Meta-analysis of Theory of Mind (ToM) impairment in bipolar disorder

E. Bora1*, C. Bartholomeusz1,2 and C. Pantelis3

1 Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne and Melbourne Health, South Carlton, VIC, Australia
2 Orygen, National Centre of Excellence in Youth Mental Health and The Centre for Youth Mental Health, The University of Melbourne, Parkville, VIC, Australia

Background. Theory of mind (ToM) dysfunction is prominent in a number of psychiatric disorders, in particular, autism and schizophrenia, and can play a significant role in poor functioning. There is now emerging evidence suggesting that ToM abilities are also impaired in bipolar disorder (BP); however, the relationship between ToM deficits and mood state is not clear.

Method. We conducted a meta-analysis of ToM studies in BP. Thirty-four studies comparing 1214 patients with BP and 1097 healthy controls were included. BP groups included remitted (18 samples, 545 BP patients), subsyndromal (12 samples, 510 BP patients), and acute (manic and/or depressed) (10 samples, 159 BP patients) patients.

Results. ToM performance was significantly impaired in BP compared to controls. This impairment was evident across different types of ToM tasks (including affective/cognitive and verbal/visual) and was also evident in strictly euthymic patients with BP (d = 0.50). There were no significant differences between remitted and subsyndromal samples. However, ToM deficit was significantly more severe during acute episodes (d = 1.23). ToM impairment was significantly associated with neurocognitive and particularly with manic symptoms.

Conclusion. Significant but modest sized ToM dysfunction is evident in remitted and subsyndromal BP. Acute episodes are associated with more robust ToM deficits. Exacerbation of ToM deficits may contribute to the more significant interpersonal problems observed in patients with acute or subsyndromal manic symptoms. There is a need for longitudinal studies comparing the developmental trajectory of ToM deficits across the course of the illness.

Received 11 July 2015; Revised 9 September 2015; Accepted 11 September 2015; First published online 12 October 2015

Key words: Bipolar disorder, manic-depressive, neurocognition, theory of mind, social cognition.

Introduction

Theory of mind (ToM), the ability to infer the mental states of self and others, is one of the most important social cognitive abilities for maintaining effective and adaptive social functioning. ToM dysfunction is a well-established finding in schizophrenia as well as developmental disorders such as autism (Bora et al. 2009; Senju, 2012; Bora & Pantelis, 2013; Ventura et al. 2015). ToM impairment can significantly contribute to psychosocial difficulties in psychiatric disorders as social cognitive abilities might be more directly related to interpersonal functioning than general neurocognitive abilities (Fett et al. 2011). In schizophrenia, studies have suggested that neuropsychological deficits and ToM deficits contribute to social dysfunction (Bora et al. 2006a; Green et al. 2012). A significant proportion of patients with bipolar disorder (BP) also have poor psychosocial functioning (Sanchez-Moreno et al. 2009; Burdick et al. 2010) and neurocognitive impairment is evident in many patients with BP even during remission (Bora et al. 2009b; Burdick et al. 2014; Trotta et al. 2014). It might be expected that ToM deficits might also be associated with BP.

Emerging literature suggests that ToM ability is also impaired in BP (Bora et al. 2005; Samamé et al. 2012). However, there are inconsistent findings regarding the persistence of ToM deficits during euthymia, as both positive and negative findings have been reported. In their meta-analyses of social cognition in BP, Samamé et al. (2012, 2013) also conducted a preliminary analysis of 9 and 11 ToM studies, respectively, in ‘remitted’ BP patients and found significant but modest impairments. However, in these meta-analyses strict euthymia criteria were not used and a number of studies comprised mixed patient samples which included subsyndromal depressive/manic or mild depression (Ioannidi et al. 2013; Lee et al. 2013; Thaler...
et al. 2013). Moreover, it is important to investigate whether ToM deficits are trait- or state-related, as it has been suggested that persistent ToM deficits are risk factors for relapse in mood disorder (Inoue et al. 2006). To date, no meta-analysis has investigated ToM deficits in acute episodes and differences between remitted patients and patients with acute or subsyndromal symptoms.

Another important point to consider is the heterogeneity of ToM as a construct. Mode (i.e. visual v. verbal), content [i.e. inferring what a person is feeling (affective) v. inferring beliefs and motivations (cognitive)] and complexity (i.e. basic v. advanced) of stimuli used in ToM tasks in previous BP studies vary greatly. These aspects of ToM can be related to relatively separate neural networks (Schurz et al. 2014) which can be more or less impaired in BP. For example, some authors have suggested that ToM for cognitive but not emotional stimuli is impaired in BP (Shamay-Tsoory et al. 2009; Montag et al. 2010; Barrera et al. 2013).

In this comprehensive meta-analysis of ToM, we aimed to investigate ToM deficits in BP, including samples with remitted patients, and patients with acute and subsyndromal symptoms, in comparison to healthy controls. This meta-analysis also aimed to explore the effect of clinical, demographic, ToM task type and neurocognitive variables on patient–control differences for ToM.

