Measuring empathy in schizophrenia: The Empathic Accuracy Task and its correlation with other empathy measures

R.J.M. van Donkersgoed a, S. de Jong b, M. aan het Rot b, L. Wunderink b, P. H. Lysaker a, I. Hasson-Ohayon c, A. Aleman d, G.H.M. Pijnenborg d

a University of Groningen, Department of Clinical Psychology and Experimental Psychopathology, Grote Kruisstraat 1/2, 9712 TS Groningen, the Netherlands
b Dijk en Duin Parnassia Groep, Department of Psychotic Disorders, Westzijde 120, 1506 GB Zaandam, the Netherlands
c University of Amsterdam, Department of Clinical Psychology, Nieuwe Achtergracht 129-B, 1018 WT Amsterdam, the Netherlands
d GGZ Friesland, Sixmastraat 2, 8932 PA Leeuwarden, the Netherlands
e Roudeboush VA Medical Center, 1481 West 10th Street, Indianapolis, IN 46202, United States of America
f Indiana University School of Medicine, 340 W 10th St #6200, Indianapolis, IN 46202, United States of America
g Department of Psychology, Bar-Ilan University, Ramat-Gan, Israel
h BCN Neuro Imaging Center, University Medical Center Groningen, Antonius Deusinglaan 2, 9713 AW Groningen, the Netherlands
i GGZ Noord-Drenthe, Department of Psychotic Disorders, Dennenweg 9, 9404 LA Assen, the Netherlands

Abstract

Introduction: Empathy is an interpersonal process thought to be impaired in schizophrenia. Past studies have mainly used questionnaires or performance-based tasks with static cues to measure empathy. We used an Empathic Accuracy Task (EAT) designed to capture the dynamic aspects of empathy by using video clips in which perceivers continuously judge emotionally charged stories. We compared individuals with schizophrenia to healthy controls and assessed correlations among the EAT and three other commonly used empathy measures.

Methods: Patients (n=92) and healthy controls (n=42) matched for age, gender and education completed the EAT, the Interpersonal Reactivity Index, the Questionnaire of Cognitive and Affective Empathy and the Faux Pas task. Differences between groups were analyzed and correlations were calculated between empathy measurement instruments.

Results: The groups differed in EAT performance, with controls outperforming patients. A moderating effect was found for emotional expressivity of the target: while both patients and controls scored low when judging targets with low expressivity, controls performed better than patients with more expressive targets. EAT performance did not correlate with questionnaire scores. Differences between patients and controls on general cognition was not significant.

Conclusions: Individuals with schizophrenia benefit less from expressivity of other people which contributes to their impaired empathic accuracy. The lack of correlation between the EAT and the questionnaires suggests a distinction between self-report empathy and actual empathy performance. To explore empathic difficulties in real life, it is important to use instruments that take the interpersonal perspective into account.

This is the author's manuscript of the article published in final edited form as:

Introduction

Empathy is commonly defined as the ability to share and understand the emotional states of others (Eisenberg and Miller, 1987; Elliot et al., 2011). It is an interpersonal phenomenon, based on the interaction between the person who is empathizing and the person whose affective state is being shared or inferred (Zaki et al., 2008). This process involves the detection and perception of multimodal social cues that are dynamic and rapidly changing (Zaki and Ochsner, 2009).

