

## How do Individuals with Asperger Syndrome Respond to Nonliteral Language and Inappropriate Requests in Computer-mediated Communication?

Gnanathusharan Rajendran,<sup>1,4</sup> Peter Mitchell,<sup>2</sup> and Hugh Rickards<sup>3</sup>

---

Computer-mediated communication in individuals with Asperger syndrome, Tourette syndrome and normal controls was explored with a program called Bubble Dialogue (Gray, Creighton, McMahon, & Cunningham (1991)) in which the users type text into speech bubbles. Two scenarios, based on Happé (1994) were adapted to investigate understanding of figure of speech and sarcasm, and a third, developed by ourselves, looked at responses to inappropriate requests (lending money and disclosing home address on a first meeting). Dialogue transcripts were assessed by 62 raters who were blind to the clinical diagnoses. Hierarchical linear modelling revealed that rated understanding of a figure of speech was predicted mainly by verbal ability and executive ability, as well as by clinical diagnosis, whereas handling inappropriate requests was predicted by age, verbal ability, executive ability and diagnosis. Notably, the Tourette comparison group showed better understanding than the Asperger group in interpreting a figure of speech and handling inappropriate requests, and differences between these groups were possibly attributable to individual differences in executive ability. In contrast, understanding sarcasm was predicted by age but not by either verbal ability, executive ability or clinical diagnosis. Evidently, there is a complicated relation between Asperger syndrome, verbal ability and executive abilities with respect to communicative performance.

---

**KEY WORDS:** Asperger syndrome, communication, Bubble Dialogue, role-play.

Individuals located on the autistic spectrum are characteristically impaired in their communicative abilities, which seems to be linked with their impaired understanding of the mind (Happé, 1994, 1995; Jolliffe, & Baron-Cohen, 1999). Impairment in communication is a serious handicap because it impedes social contact and hinders social integration, which

severely reduces the number of opportunities to learn about other minds (Woolfe, Want, & Siegal, 2002). It would therefore be valuable to gain a better understanding of autistic communicative impairment and to investigate contexts in which the impairment might be reduced or even eradicated altogether.

Extant research suggests that autistic communicative impairment might be especially noticeable in tasks where it is necessary to apprehend the mind that lies behind the literal meaning of the message. Happé (1994) and Jolliffe and Baron-Cohen (1999) found that individuals on the autistic spectrum have considerable difficulty understanding sarcasm and figures of speech. They presented 'Strange Stories', which were simple accounts of everyday events, followed by

---

<sup>1</sup> School of Education, The University of Edinburgh, St John's Land, Holyrood Road, EH8 8AQ, Edinburgh, UK.

<sup>2</sup> School of Psychology, The University of Nottingham, NG7 2RD, University Park, Nottingham, UK.

<sup>3</sup> The Queen Elizabeth Psychiatric Hospital, Mindelsohn Way, B15 2QZ, Edgbaston, Birmingham, UK.

<sup>4</sup> Correspondence should be addressed to: e-mail: Thusha.Rajendran@ed.ac.uk

questions about different motivations underlying utterances that are not literally true: “ Was it true, what X said?,” and, “ Why did X say that?”.

Varying degrees of success on the Strange Stories test (as rated by a pair of judges) could discriminate between individuals with autism who failed theory of mind tasks, individuals who passed 1st order and those who passed of 2nd order theory of mind tests. Happé (1994) also found that even individuals with autism who passed 2nd order theory of mind tasks gave incorrect mental state justifications for some of the stories, unlike the typically developing adult group who made no such errors. Jolliffe and Baron-Cohen (1999) conducted a replication of Happé’s (1994) study with two adult autistic groups: One comprised of individuals with AS the other high-functioning autistic (HFA) individuals. Jolliffe and Baron-Cohen found that both clinical groups failed to use context-appropriate mental state terms to explain the Strange Stories, and in this respect, they differed from closely matched comparison participants.

