



## Mentalizing in first-episode psychosis

Amélie M. Achim<sup>a,b,\*</sup>, Rosalie Ouellet<sup>a,c</sup>, Marc-André Roy<sup>a,b</sup>, Philip L. Jackson<sup>a,c</sup>

<sup>a</sup> Centre de Recherche de l'Institut Universitaire en Santé Mentale de Québec, Québec, QC, Canada

<sup>b</sup> Département de Psychiatrie et Neurosciences, Université Laval, Québec, QC, Canada

<sup>c</sup> École de Psychologie, Université Laval, Québec, QC, Canada

### ARTICLE INFO

#### Article history:

Received 9 December 2010

Received in revised form 7 October 2011

Accepted 11 October 2011

#### Keywords:

Recent-onset  
Schizophrenia  
Mentalizing  
Theory of mind  
Social cognition  
Emotion recognition  
Social knowledge

### ABSTRACT

Mentalizing deficits have often been observed in people with schizophrenia and a few recent studies suggest that such deficits are also present in patients with first episode psychosis (FEP). It is not clear, however, whether these mentalizing deficits in FEP can be accounted for by underlying processes such as social cue recognition, social knowledge and general reasoning. In this study, we assessed mentalizing abilities in 31 people with FEP and 31 matched controls using a novel, comprehensive mentalizing task validated through the present study. We also assessed social cue recognition, social knowledge and non-social (or general) reasoning performance in the same participants in order to determine if the mentalizing deficits in FEP can be at least partly explained by performance in these three underlying processes. Overall, the mentalizing task revealed the greatest impairment in FEP, an impairment that remained significant even after controlling for social cue recognition, social knowledge and non-social reasoning performance. Interestingly, non-social reasoning and social knowledge were both shown to contribute to mentalizing performance. In addition, social cognition measures were linked to social functioning in the FEP group, with the strongest correlation observed with mentalizing performance. Taken together, these results show that mentalizing is an aspect of social cognition that is particularly affected in FEP and might contribute to functional impairments in these patients. These deficits could be a prime target for cognitive remediation in FEP, and our results suggest that this could be done either directly or through improvement of related social and non-social cognitive skills such as social knowledge and general reasoning.

© 2012 Elsevier Ireland Ltd. All rights reserved.

### 1. Introduction

Social cognition can be defined as a group of processes that allow people to understand and interact with each other. The ability to attribute mental states to others, often referred to as mentalizing or theory of mind, is an important aspect of social cognition. In mentalizing tasks as in real life, mental states such as intentions, beliefs, knowledge or emotions are attributed based on all available sources of information about the person to whom these mental states are to be attributed and about the context in which that person evolves. Mentalizing judgments thus involve integrating several pieces of information in order to infer the appropriate mental state.

People with schizophrenia (SZ) generally present with important impairments in their ability to mentalize, i.e., to infer the mental state of a character presented in a given situation. Mentalizing deficits have been repeatedly reported in people with SZ (Sprong et al., 2007; Bora et al., 2009), along with other neurocognitive impairments (Heinrichs and Zakzanis, 1998), and it is now well established that

these deficits have a significant negative impact on social functioning and the quality of life of patients with SZ (Green, 1996; Green et al., 2000; Couture et al., 2006). However, most studies of social cognition in SZ have included patients in a chronic stage of illness and/or during active psychotic episodes (Sprong et al., 2007; Bora et al., 2009), and the mentalizing deficits observed in these patients could thus be linked to biases towards the recruitment of patients with a more chronic course of illness and a poorer outcome, effects of long-term medication use, transient perturbation due to positive symptoms, etc. In an attempt to address some of these biases, a few studies have measured mentalizing in remitted SZ patients, again revealing significant deficits in these patients, though of a lesser magnitude than those observed in acutely psychotic patients according to a recent meta-analysis (Bora et al., 2009). These studies in remitted patients suggest that mentalizing deficits represent a trait of SZ, instead of being fully linked to symptomatic states. Studies in remitted patients (e.g. Herold et al., 2002) however cannot account for the effect of long-term medication use or the long duration of psychosis and it also remains unclear when in the illness process mentalizing impairments occur. One approach to minimize the impact of illness duration, get a sample representative of diverse future outcomes and address the issue of when the deficits occur is to recruit patients early after the onset of psychosis. This is precisely what several

\* Corresponding author at: Centre de Recherche de l'Institut Universitaire en Santé Mentale de Québec (F-4500), 2601 de la Canardière, Québec, QC, Canada, G1J 2G3. Tel.: +1 418 663 5741, +1 418 663 9540.

E-mail address: [amelie.achim@crulrg.ulaval.ca](mailto:amelie.achim@crulrg.ulaval.ca) (A.M. Achim).

studies have done and these studies have also reported mentalizing deficits in people with first-episode psychosis (FEP) (Inoue et al., 2006; Bertrand et al., 2007), in line with the suggestion that these deficits could be a general characteristic of patients with SZ. The first FEP study on mentalizing by Inoue et al. (2006) relied on a single cartoon story in which three mentalizing questions were asked. This simple task revealed a greater percentage of patients than controls who responded incorrectly to one of the three questions. In a subsequent study (Bertrand et al., 2007), a mentalizing deficit in FEP was again observed using the Hinting task (Corcoran et al., 1995), which includes ten short stories from which participants have to infer a character's intentions. Other more global measures of social cognition that likely involve mentalizing abilities have also revealed significant deficits in FEP relative to healthy controls (Bertrand et al., 2007; Koelkebeck et al., 2010). With the recent focus on early intervention, and given the relationship between social cognition and social functioning, identifying and treating these mentalizing deficits early could have a significant positive impact on the functional outcome of people with SZ.

