Submitted to *Journal of Abnormal Psychology*, 16th February 2016; revisions submitted 25th April 2016 and 17th May 2016; accepted 17th May 2016.

# 'Theory of Mind' is not Theory of Emotion: A cautionary note on the Reading the Mind in the Eyes Test

Beth F. M. Oakley

University of Cambridge

Rebecca Brewer, Geoffrey Bird and Caroline Catmur

King's College London

## Author Note

Beth F. M. Oakley, Department of Psychology, University of Cambridge, UK and School of Psychology, University of Surrey, UK.

Rebecca Brewer, MRC Social, Genetic & Developmental Psychiatry Centre, King's College London, UK and School of Psychology, University of East London, UK.

Geoffrey Bird, MRC Social, Genetic & Developmental Psychiatry Centre, King's College London, UK and Institute of Cognitive Neuroscience, University College London, UK.

Caroline Catmur, Department of Psychology, King's College London, UK and School of Psychology, University of Surrey, UK.

This research was supported by the Wellcome Trust [Biomedical Vacation Scholarship to CC supporting BFMO].

Correspondence concerning this article should be addressed to Caroline Catmur, Department of Psychology, King's College London, London SE1 1UL, UK, caroline.catmur@kcl.ac.uk

Abstract

The ability to represent mental states ('Theory of Mind'; ToM) is crucial in understanding individual differences in social ability, and social impairments evident in conditions such as Autism Spectrum Disorder (ASD). The "Reading the Mind in the Eyes" Test (RMET) is a popular measure of ToM ability, validated in part by the poor performance of those with ASD. However, the RMET requires recognition of facial emotion, which is impaired in those with alexithymia, which frequently co-occurs with ASD. Thus, it is unclear whether the RMET indexes emotion recognition, associated with alexithymia, or ToM, associated with ASD. We therefore investigated the independent contributions of ASD and alexithymia to performance on the RMET. ASD and alexithymia-matched control participants did not differ on RMET performance, whereas ASD participants demonstrated impaired performance on an alternative test of ToM, the Movie for Assessment of Social Cognition (MASC). Furthermore, alexithymia, but not ASD diagnosis, significantly influenced RMET performance, but did not affect MASC performance. These results suggest that the RMET measures emotion recognition rather than ToM, and support the "alexithymia hypothesis" of emotion-related deficits in ASD.

*Keywords*: autism spectrum disorder, alexithymia, emotion recognition, social cognition, theory of mind

## General Scientific Summary

This study suggests that a highly popular test of the ability to detect what someone else is thinking, the 'Reading the Mind in the Eyes' Test, is instead a test of the ability to recognise another person's emotional expression. This is important because it suggests that patients who perform badly on this test may still be able to understand another person's mental state, and that conversely, patients who perform well on this test may still have difficulties in mental state understanding.

# ‘Theory of Mind’ is not Theory of Emotion: A cautionary note on the Reading the Mind in the Eyes Test

Questions concerning the representation of mental states (‘Theory of Mind’ or ToM) have occupied cognitive scientists for decades. The study of ToM has been particularly important in understanding individual differences in social ability, and the social impairments evident in various psychiatric disorders (Baron-Cohen, Leslie, & Frith, 1985; Corcoran, Mercer, & Frith, 1995). The difficulty associated with designing a test appropriate for detecting variance in ToM ability among typical adults prompted the development of the “Reading the Mind in the Eyes” test (RMET; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001a), which requires participants to match emotion and mental state descriptor words to images of the eye region of faces. This test has been cited over 2000 times (Web of Science, May 2016) and has been used to demonstrate gender, cultural, genetic, and personality trait influences on ToM, and to elucidate its neurobiological mechanisms (Adams et al., 2010; Stone, Baron-Cohen, Calder, Keane, & Young, 2003). Its validity is supported by the poor performance of individuals with Autism Spectrum Disorder (ASD) on the task (Baron-Cohen et al., 2001a), a group who have known ToM impairments (Castelli, Frith, Happé, & Frith, 2002).

The RMET is unusual among ToM tasks in that it includes emotional states and relies on the detection of subtle facial cues, features typically used to test emotion recognition. In contrast, the majority of ToM tasks require non-emotional mental states to be inferred from contextual information or dynamic behavioural cues (Baron-Cohen, O’Riordan, Stone, Jones, & Plaisted, 1999; Dziobek et al., 2006). The reliance of the RMET on the recognition of emotional states from faces calls into question its use as a test of ToM, and means it is possible that the RMET may instead assess emotion recognition. Importantly, if the RMET

does not index ToM but rather emotion recognition, then our conclusions about the ToM abilities of various cultural and psychiatric groups will require substantial revision.

