A new test of advanced theory of mind: The 'Strange Stories Film task' captures social processing differences in adults with autism spectrum disorders.

Kim Murray[1], Kate Johnston[1], Helen Cunane[6], Charlotte Kerr[7], Debbie Spain[2], Nicola Gillan[3], Neil Hammond[4,5], Declan Murphy[3,4,5] & Francesca Happé[2]

[1] Institute of Psychiatry, Psychology & Neuroscience
King's College London
De Crespigny Park, Denmark Hill, London SE5 8AF

[2]MRC Social Genetic and Developmental Psychiatry Research Centre
Institute of Psychiatry, Psychology & Neuroscience
King's College London
De Crespigny Park, Denmark Hill, London SE5 8AF

[3]Sackler Institute for Translational Neurodevelopment and Department of Forensic and Neurodevelopmental Sciences
Institute of Psychiatry, Psychology and Neuroscience,
King's College London
De Crespigny Park, Denmark Hill, London SE5 8AF

[4]Behavioural Genetics Clinic and Adult Autism Service,

Behavioural and Developmental Psychiatry Clinical Academic Group

South London and Maudsley NHS Foundation Trust

Maudsley Hospital,

Denmark Hill, London SE5 8AZ

[5]National Autism Unit,

South London and Maudsley NHS Foundation Trust

Bethlem Royal Hospital,

Bromley,

Kent,

BR3 3BX.

[6]Bradford Dementia Group,

School of Health Studies

University of Bradford

Richmond Road, Bradford, BD7 1DP

[7]Acute Neuro-Rehabilitation Unit

Wellington Hospital

London

NW8 9LE

No. of Tables: 5

No. of Figures: 1

No of text pages: 24 (not including acknowledgements or references)

Conflict of interest: None.

Corresponding Author: Francesca Happe

Email: francesca.happe@kcl.ac.uk

Telephone: +44 (0)20 7836 5454

Word Count: 5996

Lay Abstract

Autism Spectrum Disorders (ASD) are characterised by difficulties in understanding social situations and in particular social interactions where non-literal language (e.g. sarcasm) is used. Researchers have devised a number of measures to try and capture these subtle social differences in adults with ASD. We sought to overcome the limitations of previous measures by designing a novel task (the Strange Stories Film task) using short video clips, which were based on a well-established test of social understanding (the Strange Stories task). After piloting, the new task was shown to a group of 20 adults with ASD and their responses were compared to a group of adults without any diagnosis. The study found that the new task was superior to previous measures, and could be useful in assessing social difficulties in a clinical setting. We also consider the limitations of our study and future research directions.

Scientific Abstract

Introduction: Real-life social processing abilities of adults with autism spectrum disorders (ASD) can be hard to capture in lab-based experimental tasks. A novel measure of social cognition, the ‘Strange Stories Film task’ (SSFt), was designed to overcome limitations of available measures in the field.

Method: Brief films were made based on the scenarios from the Strange Stories task (Happé 1994) and designed to capture the subtle social-cognitive difficulties observed in ASD adults. 20 neurotypical adults were recruited to pilot the new measure. A final test set was produced and administered to a group of 20 adults with ASD and 20 matched controls, alongside established social cognition tasks and questionnaire measures of empathy, alexithymia and ASD traits.

Results: The SSFt was more effective than existing measures at differentiating the ASD group from the control group. In the ASD group the SSFt was associated with the Strange Stories task.

Conclusion: The SSFt is a potentially useful tool to identify social cognitive dis/abilities in ASD, with preliminary evidence of adequate convergent validity. Future research directions are discussed.

Key words: Autism spectrum disorder, adults, advanced theory of mind, social cognition, empathy, mentalising, and alexithymia.

Introduction

Over the last thirty years the social cognitive abilities of individuals with Autism Spectrum Disorder (ASD; American Psychiatric Association, 2013) have been widely studied, with a range of theories proposed to conceptualise observed differences in social understanding. The 'Theory of Mind' (ToM) account suggests that ASD is characterised by a fundamental difficulty in the ability to represent the mental states of others. ToM was initially assessed through 'false belief' tasks requiring *first*- ("what does Sally [mistakenly] think") and, later, *second*- order mental state attribution ("what does John [mistakenly] think that Mary thinks") (Baron-Cohen, 1989; Baron-Cohen, Leslie, & Frith, 1985; Happé, 1995). However, ceiling effects are often observed on such tasks (which typically developing 5- to 7-year-olds pass), particularly in samples of ASD adults without intellectual impairment (Bowler, 1992; White, Hill, Happé, & Frith, 2009).

To overcome the lack of sensitivity of simple false belief tasks, a number of tests of more advanced 'mentalising' have been developed. For example, Happé's 'Strange Stories' (SS; Happé, 1994) presented 24 short written vignettes (two each of 12 themes, e.g. irony, double-bluff, white lie, persuasion) requiring participants to explain the speaker's intention in cases where utterances were not literally true. A recent meta-analysis suggests that the SS are useful in differentiating ToM abilities in adults with ASD from matched controls (Chung, Barch, & Strube, 2013). However, the written format of the SS task means it does not tap the ability to process naturalistic social cues, such as facial expression and vocal intonation. Furthermore, participants are given as much time as necessary to process the material, in marked contrast to the fast-paced nature of real-life social interaction, which may explain why individuals may 'pass' the SS yet still struggle in everyday situations (Scheeren, de Rosnay, Koot, & Begeer, 2013).

The problem of assessing subtle ToM difficulties in an ecologically valid manner has led to a conceptual divide between prioritising limited, but real interactions as stimuli (Roeyers, Buysse, Ponnet, & Pichal, 2001) or diverse and theoretically tailored stimuli produced through acted interactions (Dziobek et al., 2006; Mathersul, McDonald, & Rushby, 2013) The former has been advocated by Roeyers and colleagues utilising the empathic accuracy paradigm (for more information see; Ickes, Stinson, Bissonnette, & Garcia, 1990), and has been shown to be an effective means of capturing differences in social cognitive abilities between individuals with autism and matched controls (for both adults and adolescents; Ponnet, Buysse, Roeyers, & Clercq, 2008; Ponnet, Roeyers, Buysse, De Clercq, & Van Der Heyden, 2004; Roeyers et al., 2001; Roeyers & Demurie, 2010). Whilst naturalistic in design, the nature of the stimuli (e.g. a brief conversation with a stranger) means particular linguistic constructs that people with ASD may struggle with (e.g. deception; Happe, 1994) are less likely to occur. This is especially important as incongruence between mental events and behavioural cues available to the viewer predicts poor performance on this task (Ponnet et al., 2008). The generic nature of the situation may also lead to a narrowing of the potential internal states experienced/to be inferred. A number of studies have attempted to capture social cognitive abilities using scripted social exchanges, which overcome some of these limitations, and these are outlined in Table 1.