Most studies on empathy differentiate between two aspects: affective and cognitive empathy (Michaels et al., 2014; Horan et al., 2015) which are integrated while being experienced (Ofir-Eyal et al., 2014). Affective empathy is hypothesized to be based on shared circuits in the brain: when seeing other people feel something, the same areas in the brain are activated as when feeling something yourself (Keyser and Gazzola, 2006). This makes it possible to empathize with others in an intuitive, unconscious manner. Cognitive empathy is a more conscious form of empathy and can be seen as the ability to explicitly interpret the thoughts and feelings of others (Blair, 2005; Frith and Frith, 2008). Research has shown that both affective (Bonfils et al., 2016) and cognitive aspects (Brüne, 2005; Biedermann et al., 2012; Savia et al., 2013) of empathy are impaired in people with schizophrenia. This has most commonly been measured using self-rating questionnaires. One widely used instrument is the Interpersonal Reactivity Index (IRI; Davis, 1983) with two subscales considered to measure cognitive empathy and two subscales measuring affective empathy. A more recently developed instrument, the Questionnaire of Cognitive and Affective Empathy (QCAE; Reniers et al., 2011), was developed from items of several existing self-report measures. In addition to the self-report questionnaires, performance-based tasks have been developed and used to measure empathy, for example the Reading the Mind in the Eyes Task (RMET; Baron-Cohen et al., 2001) and the Faux Pas Task (Stone et al., 1998; Baron-Cohen et al., 1999). It has been argued that these measurement instruments do not capture the dynamic process of empathy very well. Empathy appears to center around an ongoing interaction between the perceiver and the person whose affective state is being shared, not a momentary judgment by the perceiver alone (Zaki and Ochsner, 2009). Therefore it is debatable if measurement instruments with static cues such as a picture of a set of eyes (e.g., RMET), let alone questionnaires, can fully capture complex real-life empathy. One instrument that tries to account for this problem is the Empathic Accuracy Task (EAT; Zaki et al., 2008). This instrument measures cognitive empathy by judging the accuracy in which a person can estimate the emotions experienced by another person (Levenson and Ruef, 1992; Zaki et al., 2008). It consists of video clips in which one person (‘target’) describes an emotionally charged autobiographic story. The participant (‘perceiver’) has to continuously judge these stories and empathic accuracy is assessed by the extent to which the perceiver rating of each target’s emotion matches the target’s own judgement of the story’s emotional value. As empathy not only depends on the perceiver but also on the characteristics of the target, the EAT allows for the
assessment of the role of characteristics of the target, such as gender and expressivity as well as the valence of the story.

A first study using the EAT in a group of 30 patients with schizophrenia found an impairment in empathic accuracy in patients in comparison to a group of 22 healthy controls. Patients benefited less from the expressiveness of the targets than controls (Lee et al., 2011). This study found no significant correlation between the EAT and the IRI in patients, indicating a distinction between empathic accuracy and self-judgement of empathy. This distinction between objective and subjective measures of empathy was confirmed by another study in 145 patients and 45 controls (Horan et al., 2015).

The aims of the present study were 1) to assess dynamic empathy by using the empathic accuracy task in schizophrenia patients in replication of Lee et al. (2011), but in a larger sample using the Dutch version of the task (Aan het Rot and Hogenelst, 2014); 2) to assess the moderating role of the target’s gender and expressivity and the valence of the story and 3) to assess the correlation between EAT scores and scores on three commonly used empathy measurement instruments namely the IRI, QCAE and Faux Pas Task.

Methods

Participants

Ninety-three people with a diagnosis of schizophrenia or schizoaffective disorder according to DSM IV criteria (American Psychiatric Association, 2000) were included (for demographics see Table 1). These patients were recruited from six mental health care institutions in the Netherlands (GGZ Friesland, GGZ Drenthe, University Medical Centre Groningen, Lentis, Yulius, and Dimence). Part of this group was recruited from a randomized controlled trial to investigate the effect of a new metacognitive therapy (n = 70) (Van Donkersgoed et al., 2014). Participants who did not meet this inclusion criterion were approached and included in the current study as well (n=23). Exclusion criteria were: a current psychotic episode (PANSS positive symptoms average >4), IQ<70, age>18, not being able to give informed consent, medication change in the 30 days prior to assessment and comorbid neurological disorder. Diagnosis was confirmed using the Mini International Neuropsychiatric Interview (M.I.N.I.: Sheehan et al., 1998). The control group consisted of 41 people who had never received a psychiatric diagnosis with a mean age, gender distribution, handedness, and level of education similar to the patient group. They were recruited by advertisements on social media and with posters in the local area.

Assessment of the role of characteristics of the target, such as gender and expressivity as well as the valence of the story.

A first study using the EAT in a group of 30 patients with schizophrenia found an impairment in empathic accuracy in patients in comparison to a group of 22 healthy controls. Patients benefited less from the expressiveness of the targets than controls (Lee et al., 2011). This study found no significant correlation between the EAT and the IRI in patients, indicating a distinction between empathic accuracy and self-judgement of empathy. This distinction between objective and subjective measures of empathy was confirmed by another study in 145 patients and 45 controls (Horan et al., 2015).