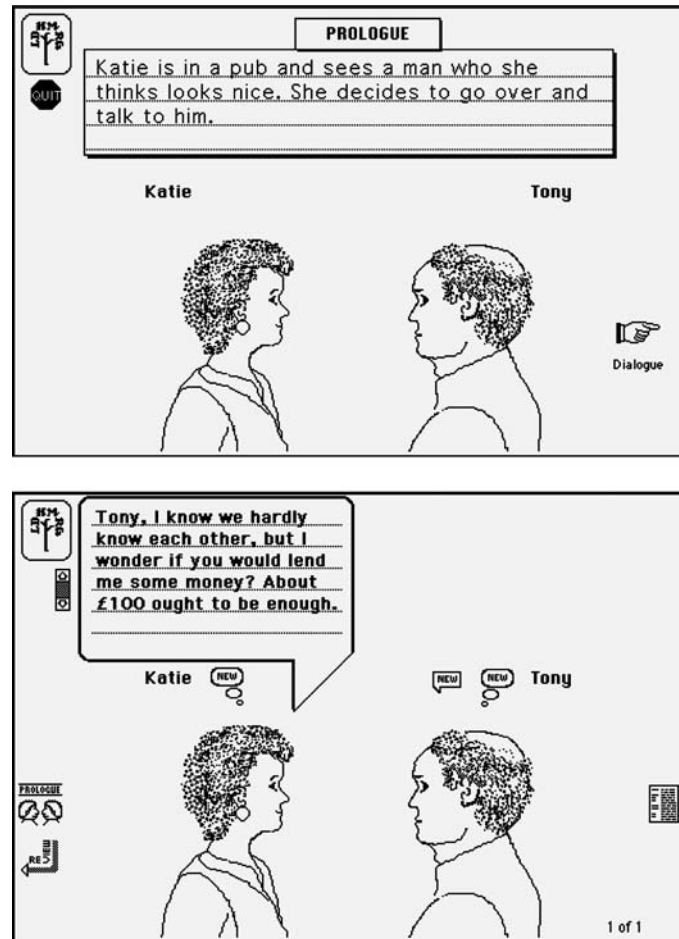
Despite the importance of these discoveries, it would not necessarily be appropriate to make assumptions about the quality of autistic communicative performance. Happé’s (1994) procedure seems to tap into a level of reflective understanding of nonliteral speech, a level that is one-step removed from the working understanding required to respond appropriately to a person who uses nonliteral speech during conversation. It is conceivable that while individuals with autism are poor at reflectively explaining what a speaker meant by a certain expression, they can nonetheless respond appropriately to a nonliteral utterance as if venturing beyond the literal interpretation. Presumably, an ability to respond appropriately to nonliteral communication is relevant to practical communicative proficiency and general social integration, so it is especially important to investigate the working understanding of nonliteral communication.

Additionally, performance at the reflective level does not necessarily translate into proficiency in a natural social context. Hadwin, Baron-Cohen, Howlin, and Hill (1996) found that although it was possible to teach individuals with autism to give correct judgments on a selection of false belief tests, there was no corresponding improvement in social functioning. The test of false belief requires a verbally explicit response to a question about what a person is thinking, and in this respect it seems to tap into understanding at a reflective level, as in Happé’s (1994) study.

The challenge, then, was to contrive a communicative exchange such that nonliteral language could be introduced whilst also taking a sensitive measure of the participant’s understanding. One way to test someone’s understanding of nonliteral language would be to say something sarcastic to them directly, and note if their response indicated an understanding of what was meant, rather than what was said. This would of course be unethical. Role-play would be ideal, except that it is notoriously difficult to coax individuals with autism to entertain the required level of make-believe. However, there is a computer program, ‘Bubble Dialogue’ (Gray, Creighton, McMahon, & Cunningham, 1991), which is well-suited for engaging individuals with autism in role-play, as explained below (see Parsons & Mitchell, 2002, for a discussion of the virtues of using the computer as a vehicle for role-play).

Bubble Dialogue presents two characters in a computer generated cartoon-strip environment (see fig. 1) and allows two players to take turns in typing text into a speech bubble above the characters’ heads. Each player ‘adopts’ one of the characters and takes the simulated conversation forward in the role of the adopted character. This computer-mediated role-play is prefaced by a ‘prologue’ which introduces the characters and the circumstance of their conversation.

Bubble Dialogue seems like a suitable tool in principle because computer-based tasks are ideal for investigating aspects of autism for the reasons suggested by Swettenham (1996): First, computers provide social and emotional distancing by acting as an interface between interactants, which might help to reduce the social anxiety that is typically experienced by individuals with autism; second, the computer accommodates the autistic need for sameness and predictability; third, it allows the individual to take control and work at his or her own speed. Additionally, there might be benefits in not conducting interactions face-to-face, given that there is no competing visual information that might cause distraction. Autistic individuals are known to have sensory and perceptual abnormalities (O’Neill & Jones, 1997), and the heavy visual processing accompanying face-to-face interaction could be at the cost of the processing needed to conduct a conversation. Moreover, many individuals with autism seem to have an affinity with computers (Prior *et al.*, 1998) and it might be possible to exploit this special interest as a vehicle for engaging them in social interaction (see Baker, Koegel, & Koegel, 1998).