Method

Study selection

We followed MOOSE and PRISMA guidelines in conducting this meta-analysis (Stroup et al. 2000; Moher et al. 2009). A literature search was conducted using the databases Pubmed, PsycINFO, ProQuest and Scopus to identify the relevant studies (January 1990 to June 2015) using the combination of key words as follows: Theory of mind, mentalizing, social cognition, bipolar disorder. The same search was performed in Google Scholar to retrieve unpublished studies. Reference lists of published reports were also reviewed for additional studies. Inclusion criteria were studies that: (1) compared ToM performance of BP patients with that of healthy controls; (2) reported sufficient data to calculate the effect sizes and standard errors of the ToM measures. Studies in pediatric samples and overlapping samples were excluded. We also contacted nine authors for unreported information and clarification (including Dr Ioannidi, Dr Van Rheenen, Dr Purcell, Dr Donohoe, Dr Lahera whose studies are included in the current meta-analysis). A total of 34 studies involving 1214 BP patients (40 samples) and 1097 healthy controls were included in the current meta-analysis (Table 1) (see Supplementary Fig. S1 for flow chart of the study selection process). Percentage of females was very similar in BP (54.7%) and healthy controls (52.2%). BP patients were significantly older than healthy controls \( [d = 0.23, 95\% \text{ confidence interval (CI)} 0.10–0.37, Z = 3.4, p < 0.001] \). We also categorized included studies into three groups: remitted (based on a strict euthymia criteria), subsyndromal (stable non-acute outpatients, including patients with mild manic and depressive symptoms who did not meet criteria for euthymia) and acute (manic and/or depressed). Strict euthymia criteria for the remission group were defined as Hamilton Depression Rating Scale (HAMD) score \(<6–10\), Young Mania Rating Scale (YMRS) score \( <6–10\) or very small mean HAMD and YMRS scores (mean ± 2 S.D. within range of euthymia criteria). Eighteen samples including 545 BP patients were classified as remitted based on these criteria. Ten samples included patients (159 BP) in acute episode (6 manic, 3 depressive, 1 mixture of manic and depressive). The remaining 12 studies included outpatients (510 BP) with subsyndromal or mild depressive or mild manic symptoms.

ToM measures

Most commonly used tasks were Reading the Mind in the Eyes Task (RMET; Baron-Cohen et al. 2001) and Faux pas recognition tasks (Stone et al. 1998). Other ToM measures included the Hinting, Happe stories, picture sequencing tasks, The Awareness of Social Inference Test (TASIT)-sarcasm, Movie for Assessment of Social Cognition (MASC) and different versions of false belief and ToM stories and ToM cartoons.

Statistical analyses

For studies that reported more than one ToM task, pooled effect size and s.e. values were calculated. We also calculated cognitive, affective, verbal and visual ToM scores. Separate task-specific analyses were also conducted if there were at least five studies reporting a particular measure; individual task analyses were possible for the RMET, Hinting task, and the Faux pas task. An additional analysis was conducted for false belief stories (story contents differ).