Insert Table 1 about here

Table 1 shows that difficulties with social cognition are consistently found in individuals with ASD when acted stimuli are employed. Convergent validity (with other social cognition tasks) was demonstrated in most cases, but not all. A consistent limitation was the lack of challenging control stimuli. Where control questions were used (usually memory based questions) performance often reached ceiling (Dziobek et al., 2006) or, in some cases, yielded group differences (Heavey, Phillips, Baron-Cohen, & Rutter, 2000). The majority of the studies also used pre-existing film material, which may have been over-dramatized or chosen because of high affective content (Golan, Baron-Cohen, Hill, & Golan, 2006). Using overdramatized material undermines ecological validity, which has been heralded as essential in understanding social cognition in ASD (Dziobek, 2012). Finally, one of the most psychometrically and theoretically robust measures to date was filmed in Germany (Dziobek et al., 2006). To the authors' knowledge, the English dubbed version has not been validated in an English speaking country with an ASD sample. Moreover, a recent Spanish dubbed version yielded differing results to the original validation paper, with the Reading the Mind in the Eyes Test (RMET; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001) outperforming the dubbed measure. Dubbing material may differentially affect performance for individuals with ASD given research suggesting preferential gaze to the mouth region in ASD during viewing of social exchanges and possible greater sensitivity to visual-auditory asynchrony (Klin, Jones, Schultz, & Volkmar, 2003; Klin, Jones, Schultz, Volkmar, & Cohen, 2002).

Past measures of social cognition tended to focus exclusively on comprehension of a social situation (Mathersul et al., 2013). Recent research has begun to ask participants to *generate* possible social responses to typical interactions (e.g. 'What would you do in this situation?'; Jameel, Vyas, Bellesi, Roberts, & Channon, 2014) and found that autistic traits are associated with fewer pro-social responses. Mutual goals that require cooperation have been argued as fundamental in the development of social cognition and young people with autism's ability to help and cooperate with

others is likely to affect their social cognitive developmental trajectory (Liebal, Colombi, Rogers, Warneken, & Tomasello, 2008). The concept of pro-social behaviour is linked to 'social acting' (Yang & Baillargeon, 2013), where beliefs are decoupled from speech for the benefit of another (e.g. white lies). Yang and Baillargeon (2013) found it was 'social acting' and not social comprehension that predicted quality of social relationships in typically developing adults (rated high or low for autistic traits). These findings highlight a gap in the social cognition literature, which is particularly pertinent to the current study. How social knowledge is applied is important, given some findings that adults with ASD can make accurate mental state inferences (Ponnet, Buysse, Roeyers, & Corte, 2005) even though abilities to maintain meaningful relationships may be limited (Palmen, Didden, & Lang, 2012).

Overall, a range of measures are currently available to assess social cognition in ASD. Limitations include a lack of appropriate control material, non-naturalistic stimuli, a trade-off between length of test and richness of open-ended response data and a focus on receptive versus expressive skills, all of which suggest new social cognition tasks are still required.

The current study introduces a new measure of theory of mind that attempts to address a number of the limitations outlined above and assess mental state attribution in a more naturalistic way in adults with ASD. To assess validity, the new task was administered to ASD and typically developing (TD) adults alongside established and widely used social cognitive tests and questionnaires about ASD-relevant traits. Social cognition clearly encompasses a range of processes, including but not limited to mental state attribution (ToM) and emotion processing, which appear to be distinct, but interdependent (Brewer, Happé, Cook, & Bird, 2015). We therefore included standard tests of both processes alongside our novel task, as well as measuring self-reported empathy and alexithymia (difficulty reflecting on and describing one's emotions); work by Bird and colleagues (Bird & Cook, 2013; Cook, Brewer, Shah, & Bird, 2013) suggests that it is the degree of frequently co-occurring alexithymia rather than ASD

itself that predicts poor affective empathy and emotion processing in those with ASD (and other clinical groups). The aim of the current study was to test the sensitivity of our new measure to capture ASD mentalising abilities and examine its relationship to existing measures and social abilities in every day life.

Method

The Strange Stories Film Task

Scripts and filming: Scripts for the Strange Stories Film Task Pilot (SSFt-P) were developed using clinical experience, research literature and personal communication from an associate of the first author who has a ASD diagnosis. The SSFt-P was based on the original Strange Stories (Happé, 1994), which used the following types of scenarios to test the ability to attribute a speaker's intention : lie, irony, double bluff, pretence, joke, appearance/reality, white-lie, persuasion, misunderstanding, forgetting, contrary emotions and idioms. For an example script and screen shots of the measure, see Appendix 1. The language used in the scripts was kept as close to everyday spoken language as possible, and complex constructions or overly sophisticated vocabulary were avoided. Three or four scripts for each theme present in Happé (1994) SS were written to enable sub-optimal clips to be deleted from the final version. In addition, ten control scripts were written. These mirrored the experimental clips in terms of length, cognitive load and linguistic sophistication. However, they required logical reasoning (e.g. economic decision making or understanding of natural phenomena) to decipher the characters' utterances or behaviour, rather than requiring attribution of mental states, akin to the control vignettes used by Fletcher et al. (1995) and White et al. (2009).

The actors were semi-professional and were recruited via online advertisement and audition. In each scene, a third person perspective shot first showed the viewer the context of the social exchange. The scenes of this initial shot were kept as sparse as possible (e.g. artwork was taken from the walls) to minimise possible distractions that might differentially distract individuals with ASD (Klin et al., 2003), but were still kept naturalistic and did not burden participants' imaginations (scenes were easy to identify as e.g., sitting room or kitchen). All speech was directed to camera and filmed in the first person (as if the viewer were in the conversation), both to reduce possible

attention biases for the viewers with ASD (Klin et al., 2003) and to provide the same sort of information available in a real-life conversation (e.g. full-face emotional expressions).

Questions

Three questions were used to assess social understanding immediately following the viewing of each clip: 1) Intention, 2) Interaction, and 3) Memory Question. The Intention question 'Why did X say that?' was taken from Happé (1994) SS, and always referred to the last speaker and final utterance of the film clip. The Interaction question asked about a possible response to the final utterance of the clip; 'If you were in Y's [other character i.e. not X] situation, what would you say next?' This question was designed to assess participants' ability to generate a response to the inferred mental state (e.g., intention) of the speaker, in order to continue the social exchange. The Memory question was used to assess potential lapses in attention or gross difficulties in memory, and always took the form of a closed question about a factual aspect of the clip, e.g. 'What instrument was X playing?'

Scoring

The scoring system for the SSFt was kept as simple as possible and was based on White et al. (2009) p.1109-1117 and Happé (1994). For the Intention question, the score given reflected how accurately the participant recognised the relevant mental states, and captured the difference between simple and more complex mental state inferences (e.g. second-order versus first-order mental state attribution), simplistic or incomplete responses, which have previously differentiated ASD from non ASD populations (Happé, 1995). Mental state language was also scored to identify whether participants used mental state words (e.g. he wants or she thinks) to describe the

actors' intentions. For the Interaction question, scoring reflected the appropriateness of the participant's suggested response to the speaker. For the Memory question, all scores were based on correctly identifying the factual information in the relevant clip.

