., Robertson, D., Cummings, E., Moore, S., Gill, M., Corvin, A., 2012. Social cognition in bipolar disorder versus schizophrenia: comparability in mental state decoding deficits. *Bipolar Disorders* 14(7), 743-748.
- Doody, G.A., Gotz, M., Johnstone, E.C., Frith, C.D., Owens, D.G.C., 1998. Theory of mind and psychoses. *Psychol Med* 28(2), 397-405.
- Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A., 2007. G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 39(2), 175-191.
- Foygel R, D.M., 2010. Extended Bayesian information criteria for Gaussian graphical models. *Adv. Neural Inf. Process. Syst.*, 604-612.
- Gong, Y., 1992. Wechsler adult intelligence scale-revised in China Version. Hunan Medical College, Changsha, Hunan/China.
- Guastella, A.J., Hermens, D.F., Van Zwieten, A., Naismith, S.L., Lee, R.S.C., Cacciotti-Saija, C., Scott, E.M., Hickie, I.B., 2013. Social cognitive performance as a marker of positive psychotic symptoms in young people seeking help for mental health problems. *Schizophr Res* 149(1-3), 77-82.
- Guloksuz, S., Pries, L., van Os, J., 2017. Application of network methods for understanding mental disorders: pitfalls and promise. *Psychol Med*, 1-10.
- Hamilton, M., 1960. A rating scale for depression. *J. Neurol. Neurosurg. Psychiatry* 23(1), 56-62.
- Hamilton, M., 1967. Development of a rating scale for primary depressive illness. *Br. J. Soc. Clin. Psychol.* 6(4), 278-296.
- Harkness, K.L., Washburn, D., Theriault, J.E., Lee, L., Sabbagh, M.A., 2011. Maternal history of depression is associated with enhanced theory of mind in depressed and nondepressed adult women. *Psychiatry Res* 189(1), 91-96.
- Ho, K.K.Y., Lui, S.S.Y., Hung, K.S.Y., Wang, Y., Li, Z., Cheung, E.F.C., Chan, R.C.K., 2015. Theory of mind impairments in patients with first-episode schizophrenia and their unaffected siblings. *Schizophr Res* 166(1-3), 1-8.
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D.S., Quinn, K., Sanislow, C., Wang, P., 2010. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am. J. Psychiat.* 167(7), 748-751.
- Isvoranu, A.-M., van Borkulo, C.D., Boyette, L.-L., Wigman, J.T.W., Vinkers, C.H., Borsboom, D., Grp, I., 2017. A Network Approach to Psychosis: Pathways Between Childhood Trauma and Psychotic Symptoms. *Schizophr. Bull.* 43(1), 187-196.
- Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr. Bull.* 13(2), 261-276.
- Kessler, R.C., Chiu, W.T., Demler, O., Walters, E.E., 2005. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch. Gen. Psychiatry* 62(6), 617-627.
- Lahera, G., Herrera, S., Reinares, M., Benito, A., Rullas, M., Gonzalez-Cases, J., Vieta, E., 2015. Hostile attributions in bipolar disorder and schizophrenia contribute to poor social functioning. *Acta Psychiatr Scand* 131(6), 472-482.
- Lauritzen, S.L., Wermuth, N., 1989. Graphical models for associations between variables, some of which are qualitative and some quantitative. *The annals of Statistics*, 31-57.