The mentalizing deficits observed in people with SZ could however result from difficulties at different points in the information processing chain that leads to mental states attributions (see Bless et al., 2004; Brunet-Gouet et al., 2011), including the ability to recognize social stimuli (social cue recognition) (Edwards et al., 2001; Kucharska-Pietura et al., 2005; Addington et al., 2006a; Kohler et al., 2010), the ability to construct and retrieve social representations (social knowledge/memory) (Cutting and Murphy, 1990; Addington et al., 2006b; Kee et al., 2009) and/or general reasoning/inferential mechanisms (Young and Bentall, 1997). The aims of the current study were thus 1) to replicate the results of mentalizing impairments in people with a FEP relative to control subjects using a novel, comprehensive mentalizing task; 2) to assess distinctively social cue recognition, social knowledge and general reasoning performance in the same two groups of participants; 3) to determine if mentalizing performance/deficits can be at least partly explained by performance in these underlying processes. These more elementary processes have not been concurrently examined in previous studies on mentalizing abilities in FEP and their assessment could provide valuable information about the pathways leading to mentalizing impairments in these patients. Since there are no standardized mentalizing tests as of yet and given that previous measures have often presented with ceiling effects and lack of sensitivity (e.g. Herold et al., 2002; Inoue et al., 2006), mentalizing was measured here with a comprehensive task developed for the purpose of this study, i.e. the Combined stories test, for which we also present initial psychometric properties.

## 2. Methods

### 2.1. Participants

Thirty-one participants with a FEP were recruited from the Clinique Notre-Dames-Victoires, a specialized outpatient clinic that offers comprehensive evaluation and treatment for young adults (18 to 35 years old) who are in the early stages of a psychosis. All patients presented with a DSM-IV SZ spectrum psychosis diagnosis (American Psychiatric Association, 1994) including SZ ( $n=23$ ), schizoaffective disorder ( $n=2$ ), delusional disorder ( $n=4$ ), and psychosis not otherwise specified ( $n=2$ ). Our decision to include patients with this range of diagnoses was based on our objective to include a sample representative of all patients with a SZ spectrum psychotic disorder and on previous reports that these diagnoses fall within the SZ spectrum when diagnoses are reassessed later in the course of the illness (Schimmelmann et al., 2005; Malla et al., 2006) or based on family studies (Kendler et al., 1995; Schimmelmann et al., 2005; Malla et al., 2006). Patients were excluded if they had a history of neurological disorder, if they presented an estimated IQ under 70 (based on the Wechsler Adult Intelligence Scale Third Edition [WAIS-III] Vocabulary and Block Design dyad; (Ringe et al., 2002)) or if they did not have an adequate understanding of French (having completed most schooling in French was considered as leading to adequate understanding whenever French was not the first language). All patients were taking a second-generation antipsychotic as their primary medication, with 17 taking quetiapine (mean dose = 770.6 mg), four taking olanzapine (mean dose = 15 mg), seven taking

risperidone either in oral (mean dose = 2.3 mg/day [three patients]) or intramuscular long-acting form (mean dose = 25 mg/2 weeks [three patients]; one patient was taking both formulations) and three taking a combination of quetiapine and another antipsychotic medication. Treatment had been initiated on average 20.9 months prior to the study (median = 13.3 months, range = 1 to 57). Though we favored patients with short illness duration (less than 24 months), we also included patients who had been followed for up to 60 months in order to get a more important sample size. Symptoms were assessed with the Positive and Negative Syndrome Scale (PANSS) (Kay et al., 1987) and level of social functioning was assessed with the Social and Occupational Functioning Assessment Scale (SOFAS) (American Psychiatric Association, 1994).

Thirty-one control participants were recruited from the community through ads in local stores, local media and through word of mouth. The exclusion criteria for the control group were the same as those of the FEP group, with the additional requirement of not presenting with a psychotic disorder or a cluster A personality disorder (as assessed through a clinical interview based on the Structured Clinical Interview for DSM-IV (SCID) (First et al., 1998)), and not having a first-degree relative with a psychotic disorder. Because we wanted to recruit control participants that were otherwise as representative as possible of the general population, we however did not exclude controls that our SCID-based assessment allowed us to identify as presenting with other axis I diagnoses. The control group thus included three participants that we identified as meeting the criteria for mild to moderate substance abuse or dependence and two participants identified as meeting the criteria for a specific phobia. None of the control subjects had received a diagnosis and none reported taking a psychoactive medication to control these symptoms.

The groups were matched in terms of age, gender and parental socio-economic background (Miller, 1991). There was no significant difference in estimated IQ between the two groups (see Table 1). After a complete description of the study, all participants signed a consent form in accordance with the local ethics committee requirements.

### 2.2. Material

#### 2.2.1. Mentalizing and non-social reasoning assessment

For the measure of mentalizing abilities, no standardized or validated tests are yet available and to the best of our knowledge, very few of the most widely used tests have been translated into French. Moreover, most of the available measures include a limited number of items and suffer from ceiling effects at least in the control group, which can be problematic when contrasting with the performance of a patient group. For these reasons and to increase the sensitivity of our measure, we herein used a combination of mentalizing stories that included items translated and adapted from several previous tests that are all well regarded and have been used often in the experimental psychology literature. These include the 'Hinting task' (Corcoran et al., 1995), the 'False Belief task' (Baron-Cohen, 1989; Frith and Corcoran, 1996), the 'Faux-pas test' (Baron-Cohen et al., 1999) and the 'Strange Stories Test' (Happé, 1994). These tests were targeted so as to have items that cover attributions of a full range of mental states,

**Table 1**  
Demographic and clinical data.

	FEP	Controls	$p^a$
<i>Demographic data</i>			
<i>n</i>	31	31	
Gender (men/women)	26/5	26/5	
Age (mean, S.D.)	24.9 (4.5)	25.2 (4.2)	NS
SES score (mean, S.D.)	50.7 (18.2)	48.6 (14.2)	NS
SES category (mean, S.D.)	3.7 (1.2)	3.6 (1.0)	NS
Estimated IQ (mean, S.D.)	100.4 (15.1)	101.8 (10.5)	NS
Education category <sup>b</sup>	4.0 (1.1)	3.3 (1.2)	0.03
<i>Clinical data</i>			
PANSS positive (mean, S.D.)	15.1 (4.8)		
PANSS negative (mean, S.D.)	16.0 (5.9)		
PANSS general (mean, S.D.)	32.0 (7.0)		
SOFAS (mean, S.D.)	58.1 (12.1)		
Patient status (outpatient/inpatient) <sup>c</sup>	28/3		
Duration of illness in months (mean/median)	20.9/13.3		

S.D. = standard deviation.

