This is the peer reviewed version of the following article: “Sheppard, E., Pillai, D., Wong, G. T.L., Ropar, D., & Mitchell, P. (2016). How easy is it to read the minds of people with Autism Spectrum Disorder? *Journal of Autism and Developmental Disorders, 46,* 1247-1254.” The final publication is available at Springer via https://link.springer.com/article/10.1007/s10803-015-2662-8

How easy is it to read the minds of people with Autism Spectrum Disorder?

*Elizabeth Sheppard[a], Dhanya Pillai[b], Genevieve T-L Wong[b], Danielle Ropar[a], Peter Mitchell[a]

[c]University of Nottingham,

University Park,

Nottingham, NG7 2RD

United Kingdom

[b]University of Nottingham Malaysia Campus,

Jalan Broga,

43500 Semenyih

Selangor Darul Ehsan,

Malaysia

*Correspondence author, Email: elizabeth.sheppard@nottingham.ac.uk; Phone: +44 115 8467436

How well can neurotypical adults' interpret mental states in people with ASD? 'Targets' (ASD and neurotypical) reactions to four events were video-recorded then shown to neurotypical participants whose task was to identify which event the target had experienced. In study 1 participants were more successful for neurotypical than ASD targets. In study 2, participants rated ASD targets equally expressive as neurotypical targets for three of the events, while in study 3 participants gave different verbal descriptions of the reactions of ASD and neurotypical targets. It thus seems people with ASD react differently but not less expressively to events. Because neurotypicals are ineffective in interpreting the behaviour of those with ASD, this could contribute to the social difficulties in ASD.

Keywords: autism spectrum disorders, emotion, facial expressions, mental states, social interaction

How easy is it to read the minds of people with Autism Spectrum Disorder?

People with autism spectrum disorder (ASD) have difficulty interpreting others' mental states (e.g. Baron-Cohen, Leslie & Frith, 1985), and this topic has been extensively researched (Rajendran & Mitchell, 2007, for a review). The purpose of the current study is to illuminate another matter: Little is known about how well people interpret mental states in those who have ASD, which could depend on the way that behaviour offers an interpretable signal of those states. This is a potentially important question, because if people have difficulty reading the mental states of those with ASD it could lead to interpersonal misinterpretations that impact negatively on social interaction (Begeer, Koot, Rieffe, Terwogt & Stegge, 2008).

Mental states are revealed through various aspects of behaviour, particularly facial expressions (Baron-Cohen, Jolliffe, Mortimore & Robertson, 1997). According to DSM-5 (APA, 2013), ASD is characterised by a lack of facial expression, though recent studies have yielded somewhat mixed support. For example, Stagg, Salvny, Hand, Cardoso and Smith (2013) video recorded children with and without ASD while describing events from their lives, and presented the muted videos to neurotypical adults. The adults rated the children with ASD as less expressive than those without. Similarly, Macdonald et al. (1989) asked individuals with and without ASD to pose facial expressions, and found that adults were less able to correctly identify the emotion expressed by those with ASD. However, Grossman, Edelson and Tager-Flusberg (2013) used videos of individuals with and without ASD retelling stories, and the two groups' facial expressions were identified equally well and rated as equally expressive by adult participants. Beadle-Brown & Whiten (2004) found that individuals with ASD were equally able to pose a variety of facial expressions as neurotypical individuals, as did Volker, Lopata, Smith and Thomeer (2009) apart from

sadness, which was identified less accurately when posed by children with ASD. Studies of automatic mimicry also suggest that individuals with ASD can be just as expressive as people without ASD. For instance, Press, Richardson and Bird (2010) presented animated facial expressions to people with and without ASD and found that both groups automatically mimicked the expressions to the same extent.

One possible reason for these differences might be the methods used to elicit the expressions, with natural expressions showing more marked group differences than posed. However, a recent study suggests that the natural reactions of people with ASD can be even more expressive than comparison participants in some circumstances. Faso, Sasson & Pinkham (2015) carried out a study in which emotions were elicited by recording and playing back narratives of life events of the participants when they had felt a particular emotion. The expressions of those with ASD were subsequently rated as more intense and less natural than those of neurotyical individuals, and identified with greater accuracy. Faso et al. suggest that people with ASD might be *different* from people without ASD in their facial expressions and that the degree of expressiveness is perhaps not the key factor.