It is noteworthy, therefore, that recent evidence suggests that apparent emotion recognition impairments in ASD are in fact due to alexithymia (a trait characterised by poor recognition of one's own emotions; Nemiah, Freyberger, & Sifneos, 1976), which frequently co-occurs with ASD (the 'alexithymia hypothesis' of emotion deficits in ASD; Bird & Cook, 2013; Cook, Brewer, Shah, & Bird, 2013). Alexithymia has been shown to predict performance on other tasks that require participants to match emotion words to faces (Grynberg et al., 2012), meaning that increased rates of alexithymia in ASD may explain the ASD impairment on the RMET if emotion recognition ability, rather than ToM, drives performance. We therefore investigated the independent contributions of ASD and alexithymia to RMET performance, compared to another ToM task that does not rely exclusively on facial emotion recognition (the Movie for the Assessment of Social Cognition, MASC; Dziobek et al., 2006). If the RMET indexes ToM, then ASD diagnostic status should predict both RMET and MASC scores. If the RMET indexes emotion recognition, however, then alexithymia, not ASD, should predict RMET performance.

## Method

### Participants

19 participants (5 female) with a clinical diagnosis of ASD and 24 (11 female) without ASD volunteered to take part in the study. 1 ASD participant had a co-morbid diagnosis of dyspraxia and 1 had previously been prescribed antipsychotic medication. Excluding these two participants did not alter the pattern of significance. Control participants had no past or present clinical diagnosis. 1 control participant did not complete the MASC, leaving the final sample at 42 participants. Based on the effect size ($d$ = 1.27) between ASD

and control participants on the RMET reported in Baron-Cohen et al. (2001a), we determined that a sample size of 19 participants per group would provide power of 0.8 to detect group differences using an independent t-test and alpha <.05. Additional control participants were recruited to ensure groups were matched on demographic variables. 6 ASD and 8 control participants met the criterion for severe alexithymia, with a score of 61 or above on the 20-item Toronto Alexithymia Scale (TAS-20; Bagby, Parker, & Taylor, 1994). The TAS-20 is a self-report scale that includes questions like "I have feelings that I can't quite identify", and "I find it hard to describe how I feel about people." Items are rated on a scale from 1 ("does not describe me") to 5 ("describes me very well"), with scores ranging between 20 and 100, and higher scores indicating more alexithymic traits. The TAS-20 has good internal consistency (α=.81) and test-retest reliability (r=.77; Bagby et al., 1994). The high alexithymia cut-off score of ≥61 was established as being 1.5 *SD* from the mean score of community samples (Parker, Taylor & Bagby, 2003).

ASD and control groups were matched according to alexithymia severity (control $M$ = 52.43, $SD$ = 13.98; ASD $M$ = 57.58, $SD$ = 11.51; $t(40) = 1.28$, $p = .207$, $d = 0.398$, 95% CI for $d$ [-0.218, 1.009]), age (control $M$ = 30.13, $SD$ = 12.21; ASD $M$ = 30.89, $SD$ = 11.86; $t(40) = 0.21$, $p = .839$, $d = 0.064$, 95% CI for $d$ [-.545, 0.671]), gender ($X^2(1) = 2.04$, $p = .153$, $r = .220$), and IQ, measured by the vocabulary and matrix reasoning sub-scales of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999; control $M$ = 108.48, $SD$ = 11.68; ASD $M$ = 109.79, $SD$ = 15.71; $t(40) = 0.31$, $p = .758$, $d = 0.096$, 95% CI for $d$ [-.512, 0.704]). As shown by the IQ scores, both the ASD and control groups were in the average range for intelligence.

ASD symptom severity for all participants was measured using the Autism Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001b). The AQ is a 50-item self-report questionnaire for assessing traits associated with the autism spectrum,

with statements like "I notice patterns in things all the time", and "I find it hard to make new friends." Statements are rated from "strongly agree" to "strongly disagree", with a resulting total score of 0-50, and higher scores indicating more autistic traits. The AQ has good internal consistency (α=0.82; Austin, 2005) and test-retest reliability (r=0.70; Baron-Cohen et al., 2001b). AQ scores were significantly higher in the ASD ($M$ = 32.74, $SD$ = 9.33) than control ($M$ = 21.91, $SD$ = 11.06) group ($t(40) = 3.38$, $p = .002$, $d = 1.049$ , 95% CI for $d$ [0.394, 1.693]). Current functioning in the ASD group was assessed using the Autism Diagnostic Observation Schedule (ADOS-G; Lord et al., 2000). Four participants did not meet ADOS criteria for ASD, but they received diagnoses from independent clinicians, were not outliers on any measure, and their exclusion did not alter the pattern of significance. The study was approved by the University of Surrey Research Ethics Committee.