SES = socio-economic status.

PANSS = Positive and Negative Syndrome Scale.

SOFAS = Social and Occupational Functioning Assessment Scale.

<sup>a</sup> The demographic variables were contrasted between groups using bilateral two sample t-tests.

<sup>b</sup> According to Hollingshead's categories, adapted for Quebec: 1 = postgraduate, 2 = bachelor degree, 3 = CEGEP degree, 4 = high school or equivalent, 5 to 7 = decreasing levels of partial high school. Note that this information was available for all control subjects but for only 21 of the FEP patients.

<sup>c</sup> The inpatients were stabilized and about to be discharged at the time of testing.

including attributions of beliefs/knowledge, of intentions/desires and of emotions. Each item consisted of a short story depicting at least two characters evolving in a specific situation. The ensuing task (called the “Combined stories test”) took approximately 30 min to complete.

A total of 20 second-order mentalizing stories were included in this Combined stories test; each of these items involved interactions between at least two characters. Second-order mentalizing was here defined as the attribution of a mental state about another character's mental state, in contrast to first-order mentalizing, which involves a mental state about the state of the physical world. Our second-order mentalizing items involved one character having a belief/knowledge, intention/desire or emotion in relation to another character's belief/knowledge, intention/desire, emotion or action/verbalization. For example, the items derived from the ‘False Belief task’ involved a character having a (false) belief about another character's belief or action (e.g. X thinks Y did something), the items from the hinting task involved a character's intention about another character's action (e.g. X wants Y to do something), the items from the faux-pas tasks involved a character's emotion about another character's action/verbalization (e.g. X is upset about something that Y said), etc. These second-order mentalizing items were purposely selected so to include a variety of mental states combinations. Of these combinations, only the ‘belief about belief’ items (i.e. the False Belief items) could be considered as purely cognitive mentalizing, whereas all other combinations included emotions and/or intentions/desires in at least one of the characters, and could thus be viewed as more affective mentalizing. All original stories were translated in French, and when necessary adapted to the reality of young adults (e.g., a faux-pas story originally taking place at the playground was presented as happening in the kitchen, while the rest of the situation, including the verbalizations, was consistent with the original). The items that were included in our test were thus selected by prioritizing those that were the easiest to adapt to the socio-cultural reality of our participants (French-Canadian young adults). The final version of the Combined stories test included six second-order mentalizing stories from each of the original test sources, except for the ‘False Belief task’ for which only two second-order false belief stories were available. This restricted number of belief items was thus compensated by also including the belief questions from ‘Faux-pas test’ (Baron-Cohen et al., 1999).

For each item, participants were asked to read the story aloud and were then asked open questions that required making inferences about the characters' mental states (mentalizing questions). The text remained in front of the participant during questioning and participants were encouraged to go back to the text whenever they felt the need (to minimize the memory load of the task). For six of the 20 mentalizing stories two mentalizing questions were asked, whereas only one mentalizing question was asked for the other 14 stories, for a total of 26 answers. The answers were taken verbatim and then scored 2, 1 or 0 point according to a pre-determined correction grid. A correct, complete answer was attributed 2 points, an incomplete answer was allowed 1 point and an incorrect answer was allowed 0 point.

Several control conditions were also integrated in the test: 1) items to test for general reasoning abilities (non-social reasoning questions); 2) items to test for the general capacities to link a mental state to a behavior (first-order inference questions); and 3) questions to control for attention and memory effects (attention/memory questions). The items testing for general reasoning capacities focused on six non-social reasoning stories (4 adapted from Happe et al. (Happe, 1994) and 2 created by us) and required drawing inferences about physical causalities. Similarly to the mentalizing questions, these items were scored 2, 1 or 0 point according to a pre-determined correction grid. The first-order inference items testing for the general capacity to link a mental state to a behavior were three simple first-order false belief stories (two from the literature (Baron-Cohen, 1989; Frith and Corcoran, 1996) and one created by us

using the same item structure). In these stories a character acts on a false belief about the state of the physical world. These questions do not involve interactions between the characters and require simpler answers. They were thus scored 1 point for a correct answer or 0 point for an incorrect answer. The rationale for scoring these items independently from the other mentalizing items is that these First-order items might require less integration of information and could simply rely on the knowledge that people act according to what they know or do not know. In addition, it is worth noting that since there were very few mistakes on these items in either group (see Results section and Table 2), including them in the mentalizing score would not have had a noticeable impact on the statistical tests. For these reasons, and to keep the mentalizing score representative of the ability to infer mental states in the context of social interactions (rather than also including a character's non-social interactions with the physical world), we chose to report these items independently from the second-order mentalizing questions. Finally, the attention/memory questions simply inquired about a detail of each story and were scored 1 point for a correct answer and 0 points for a wrong answer. These questions were not meant to measure memory or attention capacities per se, but rather to verify that our procedure (stories remained in front of the participant) was successful at minimizing the effects of potential attention or memory deficits on mentalizing performance. These items were always given after the mentalizing, the non-social reasoning or the First-order inference questions, to make sure that the participants had paid attention and were able to remember and/or seek for details of the story.

Except for the attention/memory questions, which were asked for every story, the items/stories from the other categories (i.e., the different types of mentalizing stories, the non-social reasoning stories and the first-order inference stories) were intermixed throughout the test. A practice story was also presented before the test, and a mentalizing question and a control question were then asked and participants were given feedback on their answers for that story only. This was done to familiarize them with the material and to make sure that they understood the expected type of answer. The answer to this question was not included in the test score.

In sum, the Combined stories test produces four measures, including the main mentalizing score (second-order mentalizing) as well as three control measures, namely non-social reasoning, attention/memory and first-order inference scores.