Drawing upon a procedure developed by Pillai, Sheppard and Mitchell (2012; Pillai et al., 2014; Cassidy, Ropar, Mitchell & Chapman, 2013, 2015, Wu, Sheppard & Mitchell, in press), it is possible to investigate whether people find it difficult to interpret the mental states of those with ASD. In the previous research participants (with and without ASD) were asked to guess what had happened to another, neurotypical person (the target) from that person's reaction (including their facial expression). Those with ASD were not as effective at interpreting the reactions of targets in guessing what had happened, presumably because they had difficulty inferring the mental states of the targets (Pillai et al., 2014). In the research reported here we used the same procedure except that targets were people with and without ASD, with the prediction that participants (neurotypicals) would find it difficult to interpret

the reactions specifically of targets with ASD, raising the possibility that they have difficulty interpreting the mental states of those with ASD.

Stimulus Development

The procedure replicated that devised by Pillai et al. (2012; 2014). Participants' ('targets') natural reactions to an aspect of the researcher's behaviour were surreptitiously filmed when they arrived at the testing room.

Participants (targets)

Forty males aged between 13 and 21 (M = 15.4 years) were enlisted from educational establishments and were all native speakers of English. Those with ASD (n=20) were evaluated by mental health professionals according to DSM-IV criteria (American Psychiatric Association, 2000). All targets (both ASD and neurotypical) were undertaking basic or advanced level secondary or tertiary education courses. The ASD and neurotypical targets were all Caucasian and matched for chronological age.

Procedure

A Sony DCR-SR60 video camera was placed about 1.7 meters directly opposite the target, across the table on a tripod and was switched on prior to their entering the room. Targets were told that they would be filmed while posing some facial expressions to be used as video stimuli for a later study. Following Pillai et al (2012, 2014), when each target arrived, the researcher performed one of four possible scenarios (Joke, waiting, Story & Compliments). As the real aim was to record targets' reactions to the scenario, the camera was already switched on before the target entered the room. In the Joke scenario, the researcher shared a joke with the target while in the Waiting scenario the researcher kept the target waiting while performing irrelevant activities such as texting. In the Story scenario the researcher described several mishaps which had occurred earlier that day while in the

Compliments scenario the researcher paid the target a series of three compliments. The four different scenarios were counterbalanced between targets in each group, such that five individuals in each group experienced each scenario (i.e. five individuals with ASD experienced the Joke scenario, five experienced the Waiting scenario etc.) Prior to leaving the testing room, the targets were debriefed about the aims of the study and their consent to use the video recordings of their reactions was obtained.

Editing

The video recordings were edited using Windows Live Movie Maker software according to the procedure developed by Pillai et al. (2012; 2014). A set of editing criteria was developed in order to ensure objective selection of video excerpts across the target groups. For each of the scenarios the aim was to generate video clips capturing the timeframe when the targets were likely to be most responsive. The recordings for the Joke scenario were edited from the point where the researcher delivered the punchline. The Compliments footage was edited from the end of the last spoken compliment. The Story and Waiting scenarios were edited from three seconds prior to the end of the researcher's vocalisations during the enactment. This resulted in a total of 40 edited videoclips (20 for each group with a total of 5 clips per scenario for each group) with a mean of 7.22 seconds. The videoclips were 1080 pixels in width and 720 pixels in height, presented at 25 frames per second. As in Pillai et al. (2012; 2014) targets' verbalisations (i.e. the auditory component of footage) was muted.

Study 1

Methods

Participants

Thirty participants (15 males and 15 females) aged between 18 and 35 (mean age= 23.2 years) were recruited. All were native English speakers.

Procedure

Participants were tested individually and to begin were shown four videos of the researcher enacting the scenarios, the order of which was randomised by PsychoPy2. Each video was shown only once and included audio as well as visual information. The purpose of presenting these videos was to give the participants an idea of what the targets experienced in each scenario. Participants subsequently watched the 40 target videos (one video for each target), again randomised by PsychoPy2. After each target video was presented, the scenario names appeared on the screen and participants were instructed to state which scenario the individual in the video was reacting to by pointing to the relevant scenario name on the screen (i.e. Joke, Waiting, Story Compliments). No indication was given to the participants that some of the individuals in the videos were diagnosed with ASD. Participants were then debriefed and asked if they had guessed that some of the targets had ASD and all said that they had not. The procedure for the study was subjected to ethical review and approval was obtained from the School of Psychology Ethics committee, University of Nottingham.