As an example, the scoring system for the white lie scene (see Appendix 1 for screen shots of 'white lie' clip), which was based on White et al. (2009) p.1110 is outlined below:

White Lie:

Intention Question: Why did Max say that?

Accuracy:

2 points - reference to white lie or making her feel good or not wanting to hurt Alice's feelings

1 point - response that states simple traits (e.g., he is nice, being supportive, polite) or is simply relational (e.g., he likes her). Incomplete response (e.g., offering fake praise) or solely motivational (e.g., so she won't be annoyed, avoid an argument, reassure her).

0 points – incorrect e.g. 'he thought it was good' or only 'he didn't like it', or irrelevant responses.

Mental State Language

0 points - no mental state words.

1point – simple mental state words regarding one character or another character's actions OR words that imply psychological states in social context.

2 points – meta-cognitive statements e.g. beliefs about beliefs OR intentions to affect another person's mental state e.g. he didn't want to hurt her feelings OR complex collection of mental states.

Interaction Question: 'If you were in Alice's situation, what would you say next?'

2 points – statement that acknowledges that Max's comment might not have been completely honest *and* either asks for additional clarification or additional feedback in socially appropriate manner (e.g., 'do you really mean that?'); sarcastic agreement with his opinion that implies it could be improved.

1 point –Incomplete response e.g. 'thank you', that doesn't reflect white lie.

0 points – don't know, socially inappropriate (e.g. response that sees comment as unsupportive or misses intention of white lie), or irrelevant comments.

Memory Question: "What instrument was Alice playing?"

1 point – mentions guitar.

0 points – don't know, can't remember or incorrect recall.

Similar scoring systems are described in White et al. (2009), Devine and Hughes (2013) and Castelli, Frith, Happé, and Frith (2002). Of particular importance, this type of system has been shown to be reliable in other film-based tasks (Devine & Hughes, 2013). In accordance with these systems, possible scores ranged from 0-2 for the Intention, Mental State Language and Interaction questions and 0-1 for the memory

question for each clip; maximum total scores were therefore 24, 24 and 12 respectively. Full scoring guidelines are available from the last author.

Piloting

20 neurotypical adults (10 male, 10 female) were recruited via an opportunity sample. The mean age of the sample was 28.8years (SD = 7.66). Participants were only recruited into the study if they had an Autism Quotient (AQ) score below 32 (Baron-Cohen et al., 2001) No participants who opted into the study had to be rejected from the pilot due to the presence of high ASD traits as measured by the AQ (M= 10.80 SD = 3.81 range = 6-17). Ethical approval was granted by the King's College London Psychiatry, Nursing and Midwifery Ethics Sub-Committee (PNM/10/11-22). The SSFt-p set consisted of 48 clips. Thirty-eight clips followed the themes of the 12 types of mental state vignettes presented in (Happe, 1994) Strange Stories. Ten control clips were based on physical state reasoning stories (White et al., 2009).

Scenes were then selected based on who delivered the target utterance (male or female actor), and setting (kitchen, living room, outside, in an office) with the aim of having a balanced set of scenes. Ineffective clips were also removed if: fewer than a quarter of viewers identified the whole intended meaning in response to the Intention question (6 experimental and 2 control scenes); or a new character was introduced (n=1).

The final set consisted of 12 experimental (one of each theme) and 3 control clips, where the female actor delivered the target utterance on nine occasions and the male on six. A second set of 12 viable clips remained for future research purposes.

Experimental study

Participants

A total of 40 participants were recruited into the experiment. Individuals in the ASD group (N=20) had all been assessed by a specialist adult ASD diagnostic service. The control group was recruited through an opportunity sample and advertisements in the local community detailing the research. To be included in the study, participants in the ASD group had to have a formal diagnosis of either Asperger Syndrome (N=16) or Autistic Disorder (N=4) decided by a multi-disciplinary team according to ICD-10 criteria, be aged between 18 and 65 years at the time of testing, be fluent in English, have a verbal IQ> 70, have no other neurodevelopmental or organic disorder present (e.g. head injury) and none of the following psychiatric diagnoses: schizophrenia, eating disorders, personality disorder or substance abuse/dependence. Inclusion criteria for the control group were (in addition to the criteria above excluding the ASD diagnosis and ASD structured interviews); an AQ score below 32. Demographics of the groups can be seen in Table 2.

Insert Table 2 about here

The two groups were matched for age, gender and verbal ability (the control group's scores ranged from 81-138 and the ASD group's scores ranged from 73-134). The AQ acted as a screening measure for ASD traits (primarily for exclusion of participants from the Control group), and showed a significant difference between the groups (the control group's scores ranged from 5-30, while the ASD group's scores ranged from 18-48). In all but one case, a suitable informant was available to provide developmental history information for the participant's ASD diagnosis via an ADI-R (Lord, Rutter, & Couteur, 1994). For the individual who did not have ADI-R data, diagnosis was supported by an ADOS (Lord et al., 1989). One participant in the ASD

group was unable to complete the AQ due to testing constraints. Ethical Approval for the study was granted by the National Research Ethics Service Committee – London, Westminster (13/LO/0092).

Measures

Wechsler Intelligence Scales: Verbal ability was measured using The Wechsler Abbreviated Scale of Intelligence (WASI), which is a brief, reliable and valid measure of general intelligence that is recommended for research purposes (Wechsler, 1999). In cases where a neuropsychological assessment had been completed within the NHS clinic they were recruited from, participants' verbal ability was estimated from the short form of the Wechsler Adult Intelligence Scale–III (WAIS-III; Axelrod, Ryan, & Ward, 2001). The WASI and the WAIS-III scores show good convergent validity (Wechsler, 1999). In two cases, the Wechsler Adult Intelligence Scale–IV (WAIS-IV) was used (Wechsler, 2008).

The Twenty item Toronto Alexithymia Scale (TAS-20): The TAS-20 is a self-report instrument developed to identify alexithymia traits in both clinical and non-clinical populations (Bagby, Parker, & Taylor, 1994). In adults with ASD the TAS-20 shows good test-retest reliability, convergent validity and discriminate validity (Berthoz & Hill, 2005)

The Interpersonal Reactivity Index (IRI): The IRI is a 28 item self-report questionnaire designed to test empathy as a multi-dimensional construct (Davis, 1980; Davis, 1983). The IRI has been shown to effectively discriminate ASD individuals from a matched typically developing adult sample (Rogers, Dziobek, Hassenstab, Wolf, & Convit, 2007)

The Reading the Mind in the Eyes task (RMET): The RMET is a widely-used forced choice measure designed to tap mentalising abilities (Baron-Cohen et al., 2001). Participants view 36 photographs of the eye region of a face and in each case choose from four words the one that best describes the emotion/internal state depicted. The RMET is deemed one of the most effective socio-cognitive tasks available (Pinkham et al., 2013).

The Awareness of Social Inference Test (TASIT): Participants completed the forced choice 'Emotion Recognition' subsection (Part 1) of the TASIT (McDonald, Flanagan, & Rollins, 2002). Participants view 28 short film clips, where an actor performed one of the 6 universal emotions: Anger, Sadness, Happiness, Anxiety, Surprise, Disgust, or was emotionally 'Neutral.'