Because our mentalizing test represents a new measure, even if it was created based on existing material, we also administered another well-known measure of mentalizing, namely Sarfati's cartoon task (Sarfati et al., 2003) to assess convergent validity for our test. Sarfati's cartoon task mainly assesses a character's intentions (and rarely involves a character's belief or emotion), and seems limited by ceiling effects in the control group, but it was considered to be a good option as a validating criterion because it has repeatedly been used in SZ and it does not involve verbal material which would have required a translation that could have affected the validity of the measure. In addition to assessing our test's convergent validity, internal consistency and inter-rater reliability were also considered (see results below).

### 2.2.2. Social cue recognition and social knowledge

Social cue recognition and social knowledge tests were also administered to determine whether or not our FEP patients presented with impairments in these aspects of social cognition and to determine if such impairments can partly explain the mentalizing deficits previously reported in these patients (Inoue et al., 2006; Bertrand et al., 2007).

Social cue recognition was assessed by presenting standardized emotional facial stimuli from the Ekman and Friesen series (Ekman and Friesen, 1976) and asking participants to select the corresponding emotion from a given list of labels (i.e., happy, surprise, sad, angry, disgust, fear or neutral). There were a total of 14 items, scored 1

**Table 2**  
Social cognition performance.

	FEP		Controls		Group difference
	Mean (S.D.)	Range	Mean (S.D.)	Range	
Combined stories: mentalizing (/52)	38.9 (6.8)	25–48	43.6 (4.3)	34–50	0.001
Combined stories: non-social reasoning (/12)	10.5 (1.5)	7–12	11.2 (1.1)	7–12	0.050 <sup>a</sup>
Combined stories: first-order (/3)	2.9 (0.3)	2–3	2.8 (0.4)	2–3	NS <sup>a</sup>
Combined stories: att./mem. (/29)	28.7 (0.5)	27–29	28.8 (0.4)	28–29	NS <sup>a</sup>
Social knowledge test (/14)	10.4 (1.4)	7–13	11.0 (1.8)	8–14	0.14
Social cue recognition test (/14)	10.6 (1.8)	7–13	11.0 (1.5)	8–14	NS
Sarfati's cartoon task (/28)	25.9 (1.9)	21–28	26.5 (1.3)	23–28	NS <sup>a</sup>
Combined stories: subscores from different original references					
False beliefs (/4)	3.1 (1.2)		3.7 (0.7)		0.018 <sup>a</sup>
Faux-pas false beliefs (/12)	7.9 (3.3)		9.7 (2.2)		0.005 <sup>a</sup>
Faux-pas identification (/12)	6.7 (3.0)		8.8 (2.6)		0.043 <sup>a</sup>
Hinting (/12)	9.6 (3.0)		11.45 (1.0)		0.151 <sup>a</sup>
Strange stories (/12)	9.0 (3.0)		10.1 (1.6)		0.001 <sup>a</sup>

NS = not significant.

<sup>a</sup> From non-parametric Mann–Whitney *U*-tests (rank sum tests) instead of standard *t*-tests due to non-normal distributions and/or lack of variability.

point for a correct answer or 0 points for an incorrect answer. Deficits have previously been reported in FEP for similar social cue recognition tasks (Addington et al., 2006a).

Because we could not find measures of social knowledge about mental states in the literature, the task used for this purpose was developed in our lab. The task assessed social knowledge by presenting hypothetical situations (these original situations were inspired by the social themes listed in Blair and Cipolotti (2000)) and then asking participants how people in general would feel in the situation. For example, one item was someone who learns he (or she: the way it is formulated in French does not suggest a specific gender) has been lied to. After the experimenter read the situation, participants gave open responses (most often a single word) that were later scored 1 or 0 points according to a pre-determined correction grid. The correction grid was based on a pre-test with ten healthy subjects and supplemented with several examples of answers (mostly from the current control group) that were considered correct and incorrect. There were a total of 14 items.

Several responses could be considered a correct answer but these were all related. Although the task was pre-tested in 10 healthy young adults, its psychometric properties had not been established prior to this study. It however represented an important step towards the measure of a construct that, though recognized as being an important aspect of social cognition processing (Bless et al., 2004; Brunet-Gouet et al., 2011; Green et al., 2008), is often neglected in studies of social cognition in SZ or other disorders.

### 2.3. Statistical analyses

The psychometric properties of the Combined stories test were first assessed. After examining the normality of the distributions to identify eventual ceiling effects, the following properties were determined: 1) convergent validity through a correlation of our mentalizing score with that of Sarfati's cartoon task; 2) inter-rater reliability through a correlation between initial scoring and scoring by an independent rater; 3) internal consistency through a Cronbach's alpha test performed on the scores from the different initial test sources.

Some psychometric properties of the Social knowledge test were also assessed, including 1) convergent validity with our mentalizing task; 2) inter-rater reliability through a correlation between initial scoring and scoring by an independent rater; 3) internal consistency through a KR-20 test (for dichotomous variables) performed on the scores of all 14 tests items.

Then, *t*-tests were used to contrast the two groups, FEP and control, for the normally distributed social cognition measures, whereas Mann-Whitney rank sum tests were used for the measures not meeting that condition. The between-group difference on the Combined stories test's mentalizing scores was thereafter reassessed, this time including non-social reasoning, social knowledge and social cue recognition as covariates in the analysis. Effects reaching an alpha level of 0.05, two-tailed, were considered significant.

Finally, we also explored the correlation pattern between social cognition performance on our three social cognition tasks in the FEP group and social functioning as assessed with the SOFAS, as well as with duration of illness. Because these were exploratory analyses, correlations meeting an alpha level of 0.05 were considered significant, but these results from multiple exploratory comparisons should be interpreted with caution.

## 3. Results

### 3.1. Validation of the Combined stories test and the Social knowledge test

#### 3.1.1. Combined stories test

A first step for the validation consisted of examining the normality of the score distributions in the FEP group and the control group for our Combined stories test and for Sarfati's cartoon task. There was no evidence of significant skewness for our Combined stories test but significantly skewed distributions were observed in both groups for Sarfati's cartoon task even after removing one outlier in the FEP group (skewness  $z = -2.2$  in FEP and  $z = -2.6$  in controls). The pattern of results for Sarfati's cartoon task suggested a ceiling effect for that task, which was particularly prominent in the control group with two-thirds of the group presenting with zero or a single error.