- Lee, J., Altshuler, L., Glahn, D.C., Miklowitz, D.J., Ochsner, K., Green, M.F., 2013. Social and Nonsocial Cognition in Bipolar Disorder and Schizophrenia: Relative Levels of Impairment. *Am. J. Psychiat.* 170(3), 334-341.
- Lee, L., Harkness, K.L., Sabbagh, M.A., Jacobson, J.A., 2005. Mental state decoding abilities in clinical depression. *J Affect Disord* 86(2-3), 247-258.
- Lincoln, T.M., Mehl, S., Kesting, M.-L., Rief, W., 2011. Negative symptoms and social cognition: identifying targets for psychological interventions. *Schizophr. Bull.* 37(suppl\_2), S23-S32.
- Marjoram, D., Gardner, C., Burns, J., Miller, P., Lawrie, S.M., Johnstone, E.C., 2005. Symptomatology and social inference: a theory of mind study of schizophrenia and psychotic affective disorder. *Cogn. Neuropsychiatry* 10(5), 347-359.
- McGorry, P.D., Purcell, R., Goldstone, S., Amminger, G.P., 2011. Age of onset and timing of treatment for mental and substance use disorders: implications for preventive intervention strategies and models of care. *Current opinion in psychiatry* 24(4), 301-306.
- Mellick, W., Sharp, C., 2016. Mental State Decoding in Adolescent Boys with Major Depressive Disorder versus Sex-Matched Healthy Controls. *Psychopathology* 49(1), 53-59.
- Mitchell, R.L.C., Young, A.H., 2015. Theory of Mind in Bipolar Disorder, with Comparison to the Impairments Observed in Schizophrenia. *Frontiers in psychiatry* 6, 188-188.
- Ohayon, M.M., Schatzberg, A.F., 2002. Prevalence of depressive episodes with psychotic features in the general population. *Am. J. Psychiat.* 159(11), 1855-1861.
- Premack, D., Woodruff, G., 1978. Does the chimpanzee have a theory of mind. *Behav. Brain Sci.* 1(4), 515-526.
- Purcell, A.L., Phillips, M., Gruber, J., 2013. In your eyes: Does theory of mind predict impaired life functioning in bipolar disorder? *J Affect Disord* 151(3), 1113-1119.
- Roberts, D.L., Combs, D.R., Willoughby, M., Mintz, J., Gibson, C., Rupp, B., Penn, D.L., 2014. A randomized, controlled trial of Social Cognition and Interaction Training (SCIT) for outpatients with schizophrenia spectrum disorders. *Br. J. Clin. Psychol.* 53(3), 281-298.
- Rowland, J.E., Hamilton, M.K., Vella, N., Lino, B.J., Mitchell, P.B., Green, M.J., 2012. Adaptive Associations between Social Cognition and Emotion Regulation are Absent in Schizophrenia and Bipolar Disorder. *Front. Psychol.* 3, 607-607.
- Schaafsma, S.M., Pfaff, D.W., Spunt, R.P., Adolphs, R., 2015. Deconstructing and reconstructing theory of mind. *Trends Cogn Sci* 19(2), 65-72.
- Shamay-Tsoory, S.G., Aharon-Peretz, J., 2007. Dissociable prefrontal networks for cognitive and affective theory of mind: A lesion study. *Neuropsychologia* 45(13), 3054-3067.
- Shamay-Tsoory, S.G., Shur, S., Barcal-Goodman, L., Medlovich, S., Harari, H., Levkovitz, Y., 2007. Dissociation of cognitive from affective components of theory of mind in schizophrenia. *Psychiatry Res* 149(1-3), 11-23.
- Sheehan, D., Janavs, J., Harnett-Sheehan, K., Sheehan, M., Gray, C., Lecrubier, Y., 2009. MINI: Mini International Neuropsychiatric Interview English version 6.0. 0 DSM-IV; 2009. Website: <http://www.psy.uwa.edu.au/user/andrew/MINI600.pdf>. Accessed 1.
- Strauss, G.P., Vertinski, M., Vogel, S.J., Ringdahl, E.N., Allen, D.N., 2016. Negative symptoms in bipolar disorder and schizophrenia: A psychometric evaluation of the brief negative symptom scale across diagnostic categories. *Schizophr Res* 170(2-3), 285-289.
- Thaler, N.S., Allen, D.N., Sutton, G.P., Vertinski, M., Ringdahl, E.N., 2013. Differential impairment of social cognition factors in bipolar disorder with and without psychotic features and schizophrenia. *J*

- Psychiatr Res 47(12), 2004-2010.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288.
- Valle, A., Massaro, D., Castelli, I., Marchetti, A., 2015. Theory of Mind Development in Adolescence and Early Adulthood: The Growing Complexity of Recursive Thinking Ability. *Europe's journal of psychology* 11(1), 112-124.
- Wilbertz, G., Brakemeier, E.-L., Zobel, I., Haerter, M., Schramm, E., 2010. Exploring preoperational features in chronic depression. *J Affect Disord* 124(3), 262-269.
- Young, R.C., Biggs, J.T., Ziegler, V.E., Meyer, D.A., 1978. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry* 133(NOV), 429-435.
- Zhang, Q., Li, X., Parker, G.J., Hong, X.-h., Wang, Y., Lui, S.S., Neumann, D.L., Cheung, E.F., Shum, D.H., Chan, R.C., 2016. Theory of mind correlates with clinical insight but not cognitive insight in patients with schizophrenia. *Psychiatry Res* 237, 188-195.