Methods

Participants

Ninety-three people with a diagnosis of schizophrenia or schizoaffective disorder according to DSM IV criteria (American Psychiatric Association, 2000) were included (for demographics see Table 1). These patients were recruited from six mental health care institutions in the Netherlands (GGZ Friesland, GGZ Drenthe, University Medical Centre Groningen, Lentis, Yulius, and Dimence). Part of this group was recruited from a randomized controlled trial to investigate the effect of a new metacognitive therapy (n = 70) (Van Donkersgoed et al., 2014). Participants who did not meet this inclusion criterion were approached and included in the current study as well (n=23). Exclusion criteria were: a current psychotic episode (PANSS positive symptoms average >4), IQ<70, age>18, not being able to give informed consent, medication change in the 30 days prior to assessment and comorbid neurological disorder. Diagnosis was confirmed using the Mini International Neuropsychiatric Interview (M.I.N.I.: Sheehan et al., 1998). The control group consisted of 41 people who had never received a psychiatric diagnosis with a mean age, gender distribution, handedness, and level of education similar to the patient group. They were recruited by advertisements on social media and with posters in the local area.
### Table 1. Demographic variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Patients (n=93)</th>
<th>Controls (n=41)</th>
<th>$X^2$ (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (% male)</td>
<td>67</td>
<td>78</td>
<td>$X^2=5.69$ (p=0.06)</td>
</tr>
<tr>
<td>Handedness (% right)</td>
<td>88</td>
<td>93</td>
<td>$X^2=0.62$ (p=0.43)</td>
</tr>
<tr>
<td>Level of education mean (sd)</td>
<td>4.77 (1.31)</td>
<td>5.22 (0.94)</td>
<td>$X^2=9.57$ (p=0.14)</td>
</tr>
<tr>
<td>Age in years mean (sd)</td>
<td>41 (11)</td>
<td>41 (13)</td>
<td>$t=0.09$ (p=0.93)</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>64</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>29</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Schizoaffective disorder</td>
<td>29</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Mean (min-max)</td>
<td>64</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Hospital admissions</td>
<td>13 (0-38)</td>
<td>13 (0-38)</td>
<td></td>
</tr>
<tr>
<td>Illness duration in years</td>
<td>24 (12-52)</td>
<td>24 (12-52)</td>
<td></td>
</tr>
<tr>
<td>Age first psychosis</td>
<td>3 (1-20)</td>
<td>3 (1-20)</td>
<td></td>
</tr>
<tr>
<td>Psychotic episodes</td>
<td>62.96 (30-102)*</td>
<td>62.96 (30-102)*</td>
<td></td>
</tr>
<tr>
<td>PANSS Total</td>
<td>15.11 (7-31)</td>
<td>15.11 (7-31)</td>
<td></td>
</tr>
<tr>
<td>PANSS Positive</td>
<td>15.80 (7-29)</td>
<td>15.80 (7-29)</td>
<td></td>
</tr>
<tr>
<td>PANSS Negative</td>
<td>32.82 (16-56)</td>
<td>32.82 (16-56)</td>
<td></td>
</tr>
<tr>
<td>PANSS General</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*corresponds with a Clinical Global Impression (CGI; Guy, 1976) of moderately ill (Leucht et al., 2005).

#### Instruments

**General measures**

Positive and Negative Syndrome Scale (PANSS; Kay et al., 1987). To assess symptoms, this structured interview consisting of thirty items was conducted by raters who had attended a 2-day training course at the local hospital. The items fall into three subscales: positive symptoms (Cronbach's alpha = 0.72), negative symptoms (alpha = 0.77) and general symptoms (alpha = 0.80) and were rated on a scale from one to seven. The Cronbach's alpha for the total scale was 0.88.

M.I.N.I. Plus (Sheehan et al., 1998). Diagnosis according to the DSM-IV-TR criteria was confirmed with this structured interview. The interview is divided into 26 sections; each section concerns a diagnostic category. For this assessment we used the sections on psychotic disorders, depression and bipolar disorder and substance abuse.
General cognition

**Dutch Adult Reading Test (DART; Schmand et al., 1991).** The DART tests the pronunciation of irregularly spelled words and is used to determine premorbid intelligence.

**Trailmaking test A&B (TMT; Reitan and Wolfson, 1985).** The TMT provides information on visual search, scanning, mental flexibility speed of processing and executive functions. It is part of the Halstead–Reitan Battery. The TMT consists of two parts. Part A requires an individual to draw lines sequentially connecting 25 encircled numbers distributed on a sheet of paper. Task requirements are similar for Part B except the person must alternate between numbers and letters (e.g., 1, A, 2, B, 3, C, etc.). The final score is determined by subtracting the time to complete task A from the time it took to complete task B, with higher scores indicating lower cognition (Tombaugh, 2004).