In a second step, we examined the convergent validity of our Combined stories test by assessing its correlation with Sarfati's cartoon task, which revealed a significant association when assessed across both groups while controlling for between group effects through a partial correlation ( $r(58) = 0.424$ ,  $p = 0.001$ ). Since the regression slopes were significantly different between groups ( $F(1,57) = 8.44$ ,  $p = 0.005$ ), correlations were thereafter assessed separately in each group. In the FEP group, a significant correlation was observed between the two tests ( $r(30) = 0.58$ ,  $p = 0.001$ ), although no significant

relationship could be detected in the control group ( $r(31) = 0.08$ ,  $p = 0.65$ ). Given that the absence of a correlation in the control group was more likely linked to the restricted distribution of scores on Sarfati's task (i.e., ceiling effect), the overall pattern of results suggests adequate convergent validity for our task.

A third step was to examine the inter-rater reliability of our Combined stories test. After the initial scoring (performed by AMA), we asked a research assistant that had not been previously involved in the study and was blind to initial scoring to independently re-score each item of the test for all 31 patients using the same standardized procedure, i.e. based on the verbatim of the responses and on the information provided in the correction grid. Convergent validity was not examined in the control group because the predetermined correction grid was supplemented with examples of verbal answers from that group. The protocols for 10 control subjects were nonetheless intermixed with those of the patients to keep the assistant blind to diagnosis while she scored the protocols. This step of examining the inter-rated reliability for our 31 patients, which was pivotal to determine whether our correction grid was detailed enough to allow consistent correction of the items based on the participants open verbalizations, yielded excellent inter-rater reliability ( $r(31) = 0.98$ ,  $p < 0.001$ ).

Lastly, internal consistency of the test was assessed to determine whether items borrowed from the different tasks yielded significantly heterogeneous performances. Across all participants, a good internal consistency between the sources was observed with a Cronbach's alpha of 0.81.

#### 3.1.2. Social knowledge test

The scores were also normally distributed in both groups for the social knowledge test. Convergent validity was examined via the expected relationship between social knowledge and mentalizing, through a partial correlation controlling for group. This association was found to be significant ( $r = 0.44$ ,  $p < 0.001$ ) and the regression slopes did not significantly differ between the two groups.

Inter-rater reliability was excellent ( $r = 0.97$ ,  $p < 0.001$ ), with only one patient for which one item was re-scored differently, attesting that our scoring grid allowed consistent scoring of the participant's open responses.

Internal consistency as assessed through a KR-20 test showed a relatively modest stability between the different test items (KR-20 = 0.24). This result is not too surprising given the all-or-none scoring scheme of our test, but could also reflect that our test measures several distinct components of social knowledge (ex: different mental states). Using a larger participant sample could eventually help to determine whether the current test items cluster into a certain number of components that each show higher internal consistency.

### 3.2. Social cognition in first-episode psychosis patients versus controls

Performance and between group statistics for each tasks are reported in Table 2. When contrasting the performance of the two groups separately for each test, a highly significant between group effect was observed for performance on the mentalizing questions ( $t(60) = 3.28$ ,  $p = 0.002$ , effect size  $r$  (ESr)<sup>1</sup> = 0.39), and a significant between group effect was also observed for the non-social reasoning questions ( $U = 348.0$ ,  $Z = 1.96$ ,  $p = 0.050$ , ESr = 0.25). No significant between group differences were found for the social knowledge test ( $t(60) = 1.49$ ,  $p = 0.143$ , ESr = 0.19), the social cue recognition task ( $t(60) = 0.99$ ,  $p = 0.329$ , ESr = 0.13), Sarfati's cartoon task ( $U = 389.5$ ,  $Z = 1.13$ ,  $p = 0.258$ , ESr = 0.14) or the other two control

<sup>1</sup> Note that effect sizes  $r$  vary between  $-1$  and  $1$ . An  $r = 0.50$  is considered as a large effect size (ES) large and corresponds to a Cohen's  $d = 1.15$ . An  $r = 0.30$  is considered a medium ES and corresponds to a Cohen's  $d = 0.63$ . An  $r = 0.10$  is considered a small ES and corresponds to a Cohen's  $d = 0.20$ .

**Table 3**  
Results from the univariate analysis of covariance (ANCOVA).

Source of variance	<i>F</i> (1,57)	<i>p</i>
<i>Covariates</i>		
Non-social reasoning	4.48	0.039
Social knowledge	12.0	0.001
Social cue recognition	0.20	NS
<i>Independent variable</i>		
Group	5.07	0.028

NS = not significant.

conditions from the Combined stories test, namely the first-order inference questions ( $U = 434.0$ ,  $Z = -1.07$ ,  $p = 0.283$ ,  $ESr = -0.14$ ) and memory/attention questions ( $U = 446.0$ ,  $Z = 0.64$ ,  $p = 0.523$ ,  $ESr = 0.08$ ).

In a second and more crucial step because non-social reasoning, social knowledge and social cue recognition are all suggested to contribute to higher order mentalizing capacities, we used a univariate analysis of covariance (ANCOVA) to test whether these three measures had a significant impact on mentalizing performance, and whether the between group difference on our mentalizing test was still observed after entering these three measures as covariates for the between group analysis. As shown in Table 3, the ANCOVA revealed a significant impact of non-social reasoning ( $F(1,57) = 4.48$ ,  $p = 0.039$ ) and of social knowledge ( $F(1,57) = 12.04$ ,  $p = 0.001$ ) on mentalizing performance (i.e. these variables are significant predictors of mentalizing performance), suggesting that both these variables contribute to mentalizing. However, a significant between group effect on mentalizing performance was still observed with this analysis ( $F(1,57) = 5.07$ ,  $p = 0.028$ ,  $ESr = 0.28$ ), meaning that the mentalizing deficit observed in our patients could not be fully explained by performance on the other three measures.