Meta-analyses were performed using MIX software version 1.7 on a Windows platform (Bax et al. 2006) and in R environment (OpenMetaAnalyst, Metafor) (Viechtbauer, 2010; Wallace et al. 2012). Effect sizes were weighted using the inverse variance method. A random-effects model (DerSimonian–Laird estimate) was used as the distributions of effect sizes were heterogeneous for the number of variables.
<table>
<thead>
<tr>
<th>Study</th>
<th>Sample</th>
<th>ToM</th>
<th>BP characteristics</th>
<th>State</th>
<th>Mood symptoms and euthymia criteria</th>
<th>Cognitive variables</th>
<th>Outcome</th>
</tr>
</thead>
</table>
| Andrews (2013)<sup>a</sup> | 16 BP  
20 HC | Yoni                    | 11 BP I, 5 BP II  
Age = 41.6  
Duration = 21.3 | Remitted       | EC = HAMD < 6, YMRS < 6,  
HAMD = 3.3, YMRS = 1.4 | RBANS             | BP significantly impaired |
| Barrera et al. (2013)  | 12 BP  
12 HC | Faux pas, RMET          | 7 BD II, 5 BD I  
Age = 48.2  
Duration = 23.2 | Remitted       | EC = HAMD < 7, YMRS < 8 |                       | Faux pas impaired in BP  
Not related to functioning |
| Bazin et al. (2009)    | 15 BP  
15 HC | Intention               | Age = 36.1        | Manic        | YMRS = 19.3                         |                      | Impaired compared to controls            |
| Bora et al. (2005)     | 43 BP  
30 HC | Hinting, RMET           | All BP I  
Age = 38.6  
Duration = 15.5  
26/43 history of psychosis | Remitted       | EC = HAMD < 7, YMRS < 6,  
HAMD = 1.7, YMRS = 0.7 | WCST, TMT, CPT,  
Stroop, verbal fluency, verbal memory | BP impaired in both tasks  
Lost significance when corrected for EF deficit |
| Budak (2011)<sup>a</sup> | 52 BP  
60 HC | RMET, ToM battery including FB, irony, Faux pas, | All BP I  
Age = 34.5  
Duration = 11.9  
47/52 history of psychosis | Remitted       | EC = HAMD < 6, YMRS < 6 |                       | BP significantly impaired |
| Caletti et al. (2013)  | 18 BP  
18 HC | Faux pas, RMET          | 10 BP I, 8 BP II  
Age = 42.2  
Duration = 17.5 | Remitted       | EC = HAMD < 7, YMRS < 10,  
HAMD = 4.8, YMRS = 2.5 | Planning, fluency,  
Speed, WM, verbal memory | No difference |
| Caponigro (2007)<sup>a</sup> | 19 BP  
15 HC | FB1 and FB2             | Age = 46.5        | Remitted       | EC = HAMD < 10, BRMS < 7 | IQ                       | In BP ToM impaired compared to HC |
| Cusi et al. (2012)     | 25 BP  
25 HC | RMET                    | 17 BP I, 7 BP II  
Age = 45.2  
Duration = 23.1 | Subsyndromal   | EC = YMRS < 10, HAMD = 8.1, YMRS = 2.0 |                      | ToM impaired in BP |
| Donohoe et al. (2012)  | 102 BP  
132 HC | Hinting, RMET           | Age = 44.8  
Duration = 20.6 | Subsyndromal   | IQ, WM, CPT, verbal memory       |                      | ToM impaired in BP |
| Duman (2014)<sup>a</sup> | 102 BP  
132 HC | Hinting, RMET, Faux pas | All BP I  
Age = 36.3  
Duration = 9.4 | Remitted       | EC = HAMD < 8, YMRS < 6,  
HAMD = 0.3, YMRS = 0.4 | WCST, TMT, Digit span, Stroop | No difference |
| Ibanez et al. (2012)   | 13 BP  
13 HC | Faux pas, RMET          | All BP II  
Age = 40.1 | Remitted       | EC = BDI < 6, YMRS < 6 | WM, TMT A and B | Faux pas impaired |
| Ioannidi et al. (2013) | 13 BP  
55 HC | Faux pas, FB1, Hinting  | Age = 41.9  
Duration = 14.1 | Subsyndromal   | HAMD = 7.7, YMRS = 5.7 |                      | BP impaired in faux pas recognition |