Results & Discussion

Wagner's unbiased hit rate was used which codes accurate responses weighted for biases to select any particular scenario (1993). Wagner's unbiased hit rate was calculated for each participant for each scenario/target group combination (i.e. ASD Joke, ASD Waiting, ASD Story etc.) This involves computing the square of the participant hit rate (number of times the participant correctly identified the particular scenario/target group combination out of a possible 5) divided by the product of the number of items within the category (number of targets belonging to a particular group – ASD or neurotypical -- in each scenario, in this case always 5) and the total frequency with which the category is chosen (number of times the participant responded with a particular scenario regardless of accuracy). This yields a value

between 0 and 1, where 0 indicates that the participant could not systematically identify responses to a particular scenario.

[Figure 1 near here]

Figure 1 shows mean values of Wagner's unbiased hit rate for ASD and neurotypical target videos for each scenario. Data were submitted to analysis of variance (ANOVA) and all comparisons were repeated with non-parametric equivalents (Wilcoxon signed ranks and Friedman tests), which gave rise to an identical pattern of results. Eight one-sample *t*-tests with a Bonferroni adjusted alpha level of .0063 were conducted: Mean Wagner's unbiased hit rates were significantly greater than 0 in all cases (i.e. $p < .0063$). A 2x4 (target group x scenario) repeated measures ANOVA was performed in order to examine any differences in participants' levels of success in identifying the scenarios targets were responding to. A main effect of target group was found, $F(1,29) = 37.79$, $p < .001$, $\eta_p^2 = .57$, in that participants were better at identifying the scenario when viewing the neurotypical targets than when viewing targets with ASD. A main effect of scenario was also found, $F(3,87) = 115.26$, $p < .0005$, $\eta_p^2 = .80$. Posthoc tests with Bonferroni correction revealed that participants were better in identifying targets responding to the Waiting scenario than the other three scenarios (all *ps* $< .0005$). Also, responses to the Joke scenario were easier to identify than responses to the Compliments scenarios ($p < .0005$).

A significant interaction effect between scenario and target group was found, $F(3,87) = 14.57$, $p < .0005$, $\eta_p^2 = .33$, raising the possibility that performance was better for one target group over the other, depending on the scenario. To explore this possibility, paired-samples *t*-tests were employed to compare participant performance across the two target groups for each scenario independently. Participants were better at discriminating the videos showing neurotypical compared with ASD targets reacting to the Waiting, $t(29) = 4.50$, $p < .001$, $d = .87$, Story, $t(29) = 7.43$, $p < .001$, $d = 1.56$ and Compliments scenarios, $t(29) = 5.88$, p

< .001, d = 1.14. However, there was no evidence to suggest that participants could identify responses to the Joke scenario any better depending on whether the target had ASD or was neurotypical, $t(29) = 1.50$, $p = .145$, $d = .28$.

In summary, participants' judgments effectively (and unwittingly) discriminated between those who did and did not have ASD in three of the scenarios targets faced. What is the basis of these judgments? One possibility is that in so far as targets with ASD were inexpressive (e.g. APA, 2013), the strength of the 'signal' in the facial expression of these targets was weaker and harder to interpret. We tested this prediction in Study 2 by asking a different group of participants to rate the expressiveness of the targets.

Study 2

Methods

Participants

Twenty participants (10 males and 10 females) aged between 19 and 22 (*Mean age* = 20.3 years) took part; none had participated in Study 1. All were native English speakers.

Materials and Apparatus

The stimuli in this study were the same 40 target responses used in Study 1. All target videos were shown on a 20 inch computer monitor. The videos were presented using PsychoPy2 version 1.80 which randomised their order.

Procedure

Each participant watched the 40 videos of target's reactions and judged expressiveness on a 7-point scale, whereby 1 indicated low expressiveness and 7 indicated high expressiveness. Participants made their judgment by clicking the appropriate point on a Likert scale presented in the centre of the screen following each target video. After participants made their judgment the next video started. Participants were not informed about

the different scenarios to which the targets were reacting and neither were they told that some targets had been diagnosed with ASD.