Redoing this whole set of analyses after excluding the five control subjects that were identified to fulfill DSM-IV criteria for a specific phobia or for substance abuse/dependence did not change the pattern of results, with the exception that the direct group comparison no longer reached significance for the non-social reasoning measure ( $p = 0.080$ , instead of 0.050).

### 3.3. Correlations between our social cognition tests and the clinical characteristics of the FEP patients

We also explored the pattern of correlations between performance on our tests in the FEP group and social functioning as assessed with the SOFAS. As displayed in Table 4, significant positive correlations were observed both with the mentalizing task ( $r = 0.45$ ,  $p = 0.011$ ) and with the social knowledge task ( $r = 0.37$ ,  $p = 0.038$ ), whereas the correlation with social cue recognition task did not reach significance ( $r = 0.31$ ,  $p = 0.095$ ). As for the duration of illness, a significant negative relationship was observed with the social cue recognition test ( $r = -0.36$ ,  $p = 0.049$ ), but not with mentalizing

**Table 4**  
Correlation of social cognition performance with functioning and duration of illness in patients with first episode psychosis (FEP).

	SOFAS	Duration of illness
Mentalizing ( $N = 31$ )	$r = 0.45$ $p = 0.011^*$	$r = -0.01$ NS
Social knowledge ( $N = 31$ )	$r = 0.37$ $p = 0.038^*$	$r = 0.13$ NS
Social cue recognition ( $N = 31$ )	$r = 0.31$ $p = 0.095$	$r = -0.36$ $p = 0.049^*$

NS = not significant ( $p > 0.1$ ).\* Significant at  $p < 0.05$ . But note that none of these exploratory correlations would have survived a Bonferroni correction for multiple comparisons.

( $r = -0.01$ ,  $p = 0.940$ ) or social knowledge ( $r = 0.13$ ,  $p = 0.486$ ). Note however that none of the significant correlations from these exploratory correlation analyses would have survived a Bonferroni correction for multiple comparisons; these results should thus be interpreted with caution.

## 4. Discussion

This study suggests that mentalizing is an aspect of social cognition that is particularly impaired in FEP patients, with a highly significant between group difference (Effect size  $r = 0.39$ , equivalent to a Cohen's  $d = 0.84$ ). The magnitude of the between group difference observed in this study was in fact in the same range in terms of its effect size than what has previously been reported in SZ patients with a longer duration of illness who, as was the case with the patients in this study, were outpatients or inpatients about to be discharged at the time of testing (Bora et al., 2009). Moreover, the between group difference remained significant even after controlling for performance on other measures including social cue recognition, social knowledge as well as non-social reasoning. This study thus confirms that FEP patients do present specific mentalizing impairments similar to those observed in patients with a longer duration of illness, at least for second-order mentalizing from situations of social interactions. The fact that FEP outpatients were recruited for this study (except for a few inpatients waiting to be discharged) further suggests that these second-order mentalizing deficits represent a trait of the illness and that such deficits can be observed early after illness onset.

Unlike some of the previously published SZ studies (Sprong et al., 2007; Bora et al., 2009), however, we did not observe a deficit on the first-order inference items in our FEP patients. Contrary to the second-order mentalizing items, these first-order items do not require that participants understand interactions between the story characters. Our study thus suggests that the mentalizing deficit in FEP patients is specific to mentalizing about social interactions. This observation is consistent with the previous FEP study by Inoue and collaborators (Inoue et al., 2006) that reported a significant impairment only for their second-order mentalizing question, and not for their first-order mentalizing question. Taken together, these results could reflect that first-order mentalizing abilities deteriorate with the progression of SZ or that patients with impaired first-order abilities are overrepresented in samples of people with chronic schizophrenia (which could be the case if these patients have more regular contacts with psychiatric services than those individuals that do not show such impairments). Longitudinal studies would however be required to clarify this question.

We also failed to observe a significant social cue recognition deficit in our FEP group, contrary to a few previous FEP studies. It should however be noted that the only two studies that used social cue recognition tests that were methodologically similar to the one used here, i.e. that presented a series of emotional faces and asked participants to identify the basic expressed emotion, reported effect sizes of low or medium magnitude in FEP (Edwards et al., 2001; Addington et al., 2006a). Of those two previous studies, a significant between group difference was only observed in the study that included larger groups of participants and thus had greater statistical power (Addington et al., 2006a). The significant relationship observed between social cue recognition and duration of illness in our study further suggests that these deficits are more subtle in the early stages of psychosis. Other studies using a wider range of facial expressions (i.e. not restricted to basic emotions; (Kucharska-Pietura et al., 2005)), or assessing emotion recognition through emotional prosody (Edwards et al., 2001; Kucharska-Pietura et al., 2005) in FEP have observed deficits of greater magnitude in terms of their effect sizes. Taken together, these results suggest that while no significant social cue recognition deficits were detected in the current study, social cue recognition can be affected in FEP patients. Indeed, a deficit might have been

observed in our patient group had we used a more sensitive social cue recognition measure and/or larger groups of participants.

To our knowledge, social knowledge about mental states has not previously been assessed in FEP and though we did not observe a significant deficit, performance on that test was observed to significantly contribute to mentalizing performance. This suggests that the mentalizing deficits observed in our patients could be more pronounced in patients with poorer social knowledge, a relationship that should be taken into account when considering social cognition remediation strategies, especially since some aspects of social knowledge have been found to be related to improvement in functional outcome following cognitive enhancement therapy in people with SZ (Eack et al., 2011). Taken together with these previous results, the pattern of results observed in this study argues for the importance of considering social knowledge about mental states when assessing social cognition in clinical settings.

Non-social reasoning performance, which was significantly impaired in our FEP patients, was also observed to play a role in explaining mentalizing performance, suggesting that the deficits observed on such social cognition measures are not fully independent from other non-social cognitive processes. Non-social reasoning deficits have previously been reported in SZ (e.g. Young and Bentall, 1997) and the idea that non-social neurocognitive processes can contribute to social cognition performance is not new. For example, it has been recognized by Addington et al. (2006b), who showed that social cognition acts as a mediator between basic cognition and functioning in FEP patients. Consistent with that previous study, we also observed significant positive relationships between our social cognition measures and social functioning in our FEP group, attesting to the functional significance of identifying such deficits early in the SZ illness process.