<sup>a</sup> Studies included into the meta-analysis.
<table>
<thead>
<tr>
<th>Study</th>
<th>Sample</th>
<th>ToM</th>
<th>BP characteristics</th>
<th>State</th>
<th>Mood symptoms and euthymia criteria</th>
<th>Cognitive variables</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ioannidi et al. (2015)</td>
<td>29 BP</td>
<td>Faux pas, FB</td>
<td>Age = 44.2 Duration = 16.30</td>
<td>Acute episode</td>
<td>13 manic 16 depressed</td>
<td>WCST, TMT, Digit Span, Stroop, Verbal memory</td>
<td>BP impaired in all ToM tasks</td>
</tr>
<tr>
<td>Kerr et al. (2003)</td>
<td>48 BP</td>
<td>FB1, FB2</td>
<td>Age = 43.9 Duration = 10.9</td>
<td>20 manic 15 depressive 13 remitted</td>
<td>BDI = 3.2/16.9/2.1, BRMS = 13.3/2.2/2.9</td>
<td>ToM only impaired in manic and depressed patients</td>
<td></td>
</tr>
<tr>
<td>Lahera et al. (2008)</td>
<td>75 BP</td>
<td>Happe stories</td>
<td>All BP I</td>
<td>Remitted</td>
<td>EC = HAMD &lt;7, YMRS &lt;7</td>
<td>WCST, sustained attention</td>
<td>ToM impaired in BP</td>
</tr>
<tr>
<td>Lahera et al. (2015)</td>
<td>46 BP</td>
<td>Hinting</td>
<td>Age = 48.6 42/75 history of psychosis</td>
<td>Subsyndromal</td>
<td>EC = HAMD &lt;7, YMRS &lt;7</td>
<td>ToM impaired in BP</td>
<td></td>
</tr>
<tr>
<td>Lee et al. (2013)</td>
<td>68 BP</td>
<td>TASIT 3-sarcasm</td>
<td>46 BP I, 22 BP II</td>
<td>Subsyndromal</td>
<td>HAMD = 8.1, YMRS = 3.3</td>
<td>Reasoning, WM, Speed, Verbal and Visual memory, attention</td>
<td>BP not impaired compared to HC</td>
</tr>
<tr>
<td>Martino et al. (2011)</td>
<td>81 BP</td>
<td>Faux pas, RMET</td>
<td>45 BP I, 36 BP II</td>
<td>Remitted</td>
<td>EC = HAMD &lt;8, YMRS &lt;6</td>
<td>Verbal memory Digit span, verbal Fluency, TMT</td>
<td>ToM impaired in BP</td>
</tr>
<tr>
<td>McKinnon et al. (2010)</td>
<td>14 BP</td>
<td>ToM: 1st and 2nd</td>
<td>8 BP I, 5 BP II, 1 BP NOS</td>
<td>Subsyndromal</td>
<td>HAMD = 7–15, YMRS &lt;10, HAMD = 10.8, YMRS = 3.2</td>
<td>BP impaired performance in 2nd order task</td>
<td></td>
</tr>
<tr>
<td>Montag et al. (2010)</td>
<td>29 BP</td>
<td>MASC</td>
<td>All BP I</td>
<td>Subsyndromal</td>
<td>EC = HAMD &lt;14, YMRS &lt; 5, HAMD = 6.7 YMRS = 3.4</td>
<td>AVLT</td>
<td>ToM impaired in BP</td>
</tr>
<tr>
<td>Olley et al. (2005)</td>
<td>15 BP</td>
<td>ToM stories, ToM cartoon</td>
<td>All BP I</td>
<td>Remitted</td>
<td>EC = HAMD &lt;12, YMRS &lt; 12, HAMD = 3.0, YMRS = 1.5</td>
<td>SOC,Verbal fluency, Set shifting, Stroop</td>
<td>Impaired only in verbal ToM ToM is significantly correlated with EF deficits</td>
</tr>
<tr>
<td>Ozel-Kizil et al. (2012)</td>
<td>18 BP</td>
<td>Faux pas</td>
<td>All BP I History of psychosis</td>
<td>Remitted</td>
<td>EC = HAMD &lt;7, YMRS &lt;7</td>
<td>No significant difference</td>
<td></td>
</tr>
<tr>
<td>Purcell et al. (2013)</td>
<td>26 BP</td>
<td>RMET</td>
<td>All BP I</td>
<td>Remitted</td>
<td>EC = IDS-C &lt;11, YMRS &lt;7</td>
<td>LNS</td>
<td>No difference</td>
</tr>
<tr>
<td>Study</td>
<td>Sample Size</td>
<td>Design</td>
<td>Control Group</td>
<td>Type of Test</td>
<td>Duration</td>
<td>Diagnosis</td>
<td>Outcome Measures</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------</td>
<td>--------</td>
<td>---------------</td>
<td>--------------</td>
<td>----------</td>
<td>------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Robinson (2010)*</td>
<td>39 BP</td>
<td>RMET</td>
<td>28 C</td>
<td>Age = 44.9</td>
<td>Duration = 23.3</td>
<td>Remitted</td>
<td>EC = HAMD &lt; 8, YMRS &lt; 8, HAMD = 2.1, YMRS = 1.5</td>
</tr>
<tr>
<td>Rossell &amp; Van Rheenen (2013)</td>
<td>28 BP</td>
<td>ToM stories</td>
<td>29 BP I, psychotic</td>
<td>Age = 38.3</td>
<td>Duration = 16.3</td>
<td>Manic</td>
<td>Subsyndromal</td>
</tr>
<tr>
<td>Rowland et al. (2013)</td>
<td>33 BP</td>
<td>TATIS 2.3- Sarcasm</td>
<td>58 HC</td>
<td>Age = 40.7</td>
<td>Duration = 12.5</td>
<td>Subsyndromal</td>
<td>HAMD = 4.2, YMRS = 2.9</td>
</tr>
<tr>
<td>Sakarya (2012)*</td>
<td>30 BP</td>
<td>FB1, FB2, Hinting, Faux pas</td>
<td>30 HC</td>
<td>Age = 36.3</td>
<td>Duration = 12.5</td>
<td>Remitted</td>
<td>EC = HAMD &lt; 9, YMRS &lt; 7</td>
</tr>
<tr>
<td>Sarfati &amp; Hardy-Baylé (1999)</td>
<td>10 BP</td>
<td>ToM picSeq (intention)</td>
<td>15 HC</td>
<td>Age = 33.9</td>
<td>Duration = 6.8</td>
<td>Manic</td>
<td>BP: No difference compared to HC</td>
</tr>
<tr>
<td>Shamay-Tsoory et al. (2009)</td>
<td>19 BP</td>
<td>Faux pas, RMET</td>
<td>20 HC</td>
<td>Age = 40.2</td>
<td>No history of psychosis</td>
<td>Remitted</td>
<td>EC = HAMD &lt; 9, YMRS &lt; 7</td>
</tr>
<tr>
<td>Simon et al. (2013)</td>
<td>54 BP</td>
<td>Faux pas</td>
<td>34 HC</td>
<td>Age = 35.9</td>
<td>Duration = 19</td>
<td>26 remitted</td>
<td>HAMD = 7.6, YMRS = 5.9</td>
</tr>
<tr>
<td>Thaler et al. (2013)</td>
<td>48 BP</td>
<td>Hinting, RMET</td>
<td>24 HC</td>
<td>Age = 35.9</td>
<td>Duration = 19</td>
<td>28 subsyndromal</td>
<td>HAMD = 10.1</td>
</tr>
<tr>
<td>Van Rheenen &amp; Rossell (2013)</td>
<td>49 BP</td>
<td>Picture sequencing</td>
<td>49 HC</td>
<td>Age = 38.5</td>
<td>24/48 history of psychosis</td>
<td>Subsyndromal</td>
<td>MADRS = 11.9, YMRS = 6.3</td>
</tr>
<tr>
<td>Wiener et al. (2011)</td>
<td>20 BP</td>
<td>RMET</td>
<td>40 HC</td>
<td>Age = 43</td>
<td>8 manic 12 depressed</td>
<td>Acute episode</td>
<td>EC = HAMD &lt; 15, YMRS &lt; 12</td>
</tr>
<tr>
<td>Wolf et al. (2010)</td>
<td>33 BP I</td>
<td>ToM picSeq and questions</td>
<td>29 BP I</td>
<td>Age = 47.7</td>
<td>Duration = 12.4</td>
<td>Subsyndromal Manic depressed</td>
<td>EC = HAMD &lt; 15, YMRS &lt; 12</td>
</tr>
</tbody>
</table>