**Table 1. Demographic and clinical characteristics of participants.**

	SCZ(n=35)		MDD(n=35)		BD(n=35)		HC(n=35)		<i>F/t/χ<sup>2</sup></i>	<i>p</i>
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>		
Age (year)	29.14	5.791	29.40	7.93	31.11	6.816	27.26	6.70	1.860	0.139
Gender (male:female)	14:21		15:20		16:19		16:19		0.320	0.956
Education (year)	11.66	2.51	11.60	2.64	12.74	3.73	13.26	2.49	2.820	0.041
Estimated IQ	107.74	14.21	109.32	18.04	112.20	18.58	106.94	16.47	0.659	0.578
Duration of illness (month)	23.46	19.002	31.12	24.709	47.80	45.846			5.241	0.007
Executive function	11.11	4.31	14.71	5.42	11.80	5.31	18.29	3.90	16.407	<.001
PANSS_P	16.83	5.43			13.69	5.274			2.456	0.017
PANSS_N	15.74	5.705			9.34	3.827			5.512	<.001
PANSS_G	33.83	9.694			26.66	6.029			3.717	<.001
HRSD-24			26.20	11.908	20.69	10.983			2.014	0.048
YMRS (n=22)					21.05	6.88				

Note: PANSS = Positive and Negative Syndrome Scale, PANSS\_P = the score of positive scale measured by PANSS, PANSS\_N = the score of negative scale measured by PANSS, PANSS\_G = The score of general psychopathology scale measured by PANSS, HRSD-24 = the total score of the 24-item Hamilton Rating Scale for Depression, YMRS = The score of Young manic rating scale.

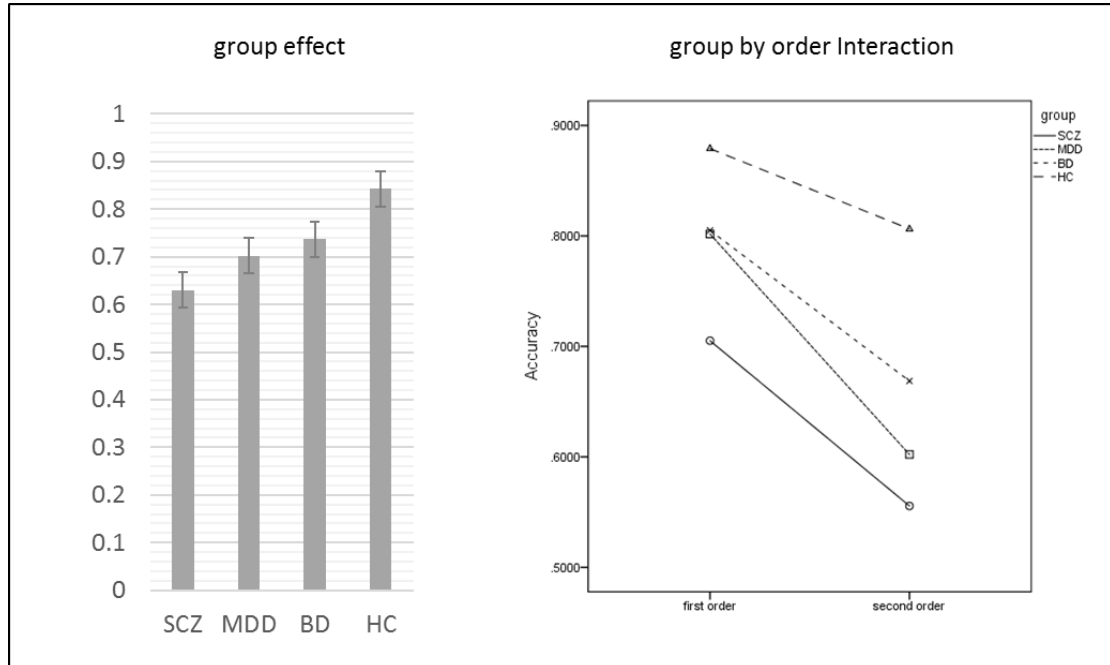
**Table 2. Comparisons of percentage of correct response under different conditions between four groups in Yoni Task**