**Fig. 1.** Prologue from the Appropriacy scenario and researcher's standard opening response. Figure from Bubble Dialogue, copyright by Harry McMahon; printed with kind permission from developers Harry McMahon, Bill O'Neil, and the Hypermedia Language Development and Research Group at the University of Ulster at Coleraine.

It is already known that individuals with autism perform well with Bubble Dialogue: Rajendran and Mitchell (2000) used the application with two adults with AS, and found that they easily adopted the role of a character in accordance with contextual information presented in the prologue. To a large extent, the participants responded appropriately to the utterances of a character played by the experimenter and their dialogue was clearly 'in-role'.

In this experiment, prologues of Bubble Dialogue were modified in order to portray a context borrowed from Happé's (1994) study that would allow the experimenter to begin his dialogue with a nonliteral utterance. How would participants respond? Would they behave as if taking the utterance literally, as we might expect from Happé's

results, or would there be signs that they interpreted nonliterally, thereby demonstrating a working understanding of nonliteral language?

In addition to investigating how participants respond to sarcasm and figures of speech, a supplementary dimension to the experimenter's dialogue was also scripted. In-role, he made a couple of socially inappropriate requests. One was for the loan of a large sum of money (£100) and another was for details of the participant's character's home address. The purpose was to begin to investigate social naivety in autistic communication. According to Frith (1989, 2003), individuals with autism characteristically lack common sense and in consequence could be socially vulnerable. Hitherto, we have no information on the relation between communicative discretion and other

aspects of communication, such as being able to interpret nonliterally. Specifically, even if individuals show a level of proficiency in responding to nonliteral communication, they might still divulge personal details or inappropriately agree to lend a large sum of money. Additionally, we do not know if the factors which affect these two areas of functioning are the same or different. Hence, the current study serves as a first attempt to investigate this possibility.

The current study combines elements of a group design with those of a single case design. Rather than using Happé's (1994) protocol, in which two judges (one of whom was the experimenter) decided whether a participant essentially understood the nonliteral utterance or not, multiple blind raters were recruited to offer a valuation. This method of using multiple raters was pioneered by Rajendran and Mitchell (2000) and allowed each participant to be entered into an analysis as a level of an independent variable. This enabled a comparison of the performance between individuals (rather than groups) using parametric statistics because each participant receiving multiple scores based on his/her typed responses. These scores were then analysed in a way which took account of each participant's characteristics (i.e., age, verbal and executive ability) as well as their clinical diagnosis. This was done by using a method of analysis (hierarchical linear modelling, HLM) which allowed the calculation of how much variance in rated scores was attributable to group membership and how much was due to individual characteristics. This in turn allowed the interplay between the individual's characteristics and their clinical label to be analysed.

The make-up of the comparison groups required special consideration. Hence, in addition to typically developing individuals, a clinical comparison group of individuals with Tourette syndrome (TS) was recruited. TS is characterised by motor and vocal tics, which may change in their frequency, complexity and severity (DSM IV-TR, APA, 2000). The disorder has an age of onset before 18 (DSM IV-TR, APA, 2000) and is most frequently seen in boys (Rickards, 1995), but unlike individuals with autism and AS, individuals with TS seem to have social awareness. However, in common with AS, it is a neurodevelopmental disorder with many individuals having obsessional behaviours; for example Robertson, Trimble, and Lees (1988) found that 37% of their group of 90 patients reported obsessive compulsive behaviour. Additionally, some individuals with TS are echolalic (Rickards, 1995) and characteristically have impaired executive abilities (Bornstein, 1990, 1991).

It is especially important to demonstrate that any autistic tendency to interpret literally is substantive rather than secondary to a failure of inhibition due to executive dysfunction (Mitchell, Saltmarsh, & Russell, 1997). If individuals with autism have a tendency to interpret literally and this is primarily due to executive dysfunction, then a similar level of performance should be witnessed in individuals with TS who also have executive dysfunction. On the other hand, if any tendency to interpret literally is primarily a consequence of a peculiarly autistic impairment in mentalising, then perhaps a tendency to be too literal will be confined to those with autism. Therefore, the TS group was included specifically to examine the possibility that difficulties with nonliteral speech result from executive impairments. Hence, we took measures of executive functioning in the participants with AS and TS and examined the extent to which diagnosis predicts performance over and above measures of executive functioning.