**Procedure**

Order of RMET and MASC administration was counterbalanced across participants. The RMET comprised 36 stimuli depicting the eye section of a face. For each stimulus, participants were required to select one of four verbal labels, presented underneath the image, that best described what the individual was thinking or feeling. Examples include items such as "upset", "excited", and "terrified". The stimuli were identical to those used in the revised RMET (Baron-Cohen et al., 2001a) and were presented in black and white at a standard size. 19 faces were male and 17 female. A gender identification control task was not used due to previous ceiling effects in comparable populations (Baron-Cohen et al., 2001a). Each stimulus was presented on a computer screen for unlimited time, however participants were instructed to respond with their best estimate if they felt unsure of the answer, to prevent them spending too long looking at each stimulus. Responses were made via a key press. RMET score was calculated as total correct responses (maximum 36). The internal

consistency of the RMET is modest, with Cronbach's alpha varying between 0.37-0.61 across cultural adaptations (Khorashad et al., 2015; Vellante et al., 2012).

The MASC involved watching a 15-minute film depicting four individuals socialising, interrupted at intervals with a mental state or control question about the section of film that had just been played. The 45 mental state questions focused on why characters were behaving in a particular way; it should be noted however that 18 of the mental state questions on the MASC are relatively more dependent on emotional mental state decoding (e.g. “what is Betty feeling?”) whereas the remaining questions measure cognitive ToM (e.g. “what is Sandra thinking?”); see Montag et al. (2010). The 21 control questions related to specific details given in the film to ensure that participants were paying attention (e.g. “what time are they meeting?”). Participants responded to each question by selecting from four possible answers and recorded their answers on an answer sheet. MASC score was calculated as total correct responses to the mental state questions (maximum 45). The mental state questions on the MASC have been shown to have a high internal consistency ($\alpha$=0.84), high intraclass correlation coefficients in the assessment of test-retest reliability (ICC=0.97), and to be sensitive to subtle ToM difficulties in participants in adults of normal intelligence (Dziobek et al., 2006). Control question score was calculated as total correct responses to the control questions (maximum 21; additional control questions were used as per Santiesteban et al., 2015, to provide more possibility for any between-group variance in control question performance to be detected, if present, and were generated as in the original version of the task, such that they assessed general comprehension and memory for the material, but not the understanding of mental states: control questions are listed in Supplementary Table 1).

## Results

One participant in the ASD group was excluded from MASC analyses due to being an outlier (> 3 *SD* from the mean) on this task.

**RMET.** The ASD (*M* = 26.32*, SD* = 3.77) and alexithymia-matched control (*M* = 26.65*, SD* = 2.99) groups did not differ on RMET score ($t(40) = 0.32$, $p = .749$, $d = 0.100$, 95% CI for *d* [-0.509, 0.707]; Figure 1a). A Bayes Factor of 0.317 suggested that the data were over 3.15 times as likely under the null hypothesis as under the alternative hypothesis.

However, when comparing alexithymic and non-alexithymic participants as two groups (regardless of ASD), the alexithymic group (*M* = 24.71, *SD* = 2.84) exhibited significantly worse RMET performance than the non-alexithymic group (*M* = 27.39, *SD* = 3.24; $t(40) = 2.63$, $p = .012$, $d = 0.861$, 95% CI for *d* [0.187, 1.524]; Figure 1b). A Bayes Factor of 4.287 provided strong support for a group difference. The alexithymic group also performed significantly worse than the control participants reported in Baron-Cohen et al. (2001a) (*N* = 225, *M* = 27.02, *SD* = 3.67; independent-samples *t* test: $t(237) = 2.31$, $p = .021$, $d = 0.638$, 95% CI for *d* [0.094, 1.180]), illustrating that the alexithymic group was also impaired compared to published normative data on control participant RMET performance.

**MASC.** In contrast to the findings for the RMET, the control group (*M* = 35.70, *SD* = 3.31) performed significantly better on the MASC than the ASD group (*M* = 31.22, *SD* = 4.60; $t(39) = 3.62$, $p = .001$, $d = 1.141$, 95% CI for *d* [0.468, 1.801]), ruling out the possibility that the ASD group had intact ToM. This significant group difference was supported by a Bayes Factor of 36.40. There was no significant difference in the total number of MASC control questions correctly answered between the ASD group and the control group (control *M* = 18.91, *SD* = 1.68; ASD *M* = 18.22, *SD* = 2.39; $t(39) = 1.09$, $p = .284$, $d = 0.342$ , 95% CI for *d* [-0.281, 0.961]).

No significant difference between alexithymia groups was found for MASC scores (alexithymic $M$ = 34.71, $SD$ = 4.27; non-alexithymic $M$ = 33.22, $SD$ = 4.58; $t(39) = 1.01$, $p$ = .318, $d$ = 0.333, 95% CI for $d$ [-0.319, 0.981]). A Bayes Factor of .475 indicated that the data were over 2.1 times as likely under the null than alternative hypothesis.