	SCZ( <i>n</i> =35)		MDD( <i>n</i> =35)		BD( <i>n</i> =35)		HC( <i>n</i> =35)		Pairwise comparisons (not controlling physical conditions)	Pairwise comparisons (controlling physical conditions)
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>		
1 <sup>st</sup> _cognitive	0.700	0.331	0.801	0.307	0.800	0.300	0.886	0.231	SCZ<HC( <i>p</i> =0.055)	
1 <sup>st</sup> _affective	0.710	0.291	0.803	0.308	0.810	0.253	0.872	0.235		
2 <sup>nd</sup> _cognitive	0.530	0.219	0.572	0.220	0.649	0.215	0.809	0.189	SCZ<HC*** MDD<HC*** BD<HC*	SCZ<HC*** MDD<HC**
2 <sup>nd</sup> _affective	0.581	0.198	0.632	0.198	0.689	0.171	0.804	0.163	SCZ<HC*** MDD<HC** BD<HC( <i>p</i> =0.057)	SCZ<HC**

Note: 1<sup>st</sup>\_cognitive = first-order cognitive ToM, 1<sup>st</sup>\_affective = first-order affective ToM, 1<sup>st</sup>\_physical = first-order physical condition, 2<sup>nd</sup>\_cognitive = second-order cognitive ToM, 2<sup>nd</sup>\_affective = second-order affective ToM, 2<sup>nd</sup>\_physical = second-order physical condition.

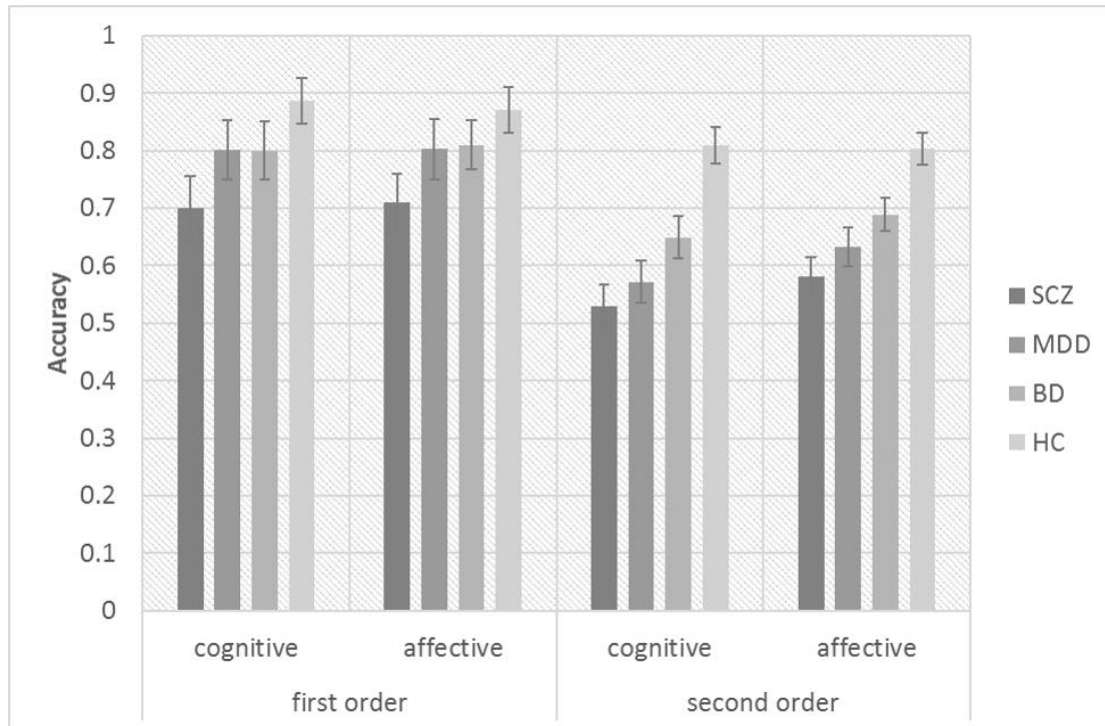
\* *p*<0.05, \*\* *p*<0.01, \*\*\* *p*<0.001