Alternatively, it could be that a tendency to interpret literally is secondary to general linguistic impairment in autism (see Happé, 1995), in which case clinical diagnosis will not contribute anything to the accuracy of prediction, over and above measures of general verbal ability. Therefore, we compared performance in the participants with autism with participants without autism who had a similar level of linguistic ability. We examined the extent to which diagnosis predicts performance over and above measures of linguistic ability.

Precisely the same arguments can be made about the relation between responding to inappropriate requests, executive dysfunction and general verbal ability. Evidently, one needs to inhibit a response to an inappropriate request, which in turn requires a certain level of executive control. The provision of a TS comparison group and the measure of executive ability is pertinent to this issue. Also, the nature of one's response to an inappropriate request might be linked with general linguistic abilities, and so a measure of this would be valuable as well.

From Happé's (1994) and Jolliffe and Baron-Cohen's (1999) studies, it seems many people with autism have severe difficulty with nonliteral language. However, those authors presented no evidence suggesting that the difficulty was related with CA, verbal and/or executive abilities. According to these researchers, even autistic individuals who pass second-order theory of mind tasks have problems with nonliteral language comprehension. Furthermore, they suggest that the problem is unique to autism and stems from a

domain-specific problem in mentalising. This study set out to update that picture by allowing participants an opportunity to demonstrate their working understanding, and discover what factors are associated with that working understanding.

## METHOD

### Participants

Nine people with AS and three with HFA were recruited. One participant with HFA was originally diagnosed with AS, but this was later changed to HFA by their clinician (according to his mother). Twelve individuals with TS and 12 typically developing participants (control group) were also recruited. The gender composition in each group was eight males and four females.

The AS and HFA participants were recruited via the Leicestershire Autism Outreach Team, or through the Leicestershire and Derbyshire Autistic Support Groups which require students and members to have formal diagnoses made by clinicians. Those who were students ( $n = 11$ , age range 11–19 years) had statements of special educational needs. For example, one student had a statement which said that his special educational needs arose from social, emotional and behavioural difficulties associated with his AS, and that his diagnosis was made through a medical referral. Furthermore, no participants were known to have a diagnosis co-morbid with any other disorder, and their statements of special educational needs corroborated this.

For the purposes of this study, the HFA individuals were combined with AS individuals to form the AS group because the nosologic status of

AS, especially in relation to autism without an accompanying learning disability remains unresolved, and it has proven diagnostically and experimentally difficult to distinguish the two (see Baron-Cohen, O'Riordan, Stone, Jones, & Plaisted, 1999, for a concise rationale).

The members of the TS group were recruited as outpatients from The Queen Elizabeth Psychiatric Hospital, Birmingham, and all met DSM IV-TR criteria (American Psychiatric Association, 2000) and were not co-morbid with AS. They were assessed by clinical interview, examination and appropriate investigations by consultant psychiatrist H.R. who is an expert in the field. Because they were outpatients, and essentially an opportunity sample, there was insufficient time to present more than one test. Therefore, a priority was to test the TS group with the BADS because they were included to address questions specifically relating to executive ability.

The control group members were selected individually to match the AS participants in terms of gender and as closely as possible in terms of age, level of education and verbal IQ (VIQ) Table I provides details of CA, VIQ, performance IQ (PIQ) and full-scale IQ (FSIQ). Four control participants were classmates of four of the individuals with AS. One was a University of Nottingham manager, and the other seven attended a local school.

All the participants were tested in a quiet space either at their place of education, place of work, hospital interview room, or home. Four one-way ANOVAs showed no significant differences between the groups on any of the variables: CA,  $F(2, 33) = 2.41$ ,  $p = .11$ ; VIQ,  $F(1, 22) = 0.06$ ,  $p = .81$ ; PIQ,  $F(1, 22) = 2.23$ ,  $p = .15$ ; FSIQ,  $F(1, 22) = 0.84$ ,  $p = .37$ , BADS profile score (see below),  $F(2, 33) = 3.04$ ,  $p = .06$ . Because the BADS profile score approached significance, a *post hoc* Tukey test was performed. This revealed a significant difference between the AS and Control groups,  $p = .049$ . Thus it was this pairwise comparison which drove the ANOVA for the three groups close to significance. Despite the difference between all three experimental groups approaching significance on BADS profile score, this is taken into account statistically in the ensuing analysis.