**Hierarchical regression analyses.** Hierarchical regression analyses were used to compare the independent contribution of alexithymia and ASD symptom severity to RMET and MASC performance, respectively. These analyses included the effect of gender in the first step as the gender balance of the ASD and Control groups was numerically (although not significantly) different, and gender has been associated with both facial emotion recognition (Forni-Santos & Osório, 2015) and ToM (e.g. Devine & Hughes, 2013). Accordingly, performance on the RMET was analysed using a hierarchical regression in which gender was entered in the first step, AQ scores were entered in the second step, and TAS-20 scores in the third step. TAS-20 scores remained a significant predictor of RMET performance (standardised $\beta$ = -.410, $p$ = .030, 95% CI for $\beta$ [-.199, -.011], $\Delta R^2$ = 11.5%), even after accounting for gender and AQ scores, but AQ scores were not a significant predictor in either the second or third steps of the regression (second step: $\beta$ = -.111, $p$ = .490, 95% CI for $\beta$ [-.124, .061], $\Delta R^2$ = 1.2%; third step: $\beta$ = .104, $p$ = .565, 95% CI for $\beta$ [-.074, .134]). When analysing performance on the MASC, gender was entered in the first step, TAS-20 scores in the second, and AQ scores in the final step. In this model neither TAS-20 ($\beta$ = -.161, $p$ = .304, 95% CI for $\beta$ [-.162, .052], $\Delta R^2$ = 2.5%) nor AQ scores ($\beta$ = -.277, $p$ = .125, 95% CI for $\beta$ [-.243, .031], $\Delta R^2$ = 5.5%) predicted MASC scores; therefore, a follow-up analysis focussed on those MASC items which do not ask about ‘feelings’ and are thus less dependent on emotion understanding, constituting a test of cognitive ToM (Montag et al., 2010). The same analysis applied to these items revealed that AQ scores were a significant predictor of performance ($\beta$ = -.357, $p$ = .046, 95% CI for $\beta$ [-.188, -.002], $\Delta R^2$ = 9.1%) even after

controlling for TAS-20 scores and gender. Full details of the hierarchical regression analyses are presented in Supplementary Table 2.

## Discussion

This study attempted to determine whether the RMET, currently the most popular test of ToM in adults, is a valid measure of ToM, or whether performance is in fact determined by emotion recognition ability. Deficits in emotion recognition are commonly associated with alexithymia, and the 'alexithymia hypothesis' of emotion-related impairments in ASD (Bird & Cook, 2013) suggests that in individuals with ASD and co-morbid alexithymia, it is alexithymia, rather than ASD per se, which impairs emotion recognition performance. We therefore investigated whether alexithymia or ASD was the better predictor of RMET performance, given the reliance of the RMET on the ability to perceive facial emotion. We also measured whether either ASD or alexithymia could predict performance on a validated ToM task (the MASC) which relies on emotion recognition to a lesser extent than the RMET. Alexithymia, rather than ASD diagnosis, predicted RMET performance, while ASD diagnosis and symptom severity, rather than alexithymia, predicted performance on the MASC task. These results suggest that the RMET may be better characterised as a test of emotion recognition than of mental state understanding.

Should converging evidence for the reliance of the RMET on emotion recognition ability be obtained, substantial revision of conclusions based on over 2000 studies is required. Such revision will be urgently needed for studies that have been used to support diagnostic or therapeutic approaches with clinical groups (e.g. Anagnostou et al., 2012; Berlim, McGirr, Beaulieu, & Turecki, 2011). In addition, the current data may explain inconsistencies in both the ToM and emotion recognition literatures regarding the abilities of clinical groups. For example, contemporary adaptations of the RMET have failed to show significant differences in performance between ASD and neurotypical participants (Back, Ropar & Mitchell, 2007;

Roeyers, Buysse, Ponnet, & Pichal, 2001), challenging the original RMET findings. The current results suggest that such inconsistent findings are likely a product of sampling variance with respect to alexithymia in ASD samples; those studies with a greater proportion of alexithymic participants in their samples of individuals with ASD would be more likely to report 'ASD-related' deficits than studies with fewer alexithymic participants within the ASD group. Within studies of typical individuals, these results are also likely to explain why poor correlations have been reported between the RMET and alternative tests of ToM, such as the Strange Stories and Faux Pas tasks (Ahmed & Miller, 2011; Spek, Scholte & van Berckelaer-Onnes, 2010).