## Figures



**Figure 1. Main effects and interaction effect on Yoni task performance**

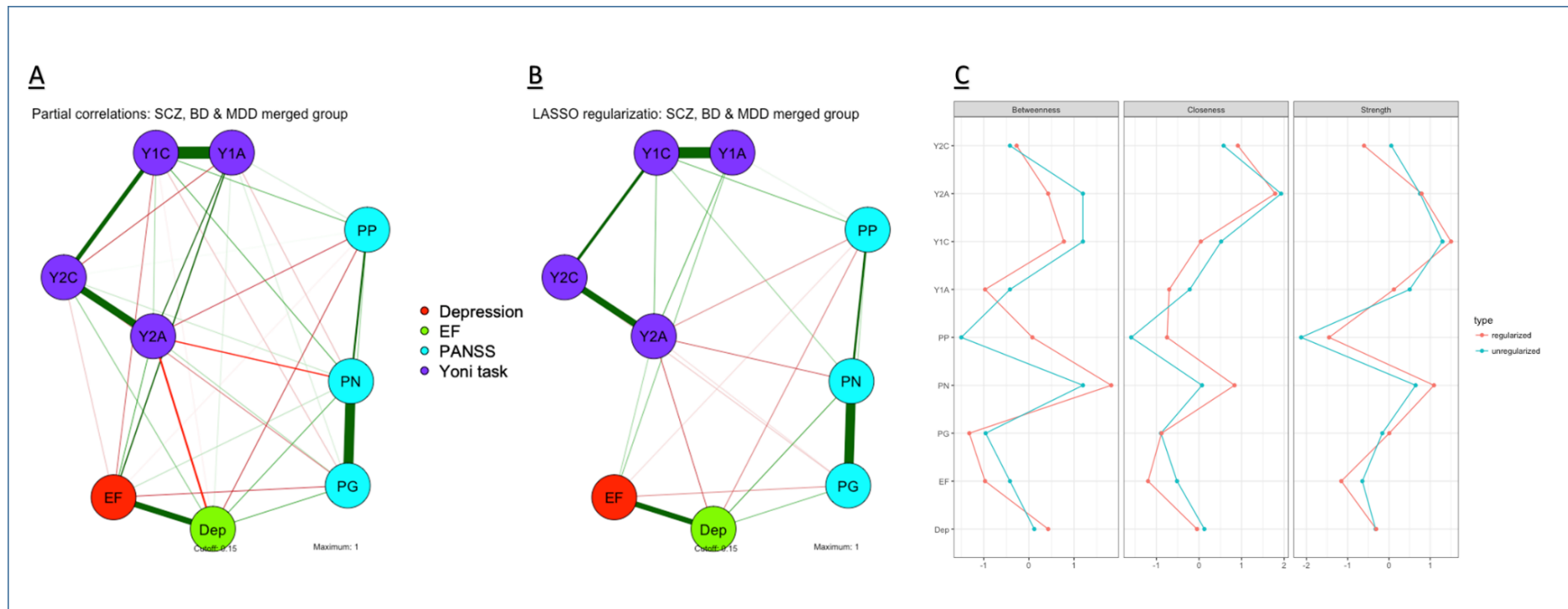
**Legend:** Main effect of group was significant, with all patient groups (SCZ, MDD and BD) showing lower accuracy on the overall performance of Yoni task compared to controls; The SCZ group had inferior performance than the BD group. The group by order interaction was significant. The threshold for significance was set at  $p < 0.05$ .



**Figure 2. Group comparisons on each condition of Yoni task.**

**Legend:** Multivariate analyses were conducted for each cognitive and affective condition of Yoni task to test for the group differences. The results showed that on the first order conditions no significant differences were found; on the second order conditions SCZ, BD and MDD groups showed poorer accuracy on cognitive conditions; SCZ and MDD groups also underperformed on the affective conditions.





**Figure 3.** Partial correlation networks and centrality measures in a merged sample including all patients. Note: Y1A = first-order affective ToM, Y1C = first-order cognitive ToM, Y2A = second-order affective ToM, Y2C = second-order cognitive ToM, EF = executive function, PP = PANSS positive syndrome, PN = PANSS negative syndrome, PG = PANSS general psychopathology, Dep = depression measured by HRSD-24

### **Role of funding source**

The funding agents had no further role in the study design; in the collection, analysis and interpretation of the data; in the writing of the manuscript; and in the decision to submit the paper for publication.