The Frith-Happé Animations (Triangles): The Triangles is a silent dynamic ToM task (Castelli et al., 2002). Participants viewed a practice animation followed by four theory of mind animations on a computer screen. The Triangles task has been shown to reliably differentiate between high-functioning ASD groups and verbal ability matched control groups.

The Strange Stories (SS): Participants completed a short form of the SS task (Fletcher et al., 1995; Happé, 1994) consisting of 8 short vignettes (two versions of the following themes: White lie, persuasion, double bluff and misunderstanding). The SS task has been shown to reliably differentiate adult ASD participants from control groups (Chung et al., 2013).

The Strange Stories Film Task (SSFt): Prior to the task, participants were informed about the nature of the task and the characters' relationship. Participants viewed 3 practice clips, two of which were experimental clips and one was a control clip, but did

not receive feedback on performance. Participants then viewed 15 clips; 12 mental state clips and three control clips, presented in a quasi-randomised order (A). Half the participants viewed order A and the other half viewed the same clips but in reversed order (B). Clips lasted no longer than 27 seconds each (M= 17.5, *SD*= 5.83) and the total running time was six minutes and 21 seconds. Participants were asked the three questions described above following each clip (including the three practice clips). Cronbach's alpha of 0.58 for the Intention, 0.42 for the mental state language (e.g. use of words like want, feel etc.) question and 0.73 for the Interaction question, suggest adequate and satisfactory levels of internal consistency for the Intention and Interaction question respectively. The control questions (Intention and Interaction) showed alpha values lower than 0.4, which might be expected since they were not designed to tap a unitary underlying construct. Intra class coefficients (ICC) were above .80 on all elements of the SSFt suggesting high levels of inter-rater reliability.

Procedure

Testing took place for all participants in a quiet room, with breaks given as needed. Participants completed the AQ, TAS-20, IRI, SS, RMET, Triangles, TASIT and the SSFt. In some cases participants chose to complete some questionnaires/tasks outside the main session.

Statistical analysis

In all cases where VIQ correlated with performance on behavioural measures of social cognition, ANCOVA was completed with VIQ as a covariate; otherwise t-tests were performed to compare mean differences. Sensitivity analysis was performed using an independent bootstrap analysis to test whether the results were robust against deviations from normal distribution (Chung et al., 2013). Alpha values were set

at <.05 and effect sizes calculated using Cohen's d (Chong & Choo, 2011). Partial Cohen's d effect sizes were calculated for the ANCOVA analyses (Cohen, 1992). Depending on the variables' distribution/correlation with VIQ, correlations/partial correlations were calculated using either Spearman's or Pearson's correlation coefficient. For the correlation analysis alpha value was reduced to <.01 to account for multiple comparisons. A Receiver Operator Characteristic (ROC) curve was performed to demonstrate the traditional social cognition measures and the new SSFt's ability to assign participants to their correct diagnostic group.

Results

Group differences on the standard social cognition tasks and questionnaires will be reported, before presenting the results from our novel film task, and its relationship to existing measures.

Table 3 shows the groups differences on the standard social cognition measures.

Insert Table 3 about here

The analyses revealed a significant group difference between the adults with ASD and the controls on the SS accuracy score, but not on the degree of mental state language used to explain behaviour (see Table 3). Accuracy and mental state language scores on the Triangles were significantly lower for the ASD group than for controls. There was a borderline significant group difference on the RMET but no significant difference on the emotion recognition subtest of the TASIT.

Table 4 shows the two groups' responses to the TAS-20 and IRI questionnaires.

Insert Table 4 about here

For the cognitive empathy subscales of the IRI, significant differences were seen between the two groups on the perspective taking subscale (see Table 4). Both the control group and individuals with ASD reported equal levels of empathic concern and fantasising. However, for the personal distress scale individuals with autism rated themselves as significantly higher (see Table 4).

The TAS-20 revealed significantly higher levels of alexithymia in the ASD than the TD group, across each of the subscales and the total scale. In addition, significantly

more of the ASD group (52.6%) reported levels of alexithymia that passed the suggested cut-off (total score > 60 ; Bagby et al., 1994) compared to the control group (20%; X^2 (1,39) = 4.51, p = .034).

Table 5 shows the groups' performance on the SSFt.

Insert Table 5 about here

Participants with ASD scored significantly lower than controls on the Intention Accuracy and Interaction questions of the SSFt experimental clips, but their Mental State Language scores were statistically equivalent. Both groups performed equally well on the Intention (Accuracy and Mental State Language) and Interaction questions on the control clips (see Table 5). No significant group differences were seen on the memory question for experimental or control clips, however, for the control memory questions this was not supported by the bootstrap analysis.

Analysis revealed a trend towards a significant association between the Intention and Interaction scores of the SSFt in the ASD group once verbal abilities had been controlled for (r = .56, p = .012). For the controls however this association was statistically significant (r = .62, p = .004). Fischer r-to-z transformation revealed that these two coefficients were not statistically different however (z = -.27, p =.79).

Insert Figure 1 about here

The ROC curve in Figure 1 demonstrates each social cognition measure's ability to accurately assign the participants to their respective group. Only measures in which there was a significant mean difference between the two groups were included. Mental state language scores did not differentiate correct from incorrect responses so were not included. The AUC values and corresponding 95% confidence intervals for the scales were .87 (.76 - .98) for the SSFt Interaction scores, .78 (.63 - .93) for the SSFt Intention accuracy scores, .72 (.56 – .88) for the SS Accuracy score, .71 (.55 - .88) for the RMET and .69 (.53 - .86) for the Triangles accuracy score. Of note, all of the confidence intervals overlapped. The RMET was not included in the figure as it had a missing data point.

The SSFt convergent validity

Partial correlations (controlling for verbal ability) were performed and revealed the following in the ASD group. First, the correlation between the Intention Accuracy score on the SSFt and the Accuracy score on the SS was significant, (r = .61, p = .006). The Mental State Language scores however, did not correlate significantly between the SS and the SSFt within this group (r_s <.40). The Intention scores (Accuracy and Mental State Language) did not correlate with the corresponding scores from the Triangles task (r < *.40).* Finally, the SSFt accuracy score did not significantly correlate with the RMET (r < *.40).*

For the control group, the Intention scores (Accuracy and Mental State Language) did not correlate with the SS's Accuracy (r_s < .40) and Mental State Language (r < .40) scores, respectively. Similarly, no association was revealed between the Accuracy

score on the SSFt and the RMET (r_s < .40). The relationships between the SSFt Intention scores (Accuracy and Mental State Language) and the corresponding scores on the Triangles task were substantial, although they missed the significance level of .01 set here (r_s= .40, p = .084 and r = .54, p = .015, respectively).

SSFt association with childhood ASD symptoms and self-reported ASD traits, empathy and alexithymia.