**Digit Symbol Test (part of the Wechsler Adult Intelligence Scale; Wechsler, 1995).** This test evaluates the recognition and recoding of visual information. The test consists of several rows of paired boxes with a digit in the top box and an empty space in the box below. At the top of the page is shown which symbols are paired to the digits. The participant has to fill in as many symbols in the empty boxes within 90 seconds.

Empathy measures

**Empathic Accuracy Task (EAT):** To measure empathic accuracy we used a Dutch version of the EAT developed by Zaki et al. (2008). This instrument is considered to measure cognitive empathy. A shorter version than the original Dutch task described by (Aan Het Rot and Hogenelst, 2014) was used, this was necessary to keep the total assessment battery under two hours. The original task was shortened by selecting four out of the ten original videos. Participants were required to continuously rate the valence (positive-negative) of the videos in which a target tells a personal story, using a turning device. Scores of the participants are correlated with the target’s own ratings, leading to an index of empathic accuracy. Level of expressivity of the targets is based on their score on the Berkeley Expressivity Questionnaire (BEQ; Gross et al., 1995).

**Interpersonal Reactivity Index (IRI; Davis, 1983):** The IRI is a questionnaire intended to measure self-reported empathy and consists of 28 statements. The participant has to indicate whether the statement applies to him/her on a six point Likert scale. The four subscales of the IRI are Perspective Taking, Fantasy (together commonly labelled the Cognitive Empathy Scale, Cronbach’s alpha = 0.66), Empathic Accuracy and Personal Distress (Affective Empathy, alpha = 0.56), with higher scores indicating greater self-reported empathy. The Cronbach’s alpha for the total scale was 0.75.

Empathy measures

**Empathic Accuracy Task (EAT):** To measure empathic accuracy we used a Dutch version of the EAT developed by Zaki et al. (2008). This instrument is considered to measure cognitive empathy. A shorter version than the original Dutch task described by (Aan Het Rot and Hogenelst, 2014) was used, this was necessary to keep the total assessment battery under two hours. The original task was shortened by selecting four out of the ten original videos. Participants were required to continuously rate the valence (positive-negative) of the videos in which a target tells a personal story, using a turning device. Scores of the participants are correlated with the target’s own ratings, leading to an index of empathic accuracy. Level of expressivity of the targets is based on their score on the Berkeley Expressivity Questionnaire (BEQ; Gross et al., 1995).

**Interpersonal Reactivity Index (IRI; Davis, 1983):** The IRI is a questionnaire intended to measure self-reported empathy and consists of 28 statements. The participant has to indicate whether the statement applies to him/her on a six point Likert scale. The four subscales of the IRI are Perspective Taking, Fantasy (together commonly labelled the Cognitive Empathy Scale, Cronbach’s alpha = 0.66), Empathic Accuracy and Personal Distress (Affective Empathy, alpha = 0.56), with higher scores indicating greater self-reported empathy. The Cronbach’s alpha for the total scale was 0.75.
Questionnaire of Cognitive and Affective Empathy (QCAE; Reniers et al., 2011): This self-report questionnaire was recently developed to measure cognitive and affective empathy using 31 items on a four point Likert-scale, with higher scores indicating greater self-reported empathy. The questionnaire consists of five scales: Perspective Taking and Online Simulation (Cognitive Empathy, alpha = 0.82) and Emotion Contagion, Proximal Responsivity and Peripheral Responsivity (Affective Empathy, alpha = 0.79). The development of this questionnaire was based on factor analysis of items from other well-known empathy questionnaires. The QCAE includes six items that also appear on the IRI. The Cronbach’s alpha for the total QCAE was 0.83.

Faux-Pas Task (Stone et al., 1998; Baron-Cohen et al., 1999). This performance-based task consists of ten read aloud stories, describing interpersonal, everyday situations. Some of these stories contain a ‘faux pas’: a speaker in the story says something without considering if it is something that the listener might not want to hear. The respondent has to detect these mistakes. Cognitive Empathy is measured using the amount of correctly detected Faux Pas mistakes and Affective Empathy is measured by the amount of correctly answered emotion question paired with every faux pas (“How does person X feel?”).