The present data suggest that the RMET measures emotion recognition rather than ToM. However, it might be argued that the process of emotion recognition could be defined as a form of mental state inference. Under this account, the ability measured by the RMET might be most precisely described as 'emotion recognition', but the test would still index ToM. The link between ToM and emotion recognition is relatively under-investigated: emotion recognition is moderately related to ToM tasks in non-clinical and ASD populations in some samples (Dyck, Ferguson & Shochet, 2001), but not in others (Bora et al., 2005) and may depend on the particular ToM tasks selected. However, clinical dissociations between emotion recognition and ToM performance lend additional support to the proposal that emotion recognition and mental state inference are discrete cognitive processes. Young adults diagnosed with conduct disorder (especially with callous-unemotional traits) display poor performance on the RMET and other emotion recognition tasks, when compared to the typical population (Fairchild et al., 2009; Sharp, 2008). Despite poor emotion recognition performance, the same group tend to perform at average level on alternative ToM tasks and have intact mentalising capacity (O'Nions et al., 2014). Conversely, recent research into social functioning in dementia has found the reverse: evidence of impaired ToM with

simultaneously intact emotion recognition during a video vignette task (Freedman et al., 2013). These and similar results would argue against the suggestion that emotion recognition involves the same cognitive processes as ToM.

Our data support the majority of the literature in suggesting that alexithymia is not strongly associated with ToM performance. Moriguchi and colleagues (2006) reported a greater deficit on mentalising measured via an animated shapes task in a non-clinical sample of individuals with high alexithymia compared to low alexithymia. However, most research contradicts this finding across a number of ToM tasks, including the Strange Stories and false belief tasks (Lane et al., 2015; Milosavljevic et al., 2016). A recent paper (Gökçen, Frederickson, & Petrides, in press) demonstrated that although higher TAS-20 scores in a non-clinical sample were correlated with poorer performance on the MASC, alexithymia did not predict ToM performance once autistic traits were controlled for, a finding that is in line with our current results.

One limitation of the current study is that although the control and ASD groups did not differ significantly on alexithymia severity, the ASD group still showed higher alexithymia scores than the control group, with a moderate effect size. Nonetheless, the fact that there was no significant difference in RMET performance between the control and ASD groups, despite a higher mean score for alexithymia in the ASD group, suggests that a more closely matched sample might show even more similar RMET performance than found in our participants. Another limitation that should be considered is the small sample size, particularly with regard to the ASD group. Finally, our sample included a relatively low proportion of participants who met the criterion for severe alexithymia in both the control (33.33%) and ASD (31.58%) groups. Future research could investigate RMET performance in matched groups with higher rates of alexithymia, ideally with an equal proportion of participants scoring above and below the cut-off of ≥61 in each group.

Overall, the present study demonstrates that the RMET indexes emotion recognition, rather than ToM ability. It also provides further support for the alexithymia hypothesis of emotion-related impairments in ASD (Bird & Cook, 2013), which posits that, where emotion recognition deficits are observed in ASD, they are attributable to co-occurring alexithymia, rather than to ASD *per se*. We therefore urge caution when using the RMET to measure mental state understanding and encourage researchers to control for alexithymia when using tasks with an emotional component to measure social cognition.

References

Adams, R. B., Jr., Rule, N. O., Franklin, R. G., Jr., Wang, E., Stevenson, M. T., Yoshikawa, S., . . . Ambady, N. (2010). Cross-cultural reading the mind in the eyes: an fMRI investigation. *J Cogn Neurosci, 22*(1), 97-108.

Ahmed, F., & Miller, S. (2011). Executive function mechanisms of theory of mind. *Journal of Autism and Developmental Disorders, 41*(5), 667-678.

Anagnostou, E., Soorya, L., Chaplin, W., Bartz, J., Halpern, D., Wasserman, S., ... & Hollander, E. (2012). Intranasal oxytocin versus placebo in the treatment of adults with autism spectrum disorders: a randomised controlled trial. *Molecular Autism, 3*(1), 16.

Austin, E. (2005). Personality correlates of the broader autism phenotype as assessed by the Autism Spectrum Quotient. *Personality and Individual Differences, 38*(2), 451-460.

Back, E., Ropar, D., & Mitchell, P. (2007). Do the eyes have it? Inferring mental states from animated faces in autism. *Child Development, 78*(2)*,* 397-411.

Bagby, R., Parker, J., & Taylor, G. (1994). The twenty-item Toronto Alexithymia Scale--I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*(1)*,* 23-32.

Baron-Cohen, S., Leslie, A., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21,* 37-46.

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. *Journal of Autism and Developmental Disorders, 29*(5)*,* 407-418.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001a). The "Reading the Mind in the Eyes" Test Revised version: A study with normal adults, and adults with

Asperger Syndrome or High-functioning Autism. *Journal of Child Psychology and Psychiatry, 42*(2)*,* 241-251.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001b). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31 (1), 5–17.