### Materials

The main hardware was an Apple Macintosh 500 series powerbook, installed with Bubble Dialogue (Gray *et al.*, 1991) and Hypercard.

Table I. Participants' Characteristics

Participant <sup>a</sup>	CA	VIQ	PIQ	FSIQ	BADS
AS					
<i>M</i>	16.46	110.25	93.25	102	14.08
<i>SD</i>	6.81	22.47	22.77	23.13	4.89
<i>Range</i>	11–36	73–148	58–128	70–139	7–21
Control					
<i>M</i>	16.78	111.83	104.5	109	17.83
<i>SD</i>	7.38	14.28	14.44	13.03	2.52
<i>Range</i>	12–39	86–137	76–123	90–133	13–22
TS					
<i>M</i>	23.85	-	-	-	16
<i>SD</i>	12.65	-	-	-	3.39
<i>Range</i>	11–47	-	-	-	12–21

<sup>a</sup> $n = 12$  in each group.

## Measures

There are features of autism which fall outside Wing and Gould's (1979) triad of impairments, such as the autistic individual's need for sameness, difficulty switching attention, a tendency to perseverate and a lack of impulse control. These particular features are similar to those shown by individuals with frontal lobe brain lesions in dysexecutive syndrome (Baddeley & Wilson, 1988). Therefore all 36 participants were tested with the Behavioural Assessment of the Dysexecutive syndrome (BADS—Wilson, Alderman, Burgess, Emslie, & Evans, 1996), which according to Evans, Chua, McKenna and Wilson (1997) is an ecologically valid test of executive function, assessing the everyday difficulties associated with Dysexecutive syndrome. The BADS is composed of six subtests and assesses a range of cognitive functions representative of executive abilities such as cognitive flexibility (Rule Shift Cards), novel problem solving (Action Programme), planning (Key Search, Zoo Map, Action Programme and Modified Six Elements), judgement and estimation (Temporal Judgement), and behavioural regulation (Key Search, Zoo Map, Modified Six Elements). Some of the subtests have multiple components: Planning, problem solving and monitoring behaviour are required for the Zoo Map and Modified Six Elements (Norris & Tate, 2000). For each subtest a summary profile is obtained (with a maximum of 4 and minimum of 0), and these are summed to produce an overall profile score out of 24. From the BADS handbook (Wilson *et al.*, 1996), a score of between 16 and 20 is classified as average, and the mean score from the 78 brain injured patients, from whom the normative values were derived, is 14.03.

Additionally, the AS group and the control group were tested with the Wechsler Abbreviated Scales of Intelligence (WASI – Wechsler, 1999) which consists of four subtests: Two assess receptive VIQ and two assess PIQ—also known as non verbal IQ. Both scales combine to give a FSIQ score.

## Bubble Dialogue Scenarios

The Sarcasm and Figure of Speech vignettes from Happé (1994) were identified as the ones most easily adaptable for role-play in Bubble Dialogue. A prologue can be created to introduce the users to a scenario before they start role-playing, and those for the Figure of Speech and Sarcasm scenarios were almost identical to Happé's (1994) vignettes. The few

word changes enabled the scenario to fit in the prologue box, which only accommodates a limited number of characters. Unlike Happé's study, however, participants were not asked comprehension and justification questions. Instead, the participants role-played one of the characters involved in the stories.

## Procedure

The prologue for the Figure of Speech scenario was, "Daniel and Ian see Mrs Thompson coming out of the hairdresser's one day. She looks a bit funny because the hairdresser has cut her hair much too short". After reading the prologue, the researcher (G.R.) took on the role of Daniel by typing into a speech bubble appearing above Daniel's head, "She must have been in a fight with a lawnmower".

The participant was then asked to type whatever s/he wanted in Ian's bubble. If the participant seemed unsure how to respond, the researcher stated that there was not a right or wrong thing to type. Note that the participant was not directed in any way regarding how or what to type. The Bubble Dialogue program was neither couched as a game, nor were the participants told to role-play the scenarios as if it they were real-life.