Procedure
Both patients and controls gave their written informed consent for the use of their data for research before the assessment took place. Approval for the assessment of the patients was given by the local medical ethical committee (number METc2013.124 and METc2014.279) and for the healthy controls by the ethical committee of Psychology at the University of Groningen (ECP research code: ppc-013-109). Assessment was conducted by a trained assessor with at least a BSc. in Psychology. Patients were assessed with the MINI Plus to confirm diagnosis and the PANSS interview to assess symptoms. Both patients and controls completed the DART, TMT and DST, the IRI and QCAE and were assessed with the Faux Pas and the EAT (see Instruments).

Analysis
Scores on the three measures of cognition correlated highly and were therefore combined using Z-scores, into one measurement for general cognition. Differences between groups on general cognition were assessed using a t-test. The EAT data were analyzed using multilevel models. The first model tested for overall group differences between patients and controls. Subsequent models examined whether the valence of the videos, the gender of the target, or the expressivity of the target moderated the main result.

Independent-samples t-tests with a two-tailed significance level of p<0.05 were performed to assess the differences between groups on the other empathy

Faux-Pas Task (Stone et al., 1998; Baron-Cohen et al., 1999). This performance-based task consists of ten read aloud stories, describing interpersonal, everyday situations. Some of these stories contain a ‘faux pas’: a speaker in the story says something without considering if it is something that the listener might not want to hear. The respondent has to detect these mistakes. Cognitive Empathy is measured using the amount of correctly detected Faux Pas mistakes and Affective Empathy is measured by the amount of correctly answered emotion question paired with every faux pas (“How does person X feel?”).

Procedure
Both patients and controls gave their written informed consent for the use of their data for research before the assessment took place. Approval for the assessment of the patients was given by the local medical ethical committee (number METc2013.124 and METc2014.279) and for the healthy controls by the ethical committee of Psychology at the University of Groningen (ECP research code: ppc-013-109). Assessment was conducted by a trained assessor with at least a BSc. in Psychology. Patients were assessed with the MINI Plus to confirm diagnosis and the PANSS interview to assess symptoms. Both patients and controls completed the DART, TMT and DST, the IRI and QCAE and were assessed with the Faux Pas and the EAT (see Instruments).

Analysis
Scores on the three measures of cognition correlated highly and were therefore combined using Z-scores, into one measurement for general cognition. Differences between groups on general cognition were assessed using a t-test. The EAT data were analyzed using multilevel models. The first model tested for overall group differences between patients and controls. Subsequent models examined whether the valence of the videos, the gender of the target, or the expressivity of the target moderated the main result.

Independent-samples t-tests with a two-tailed significance level of p<0.05 were performed to assess the differences between groups on the other empathy
measures. Mann-Whitney tests (p<0.05 two-tailed) were performed to examine the differences on not normally distributed scales. Spearman correlations were calculated between the person-level mean EAT scores and the other empathy instruments scores.

Results

General cognition

Scores on the three measures of general cognition (DART, TMT and DST, see instruments) were available for 88 patients and 18 healthy controls. The three measures correlated highly and were therefore combined into one cognition measure, using Z-scores. Differences between groups were not significant (t=0.65; p=.51). Differences in empathic accuracy between participants with and without cognition scores were not significant either (healthy group: F= 0.53, p=0.47; patient group: F=0.46, p=0.49).

Empathic Accuracy Task

The mean Empathic Accuracy score over all participants was 0.38 with a range from -1.00 to 1.00. The overall difference between patients (mean correlation 0.34) and controls (mean correlation 0.52) was significant, F(1,114)=7.71, p=0.007, d=0.52.

A second multilevel analysis was performed with level 1 = valence and level 2 = group. The main effect of group was significant (F(1,114) = 5.83, p<0.02, d=0.45) and the main effect of valence was significant (F(1,112)=188.00, p<0.001, d=2.59). The interaction between group and valence was not significant (F(1,112)= 0.13, p=.71) indicating that group difference was not moderated or confounded by the valence of the videos.

A third multilevel analysis was performed with level 1 = gender of the target and level 2 = group. A main effect for group was still found (F(1,114) = 7.78, p=0.007, d=0.52), but no effect was found for gender (F(1,113)= 0.38, p=0.53, d=0.12) or the interaction between group and gender (F(1,113) = 0.45, p=0.50), indicating that group difference was not moderated by the gender of the target.

When finally group (L2) and target expressivity (L1) were entered into the model, the main effect for group remained significant F(1,114) = 4.61, p<0.04, d=0.40. Significant effects for Expressivity F(1,385) = 120.58, p<0.0001, d=1.12, and the interaction (group x expressivity) F(1,385) = 7.23, p=0.008 were found as well. Group difference in empathic accuracy was not significant for targets with lower expressivity (t(385) = -0.16, p=0.86, d=0.02), but significant for targets with higher expressivity (t(385) = 3.71, p=0.0003, d=0.38) with patients scoring lower on empathic accuracy than controls (see Figure 1).