Results & Discussion

Figure 2 shows the mean expressiveness ratings for videos of ASD and neurotypical targets in the four scenarios. Data were submitted to analysis of variance and the analyses were repeated using non-parametric equivalents that gave rise to the same effects, apart from some minor differences in relation to the effect of scenario which are described below. A 2x4 (target group x scenario) repeated measures ANOVA was performed in order to determine whether participants rated targets with ASD as less expressive than neurotypicals. A main effect of target group showed that participants rated neurotypical targets as more expressive than ASD targets, $F(1,19) = 15.94$, $p = .001$, $\eta_p^2 = .46$. There was also an interaction between target group and scenario, $F(3,57) = 12.21$, $p < .001$, $\eta_p^2 = .39$, raising the possibility that the group difference in rated expressiveness did not maintain consistently across the four scenarios. Four post-hoc paired samples *t*-tests (with a Bonferroni corrected alpha of .0125) were conducted to compare the expressiveness ratings for the two target groups in each scenario. Participants rated targets with ASD as less expressive than neurotypical targets for the Compliments scenario only, $t(19) = -4.89$, $p < .001$, $d = 1.46$.

[Figure 2 near here]

The ANOVA also yielded a main effect of scenario, $F(3,57) = 24.10$, $p < .001$, $\eta_p^2 = .56$ (Greenhouse-Geisser corrected). Post-hoc Bonferroni comparisons indicated that participants judged targets experiencing the Waiting scenario as less expressive than those experiencing the Joke, Compliments (ps < .001) and Story scenarios ($p < .005$). Participants also judged targets experiencing the Story scenario to be less expressive than those experiencing the Joke ($p < .001$). The non-parametric equivalent tests (Wilcoxon) found significant differences between all pairs of scenarios. In other words in addition to the above

effects, participants judged targets experiencing the Story scenario as less expressive than those experiencing the Compliments scenario, and those experiencing the Compliments scenario as less expressive than those experiencing the Joke (all $p < .05$).

In summary, targets with ASD were adjudged to be as expressive as neurotypicals, except in the case of the compliment scenario. The finding adds to literature which suggests that high functioning people with ASD are often at least as expressive as those who do not have ASD (e.g. Faso et al., 2015). It seems, therefore, that we cannot explain participants' difficulty in Study 1 in guessing the waiting and story scenarios by saying that targets with ASD were less expressive; rather, they were expressive to the same degree as targets without ASD. Despite targets with ASD being expressive in most cases, it seems that the form of their expressions were not easily interpretable to participants and perhaps were atypical. To explore this possibility, participants in Study 3 were asked to describe the responses of the targets, with the prediction that the two target groups would elicit different kinds of description.

Study 3

Methods

Participants

Twenty participants (10 males and 10 females) aged between 18 and 32 ($M = 21.8$ years) took part. All spoke English as their first language.

Materials and Apparatus

The stimuli were the same 40 target responses used in Study 1. All target videos were shown on a 20 inch computer monitor. The videos were presented using PsychoPy2 version 1.80 which randomised their order. Participants' verbal responses were captured using a Blue

Yeti USB microphone connected to a separate laptop, on which Audacity software was running to record the verbalisations.

Procedure

After watching each of the 40 target videos, participants were prompted by a written message on the screen to describe the response of the target. In order to avoid steering participants towards a particular type of description, they were simply asked to 'describe the response' . Participants were given unlimited time to provide their verbal response after which they were required to press the spacebar to move to the next video. Participants were not informed about the different scenarios to which the targets were reacting and neither were they told that some of the targets had ASD.

Results & Discussion

Responses were transcribed verbatim from the audio files for theme identification prior to analysis. The contents of participants' verbal descriptions fell broadly into three main categories. The first of these were responses which referred to internal states, including mental states, emotions, knowledge states etc. Examples include: 'Not in a good mood, a bit grumpy', 'I think he doesn't know what's happening', 'Nervous and a little sense of guilt'. The second category included responses which referred to overt behaviour, such as specific facial expressions, body movements, eye movements. For example, 'Okay so, he just raised an eyebrow and blinked... that's pretty much it...', 'Why is he touching his face like he has a beard over there? He's touching his face twice.' 'He seems to be laughing.' The third category included responses which involved guessing the scenario or event experienced by the person in the video, whether correct or incorrect. Examples include 'I think someone complimented this person,' 'Probably listening to a sermon' and 'Someone cracked a joke'.