Within the ASD group, partial correlations revealed no significant associations between the SSFt Intention Accuracy or Interaction scores and the ADI-R Reciprocal Social Interaction (r_s < .40) and Communication (r < .40), or the AQ in the ASD group (r < .40). The Intention Mental State Language score of the SSFt correlated negatively with the ADI-R communication domain (higher scores on the ADI-R indicate higher levels of ASD symptoms) although it did not reach the .01 significance level set here (r = - .47, p = .050).

For the control group, the AQ and the SSFt Intention Accuracy score revealed a substantial negative correlation although the .01 significance level was not met (r = -.50, p = .025), while the Intention Mental State Language score showed a significant negative association with the AQ (r = -.59, p = .006).

For the ASD group, partial correlation analysis (controlling for verbal ability) revealed no association between the SSFt Intention Accuracy scores and the IRI PT domain (r >.40). However, the Interaction question and the EC domain of the IRI showed a substantial partial correlation, but it did not meet the .01 significance level set here (r = .44 p = .067).

For the Control group the Accuracy score on the SSFt showed a substantial correlation with the PT subscale of the IRI, but it did not meet the .01 significance set here (r = .48, p = .032). Partial correlation (controlling for verbal ability) revealed no association between the IRI EC and the Interaction question of the SSFt (r_s < .40).

No significant associations were found in either group between alexithymia traits and performance on the SSFt (all r < .40).

Discussion

Overall, the SSFt was shown to be effective at discriminating between adults with and without a diagnosis of autism. Adults with ASD had lower scores, indicating difficulties with social cognition that could not be explained by general cognitive factors (e.g. verbal ability) and were specific to understanding the intentions behind nonliteral language in communication. The SSFt was superior to existing, well-evidenced measures of social cognition/emotion recognition in its ability to discriminate ASD from matched controls. The finding that the control group's performance was not undermined by ceiling effects (alongside the borderline significant association with questionnaire measures of autistic traits/empathy) suggests that the SSFt may also be useful for measuring individual differences in social cognitive ability in the general population. The development of a forced- choice paradigm that could be used online would facilitate this research and increase its scope for reaching more diverse samples (age, geographical location etc).

Perspective taking on the IRI and ASD traits (measured by the AQ) substantially correlated with the SSFt only in the control group. This might reflect differences in self-reflection in the ASD versus control group although this cannot be answered from this research. Future research including informant rated measures of perspective taking (Demurie, De Corel, & Roeyers, 2011) would help fill this gap in the literature. Informant based (retrospective) childhood ASD symptoms did not significantly correlate with performance on the SSFt again pointing to the benefits of *current* informant-rated autistic traits in future research. Also childhood ASD symptoms may not be a helpful correlate of adult social cognitive abilities due to the developmental nature of social cognition (Happé & Frith, 2013)

While the Intention question of the SSFt was effective in differentiating the two groups and replicated social cognitive differences observed in previous research using advanced theory of mind tasks, the Interaction question (the novel element) of this

social cognition paradigm yielded higher levels of sensitivity without compromising specificity. The ability to infer what others may be thinking may be *necessary but not sufficient* for generation of neurotypical social interaction in individuals with ASD. This notion fits Yang and Baillargeon's (2013) suggestion that it is the lack of 'social acting' that is most relevant to peer relation difficulties seen in adults with ASD traits. ASD participants who may comprehend why an individual is using figurative language in the SSFt (e.g. not to hurt the other's feelings), may still have a different appraisal of its usefulness and hence generate different possible subsequent responses (e.g. why did you say it's good when you clearly don't think that?). The Interaction question also involves generativity, which is among the executive functions suggested to be impaired in ASD (Channon, Crawford, Orlowska, Parikh, & Thoma, 2013; Hill, 2004). In future research with the SSFt, it would be useful to include measures of executive function to examine the role of (non-social) generativity in performance (Dziobek et al., 2006).

Alexithymia has received considerable interest as an independent but frequently co-occurring condition reported by those with ASD. Bird & Cook (2013) report evidence that it is alexithymia that explains emotion-recognition difficulties in individuals rather than autism per se. In the current sample, alexithymia was elevated in the ASD group, but there was no significant relationship between alexithymia and performance on the SSFt. The SSFt focuses primarily on recognition of propositional mental states (e.g. beliefs, intentions) rather than emotion processing, which may explain the lack of association (Lockwood, Bird, Bridge, & Viding, 2013). In line with this, Brewer et al. (2015) argue that such a fractionation of abilities is evidence that social cognition may depend not on a single or unified system but on distinct, albeit inter-dependant, cognitive processes.

This study was not without its limitations. The exploratory nature of the study, focused on the design and inclusion of a completely novel task, meant that many variables were included. To minimise the number of statistical comparisons, and hence likelihood of type 1 error, we tested a priori predictions for most variables, but

used 2-tailed probabilities to be conservative. A larger sample size would be desirable in future work; we may have lacked power to find smaller effects and some substantial correlations did not reach significance. Missing data is likely to have affected findings in such a small sample. The SSFt itself was limited for a number of reasons. Firstly, relatively low inter-item reliability suggests that the measure may not assess a single underlying construct (Devine & Hughes, 2013). However, the test was designed to have items with varying levels of difficulty (e.g. first and second order ToM), and this is likely to have added to the somewhat low rates of internal consistency. Minimal variance in the memory questions (in particular the control clips) resulted in an observed difference between the groups and this impacted their utility.. Finally, the theory of mind impairments demonstrated here on our novel task may not be specific to ASD; a wealth of literature exists evidencing individual differences in theory of mind as central to various clinical presentations (e.g. Schizophrenia; Chung et al., 2013; Pinkham et al., 2013; Sparks, McDonald, Lino, O'Donnelle, & Green, 2010). Further studies should include the use of alternative clinical samples to explore the use of the SSFt as a viable measure of social cognition across clinical presentations.

Further examination of participants' 'propensity vs. ability' (Vivanti, 2015) when answering the SSFt would also be of interest in future research. The current study was not designed to distinguish these two aspects of task performance. The use of more open-ended questions may go some way in delineating participants' internal drives to engage in the task and their social cognitive ability. Moreover, eye tracking studies, which have revealed differences in those with ASD in both implicit drives to engage in social stimuli (e.g. attending to actors faces vs. objects on screen (Klin et al. 2003) and in cases where explicit question scores are comparable to controls (Senju, Southgate, White, & Frith, 2009), could also shed light on the 'propensity vs. ability' distinction (Vivanti, 2015).

The development of the measure may also be conceptually limited by the 'methodology of consensus' (Johnston, Miles, & McKinlay, 2008). This criticism applies

to all social cognition measures using actors (see Table 1) and agreement between (neurotypical) raters to score responses, and so is not unique to the current research. However, it questions the objectivity of the measure and calls into question the pursuit of objectivity in this line of research (see Johnston, et al., (2008) for an insightful yet critical appraisal). Leading from this Milton (2012) argues that the ToM hypothesis of social cognition places the social deficit within the individual, which misrepresents the relational context within which social exchanges occur. He uses the term 'double empathy problem' to highlight that 'the social difficulty' is bi-directional in so much as it resides in both the ASD individual and those without the diagnosis. Such theoretical critiques raise interesting considerations, with regard to the nature of and direction of future research in the field of social cognition where the focus is not restricted to the observer's 'abilities' but expressivity of the agents (Zaki, Bolger, & Ochsner, 2008) and relationships between individuals. What appears to be relatively uncontentious is that novel ways of presenting interaction between agents, examining contextual effects, and the using of tools that reflect real life interactions are important in assessing social cognition (Dziobek, 2012); this piece of research is a small step in that direction.