The participant and researcher continued their dialogue until they could not take it further or until they had to leave for their next lesson or return home. The same procedure was followed for all the scenarios.

The prologue for the Sarcasm scenario<sup>1</sup> was, "Sarah and Tom are on a picnic. It is Tom's idea, he says it is going to be a lovely day. But just as they are unpacking the food, it starts to rain, and soon they are both soaked". After reading the prologue the

<sup>1</sup>A point about terminology is perhaps needed to clarify the meaning of *sarcasm* and *irony*. The Oxford English Dictionary says that ironic utterances are thought to include "the use of words to express something other than and especially opposite of the literal meaning of a sentence," whereas sarcasm depends for its effect on "bitter, caustic, and other ironic language that is usually directed against an individual" (Gibbs, 1996, p. 3). According to Davidson (1996) sarcasm is intended to be hurtful or scornful. A common way of being sarcastic is through the use of irony, which involves this use of words to convey the opposite of literal meaning of the words. *That was a pretty stupid thing to do!* is sarcastic; *that was intelligent of you!* is both ironic and sarcastic. However, something that is ironic need not be sarcastic such as saying, "They tell me you're a slow runner" (Gibbs, 1986, p. 3) to someone who has just won the 100 m sprint. Kreuz and Glucksberg (1989) use the term sarcastic irony and define it as "the use of counterfactual statements to express disapproval, usually with intent to hurt or wound" (p. 374). Thus, Happé's (1994) sarcasm scenario should be classified more precisely as sarcastic irony.

following test questions were asked to check that the participants understood what was happening in the scenario and hence why this might provoke a sarcastic response (comprehension questions were asked in the Sarcasm scenario because a sarcastic comment could only be understood if the participant understood the discrepancy between the prediction and what actually happened): Whose idea was it for the picnic? What kind of day does Tom say it is going to be for the picnic? What happens just as Tom and Sarah are unpacking the food? Once the participants answered the questions correctly (which they all did), the researcher took on the role of Sarah and typed, "Oh yes, a lovely day for a picnic alright". The participant then role-played Tom.

A third scenario (not based on Happé, 1994) was designed to investigate socially appropriate behaviour based upon first meeting someone. The 'Appropriacy' scenario (see Fig. 1) was designed to investigate how participants respond to an inappropriate request and was 'Making a friend' taken from Rajendran and Mitchell (2000). The opening exchange was read by the participant and then the character Katie, played by G.R., continued the exchange by saying, "Tony, I know we hardly know each other, but I wonder if you would lend me some money? About £100 ought to be enough." (Note that Katie explicitly states in the first typed sentence that two characters have only just met). The participant then took on the role of Tony. Later in the exchange, Katie inquires about Tony's address with the following question, or something very similar: "Where exactly do you live Tony? Or "I would like to know where you live."

Most of the participants were tested on three separate occasions, once for the 3 Bubble Dialogue scenarios, once for the BADS and another time for the IQ test, with each session lasting about 1 hour. The testing orders varied for all participants, except that the AS group were tested with the WASI in their final session. The TS group, being outpatients, could only participate in one session and so were tested with Bubble Dialogue and the BADS. The three Bubble Dialogue scenarios were given in different orders.

### Rating Bubble Dialogue Scripts

The Bubble Dialogues for all 36 participants were transcribed and collated into 72-page rating booklets. Sixty-two of these were created, one for each of the blind raters (see below). The transcriptions looked like play scripts and were randomly

ordered into nine different template booklets (i.e., three booklets per scenario). The only proviso was that the participant groups were represented equally in the first transcript.

Sixty-two 'blind raters' were recruited to assess the transcripts. The raters were either prospective undergraduates, current undergraduates, postgraduates or research assistants at The University of Nottingham and were essentially an unpaid opportunity sample. Sixteen individuals rated all the Sarcasm transcripts, 24 rated all the Figure of Speech transcripts and 22 rated all the Appropriacy transcripts. Each rater took between 30 and 45 minutes to complete a booklet.

The raters were asked to rate the responses of characters played by the participants. For the Figure of Speech and Sarcasm scenarios, they were asked to rate 'How well does Ian understand figurative language?' and 'How well does Tom understand sarcasm?' respectively. They were asked to circle a line on a 6 point bi-polar scale, with 'No understanding' at one end and 'Good understanding' at the other.