#### 4.1. Strengths and limitations of our study

One of the strengths of this study is that we relied on a well-controlled, comprehensive mentalizing task for which we observed good psychometric properties. Mentalizing tasks used in the literature often suffer from ceiling effects at least in the control group. The inclusion of more items, along with the open questions strategy and the scoring scheme on a scale from zero to two points per item, allowed additional variability in scores and made it possible to perform parametric analyses on our mentalizing data, which can be more informative than relying on criterion-based measures. Because no standard measure of mentalizing is yet available, reporting such psychometric properties could also be an important step toward identifying adequate measures for future studies and eventually for clinical evaluations.

One limitation of the study is that our mentalizing task relied purely on verbal descriptions and written presentation of the character's verbalizations and thus did not call for much social recognition processing. This may have restricted our ability to observe a link with our independent social cue recognition measure. It would thus be interesting to assess the relationship between mentalizing and social cue recognition using a mentalizing task that is more ecological, i.e. that relies on a more complete set of social cues, including perceptual cues.

Another limitation that could eventually be circumscribed by using a more ecological task is the fact that social cue recognition and social knowledge were assessed with material that was independent from the material presented in the context of the mentalizing task (i.e. the combined stories). The relationship between the two underlying social cognition processes of interest, social cue recognition and social knowledge, could thus only be tested indirectly. A more direct assessment of this relationship would have required using the same material as in the mentalizing test for the evaluation of social cue recognition and social knowledge.

It could also have been interesting to widen the focus of the social cue recognition and the social knowledge tests, which were restricted in terms of the range of mental states being assessed, (i.e. focused primarily on emotions). Knowledge of other mental states such as intentions/desires, knowledge/beliefs and/or other more complex emotions were thus overlooked. General social knowledge questions targeting these other types of mental states could certainly be added to the social knowledge task in the future. Other aspects of social knowledge that are not specifically related to mental states (e.g. social schemas that relate more to standard action sequences) might also have an influence on mentalizing and could also be interesting to include in future studies. For social cue recognition, though a person's knowledge or intentions is less directly assessed through a contextual observation of social cues, the ability to identify eye gaze direction, certain body movements or certain aspects of prosody can contribute to real life capacities to infer mental states in others. Adding some probes to assess these abilities in future studies could thus provide more complete information on the range of difficulties leading to mentalizing impairments in FEP patients.

Another limitation of this study is that we did not include a complete set of neuropsychological tests, which could have contributed to highlight potential relationships with our social cognition measures.

As for patient sampling, recruiting patients from a first-episode psychosis clinic offers several advantages, including having a patient sample that is more representative of the diversity of future outcomes. Our decision to include patients with durations of illness of up to 5 years (range 1 to 57 months) could however have limited the interpretation of our results as being specific to first-episode samples, though the main mentalizing measure showed no evidence of a relationship with illness duration in our patients.

#### 4.2. Conclusion

Even after controlling for other social and non-social cognitive abilities, mentalizing as assessed with our comprehensive task was significantly affected in our FEP group. Although our social cognition measures showed positive correlations with FEP patients' social functioning, mentalizing showed the strongest and most significant correlation, attesting to its importance and the need to assess the presence of these deficits and offer targeted interventions (i.e. interventions focused on the specific social cognition difficulties of each patient) early in the treatment to favor a positive outcome. This seems particularly relevant among FEP patients, as they are usually young adults who are at important life junctures, such as choosing and establishing a career, founding a family, etc., when the psychosis occurs. Preventing negative impacts of the illness and promoting a more positive functional outcome for these patients through remediation of mentalizing difficulties could have a major positive influence on these people's lives.

#### Acknowledgments

This work was supported by a NARSAD Young Investigator Award to PLJ, a Canadian Institutes of Health Research (CIHR) operating grant (#MOP-77647) to MAR, and a CIHR postdoctoral fellowship to AMA. PLJ was supported by a salary grants from the Fonds de la Recherche en Santé du Québec (FRSQ) and the CIHR. RO was supported by studentships from CIHR and FRSQ. We would like to thank the clinical and research team from Clinique Notre-Dame des Victoires for their interest and support with this study, and Marie-Audrey Lavoie who re-scored the mentalizing task for the inter-rater validation.

#### References

- Addington, J., Saeedi, H., Addington, D., 2006a. Facial affect recognition: a mediator between cognitive and social functioning in psychosis? *Schizophrenia Research* 85, 142–150.