The number of verbal descriptions that fell into these three categories was coded by an independent rater who was unaware of the aims of the study. Verbalisations that fell into

more than one category were coded into both categories. Some verbalisations involved references to multiple examples from the same category (e.g. a verbalisation that referred to several mental states) and these were also counted and contribute to the means reported. A second rater coded a subset of 160 responses and this gave rise to Kendall's tau_b correlations of 0.78 for internal states, 0.79 for overt behaviours and 0.83 for events. Given the acceptable reliability, ratings from the first rater were used in the analysis. For each participant, the mean number of times each category was referred to per video was calculated for the two target video groups. Figure 3 shows the mean number of descriptions per video per category for the two target groups.

[Figure 3 near here]

The data were analysed using analysis of variance, but repeated using non-parametric equivalents which gave rise to the same effects. A 2x3 (target group x category) repeated measures ANOVA showed a main effect of category, $F(2,38) = 36.01$, $p < .001$, $\eta_p^2 = .66$. Post hoc Bonferroni pairwise comparisons revealed a larger number of verbal descriptions relating to internal states than either overt behaviours or references to events (both ps $< .001$) while there was no difference in the mean number of references to overt behaviours and events.

There was no main effect associated with target group but there was a significant interaction between category and target group, $F(2,38) = 8.32$, $p = .001$, $\eta_p^2 = .31$, raising the possibility that differences between target groups did exist but not consistently across all categories of description. Post-hoc paired samples *t*-tests were conducted (with a Bonferroni corrected alpha level of .016) between the two target groups for each category of verbal description independently. Participants produced more verbal descriptions referring to internal states for targets with ASD than for neurotypical targets, $t(19) = 3.04$, $p = .007$, d

= .36, but there was no difference between target groups for the other two categories of description.

In summary, while the kinds of descriptions that participants offered to describe behaviour were largely similar for the two target groups, one difference stood out: Ironically, participants were more likely to refer to internal states when describing targets with ASD than when describing targets without ASD. Perhaps it was not straightforward to interpret how targets with ASD were reacting, causing participants to conjecture over their possible internal states to a greater extent than for the targets who did not have ASD.

General Discussion

The three studies presented in this paper investigated whether neurotypical adults could infer what scenario targets with and without ASD had experienced based on muted video recordings of their behavioural reactions. As predicted, participants found it more difficult to interpret the behaviour (and by implication the mental state) of targets with ASD at least in some of the scenarios. This cannot be explained by saying that people with ASD are inexpressive. Although Study 2 showed that people with ASD were less expressive than neurotypical individuals in response to compliments from the experimenter, the two groups of targets were adjudged to be equally expressive for three other scenarios; and in two of those, participants in Study 1 nevertheless had difficulty interpreting the behaviour specifically of targets with ASD. Study 3 added to this picture by showing that participants were inclined to describe the behaviour of targets with ASD differently from those without ASD.

Generally, the results suggest that people have difficulty interpreting the mental states of those with ASD. The results are consistent with those of Faso et al. (2015) in suggesting that the expressions of those with ASD may be different from those of neurotypical

individuals. However, in Faso et al.'s study this resulted in their expressions being identified more accurately whereas in the current research the reactions of those with ASD were interpreted less accurately. One possible reason for these differences was that in Faso et al.'s (2015) study the method aimed at equating the emotions (internal states) felt by the two target groups. The emotions were elicited by each target hearing their own description of an event in their life, which induced that emotion (so these events could have differed widely between the targets). In our study, the reactions were always elicited by the same stimuli (the four scenarios) and it is not necessarily the case that all targets experienced the same inner states in response to these scenarios. Therefore it is at least possible that the four scenarios induced different kinds of mental states in people with and without ASD, resulting in differing reactions. Future research could ask targets about the mental states they experienced in the scenarios to get some indication of whether the target groups do differ in how they interpret the situations – although there are notable challenges associated with asking targets with ASD to reflect on and describe their own mental states.