For the Appropriacy scenario, two rating scales were developed, one asking, 'How appropriate was Tony's response when asked for money?' and the other asking, 'How appropriate was Tony's response when asked about his home address?' For these scales, raters were asked to circle a line on a 6 point bi-polar scale, with 'Appropriate' at one end and 'Inappropriate' at the other.

The raters were deliberately not given examples of what constituted an understanding of figure of speech and sarcasm; nor were they directly told what would count as an appropriate or inappropriate response. Hence, the basis of interpretation was left to their own judgement.

### Data Structure and Background to Analyses

The rated scores from the Bubble Dialogues ranged from 1 to 6. One way to analyse these data would have been to calculate a mean rating for each participant, and then use ANOVA to compare groups whilst entering age, verbal score and/or executive score as covariates. However, that method would be less than ideal because we would be discarding information relating to the variance associated with the mean rating for each participant. By taking into account the variance as well as the mean score, we could test whether participants significantly differed from each individual by individual. Thus, we needed an analytic tool in which participants (plus their associated rated

means and variance) could be subsumed within diagnostic groups, whilst statistically taking into account such factors as their CA, their verbal ability and their executive ability. Unfortunately, ANOVA and ANCOVA are not capable of handling data that are hierarchically structured in this way.

Hence, hierarchical linear modelling was used (HLM—Bryk & Raudenbush, 1992, 2002; HLM/5 software Bryk, Raudenbush, & Cogden, 2001; see Steiger, Gauvin, Jabalpurwala, Séguin, & Stotland, 2000, for a concise description of HLM). HLM is a type of general linear model used in multiple regression, but accepts unbalanced hierarchically structured data. HLM is ideally suited for analysing a hybrid of case and group study design, using multiple raters because different models can be tested to see which combination of variables provides the best fit. A measure of model fitness is given by the deviance statistic: The higher the deviance, the poorer the fit. The deviance is not interpretable directly, but differences in deviances between models have a chi-squared distribution and so can indicate a significantly improved model fit. An indicator of the proportion of variance accounted for by between-participant variability is given by the intraclass correlation coefficient ( $\rho$ ). Reliability estimates of the models (which range from 0 to 1) indicate how reliable the sample means are as indicators of the true mean. In this study the reliability estimate gives an indication of the level of agreement between raters of the true rating of the participant.

In HLM the dependent variable must be at Level 1. So in this study, the Level 1 model represents the ratings each participant received from each rater and Level 2 represents the participant's clinical diagnosis, their CA, verbal ability or executive ability. Bryk and Raudenbush (1992) recommend a "step up" approach in which variability at Level 1 is modelled first and then Level 2 is considered using higher order predictors. Furthermore, the coefficients generated in HLM can be used to create a regression line for one group and each participant's performance can be compared with this idealised line. Hence, we can see if group differences are attributable to only a few or to many individuals.

## RESULTS

The level of agreement between the raters was high as indicated by the large reliability estimates for each of the models run (ranging from 0.87 to 0.94);

Table II displays the mean rated scores for each experimental group, with higher scores showing better understanding of the two nonliteral speech scenarios, and better responses to inappropriate requests<sup>2</sup> An example of a response to, "She must have been in a fight with a lawnmower", in the Figure of Speech scenario, is "A lawnmower more like some shears". Subjectively, the participant with AS who made this response seems to show a good understanding of the figure of speech because of their use of a semantically related (garden) item. This is borne out by the fact that this person received the highest rating of any of the group with AS, and was the third highest of the combined control and AS groups. An example of a response to, "Oh yes, a lovely day for a picnic alright", in the Sarcasm scenario, is "well I didn't know it was going to rain did I but we could have it another day if thats alright by you". This participant with AS who made this response also seems to have understood the nonliteral comment. An example of a response to the Appropriacy scenario, "Tony, I know we hardly know each other, but I wonder if you would lend me some money. About £100 ought to be enough." is "Katie, I really do not think it is a good idea to lend you some money for the reason that you state". Although, on the surface this may seem an appropriate response, it seems rather formulaic. This may be why the response, which although was second highest in the AS group, was rated seventh highest of the combined AS and control groups.

In the first part of the Results section we compare the performance of individuals with AS with the control group using HLM modeling. The second part is analogous, except that individuals with AS are compared with the TS group. The analysis was split in this way to allow for a clearer interpretation the effects of age, verbal ability and executive ability between the groups.