The current study developed a novel, dynamic, video-based measure to assess social cognitive abilities. This study provides clinicians and researchers with a sensitive tool to assess attribution of mental states relevant to everyday communication and interaction.

Acknowledgements

The research team would like to acknowledge and thank the study participants and staff members within the Adult Autism Services who facilitated recruitment. The research team would also like to thank the Behavioural and Developmental Clinical Academic Group for approving this research.

D.S is funded by a National Institute for Health Research (NIHR) Clinical Doctoral Research Fellowship (CDRF - 2012 - 03 - 059). This research was independently funded as part of the DClinPsy studies of the first author. There are no conflicts of interest to declare.

References

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders*. Arlington, VA: American Psychiatric Publishing.

Axelrod, B. N., Ryan, J. J., & Ward, L. C. (2001). Evaluation of seven-subtest short forms of the Wechsler Adult Intelligence Scale-III in a referred sample. *Archives of Clinical Neuropsychology, 16*(1), 1-8.

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (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.

Barnes, J. L., Lombardo, M. V., Wheelwright, S., & Baron-Cohen, S. (2009). Moral Dilemmas Film Task: a study of spontaneous narratives by individuals with autism spectrum conditions. *Autism Research, 2*(3), 148-156.

Baron-Cohen, S. (1989). The Autistic Child's Theory of Mind: a Case of Specific Developmental Delay. *Journal of Child Psychology and Psychiatry, 30*(2), 285-297.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind" ? *Cognition, 21*(1), 37-46.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). 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.

Berthoz, S., & Hill, E. L. (2005). The validity of using self-reports to assess emotion regulation abilities in adults with autism spectrum disorder. *European Psychiatry, 20*(3), 291-298.

Bird, G., & Cook, R. (2013). Mixed emotions: the contribution of alexithymia to the emotional symptoms of autism. *Transl Psychiatry, 3*, e285.

Bowler, D. M. (1992). "Theory of Mind" in Asperger's Syndrome. *Journal of Child Psychology and Psychiatry, 33*(5), 877-893.

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.

Channon, S., Crawford, S., Orlowska, D., Parikh, N., & Thoma, P. (2013). Mentalising and social problem solving in adults with Asperger's syndrome. *Cognitive Neuropsychiatry, 19*(2), 149-163.

Chong, S. F., & Choo, R. (2011). Introduction to Bootstrap. *Proceedings of Singapore Healthcare, 20*(3), 236-240.

Chung, Y. S., Barch, D., & Strube, M. (2013). A Meta-Analysis of Mentalizing Impairments in Adults With Schizophrenia and Autism Spectrum Disorder. *Schizophrenia Bulletin*.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

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.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113-126.

Demurie, E., De Corel, M., & Roeyers, H. (2011). Empathic accuracy in adolescents with autism spectrum disorders and adolescents with attention-deficit/hyperactivity disorder. *Research in Autism Spectrum Disorders, 5*(1), 126-134.

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.

Dziobek, I. (2012). Comment: Towards a More Ecologically Valid Assessment of Empathy. *Emotion Review, 4*(1), 18-19.

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., Convit, A. (2006). Introducing MASC: A Movie for the Assessment of Social Cognition. *Journal of Autism and Developmental Disorders, 36*(5), 623-636.

Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. J., & Frith, C. D. (1995). Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition, 57*(2), 109-128.

Golan, O., Baron-Cohen, S., Hill, J. J., & Golan, Y. (2006). The "Reading the Mind in Films" Task: Complex emotion recognition in adults with and without autism spectrum conditions. *Social Neuroscience, 1*(2), 111-123.

Happé, F. (1994) An advanced test of theory fo 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*(2), 129-154.

Happé, F., & Frith, U. (2013). Annual Research Review: Towards a developmental neuroscience of atypical social cognition. *Journal of Child Psychology and Psychiatry*, n/a-n/a.

Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66*(3), 843-855. Retrieved from http://europepmc.org/abstract/MED/7789204

Heavey, L., Phillips, W., Baron-Cohen, S., & Rutter, M. (2000). The Awkward Moments Test: A Naturalistic Measure of Social Understanding in Autism. *Journal of Autism and Developmental Disorders, 30*(3), 225-236.

Hill, E. L. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences, 8*(1), 26-32.

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads [Press release]

Jameel, L., Vyas, K., Bellesi, G., Roberts, V., & Channon, S. (2014). Going 'Above and Beyond': Are Those High in Autistic Traits Less Pro-social? *Journal of Autism and Developmental Disorders*, 1-13.

Johnston, L., Miles, L., & McKinlay, A. (2008). A critical review of the eyes test as a measure of social-cognitive impairment. *Australian Journal of Psychology, 60*(3), 135-141.

Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: lessons from autism. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1430), 345-360

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry, 59*(9), 809-816.

Lockwood, P. L., Bird, G., Bridge, M., & Viding, E. (2013). Dissecting empathy: high levels of psychopathic and autistic traits are characterised by difficulties in different social information processing domains. *Frontiers in Human Neuroscience, 7*.

Lord, C., Rutter, M., & Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*(5), 659-685.

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., & Schopler, E. (1989). Austism diagnostic observation schedule: A standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders, 19*(2), 185-212.

Mathersul, D., McDonald, S., & Rushby, J. A. (2013). Understanding advanced theory of mind and empathy in high-functioning adults with autism spectrum disorder. *Journal of Clinical and Experimental Neuropsychology, 35*(6), 655-668.

McDonald, S., Flanagan, S., & Rollins, J. (2002). *The Awareness of Social Inference Test*. Sydney: Harcourt Assessment.

Milton, D. E. M. (2012). On the ontological status of autism: the 'double empathy problem'. *Disability & Society, 27*(6), 883-887.

Palmen, A., Didden, R., & Lang, R. (2012). A systematic review of behavioral intervention research on adaptive skill building in high-functioning young adults with autism spectrum disorder. *Research in Autism Spectrum Disorders, 6*(2), 602-617.

Pinkham, A. E., Penn, D. L., Green, M. F., Buck, B., Healey, K., & Harvey, P. D. (2013). The Social Cognition Psychometric Evaluation Study: Results of the Expert Survey and RAND Panel. *Schizophrenia Bulletin*, sbt081.

Ponnet, K., Buysse, A., Roeyers, H., & Clercq, A. (2008). Mind-Reading in Young Adults with ASD: Does Structure Matter? *Journal of Autism and Developmental Disorders, 38*(5), 905-918.