The results from Study 2 add to the mixed picture presented by previous studies regarding whether individuals with ASD are less expressive than those without ASD. While some previous studies have reported those with ASD are viewed as less expressive than neurotypical comparison individuals (McDonald et al., 1989; Stagg et al., 2013), others have suggested participants with ASD are equally or even more expressive (Faso et al., 2015; Press et al., 2010). Given that in our study the reactions of those with ASD were rated as less expressive in one condition (the Compliments scenario) but not the others, this seems to support the conclusion that whether or not they appear expressive may depend on the particular context. Previous research using posed expressions supports the notion that those with ASD produce facially expressed emotions similarly to those without ASD (e.g. Beadle-Brown & Whiten, 2004; Volker et al., 2009). Therefore it may be that the particular

circumstances determine the extent to which they are expressive and indeed whether they are as expressive as people without ASD. At present, it is not clear why those with ASD were less expressive particularly in response to being paid compliments, and future research could usefully determine what kinds of circumstances are likely to elicit an inexpressive response from people with ASD and why.

Study 3 demonstrated that, when asked to describe the reaction of the target in each video, neurotypical adults most frequently refer to internal states – more so than to actual observable behaviour. These findings imply that at least for neurotypical adults, there is a strong tendency to explain observed behaviour with reference to possible associated mental states, even in the absence of a specific demand to do so (participants were not asked to identify or guess the mental state of the targets). Moreover, participants made even more references to internal states of the targets when describing the reactions of those with ASD than those without. While initially this result seemed surprising, it is perhaps consistent with research which has suggested that people may engage most readily in effortful mentalising under situations of high ambiguity (Jenkins & Mitchell, 2009). In more straightforward and unambiguous situations, observable actions may be understood through application of behavioural rules. If the reactions of those with ASD are more unusual or ambiguous, this may have led to increased efforts to guess their associated mental states.

If it were common to misinterpret the mental states of those with ASD, this could lead to confusing social interaction and consequent negative experiences. In turn, this could cause people with ASD to want to avoid social interaction and it could cause those without ASD to want to avoid interacting with people who have ASD. People with ASD might thus face a double jeopardy: They start out with a disadvantage in understanding other minds (e.g. Baron-Cohen et al., 1985) and they subsequently face a barrier to social interaction due to their mental states being misinterpreted by others, as suggested by the current findings, thus

excluding them from the kinds of experience that would permit learning about others' minds (Mitchell, 2015). Perhaps programmes of intervention should focus not only on helping people with ASD to understand others' minds but should also focus on helping those who frequently encounter people with ASD to improve their ability in interpreting the behaviour of those with ASD.

Compliance with ethical standards

The authors declare that they have no conflict of interest. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study. For the video targets, additional informed consent was obtained after debriefing for use of the videos for the research.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th Edition). Washington DC: American Psychiatric Association.

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

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or asperger syndrome. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 38*(7), 813-822.

Beadle-Brown, J.D. & Whiten A. (2004). Elicited imitation in children and adults with autism: Is there a deficit? *Journal of Intellectual and Developmental Disability*, *29*, 147-163.

Begeer, S. Koot, H.M., Rieffe, C., Terwogt, M. & Stegge, H. (2008). Emotional competence in children with autism: Diagnostic criteria and empirical evidence. *Developmental Review*, *28*, 342-369.

Cassidy, S., Ropar, D., Mitchell, P., & Chapman, P. (2015). Processing of spontaneous emotional responses in adolescents and adults with Autism Spectrum Disorders: effect of stimulus type. *Autism Research*, *8,* 534-544.

Cassidy, S., Ropar, D., Mitchell, P. & Chapman, P. (2013). Can adults with autism spectrum disorders infer what happened to someone from their emotional response? *Autism Research, 7*, 112-123.

Faso, D.J., Sasson, N.J. & Pinkham, A.E. (2015). Evaluating posed and evoked facial expressions of emotion from adults with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders, 45,* 75-89.

Grossman, R. B., Edelson, L. R., & Tager-Flusberg, H. (2013). Emotional facial and vocal expressions during story retelling by children and adolescents with high-functioning autism. *Journal of Speech, Language and Hearing Research, 56,* 1035-1044.

Jenkins, A. C., & Mitchell, J. P. (2009). Mentalizing under uncertainty: Dissociated neural responses to ambiguous and unambiguous mental state inferences. *Cerebral Cortex, 20,* 404-410.