### **Contributors**

YYW collected, analyzed the data and wrote up the first draft of the manuscript. YMZ, KN, XT, HWS, SSYL administered behavioural data and clinical rating for patients. SSYL, EFC and JS commented critically to the manuscript. YW and RCKC designed the study, interpreted the data and commented critically to the manuscript. All authors contributed to and have approved the final manuscript.

### **Conflict of Interest**

none

### **Acknowledgement**

This study was supported by the National Basic Research Programme of China (Precision psychiatry Programme, 2016YFC0906402), the National Science Fund China (81571317), the Beijing Training Project for the Leading Talents in Science and Technology (Z151100000315020), and the Beijing Municipal Science & Technology Commission Grant (Z161100000216138) to Raymond Chan. Yi Wang was supported by the National Science Fund China (31400884) and China Scholarship Council. John

Suckling receives support from the United Kingdom National Institute for Health Research Cambridge Biomedical Research Centre. Yan-yu Wang was supported by the Natural Science Foundation of Shandong Province (ZR2017LC023). This research was also supported by the CAS Key Laboratory of Mental Health, Institute of Psychology.

## Supplementary Materials

### Illustration of the Yoni task

The Yoni task is a computerized task capturing cognitive and affective ToM. There are three conditions: the cognitive condition involves inferences about Yoni's thinking, the affective condition involve inferences about Yoni's emotional states, while the physical condition serves as a control condition. On each condition, there are first-and second- order inferences. In the first-order inference, participants were asked to infer about Yoni's mental states, in the second-order inference, participants need to infer Yoni's beliefs/feelings about another character.

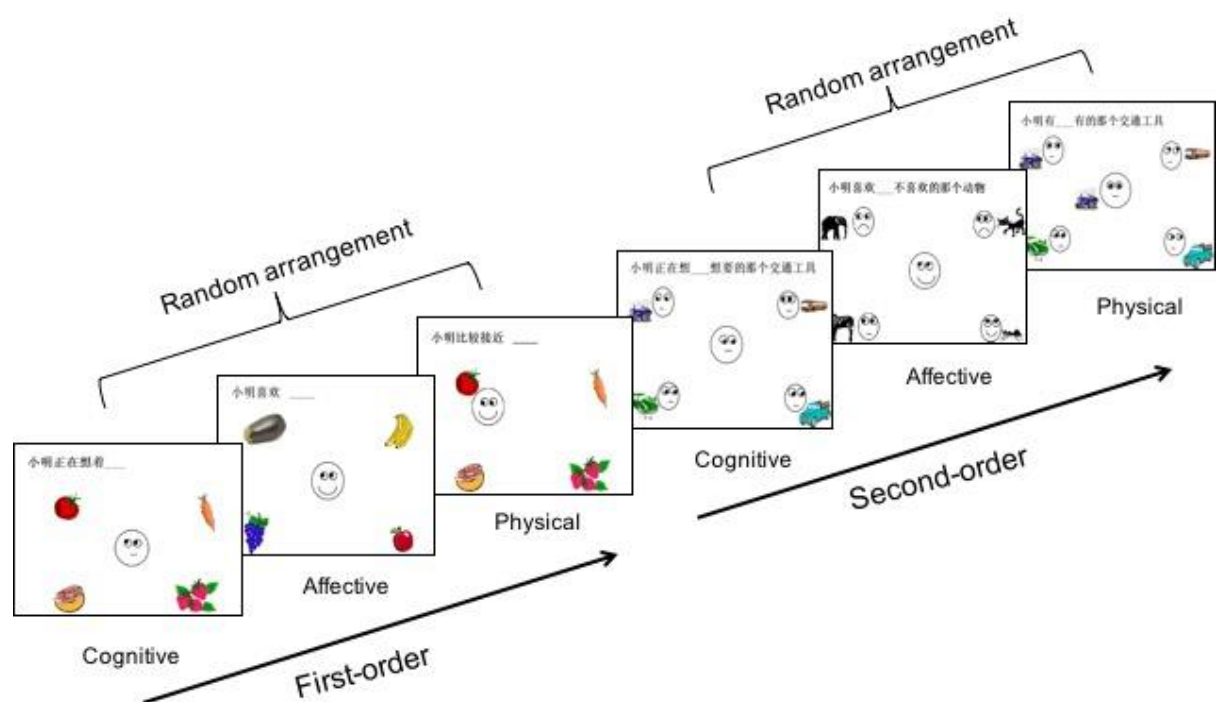
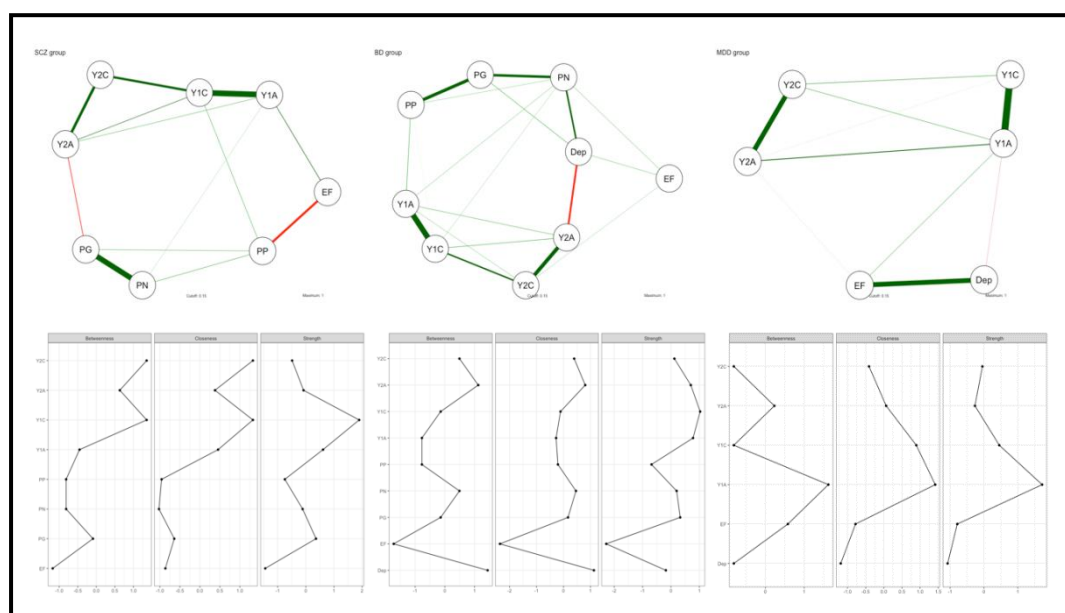