BP, Bipolar disorder; HC, healthy controls; EC, euthymia criteria; CPT, Continuous performance test; MASC, Movie for the Assessment of Social Cognition; FB, false belief; SOC, Stockings of Cambridge; ID/ED, intradimensional/extradimensional shift task; EF, executive function; TMT, Trail making task; WM, working memory; WCST, Wisconsin Card Sorting Test; RMET, Reading the Mind in the Eyes Task RBANS, Repeatable Battery for the Assessment of Neuropsychological Status; YMRS, Young Mania Rating Scale; HAMD, Hamilton Depression Rating Scale; BRMS, Bech-Rafaelsen Mania Scale; MADRS, Montgomery–Åsberg Depression Rating Scale; IDS-C, Inventory for Depressive Symptomatology – Clinician; BDI, Beck Depression Inventory.

* Unpublished thesis or conference paper.
Homogeneity of the distribution of weighted effect sizes was tested with $I^2$ and $Q$ tests. Tau squared ($\tau^2$), an estimate of between-study variance was used as a measure of heterogeneity in the random-effects model.

Publication bias was assessed by Egger’s test and the Fail-safe $N$ test. Egger’s test relies on the theory that studies involving small sample sizes would be more likely to be reported for significant rather than negative findings, while large-scale studies would be more likely to be published regardless of the significance of the findings. Fail-safe $N$ test involves computing a combined $p$ value for all studies included in the meta-analysis, and calculating how many additional studies with a zero effect (average $z$ of zero) would be necessary to create a non-significant $p$ (Rosenthal & Rosnow, 1991).

Meta-regression analyses were conducted for age, gender, duration of education, duration of illness, age of onset, YMRS, HAMD and cognitive impairment in BP-control comparisons whenever at least 10 studies reported these variables. As a measure of cognitive impairment, we used global cognition [current IQ or mean effect sizes calculated as the average of effect sizes of available cognitive domains (verbal, visual memory, working memory, attention, reasoning and problem solving, processing speed); Nuechterlein et al. 2004]. Premorbid IQ and control measures for ToM were not included. Another potential important moderator is the effect of history of psychosis on ToM impairment. Four of the studies have reported separate data to calculate effect sizes for the ToM differences between BP patients with and without history of psychosis. We conducted a preliminary analysis to explore the effect of history of psychosis on ToM based on these studies (Bora et al. 2005; Lahera et al. 2008; Lee et al. 2013; Thaler et al. 2013). Meta-regression analyses (weighted generalized least squares regressions) were conducted using SPSS software (SPSS Inc., USA). Meta-regression analyses performed with a random-effects model were conducted using the restricted-information maximum likelihood method with a significance level set at $p < 0.05$. Subgroup analyses were conducted for patient group [remitted, subsyndromal, acute (and manic only)], subtype (BP I), peer review status (journal articles v. theses/conference papers). The $Q_{bet}$ test was used to compare effect sizes of subgroups.

Results

ToM in BP

ToM performance of patients with BP was significantly impaired compared to controls (Cohen’s $d = 0.63$) (Fig. 1). ToM impairment in BP I patients in comparison to healthy controls were very similar to the primary analysis ($d = 0.68$, 95% CI 0.50–0.87, $Z = 7.3$, $p < 0.001$).

Distribution of effect sizes was heterogeneous ($I^2 = 36.5\%$, $p = 0.02$) but the magnitude of this heterogeneity was very small in the random-effects model ($I^2 = 0.04$). There was some evidence of publication bias but the fail-safe number was very high ($n = 1712$), suggesting that ToM impairment in BP is a robust effect (Table 2).

In 17 studies that assessed general cognition, there was a significant impairment in BP ($d = 0.57$, 95% CI 0.45–0.69, $Z = 9.4$, $p < 0.001$). The severity of ToM impairment in this subset of studies was similar ($d = 0.55$, 95% CI 0.42–0.68, $Z = 8$, $p < 0.001$) to general cognitive impairment, as well as ToM deficit in the main analysis ($d = 0.63$). There was no significant difference for ToM impairment between seven studies that were not peer-reviewed and others ($d = 0.62$ v. 0.63, $Q_{bet} = 0.20$, $p > 0.05$).