Berlim, M. T., McGirr, A., Beaulieu, M. M., & Turecki, G. (2012). Theory of mind in subjects with major depressive disorder: is it influenced by repetitive transcranial magnetic stimulation? *World J Biol Psychiatry*, *13*(6), 474-479.

Bird, G., & Cook, R. (2013). Mixed emotions: The contribution of alexithymia to the emotional symptoms of autism. *Translational Psychiatry, 3,* 1-8.

Bora, E., Vahip, S., Gonul, A., Akdeniz, F., Alkan, M., ... & Eryavuz, A. (2005). Evidence for theory of mind deficits in euthymic patients with bipolar disorder. *Acta Psychiatrica Scandinavica, 112*(2), 110-116.

Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger Syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain, 125*(8), 1839-1849.

Cook, R., Brewer, R., Shah, P., & Bird, G. (2013). Alexithymia, not autism, predicts poor recognition of emotional facial expressions. *Psychological Science, 24*(5)*,* 723-732.

Corcoran, R., Mercer, G., & Frith, C. D. (1995). Schizophrenia, symptomatology and social inference: investigating “theory of mind” in people with schizophrenia. *Schizophrenia Research, 17*(1), 5-13.

Devine, R. T. & Hughes, C. (2013). Silent films and strange stories: theory of mind, gender, and social experiences in middle childhood. *Child Development, 84*(3), 989-1003.

Dyck, M., Ferguson, K., & Shochet, I. (2001). Do autism spectrum disorders differ from each other and from non-spectrum disorders on emotion recognition tests. *European Child and Adolescent Psychiatry, 10*, 105-116.

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., Kessler, J., Woike, J., Wolfe, O., Convit, A. (2006). Introducing MASC: a movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders, 36(5),* 623-636.

Fairchild, G., Van Goozen, S., Calder, A., Stollery, S., & Goodyer, I. (2009). Deficits in facial expression recognition in male adolescents with early-onset or adolescent-onset conduct disorder. *Journal of Child Psychology and Psychiatry, 50*(5), 627-636.

Forni-Santos, L. & Osório, F. L. (2015). Influence of gender in the recognition of basic facial expressions: A critical literature review. *World Journal of Psychiatry, 5*(3), 342-351.

Freedman, M., Binns, M., Black, S., Murphy, C., & Stuss, D. (2013). Theory of mind and recognition of facial emotion in dementia: a challenge to current concepts. *Alzheimer Disease and Associated Disorders, 27*(1), 56-61.

Gökçen, E., Frederickson, N., & Petrides, K. V. (in press). Theory of Mind and Executive Control Deficits in Typically Developing Adults and Adolescents with High Levels of Autism Traits. *Journal of Autism and Developmental Disorders.*

Grynberg, D., Chang, B., Corneille, O., Maurage, P., Vermeulen, N., Berthoz, S., & Luminet, O. (2012). Alexithymia and the processing of emotional facial expressions (EFEs): systematic review, unanswered questions and further perspectives. *PLoS One*, *7*(8), e42429.

Khorashad, B., Baron-Cohen, S., Roshan, G., Kazemian, M., ... & Afkhamizadeh, M. (2015). The "Reading the Mind in the Eyes" Test: Investigation of Psychometric Properties and Test-retest Reliability of the Persian version. *Journal of Autism and Developmental Disorders, 45*(9), 2651-2666.

Lane, R., Hsu, C., Locke, D., Ritenbaugh, C., & Stonnington, C. (2015). Role of theory of mind in emotional awareness and alexithymia: Implications for conceptualisation and measurement. *Consciousness and Cognition, 33*, 398-405.

Lord, C., Risi, S., Lambrecht, L., Cook, E., Leventhal, B., DiLavore, P., Pickles, A., & Rutter, M. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3)*,* 205-223.

Milosavljevic, B., Leno, V. C., Simonoff, E., Baird, G., Pickles, A., ... & Happé, F. (2016). Alexithymia in adolescents with autism spectrum disorder: Its relationship to internalising difficulties, sensory modulation and social cognition. *Journal of Autism and Developmental Disorders, 46*(4), 1354-1367.

Montag, C., Ehrlich, A., Neuhaus, K., Dziobek, I., Heekeren, H. R., Heinz, A., & Gallinat, J. (2010). Theory of mind impairments in euthymic bipolar patients. *Journal of Affective Disorders, 123*(1-3), 264-269.

Moriguchi, Y., Ohnishi, T., Lane, R., Maeda, M., Mori, T., ... & Komaki, G. (2006). Impaired self-awareness and theory of mind: an fMRI study of mentalizing in alexithymia. *NeuroImage, 32*(3), 1472-1482.

Nemiah, J., Freyberger, H., & Sifneos, P. (1976). Alexithymia: A view of the psychosomatic process. In O. W. Hill (Ed.) *Modern Trends in Psychosomatic Medicine* (pp. 430-439). London, England: Butterworths.