Figure S1. Illustration of the Yoni task.

## Network analysis for each patient group

We repeated the network construction and analysis for each patient group. The variables we used to construct network in each patient group are as follow: for **SCZ group**, PANSS (3, positive, negative, general scores), Yoni (4, accuracy on 1<sup>st</sup> order cognitive/affective, 2<sup>nd</sup> order cognitive/affective) and Executive Function (1, Letter-Number Span total passed item number); for **MDD group**, HRSD-24 total score was adopted instead of the PANSS; for **BD group**, both PANSS and HRSD-24 were included for the clinical symptoms.

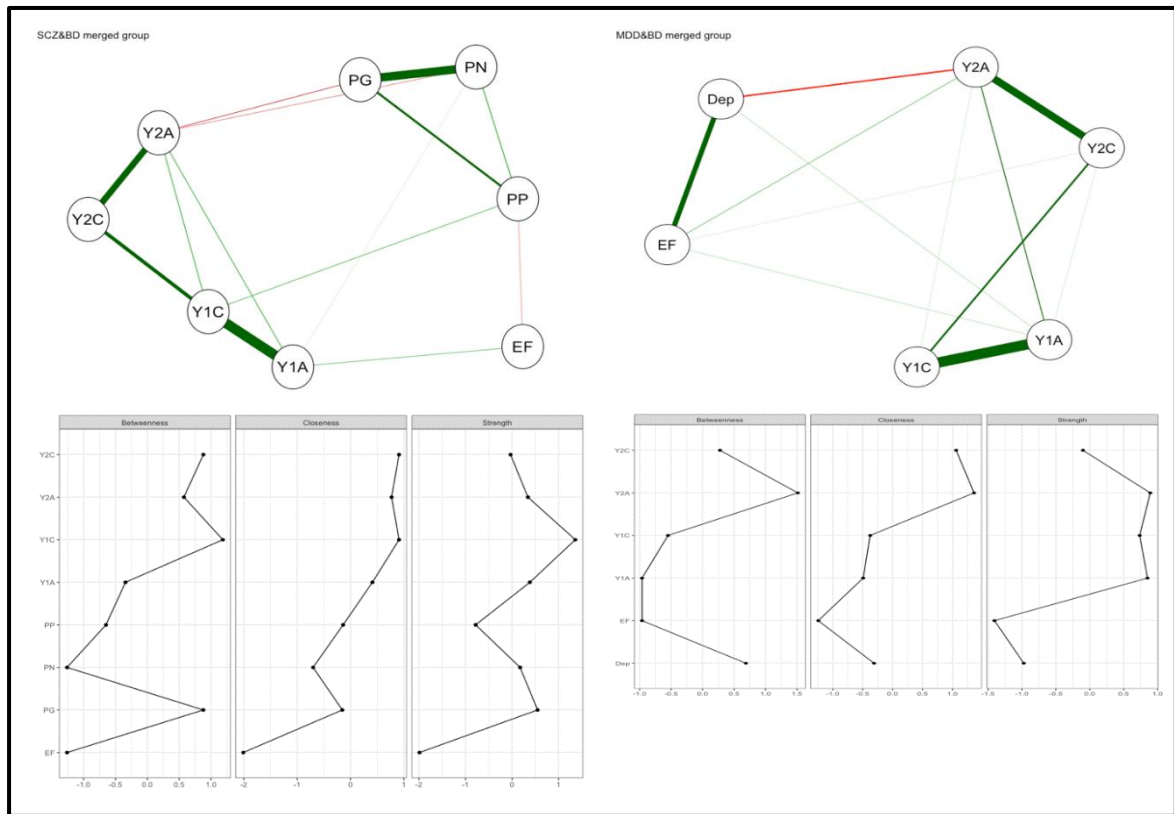


**Figure S2. Network and centrality measures in each patient groups.** Note: Y1A = first-order affective ToM, Y1C = first-order cognitive ToM, Y2A = second-order affective ToM, Y2C = second-order affective ToM, EF = executive function, PP = PANSS positive syndrome, PN = PANSS negative syndrome, PG = PANSS general psychopathology, Dep = depression measured by HRSD.

As shown in Figure S2, in patients with Schizophrenia, the results showed that performance on first order cognitive condition (Y1C) was related to PANSS positive symptom scores. Performance on the first order affective condition (Y1A) was related to executive function. Accuracy on the second order affective condition (Y2A) was negatively related to PANSS general scores. In terms of the centrality, first order cognitive ToM had the highest betweenness, closeness and strength in SCZ patient group. For BD group, as shown in the Figure S1, the first order affective condition (Y1A) was related to PANSS positive and negative scores; the first order cognitive condition (Y1C) was related to PANSS negative scores. Accuracy on the second order cognitive condition (Y2A) was positively associated with EF; accuracy of the second order affective condition (Y2A) was negatively related to HRSD total scores. In the BD patients, the depressive symptoms measured by HRSD have the highest betweenness and closeness, Y1C has the highest strength. The network for patients with MDD was shown in the Figure S1. The results show that performance on the first order affective condition (Y1A) was negatively associated with the total scores of HRSD-24 and positively associated with performance on Letter Number Span Test.

## **Network Analysis for Merged Samples**

In the merged sample of SCZ and BD, we illustrate the associations between PANSS, ToM and EF in Figure S3. Accuracy of first order affective ToM (Y1A) is positively associated with EF, and PANSS negative symptoms; accuracy of first order cognitive ToM (Y1C) is positively associated with PANSS positive symptoms; accuracy of second order affective ToM (Y2A) is negatively associated with PANSS general and negative symptoms. In the merged sample including patients with BD and MDD, accuracy of the second order affective ToM condition was positively associated with EF and negatively associated with HRSD total scores (see Fig. S3). At the same time, we observed strong associations between accuracy of four conditions of Yoni task in both merged groups. In terms of the centrality measures of the nodes, we found that in SCZ & BD merged sample, the first order cognitive ToM (Y1C) had the highest betweenness, closeness and strength; in the MDD & BD merged group, the second order affective ToM (Y2A) presented the highest betweenness, closeness and strength.



**Figure S3. Network and centrality measures in merged samples.** Note: Y1A = first-order affective ToM, Y1C = first-order cognitive ToM, Y2A = second-order affective ToM, Y2C = second-order affective ToM, EF = executive function, PP = PANSS positive syndrome, PN = PANSS negative syndrome, PG = PANSS general psychopathology, Dep = depression measured by HRSD.