ToM and mood state

ToM impairment was also evident in euthymic patients in comparison to controls ($d = 0.50$) (Table 2). Distribution of the effect sizes in euthymic patients was homogeneous in the random-effects model. ToM deficit in subsyndromal BP was also significant ($d = 0.72$) and there were no significant differences between remitted and subsyndromal patients ($Q_{bet} = 0.97$, $p = 0.32$). By contrast, ToM deficit was much more robust in acute BP patients ($d = 1.23$) and ToM impairment in these patients were significant in comparison to remitted ($Q_{bet} = 30.9$, $p < 0.001$) and subsyndromal ($Q_{bet} = 23.7$, $p < 0.001$) BP patients. When analysis for acute BP patients were repeated only for manic patients, the effect size for ToM impairment in manic patients was large ($d = 1.31$) (Table 2).

ToM and task type

Verbal v. visual

ToM impairments in BP patients compared to controls on verbal ($d = 0.58$, 95% CI 0.48–0.69, $Z = 10.7$, $p < 0.001$, $r^2 = 0$) and visual ($d = 0.58$, 95% CI 0.41–0.74, $Z = 7.0$, $p < 0.001$, $r^2 = 0.08$) tasks were very similar.

Cognitive v. affective

Both cognitive ToM ($d = 0.68$, 95% CI 0.56–0.79, $Z = 11.4$, $p < 0.001$, $r^2 = 0.03$) and affective ToM ($d = 0.46$, 95% CI 0.28–0.65, $Z = 4.9$, $p < 0.001$, $r^2 = 0.07$) were significantly impaired in BP. In the 12 studies that assessed both cognitive ($d = 0.65$) and affective ($d = 0.46$) ToM, between-group differences were not significant ($Q_{bet} = 1.17$, $p > 0.05$).

Individual task analyses

In individual task analyses, there were significant impairments in the Faux pas task ($d = 0.57$), The Hinting task ($d = 0.47$), the false beliefs tasks ($d = 0.53$).
and the RMET ($d = 0.50$) in comparison to healthy controls (Table 2). There was significant heterogeneity in the distribution of the effect sizes only in the RMET. It was also possible to conduct individual task analyses for two ToM measures in remitted BP patients: Both RMET ($d = 0.40$) and Faux pas recognition ($d = 0.50$) performances were impaired in BP in comparison to controls. There was evidence for publication bias for the Faux pas task, but the fail-safe number was very high ($n = 116$) suggesting that Faux pas recognition impairment in BP was a real effect.

**Effect of history of psychosis on ToM impairment**

There was a non-significant trend-level effect of history of psychosis on ToM performance in BP. BP patients with history of psychosis tended to perform poorer than BP without history of psychosis ($d = 0.25$, 95% CI $-0.03$ to $0.53$, $p = 0.08$).

**Meta-regression analyses**

Higher YMRS scores were significantly related to the severity of ToM deficit ($Z = 2.90$, $p = 0.004$). The relationship between HAMD and ToM impairment was less pronounced and was not statistically significant ($Z = 1.59$, $p = 0.11$). Global cognitive impairment was significantly associated with ToM impairment in BP ($Z = 4.19$, $p < 0.0001$). However, to explore possible effects of clinical and demographical variables on ToM which can be masked by acute symptoms, we conducted the following analyses in non-acute patients only (remitted and subsyndromal). There was again a significant effect of global cognitive impairment ($Z = 2.90$, $p = 0.004$) on ToM deficit. There was no significant effect of gender ($Z = 0.15$, $p = 0.89$), age ($Z = 0.62$, $p = 0.63$), education ($Z = 0.47$, $p = 0.64$), illness duration ($Z = 0.30$, $p = 0.77$) and age of onset of illness ($Z = 1.08$, $p = 0.28$) on ToM in BP.

---

**Fig. 1.** Forest plot for ToM differences between bipolar disorder and healthy controls.
Mania-related functional changes in the ventral prefrontal cortex which has a critical role in reward, decision making and ToM, could potentially explain these differences. The frontal cortex, which has a critical role in reward, decision making, and ToM dysfunction during manic episodes may be a medium-sized impairment. ToM deficits in BP persist in remitted patients. Our findings showed that ToM ability was significantly impaired in BP in comparison to healthy participants. ToM dysfunction was significantly more pronounced in acute episodes but was also evident in remitted patients.

The effect size of ToM dysfunction ($d = 0.63$) in BP suggests a medium-sized impairment. ToM dysfunction in euthymic BP patients ($d = 0.50$) was comparable to most other neuropsychological deficits observed in the literature (Bora et al. 2009b). Acute phases of BP, especially manic episodes, was associated with exacerbation of ToM deficits and YMRS scores were significantly related to the severity of ToM impairment. Differences between manic and remitted patients were relatively large compared to most traditional neuropsychological domains in which evidence suggests only subtle effects of mood state on cognition (Kurtz & Gerraty, 2009). However, similar to ToM performance, decision making, impulsivity, insight and some aspects of neurocognition (i.e. commission errors in the continuous performance test) are significantly more impaired during mania in comparison to remission (Bora et al. 2006b; Clark & Sahakian, 2008). Mania-related functional changes in the ventral prefrontal cortex which has a critical role in reward, decision making and ToM, could potentially explain these findings (Rubinsztein et al. 2001). It could be argued that ToM dysfunction during manic episodes may be related to thought disorder, psychosis, and/or impairment in functioning. However, more studies are needed to be conclusive on this topic.