O'Nions, E., Sebastian, C., McRory, E., Chantiluke, K., Happé, F., & Viding, E. (2014). Neural bases of theory of mind in children with autism spectrum disorders and children with conduct problems and callous-unemotional traits. *Developmental Science, 17*(5), 786-796.

Parker, J., Taylor, G., & Bagby, R. (2003). The 20-Item Toronto Alexithymia Scale--III. Reliability and factorial validity in a community population. *Journal of Psychosomatic Research, 55*(3), 269-275.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have theory of mind? *Behavioural and Brain Sciences, 1*(4), 515-526.

Roeyers, H., Buysse, A., Ponnet, K., & Pichal, B. (2001). Advancing advanced mind-reading tests: Empathic accuracy in adults with a pervasive developmental disorder. *Journal of Child Psychology and Psychiatry, 42*(2), 271-278.

Santiesteban, I., Bird, G., Tew, O., Cioffi, M. C., & Banissy, M. J. (2015). Mirror-touch synaesthesia: Difficulties inhibiting the other. *Cortex, 71*, 116-121.

Sharp, C. (2008). Theory of mind and conduct problems in children: Deficits in reading the "emotions of the eyes". *Cognition and Emotion, 22*(6), 1149-1158.

Spek, A., Scholte, E., & van Berckelaer-Onnes, I. (2010). Theory of mind in adults with HFA and Asperger Syndrome. *Journal of Autism and Developmental Disorders, 40,* 280-289.

Stone, V. E., Baron-Cohen, S., Calder, A., Keane, J., & Young, A. (2003). Acquired theory of mind impairments in individuals with bilateral amygdala lesions. *Neuropsychologia, 41*(2), 209-220.

Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., ... & Preti, A. (2013). The "Reading the Mind in the Eyes" Test: Systematic review of psychometric properties and a validation study in Italy. *Cognitive Neuropsychiatry, 18*(4), 326-354.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonia: The Psychological Corporation.

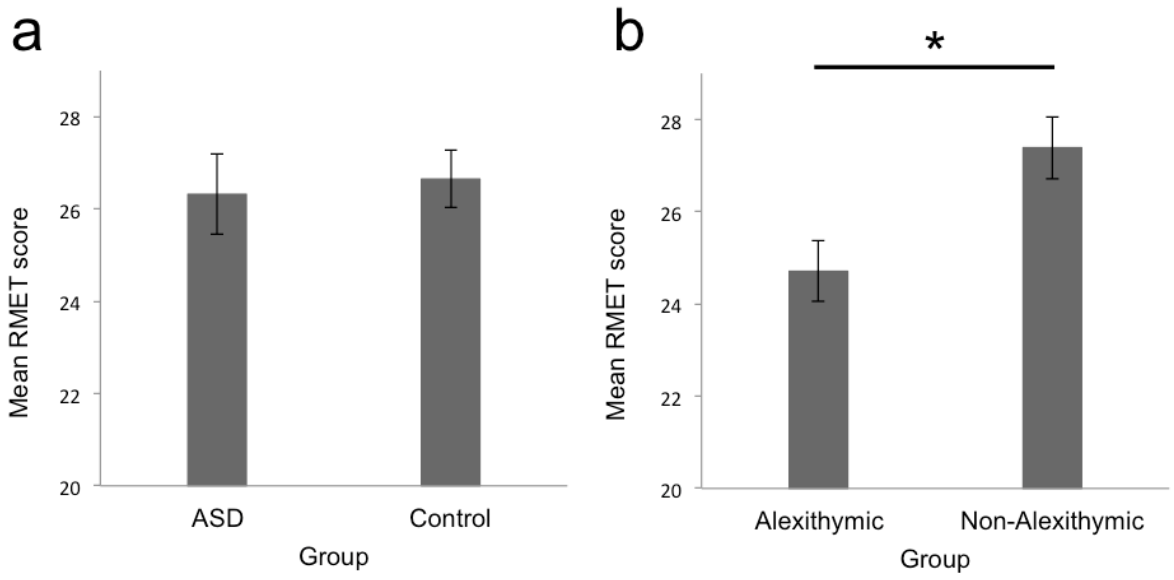


*Figure 1*. RMET performance was unaffected by ASD (a), but was negatively impacted by alexithymia (b). Error bars indicate standard error of the mean. Asterisk indicates significant difference at $p < .05$.

Supplementary Material

## Supplementary Methods

### MASC Control Questions

The control questions for the MASC were as used in Santiesteban et al. (2015) and are listed in Supplementary Table 1.