Spearman correlations were calculated between the person-level mean EAT scores and the other empathy instruments scores.

Results

General cognition

Scores on the three measures of general cognition (DART, TMT and DST, see instruments) were available for 88 patients and 18 healthy controls. The three measures correlated highly and were therefore combined into one cognition measure, using Z-scores. Differences between groups were not significant (t=0.65; p=.51). Differences in empathic accuracy between participants with and without cognition scores were not significant either (healthy group: F= 0.53, p=0.47; patient group: F=0.46, p=0.49).

Empathic Accuracy Task

The mean Empathic Accuracy score over all participants was 0.38 with a range from -1.00 to 1.00. The overall difference between patients (mean correlation 0.34) and controls (mean correlation 0.52) was significant, F(1,114)=7.71, p=0.007, d=0.52.

A second multilevel analysis was performed with level 1 = valence and level 2 = group. The main effect of group was significant (F(1,114) = 5.83, p<0.02, d=0.45) and the main effect of valence was significant (F(1,112)=188.00, p<0.001, d=2.59). The interaction between group and valence was not significant (F(1,112)= 0.13, p=.71) indicating that group difference was not moderated or confounded by the valence of the videos.

A third multilevel analysis was performed with level 1 = gender of the target and level 2 = group. A main effect for group was still found (F(1,114) = 7.78, p=0.007, d=0.52), but no effect was found for gender (F(1,113)= 0.38, p=0.53, d=0.12) or the interaction between group and gender (F(1,113) = 0.45, p=0.50), indicating that group difference was not moderated by the gender of the target.

When finally group (L2) and target expressivity (L1) were entered into the model, the main effect for group remained significant F(1,114) = 4.61, p<0.04, d=0.40. Significant effects for Expressivity F(1,385) = 120.58, p<0.0001, d=1.12, and the interaction (group x expressivity) F(1,385) = 7.23, p=0.008 were found as well. Group difference in empathic accuracy was not significant for targets with lower expressivity (t(385) = -0.16, p=0.86, d=0.02). The group difference was significant for videos of targets with higher expressivity (t(385) = 3.71, p=0.0003, d=0.38) with patients scoring lower on empathic accuracy than controls (see Figure 1).
Figure 1. Lower and higher BEQ scores in patients vs controls.

Differences between groups on IRI, QCAE and Faux Pas

Differences between groups were found for the total IRI score and the cognitive and affective subscales (see Table 2). Controls outperformed patients on both cognitive and affective empathy. Further examination shows that the significance of the group difference in Affective Empathy was primary due to a difference on the Personal Distress (PD) scale ($t=3.58, p<0.001$); no significant difference was found on the Empathic Concern Scale ($t=1.01, p=0.32$). Patients scored significantly lower (mean 13.31, sd 3.85) than controls (mean 15.98, sd 4.28) on Personal Distress. The group difference on the Cognitive Scale of the IRI was primary due to a group difference on the Fantasy scale ($t=3.19, p<0.002$); no significant difference was found on the Perspective Taking scale ($t=0.54, p=0.59$). Patients scored significantly lower (mean 14.75, sd 4.60) than controls (mean 17.54, sd 4.79) on the Fantasy subscale. No significant group difference was found for Total QCAE or QCAE Affective Empathy. However a difference was found for the Cognitive Empathy Scale, primarily due to a difference on the Perspective Taking (PT) scale ($t=2.41, p=0.02$); there was no significant difference on the Online Simulation scale ($t=1.00, p=0.32$). Patients scored lower (mean 28.28, sd 5.38) than controls (mean 30.4, sd 4.03) on PT. No significant differences between groups were found on the Faux Pas Cognitive or Affective Scales.
Correlation among empathy measures

The EAT (considered to measure cognitive empathy) did not significantly correlate with the cognitive or affective subscales of the IRI and the QCAE as shown in Table 3 and Figure 2. All cognitive and affective subscales of the IRI and QCAE correlated with each other with one exception: the IRI affective and the QCAE cognitive scale did not correlate. The Faux Pas cognitive and affective scale correlated with each other. The cognitive scale of the Faux Pas also correlated with the EAT.