It is evident that ToM deficits in BP persist in remission. Thus, it could be argued that ToM deficits might be phenotypic risk markers of BP and may contribute to the social dysfunction that is observed in many people with BP during remission. There is sufficient evidence supporting such a link between social impairment and ToM deficits in schizophrenia (Bora et al. 2006a; Green et al. 2012). However, very few studies have examined the link between ToM dysfunction and social functioning in BP (Lahtera et al. 2012; Caletti et al. 2013; Lee et al. 2013; Purcell et al. 2013). There is evidence suggesting that social cognitive training programs might be beneficial to improve these abilities in schizophrenia (Kurtz & Richardson, 2012). The inclusion of social cognition training in cognitive remediation strategies in BP can potentially improve social functioning in this disorder.

Magnitudes of both social and non-social cognitive impairment in BP were similar and relatively modest compared to similarly larger effect sizes for both domains observed in meta-analyses of schizophrenia studies (Dickinson et al. 2007; Bora et al. 2009a). This finding does not support the notion of BP is associated more severe impairment in non-social than social cognition while schizophrenia is associated with the opposite pattern (Lee et al. 2013). However, another important subject is the origin of ToM deficits in BP. It could be argued that ToM impairments in BP are just epiphenomena of neurocognitive impairment, or

### Table 2. Mean weighted effect sizes for ToM differences between patients with bipolar disorder and healthy controls

<table>
<thead>
<tr>
<th>Test</th>
<th>BP-HC</th>
<th>N</th>
<th>BP</th>
<th>HC</th>
<th>d</th>
<th>95% CI</th>
<th>Z</th>
<th>p</th>
<th>Q test (p)</th>
<th>$r^2$</th>
<th>Bias (p)</th>
<th>Fail-safe N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ToM</td>
<td>34</td>
<td>1214</td>
<td>1097</td>
<td>0.63</td>
<td>0.52–0.74</td>
<td>11.0</td>
<td>&lt;0.001</td>
<td>0.02</td>
<td>0.04</td>
<td>0.04</td>
<td>1712</td>
<td></td>
</tr>
<tr>
<td>RMET</td>
<td>14</td>
<td>528</td>
<td>514</td>
<td>0.50</td>
<td>0.29–0.71</td>
<td>4.7</td>
<td>&lt;0.001</td>
<td>&lt;0.01</td>
<td>0.08</td>
<td>0.69</td>
<td>181</td>
<td></td>
</tr>
<tr>
<td>Faux pas</td>
<td>11</td>
<td>383</td>
<td>330</td>
<td>0.57</td>
<td>0.40–0.73</td>
<td>6.8</td>
<td>&lt;0.001</td>
<td>0.43</td>
<td>0</td>
<td>0.02</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td>Hinting</td>
<td>7</td>
<td>379</td>
<td>329</td>
<td>0.47</td>
<td>0.28–0.66</td>
<td>4.9</td>
<td>&lt;0.001</td>
<td>0.22</td>
<td>0.02</td>
<td>0.14</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>FB stories</td>
<td>6</td>
<td>220</td>
<td>185</td>
<td>0.53</td>
<td>0.30–0.76</td>
<td>4.5</td>
<td>&lt;0.001</td>
<td>0.32</td>
<td>0.01</td>
<td>0.76</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>Remitted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ToM</td>
<td>18</td>
<td>545</td>
<td>475</td>
<td>0.50</td>
<td>0.35–0.66</td>
<td>6.0</td>
<td>&lt;0.001</td>
<td>0.08</td>
<td>0.04</td>
<td>0.80</td>
<td>242</td>
<td></td>
</tr>
<tr>
<td>RMET</td>
<td>10</td>
<td>337</td>
<td>273</td>
<td>0.40</td>
<td>0.15–0.65</td>
<td>3.2</td>
<td>0.004</td>
<td>0.03</td>
<td>0.08</td>
<td>0.67</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Faux pas</td>
<td>9</td>
<td>269</td>
<td>248</td>
<td>0.50</td>
<td>0.28–0.72</td>
<td>4.4</td>
<td>&lt;0.001</td>
<td>0.21</td>
<td>0.03</td>
<td>0.08</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Acute</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ToM</td>
<td>10</td>
<td>159</td>
<td>172</td>
<td>1.23</td>
<td>1.01–1.45</td>
<td>11.0</td>
<td>&lt;0.001</td>
<td>0.62</td>
<td>0</td>
<td>0.05</td>
<td>312</td>
<td></td>
</tr>
<tr>
<td>ToM (manic)</td>
<td>6</td>
<td>91</td>
<td>143</td>
<td>1.31</td>
<td>1.0–1.62</td>
<td>8.3</td>
<td>&lt;0.001</td>
<td>0.37</td>
<td>0.01</td>
<td>0.21</td>
<td>115</td>
<td></td>
</tr>
<tr>
<td>Subsyndromal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ToM</td>
<td>12</td>
<td>510</td>
<td>528</td>
<td>0.72</td>
<td>0.45–1.0</td>
<td>5.1</td>
<td>&lt;0.001</td>
<td>&lt;0.01</td>
<td>0.17</td>
<td>0.03</td>
<td>300</td>
<td></td>
</tr>
</tbody>
</table>