Ponnet, K., Buysse, A., Roeyers, H., & Corte, K. (2005). Empathic Accuracy in Adults with a Pervasive Developmental Disorder During an Unstructured Conversation with a Typically Developing Stranger. *Journal of Autism and Developmental Disorders, 35*(5), 585-600.

Ponnet, K. S., Roeyers, H., Buysse, A., De Clercq, A., & Van Der Heyden, E. (2004). Advanced Mind-Reading in Adults with Asperger Syndrome. *Autism, 8*(3), 249-266.

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.

Roeyers, H., & Demurie, E. (2010). How impaired is mind-reading in high-functioning adolescents and adults with autism? *European Journal of Developmental Psychology, 7*(1), 123-134.

Rogers, K., Dziobek, I., Hassenstab, J., Wolf, O., & Convit, A. (2007). Who Cares? Revisiting Empathy in Asperger Syndrome. *Journal of Autism and Developmental Disorders, 37*(4), 709-715.

Scheeren, A. M., de Rosnay, M., Koot, H. M., & Begeer, S. (2013). Rethinking theory of mind in high-functioning autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 54*(6), 628-635.

Sparks, A., McDonald, S., Lino, B., O'Donnelle, M., & Green, M. J. (2010). Social cognition, empathy and functional outcome in schizophrenia. *Schizophrenia Research, 122*(1-3), 172-178.

Wechsler, D. (1999). *Wechsler Abbreviated Scales of Intelligence.* San Antonio, TX: Harcourt Assessment.

Wechsler, D. (2008). WAIS-IV administration and scoring manual. . San Antonio, TX: Psychological Corporation.

White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the Strange Stories: Revealing Mentalizing Impairments in Autism. *Child Development, 80*(4), 1097-1117.

Yang, D. J., & Baillargeon, R. (2013). Brief Report: Difficulty in Understanding Social Acting (But Not False Beliefs) Mediates the Link Between Autistic Traits and Ingroup Relationships. *Journal of Autism and Developmental Disorders, 43*(9), 2199-2206.

Zaki, J., Bolger, N., & Ochsner, K. (2008). It Takes Two: The Interpersonal Nature of Empathic Accuracy, *Psychological Science, 19*(4), 399-404.

Appendix: White lie example clip:

Third person perspective of Max and Alice sitting in the living room across from each other and Alice holding a guitar about to play:

Focus on Alice from Max’s perspective: (looking nervous) ‘I’ve been working on this for ages and I think I have finally got it. I think my songs gonna end like this…. (strums badly played chord then sings out of tune) ooo ooo ooo yeah’ (looks expectant at camera)

Focus on Max from Alice’s perspective: (nods head encouragingly and half smiles)

‘Well done Alice… that sounds really good’

Author	Test	Stimuli	Question type	Participants	Relevant findings	Strengths	Limitations
Heavey et al., (2000)	AMT	UK advertisements (7) and TV series clip (1).	FC ER FC memory Open-ended interview regarding intentions of characters.	Adults: 16 ASD 15 Controls*	ASD <Controls, including some Memory questions. Intention questions yielded greater effects than FC ER questions. Only controls performance on AMT related to the SS and IQ. No group response latency difference	Open ended questions. Convergent validity	ASD group struggled with memory questions. Complex coding system for intentionality. 45-120 second long clips. Overacted/dramatic stimuli. No control clips.
Golan et al., (2006)	RMFT	22 short film clips from feature films.	FC ER	Adults: 22 ASD 22 Controls*	ASD < Controls Performance on RMFT related to VIQ, AQ and CMFVB	Replicated with child version Complex emotions. Convergent Validity	No control clips/questions. Consensus decided emotions.
Dziobek et al., (2006)	MASC	15min video of 4 characters preparing for a party. Film stopped for each question(46 times).	Open ended concerning characters' thoughts, feelings and intentions. Memory.	Adults: 19 ASD 20 Controls*	MASC group difference > Eyes, SS and ER task. ASD=Controls on Memory Questions. No association with MASC and VIQ MASC associated with SS and ADI-R No association between Eyes, ER or SS tasks.	Open questions Tailored stimuli Range of linguistic concepts Convergent validity Re-test reliability Replicated with FC version.	45min administration time. Non-English speaking. Trained rater required for scoring Basic control questions.

Table 1: Characteristics of current dynamic social cognition task

Table 1: *Age, gender and IQ matched AMT: Awkward Moments Test, RMFT: Reading the Mind in the Films Task, MASC: A Movie for the Assessment of Social Cognition, FC: Forced choice, ER: Emotion Recognition.

Table 1 cont: Characteristics of current dynamic social cognition tasks.

Author	Test	Stimuli	Scoring/ Question type	Participants	Relevant findings	Strengths	Limitations
Barnes et al., (2009)	MDFT	4 film clips from American TV show 'House'.	Use of mental state words in narrative description of task, length of description, type of mental states used.	Adults: 28 ASD 28 Controls*	Lower frequency of mental state references in ASD narratives and shorter overall. VIQ correlated with performance only for ASD. Empathy scores correlated with only controls' performance on MDFT.	Open questions. Convergent validity.	No intention questions. Dramatised stimuli.
Mathersul et al., (2013)	TASIT: part 2 and 3.	31 self-contained clips of ambiguous social interchanges.	FC regarding thoughts, feelings (ER) and intentions of characters.	Adults: 40 ASD 37 Controls*	ASD < Controls, but not on ER questions. VIQ did not correlate with performance on TASIT. Only self-reported cognitive empathy predicted by TASIT independent of group.	Large sample. Convergent validity Bespoke clips	No control clips or questions. Lengthy administration (60-75mins).

Table 1. *Age, gender and IQ matched, MDFT: Moral Dilemmas Film Task , TASIT: The Awareness of Social Inference Test, FC: Forced choice, ER: Emotion Recognition

Table 2 Participant characteristics: Mean (SD)

	ASD N=20	Control N=20	t_{df}	p-value	*d*	95% mean Difference CI.
Age in years	30.60 (6.52)	30.65 (6.27)	$.025_{38}$	.980	0.01	-3.82 - 4.00[a]
Gender (F:M)	0:20	1:19		1.00[c]	-	-
VIQ	105.05 (17.01)	111.25 (11.52)	$1.35_{33.4}$	.186	0.42	-2.89 - 14.90[a]
AQ	34.21 (7.42)[b]	15.55 (7.21)	-7.96_{37}	.000	-2.55	-23.41 - -13.91

Table 2 : [a]bootstrap derived confidence intervals [b]ASD N=19 [c] Fisher's Exact test AQ= Autism Quotient

Table 3: Results by group for traditional tests of social cognition: Mean (SD)

	ASD N=20	Controls N=20	t_{df}	*p-value*	*d*	*95% CI*
Strange Stories						
Accuracy *(max=16)*	11.60 (2.39)	13.35 (1.73)	2.65_{38}	*.012*	*.84*	*.51 – 3.01*[a]
Mental State Language *(max=16)*	11.85 (1.87)	12.50 (1.67)	1.16_{38}	*.254*	*.37*	*-.42 - 1.74*[a]
RMET						
Accuracy (max = 36)	25.00 (4.08)	27.68 (4.27)[b]	2.01_{37}	*.052*	*.64*	*.06 - 5.29* [a]
Triangles						
Accuracy *(max=8)*	3.70 (2.20)	5.20 (2.21)	2.15_{38}	*.038*	*.68*	*.09 - 2.91*
Mental State Language *(max=8)*	1.90 (1.41)	3.55 (1.90)	3.11_{38}	*.004*	*.98*	*.58 – 2.65* [a]
TASIT Part 1						
Accuracy (max = 28)	23.75 (2.36)	24.75 (1.92)	1.47_{38}	*.149*	*.47*	*-.27 - 2.33*[a]

Table 3. [a]bootstrap derived confidence intervals. [b] N=19.