Table 2. Differences between groups on empathy.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Controls (n=41) Mean (sd)</th>
<th>Patients (n=93) Mean (sd)</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRI Total</td>
<td>61.8 (14.08)</td>
<td>55.32 (10.81)</td>
<td>2.63 (p&lt;0.02)</td>
</tr>
<tr>
<td>IRI Cognitive empathy</td>
<td>30.88 (8.00)</td>
<td>27.68 (6.91)</td>
<td>2.35 (p&lt;0.02)</td>
</tr>
<tr>
<td>IRI Affective empathy</td>
<td>30.93 (7.08)</td>
<td>27.65 (5.77)</td>
<td>2.83 (p&lt;0.01)</td>
</tr>
<tr>
<td>OCAE Total</td>
<td>90.37 (9.47)</td>
<td>87.85 (11.16)</td>
<td>1.26 (p=0.21)</td>
</tr>
<tr>
<td>OCAE Cognitive empathy</td>
<td>57.79 (6.93)</td>
<td>54.78 (8.42)</td>
<td>2.00 (p&lt;0.05)</td>
</tr>
<tr>
<td>OCAE Affective empathy</td>
<td>32.59 (4.78)</td>
<td>33.06 (5.01)</td>
<td>0.52 (p=0.61)</td>
</tr>
<tr>
<td>Faux Pas Cognitive Faux</td>
<td>4.36 (0.76)</td>
<td>4.01 (1.08)</td>
<td>-1.48 (p=0.14)</td>
</tr>
<tr>
<td>Pas Affective</td>
<td>2.31 (1.26)</td>
<td>2.26 (1.65)</td>
<td>-0.25 (p=0.81)</td>
</tr>
<tr>
<td>EAT score (mean across video clips)</td>
<td>0.52</td>
<td>0.34</td>
<td>7.71 (p&lt;0.007)</td>
</tr>
</tbody>
</table>

IRI = Interpersonal Reactivity Index, OCAE = Questionnaire of Cognitive and Affective Empathy, EAT = Empathic Accuracy Task.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Controls (n=41) Mean (sd)</th>
<th>Patients (n=93) Mean (sd)</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRI Total</td>
<td>61.8 (14.08)</td>
<td>55.32 (10.81)</td>
<td>2.63 (p&lt;0.02)</td>
</tr>
<tr>
<td>IRI Cognitive empathy</td>
<td>30.88 (8.00)</td>
<td>27.68 (6.91)</td>
<td>2.35 (p&lt;0.02)</td>
</tr>
<tr>
<td>IRI Affective empathy</td>
<td>30.93 (7.08)</td>
<td>27.65 (5.77)</td>
<td>2.83 (p&lt;0.01)</td>
</tr>
<tr>
<td>OCAE Total</td>
<td>90.37 (9.47)</td>
<td>87.85 (11.16)</td>
<td>1.26 (p=0.21)</td>
</tr>
<tr>
<td>OCAE Cognitive empathy</td>
<td>57.79 (6.93)</td>
<td>54.78 (8.42)</td>
<td>2.00 (p&lt;0.05)</td>
</tr>
<tr>
<td>OCAE Affective empathy</td>
<td>32.59 (4.78)</td>
<td>33.06 (5.01)</td>
<td>0.52 (p=0.61)</td>
</tr>
<tr>
<td>Faux Pas Cognitive Faux</td>
<td>4.36 (0.76)</td>
<td>4.01 (1.08)</td>
<td>-1.48 (p=0.14)</td>
</tr>
<tr>
<td>Pas Affective</td>
<td>2.31 (1.26)</td>
<td>2.26 (1.65)</td>
<td>-0.25 (p=0.81)</td>
</tr>
<tr>
<td>EAT score (mean across video clips)</td>
<td>0.52</td>
<td>0.34</td>
<td>7.71 (p&lt;0.007)</td>
</tr>
</tbody>
</table>

IRI = Interpersonal Reactivity Index, OCAE = Questionnaire of Cognitive and Affective Empathy, EAT = Empathic Accuracy Task.

**Table 3.** Spearman correlations in patient sample between QCAE, IRI, Faux Pas, and EAT.

**Figure 2.** Schematic representation of the correlations among the various empathy instruments in schizophrenia patients.
Discussion

To capture empathy as a complex social-cognitive process, involving dynamic cues in the interaction between perceiver and the person whose affective state is being shared (Zaki and Ochsner, 2009; Lee et al., 2011), we used the Empathic Accuracy Task in a large sample of schizophrenia patients. We compared EAT scores of patients to healthy controls and assessed the correlation of the EAT with three other widely used empathy measurements.