- Addington, J., Saeedi, H., Addington, D., 2006b. Influence of social perception and social knowledge on cognitive and social functioning in early psychosis. *British Journal of Psychiatry* 189, 373–378.
- American Psychiatric Association, 1994. *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. American Psychiatric Association Press, Washington D.C.
- Baron-Cohen, S., 1989. The autistic child's theory of mind: a case of specific developmental delay. *Journal of Child Psychology and Psychiatry* 30, 285–297.
- Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., Plaisted, K., 1999. Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *J Autism Dev Disord* 29, 407–418.
- Bertrand, M.C., Sutton, H., Achim, A.M., Malla, A.K., Lepage, M., 2007. Social cognitive impairments in first episode psychosis. *Schizophrenia Research* 95, 124–133.
- Blair, R.J., Cipolletti, L., 2000. Impaired social response reversal. A case of 'acquired sociopathy'. *Brain* 123 (6), 1122–1141.
- Bless, H., Fiedler, K., Strack, F., 2004. *Social Cognition: How Individuals Construct Social Reality*. Taylor & Francis, New York.
- Bora, E., Yucel, M., Pantelis, C., 2009. Theory of mind impairment in schizophrenia: meta-analysis. *Schizophrenia Research* 109, 1–9.
- Brunet-Gouet, E., Achim, A.M., Vistoli, D., Jackson, P.L., Hardy-Baylé, M.-C., Passerieux, C., 2011. The study of social cognition with neuroimaging as a means to explore future directions of deficit evaluation in schizophrenia? *Psychiatry Research* 190, 23–31.
- Corcoran, R., Mercer, G., Frith, C.D., 1995. Schizophrenia, symptomatology and social inference: investigating "theory of mind" in people with schizophrenia. *Schizophrenia Research* 17, 5–13.
- Couture, S.M., Penn, D.L., Roberts, D.L., 2006. The functional significance of social cognition in schizophrenia: a review. *Schizophrenia Research* 32, S44–S63.
- Cutting, J., Murphy, D., 1990. Impaired ability of schizophrenics, relative to manics or depressives, to appreciate social knowledge about their culture. *British Journal of Psychiatry* 157, 355–358.
- Eack, S.M., Pogue-Geile, M.F., Greenwald, D.P., Hogarty, S.S., Keshavan, M.S., 2011. Mechanisms of functional improvement in a 2-year trial of cognitive enhancement therapy for early schizophrenia. *Psychological Medicine* 41, 1253–1261.
- Edwards, J., Pattison, P.E., Jackson, H.J., Wales, R.J., 2001. Facial affect and affective prosody recognition in first-episode schizophrenia. *Schizophrenia Research* 48, 235–253.
- Ekman, P., Friesen, W.V., 1976. *Pictures of Facial Affect*. Consulting Psychologists Press, Palo Alto.
- First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W., 1998. *Structured clinical interview for DSM-IV patient edition (SCID-I/P V and SCID-I/NP Version 2.0)*. Biometric Research Department, New York State Psychiatric Institute, New York, NY.
- Frith, C.D., Corcoran, R., 1996. Exploring 'theory of mind' in people with schizophrenia. *Psychological Medicine* 26, 521–530.
- Green, M.F., 1996. What are the functional consequences of neurocognitive deficits in schizophrenia? *The American Journal of Psychiatry* 153, 321–330.
- Green, M.F., Kern, R.S., Braff, D.L., Mintz, J., 2000. Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the "right stuff"? *Schizophrenia Bulletin* 26, 119–136.
- Green, M.F., Penn, D.L., Bentall, R., Carpenter, W.T., Gaebel, W., Gur, R.C., Kring, A.M., Park, S., Silverstein, S.M., Heinssen, R., 2008. Social cognition in schizophrenia: an NIMH workshop on definitions, assessment, and research opportunities. *Schizophrenia Bulletin* 34, 1211–1220.
- Happe, F.G., 1994. An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders* 24, 129–154.
- Heinrichs, R.W., Zakzanis, K.K., 1998. Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology* 12, 426–445.
- Herold, R., Tenyi, T., Lenard, K., Trixler, M., 2002. Theory of mind deficit in people with schizophrenia during remission. *Psychological Medicine* 32, 1125–1129.
- Inoue, Y., Yamada, K., Hirano, M., Shinohara, M., Tamaoki, T., Iguchi, H., Tonooka, Y., Kanba, S., 2006. Impairment of theory of mind in patients in remission following first episode of schizophrenia. *European Archives of Psychiatry and Clinical Neurosciences* 256, 326–328.
- Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin* 13, 261–276.
- Kee, K.S., Horan, W.R., Salovey, P., Kern, R.S., Sergi, M.J., Fiske, A.P., Lee, J.H., Subotnik, K.L., Nuechterlein, K., Sugar, C.A., Green, M.F., 2009. Emotional intelligence in schizophrenia. *Schizophrenia Research* 107, 61–68.
- Kendler, K.S., Neale, M.C., Walsh, D., 1995. Evaluating the spectrum concept of schizophrenia in the Roscommon Family Study. *American Journal of Psychiatry* 152, 749–754.
- Koelkebeck, K., Pedersen, A., Suslow, T., Kueppers, K.A., Arolt, V., Ohrmann, P., 2010. Theory of mind in first-episode schizophrenia patients: correlations with cognition and personality traits. *Schizophrenia Research* 119, 115–123.
- Kohler, C.G., Walker, J.B., Martin, E.A., Healey, K.M., Moberg, P.J., 2010. Facial emotion perception in schizophrenia: a meta-analytic review. *Schizophrenia Bulletin* 36, 1009–1019.
- Kucharska-Pietura, K., David, A.S., Masiak, M., Phillips, M.L., 2005. Perception of facial and vocal affect by people with schizophrenia in early and late stages of illness. *British Journal of Psychiatry* 187, 523–528.
- Malla, A., Norman, R., Schmitz, N., Manchanda, R., Bécharé-Evans, L., Takhar, J., Haricharan, R., 2006. Predictors of rate and time to remission in first-episode psychosis: a two-year outcome study. *Psychological Medicine* 36, 649–658.
- Miller, D.C., 1991. Hollingshead's two-factor index of social position. In: *Handbook of research design and social measurement*, 5th ed. Sage Publications, Newbury Park, pp. 351–359.
- Ringe, W.K., Saine, K.C., Lacritz, L.H., Hynan, L.S., Cullum, C.M., 2002. Dyadic short forms of the Wechsler Adult Intelligence Scale III. *Assessment* 9, 254–260.
- Sarfati, Y., Brunet, E., Hardy-Bayle, M.C., 2003. Comic-Strip Task: Attribution of intentions to Others. *Service de Psychiatrie Adulte, Hôpital de Versailles, Le Chesnay, France*.
- Schimmelmann, B.G., Conus, P., Edwards, J., McGorry, P.D., Lambert, M., 2005. Diagnostic stability 18 months after treatment initiation for first-episode psychosis. *Journal of Clinical Psychiatry* 66, 1239–1246.
- Sprong, M., Schothorst, P., Vos, E., Hox, J., van Engeland, H., 2007. Theory of mind in schizophrenia: meta-analysis. *British Journal of Psychiatry* 191, 5–13.
- Young, H.F., Bentall, R.P., 1997. Probabilistic reasoning in deluded, depressed and normal subjects: effects of task difficulty and meaningful versus non-meaningful material. *Psychological Medicine* 27, 455–465.