Macdonald, H., Rutter, M., Howlin, P., Rios, P., Le Couter, A., Evered, C., & Folstein, S. (1989). Recognition and expression of emotional cues by autistic and normal adults. *Journal of Child Psychology and Psychiatry,30(6),* 865-877.

Mitchell, P. (2015). Mindreading as a transactional process: Insights from autism. In V. Slaughter & M. de Rosnay (eds), *Environmental Influences on ToM Development*. Hove: Psychology Press.

Pillai, D., Sheppard, E., & Mitchell, P. (2012). Can People Guess What Happened to Others from Their Reactions? *PLoS ONE 7(11)*: e49859. doi:10.1371/journal.pone.0049859

Pillai, D., Sheppard, E., Ropar, D., Marsh, L., Pearson, A., & Mitchell, P. (2014). Using other minds as a window onto the world: Guessing what happened from clues in behavior. *Journal of Autism and Developmental Disorders*, *44,* 2430-2439.

Press, C., Richardson, D., & Bird, G. (2010). Intact imitation of facial actions in autism spectrum conditions. *Neuropsychologia, 48,* 3291-3297.

Rajendran, G., & Mitchell, P. (2007). Cognitive Theories of Autism. *Developmental Review, 27(2),* 224-260.

Volker, M. A., Lopata, C., Smith, D. A., & Thomeer, M. L. (2009). Facial encoding of children with high-functioning autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities, 24(4),* 195-204.

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behaviour. *Journal of Nonverbal Behavior, 17(1),* 3-28.

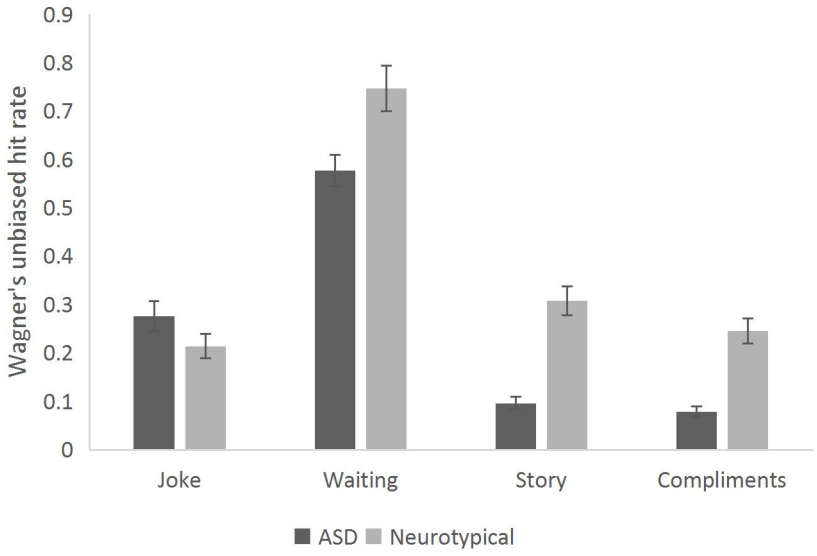
Wu, W., Sheppard. E. & Mitchell, P. (2016). Being Sherlock Holmes: Can we sense empathy from a brief sample of behavior? *British Journal of Psychology* (in press).