Results showed reduced overall empathic accuracy performance in patients in comparison to healthy individuals. The valence of the stories or the gender of the target had no influence on this result, but the expressivity of the target did. With less expressive targets, the schizophrenia patients and the control group scored similarly low on empathic accuracy. In contrast, with more expressive targets, the controls performed better on empathic accuracy than the patients. Patients apparently benefit less from the expressiveness of more expressive persons. A previous study with a smaller participant group using the English version of the EAT found the same result (Lee et al., 2011).

A possible explanation for this effect is the impairment in emotion recognition in people with schizophrenia (Feingold et al., 2016). Patients with schizophrenia experience problems in reading facial expressions (Kohler et al., 2010) and recognizing emotional prosody (Leitman et al., 2005; Petkova et al., 2014). Healthy people benefit from explicit emotional cues in expressive persons and patients may miss these cues and therefore have less information to base their estimation of the emotional state of the other on. Further research is necessary to determine if problems in empathic accuracy are based on problems in basic emotion recognition. The current study found differences in cognitive and affective empathy on the self-report measures between groups, with controls outperforming patients, as did previous studies (Smith et al., 2012; Michaels et al., 2014; Horan et al., 2015). These questionnaires did not correlate with the empathic accuracy task, which is in line with other studies with schizophrenia patients (Horan et al., 2015) and in multiple healthy samples (Levenson and Ruef, 1992; Ickes et al., 1990; Zaki et al., 2008). It seems that these questionnaires do not measure the same aspects of empathy as the EAT, as they measure one’s own view of one’s empathic abilities, which can be distorted. Insight is impaired in many people with schizophrenia (see for latest review: Elové & Conus, 2016). It is possible that the view of schizophrenia patients of their own empathic performances is not accurate. Furthermore, self-report measurements are prone to different biases including intrusive symptoms biases, cognitive status biases and values and social comparison biases (McGurk et al., 2000; Hendryx et al., 2001; Bromley and Brekke, 2010; Patterson et al., 2001).

Taken together, it may be best to see empathy as a multi-faceted construct encompassing multiple overlapping domains including the basic interpretation of emotional cues, the dynamic integration of these cues, affective and cognitive pathways and trait empathy as measured with self-report questionnaires. To understand empathic difficulties, it is important to account for these different aspects, especially the gap between one’s belief and one’s performance in
empathy (Devlin et al., 2014; Zaki et al., 2008). It is possible that more basic elements underlying empathy are impaired in patients with schizophrenia, while their subjective experience of empathy does not change. Future research is necessary to identify the distinctions and overlap between the elements of the empathy construct.

In addition, to understand empathic difficulties in real life, it is important to take an interpersonal perspective of the construct. Understanding the empathic difficulties among persons with schizophrenia that affect their social well-being (Ofir-Eyal et al., 2014) as well as their possible benefits from psychotherapy (Hasson-Ohayon et al., 2017) will allow to better tailor interventions to improve empathy.

The current study has several limitations. Reduced empathy performance did not seem to be part of a generalized cognitive deficit, as differences between patients and controls on a measure of general cognition was not significant. However, data on general cognition was only available for half of the healthy participants. Furthermore, most schizophrenia patients in this study used antipsychotic medication. It is not clear if the use of medication has any influence on empathic accuracy; more information needs to come from further studies on empathy in people with first onset psychosis who do not use medication. It must also be noted that we defined empathy in this study as ‘the accurateness in which someone can understand another person’s feelings’. We did not measure empathy in the sense of ‘care for the other’. Furthermore, empathic accuracy in patients might be influenced by problems with sustained attention or motor abilities, although no differences on general cognition between groups was found and a previous study on empathic accuracy found that patients tracked a dynamically moving non-social visual stimulus with high accuracy (Lee et al., 2011).

Conclusion

Individuals with schizophrenia benefit less from the emotional expressivity of other persons than controls, which contributes to their impaired empathic accuracy. The lack of correlation between the EAT and the questionnaires suggests a discrepancy between subjectively experienced empathy and actual empathy performance. To understand empathic difficulties of people with schizophrenia in real life, it is important to take a dynamic, interpersonal perspective of the construct. The Empathic Accuracy Test can be a useful instrument to measure empathy in an ecologically valid way.