Table 4: Results by group for IRI and TAS-20: Mean (SD)

	ASD N=19	Controls N=20	t_{df}	p-value	d	95% CI
IRI subscales *[max=28]*						
Perspective Taking	13.21 (6.07)	17.75 (4.79)	2.60_{37}	.013	0.83	1.00 - 8.08
Fantasising	13.05 (5.58)	15.90 (4.72)	1.72_{37}	.093	0.55	-.50 -6.20
Empathic Concern	17.42 (4.10)	17.75 (2.59)	0.30_{37}	.765	0.10	-1.80 - 2.51[a]
Personal Distress	14.42 (5.71)	10.25 (4.28)	-2.59_{37}	.014	-0.83	-7.43 - -.91
TAS-20:						
Identify Emotions *[max=35]*	20.58 (5.98)	15.60 (6.02)	-2.59_{37}	.014	-0.83	-8.88 - -1.08
Describe Emotions [max=25]	17.95 (3.46)	12.95 (5.27)	-3.52_{33}	.001	-1.13	-7.71 - -2.18[a]
Externally Orientated Thinking [max=40]	23.05 (4.48)	18.05 (4.44)	-3.50_{37}	.001	-1.12	-7.90 - -2.10
Total *[max=100]*	61.58 (10.07)	46.60 (11.10)	-4.41_{37}	.000	-1.41	-21.87 - -8.09

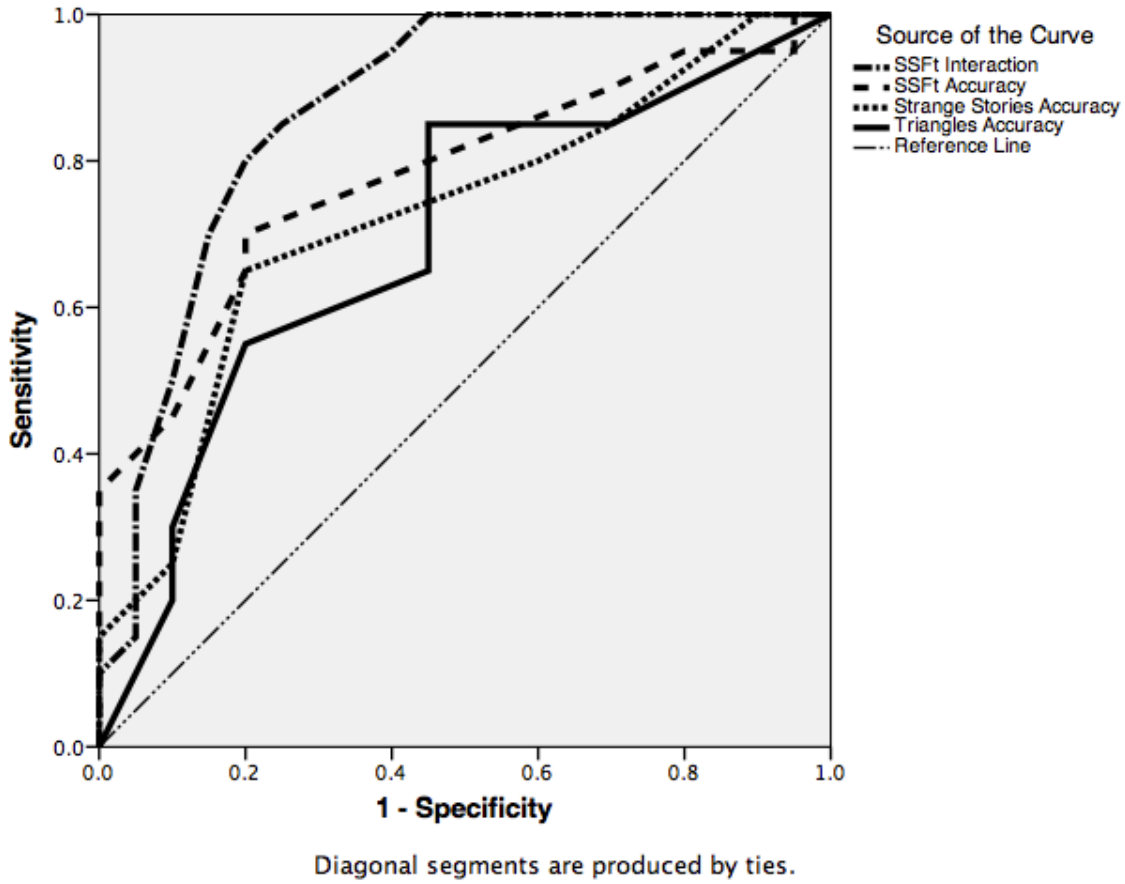
Table 4: [a]bootstrap derived confidence intervals

Table 5: Results by group for The Strange Stories Film task : *M (*SD)

The Strange Stories Film task	ASD N=20	Controls N=20	t_{df}	F_{df}	p-value	d	95% CI
Experimental Clips							
Accuracy *(max=24)*	15.50 (3.55)	18.80 (2.33)		9.55_{37}	.004	1.00	.96 – 4.62
Mental State Language *(max=24)*	12.15 (2.56)	13.75 (2.45)		2.43_{37}	.128	.56	-.35 – 2.70
Interaction *(max=24)*	10.95 (3.46)	16.95 (4.14)		22.06_{37}	.000	1.52	3.01 - 7.59
Memory *(max=12)*	11.55 (0.69)	11.85 (0.37)		2.32_{37}	.137	.49	-.08 - .66 [a]
Control Clips							
Accuracy (max=6)	4.05 (1.39)	4.40 (0.99)		$.16_{37}$	.690	.13	-.52. - .83[a]
Mental State Language (max=6)	0.90 (0.85)	1.25 (1.16)		$.68_{36}$	.415	.27	-.46 - .89[a]
Interaction (max=6)	5.50 (0.89)	5.70 (0.65)	0.81_{38}		.423	.26	-.29 – .70[a]
Memory (*max = 3)*	2.95 (0.22)	3.00 (0.00)	1.45_{19}		.163	.46	.04 – .26[a]

Table 5. [a]bootstrap derived confidence intervals.

Figure 1


Source of the Curve
SSFt Interaction
SSFt Accuracy
Strange Stories Accuracy
Triangles Accuracy
Reference Line
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0
Diagonal segments are produced by ties.