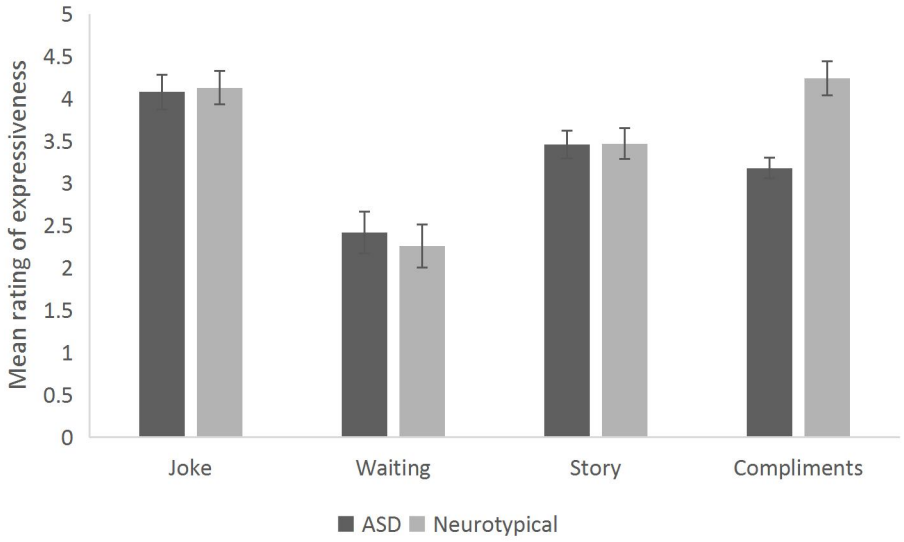
Figure Captions

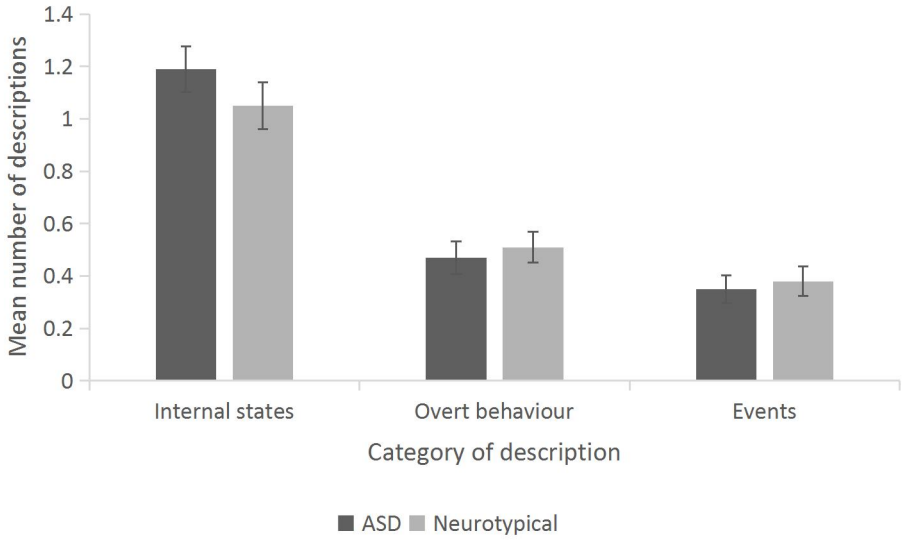
Figure 1. Wagner's unbiased hit rate for ASD and neurotypical targets for Joke, Waiting, Story and Compliments scenarios (error bars show the standard error)

Figure 2. Mean expressiveness ratings (out of 7) for ASD and neurotypical targets in each scenario (error bars show the standard error)

Figure 3. Mean number of verbal descriptions per video involving internal states, overt behaviour and events for videos of ASD and neurotypical targets (error bars show the standard error)


Wagner's unbiased hit rate
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Joke
Waiting
Story
Compliments
ASD
Neurotypical


Mean rating of expressiveness
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Joke
Waiting
Story
Compliments
ASD
Neurotypical


1.4
1.2
1
0.8
0.6
0.4
0.2
0
Mean number of descriptions
Internal states
Overt behaviour
Events
Category of description
ASD
Neurotypical

Author Note

Elizabeth Sheppard, University of Nottingham Malaysia Campus. Dhanya Pillai, University of Nottingham Malaysia Campus. Genevieve Tze-Lyn Wong, University of Nottingham Malaysia Campus. Danielle Ropar, University of Nottingham. Peter Mitchell, University of Nottingham.

Elizabeth Sheppard is now at University of Nottingham

We are grateful to all the participants who contributed to this research. Thanks also to Yvonne Teoh for help with the data analysis.

Correspondence concerning this article should be addressed to Elizabeth Sheppard, School of Psychology, University of Nottingham, University Park, NG7 2RD, Phone: +44 115 8467346, Email: elizabeth.sheppard@nottingham.ac.uk

ES involved in conception and design of 3 studies, data analysis and write-up; DP involved in conception and design of study 1, creation of videos, data collection and analysis of study 1; GTLW involved in conception, design, data collection and analysis of studies 2&3; DR involved in conception and design of study 1 and creation of videos; PM involved in conception and design of 3 studies, data analysis and write-up