CI, Confidence interval; BP, bipolar disorder; HC, healthy controls; d, Cohen’s d; RMET, Reading the Mind in the Eyes Task; FB, false belief.

### Discussion

The current meta-analysis investigated ToM performance in a large sample of BP patients in comparison with healthy controls. Our findings showed that ToM ability was significantly impaired in BP in comparison to healthy participants. ToM dysfunction was significantly more pronounced in acute episodes but was also evident in remitted patients.

The effect size of ToM dysfunction ($d = 0.63$) in BP suggests a medium-sized impairment. ToM dysfunction in euthymic BP patients ($d = 0.50$) was comparable to most other neuropsychological deficits observed in the literature (Bora et al. 2009b). Acute phases of BP, especially manic episodes, was associated with exacerbation of ToM deficits and YMRS scores were significantly related to the severity of ToM impairment. Differences between manic and remitted patients were relatively large compared to most traditional neuropsychological domains in which evidence suggests only subtle effects of mood state on cognition (Kurtz & Gerraty, 2009). However, similar to ToM performance, decision making, impulsivity, insight and some aspects of neurocognition (i.e. commission errors in the continuous performance test) are significantly more impaired during mania in comparison to remission (Bora et al. 2006b; Clark & Sahakian, 2008). Mania-related functional changes in the ventral prefrontal cortex which has a critical role in reward, decision making and ToM, could potentially explain these findings (Rubinsztein et al. 2001). It could be argued that ToM dysfunction during manic episodes may be...
at least there might be partial overlap in deficits in neurocognition and social cognition (Bora et al. 2009c). Our findings suggest that neurocognitive deficits may significantly contribute to ToM deficits in BP given that meta-regression analyses found a significant relationship between ToM and general cognitive impairment. A number of previous studies also proposed that ToM deficits in BP might be secondary to cognitive deficits (Caletti et al. 2013; Lee et al. 2013; Thaler et al. 2013; Bora et al. 2005). For example, a previous study in euthymic BP also supports this notion, as the significant ToM dysfunction in this patient population was no longer significant following correction for working memory deficits (Bora et al. 2005). ToM impairment observed in schizophrenia can also be partly explained by neurocognitive deficits (Sergi et al. 2007; Bora et al. 2009a–c). However, in schizophrenia, it has been shown that neurocognition and social cognition are partly overlapping but relatively unique constructs. In BP, there is a need for further studies investigating the separability of neurocognition and social cognition.

Some authors suggest that BP is only associated with impairment in ‘cognitive’ but not in ‘affective’ ToM (Shamay-Tsoory et al. 2009; Montag et al. 2010). However, findings from the current meta-analysis do not support this suggestion, as BP was associated with deficits in both types of tasks. While the effect size for impairment in ‘cognitive’ ToM was moderately higher than ‘affective’ ToM, the difference between these two task types was not statistically significant. Contradictory findings might be related to low power of past studies. Donohoe et al. (2012), who had a considerably larger sample size than other past studies, was able to demonstrate a similar degree of impairment in both affective and cognitive ToM. Moreover, differences in some studies can reflect differences in task difficulties rather than task type. Another consideration for future research is the potential effect of history of psychosis on ToM deficits in BP. History of psychosis in BP has been associated with more severe cognitive deficits, especially in some executive abilities (planning and working memory) (Bora et al. 2010). However, we were not able to investigate the relationship between ToM impairment and psychosis in BP as only a few studies investigated the relationship between these variables.

In conclusion, ToM abilities are impaired in BP. ToM dysfunction is more pronounced during acute episodes, but deficits are also present in remission, suggesting that ToM impairment might be a trait-marker of BP. There is a need for first-episode studies and longitudinal studies comparing the developmental course of ToM and other social cognition abilities in individuals at risk for BP.

Supplementary material
For supplementary material accompanying this paper visit http://dx.doi.org/10.1017/S0033291715001993.

Acknowledgements
Prof Christos Pantelis was supported by a NHMRC Senior Principal Research Fellowship (ID 628386).

Declaration of Interest
Over the last 2 years, Christos Pantelis has participated on Advisory Boards for Janssen-Cilag and Lundbeck. He has received honoraria for talks presented at educational meetings organized by AstraZeneca, Shire, Janssen-Cilag and Lundbeck. The remaining authors have no conflict of interests to report.

References


and theory of mind in euthymic bipolar disorder. European Psychiatry 28 (Suppl. 1), 1262.