### Supplementary Table 1

*MASC Control Questions*

| Location in film after ToM question no. | Question text | Answer text | | | | Correct |
|---|---|---|---|---|---|---|
| | | A | B | C | D | |
| 3 | Who did Betty say Sandra thought was cute? | Pete | Cliff | Michael | David | B |
| 6 | How many times has Sandra met Cliff before now? | Never | Once | Twice | Many times | B |
| 6 | What time are they meeting? | 7pm | 8pm | 9pm | 7:30pm | B |
| 8 | What did Cliff say *wasn't* a reason for coming? | Drinks | The meal | To flirt | To get the money back from Michael | C |
| 9 | What did Cliff ask for? | A cola | Water | Orange juice | Beer | A |
| 12 | How long were Cliff and his ex together? | 4 years | 3 years | 2 years | 4 months | A |

| | | | | | | |
|---|---|---|---|---|---|---|
| 12 | What country did Sandra's ex go to? | The Netherlands | Sweden | Denmark | Germany | B |
| 13 | How did Cliff likely shave in Sweden? | Outdoors with an electric shaver | As usual, in the bathroom | With a razor and cold water | In his hotel room | C |
| 14 | Who has owned pets? | Michael and Betty | Michael and Cliff | Cliff and Sandra | Betty and Cliff | C |
| 20 | Which character *hasn't* met two of the other characters before this point? | Sandra | Betty | Cliff | Michael | B |
| 26 | Who does Sandra look at during the toast? | Betty | Cliff | Michael | Paulie | B |
| 29 | Apart from Sandra, who else enjoys motorcycles? | Betty | Cliff | Michael | Paulie | A |
| 30 | How many cups of cream go in the recipe? | 1 | 2 | 3 | 4 | B |
| 33 | What kind of pasta sauce are the four characters preparing? | A sauce with sardines | A sauce with ground meat | A sauce with red peppers | A sauce with salmon | C |
| 37 | How many characters *don't* drink alcohol? | 2 | 1 | 0 | 3 | C |
| 38 | Which chips does Betty have to play? | The white chips | She can pick any color | The black chips | The same chips that Cliff played | A |

| | | | | | | |
|---|---|---|---|---|---|---|
| 39 | How many people is the game intended for? | 1 | 3 | 5 | 4 or 2 | D |
| 44 | Who is likely to have drunk more alcohol by the time they go to sleep? | Michael | Cliff | Sandra | Betty | D |
| 45 | What are Cliff's favourite leisure time activities? | Playing sports | Engaging in various cultural activities | Going to parties | Reading books | B |
| 45 | Which of the four characters is involved in a relationship? | None of them | Cliff and Sandra | Michael | Betty | A |
| 45 | What was the weather like on that evening? | Cold and dry | Mild and overcast | Rainy | Cold and snowy | A |

**Supplementary Results**

**Supplementary Table 2**

*Hierarchical Regression Analyses*

| **a** | | **RMET performance** | | | |
|---|---|---|---|---|---|
| **Step** | **Predictor** | ***B*** | ***p*** | $R^2$ | $\Delta R^2$ |
| 1 | Gender | -.104 | .511 | 1.1% | 1.1% |
| 2 | Gender | -.094 | .556 | 2.3% | 1.2% |
| | ASD severity | -.111 | .490 | | |
| 3 | Gender | -.027 | .864 | 13.8% | 11.5% |
| | ASD severity | .104 | .565 | | |
| | Alexithymia | -.410 | .030 | | |
| **b** | | **MASC mental state performance** | | | |
| **Step** | **Predictor** | ***B*** | ***p*** | $R^2$ | $\Delta R^2$ |
| 1 | Gender | .320 | .041 | 10.2% | 10.2% |
| 2 | Gender | .353 | .028 | 12.7% | 2.5% |
| | Alexithymia | -.161 | .304 | | |
| 3 | Gender | .346 | .029 | 18.2% | 5.5% |
| | Alexithymia | -.012 | .949 | | |
| | ASD severity | -.277 | .125 | | |
| **c** | | **MASC cognitive ToM performance** | | | |
| **Step** | **Predictor** | ***B*** | ***p*** | $R^2$ | $\Delta R^2$ |
| 1 | Gender | .347 | .026 | 12.0% | 12.0% |
| 2 | Gender | .363 | .025 | 12.6% | 0.6% |
| | Alexithymia | -.077 | .621 | | |
| 3 | Gender | .353 | .023 | 21.7% | 9.1% |
| | Alexithymia | .116 | .513 | | |
| | ASD severity | -.357 | .046 | | |

*Note.* Regression analyses predicting (a) RMET performance (including gender in the first step, ASD symptom severity in the second step, and alexithymia in the third step); (b) performance on all mental state MASC questions (including gender in the first step, alexithymia in the second step, and ASD symptom severity in the third step); and (c) performance on pure cognitive ToM MASC questions (including gender in the first step, alexithymia in the second step, and ASD symptom severity in the third step).