### Comparisons between AS and the Control Group

We investigated whether individuals with AS have problems understanding nonliteral speech and

<sup>2</sup>For the Sarcasm scenario, Rater 2 omitted rating participant 5. Therefore this rating was included as the mean score given to participant 5 by the other 15 raters. Additionally, the rating scale for the Sarcasm scenario for participant number 25 was incorrectly labelled, so two analyses were performed, one including and one excluding participant 25. There was virtually no difference between the analyses, so participant 25's scores were included in all the reported analyses.



**Table II.** The Mean Rated Scores (Standard Deviations in Parentheses) for Each Participant Group, with Higher Scores Showing Better Understanding of the Two Nonliteral Speech Scenarios, and Better Responses to Inappropriate Requests

Group	Figure of Speech	Sarcasm	Appropriacy-lending money	Appropriacy-disclosing home address <sup>a</sup>	Appropriacy-combined scales <sup>a</sup>
AS	3.65 (1.69)	3.68 (1.70)	3.34 (1.68)	3.05 (1.54)	6.48 (2.72)
Control	4.27 (1.40)	4.07 (1.65)	4.34 (1.41)	3.74 (1.53)	8.08 (2.54)
TS	4.41 (1.47)	4.13 (1.61)	4.03 (1.66)	3.42 (1.54)	7.77 (2.61)

<sup>a</sup> For the Appropriacy—1 participant in the AS group and 3 in the TS group did not continue the dialogue far enough to give responses when asked for their home address, and were not rated.

inappropriate requests that are independent of the characteristics of their general verbal and executive abilities. Specifically, is there a significant difference at the group level after verbal and executive ability are statistically taken into account, with the control group rated as showing better understanding? Raw verbal score, as opposed to verbal IQ, was used in the ensuing analysis, so that age could be entered a separate variable.

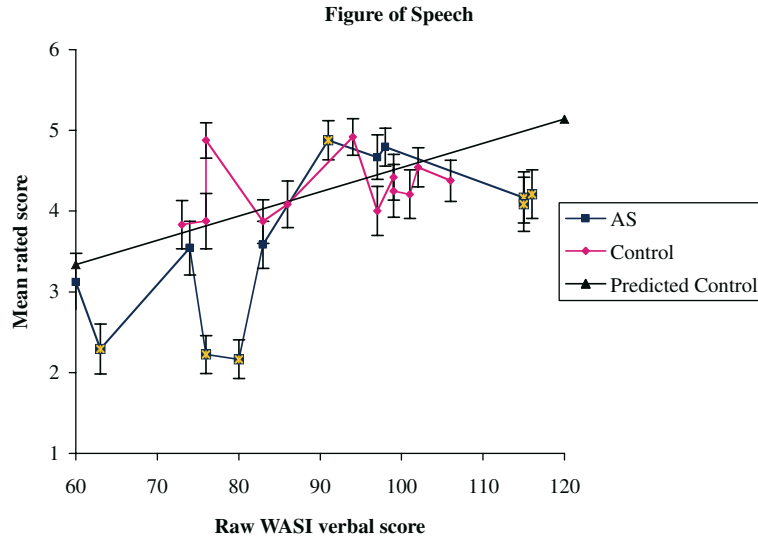
From the Figure of Speech scenario analysis, the intraclass correlation coefficient showed that 21% of the variance in rated score is between participants (and consequently 79% is within participants). This established, the fixed effect of age was entered as a Level 2 predictor in the next model and this model showed that age did not significantly predict rated score ( $t(22) = 1.08, p = .30$ ). Then raw verbal score was included in the next model and its coefficient was found to be significant ( $t(22) = 3.58, p = .002$ ). Hence, group was then added to the model and this further improved the fit to the data, as indicated by a significant reduction in the deviance:  $\chi^2(1) = 5.28, p = .02$ . BADS profile score was included in the next model, though its coefficient was found to be non-significant. Therefore, the final model included both raw verbal score and group and accounted for approximately 53% of the variance. A model which included only the verbal score as a predictor accounted for approximately 38% of the between-participant variance. A model which used only group accounted for about 19% of the variance. In other words, both verbal ability and clinical diagnosis predict the ratings participants received for their understanding of a figure of speech. Notably, clinical diagnosis accounted for variance in rated understanding of figure of speech independently of verbal ability.

Fig. 2 suggests a distinct lack of uniformity in rated understanding within the AS group. In order to describe this state, we plotted the relation between

rated score and verbal ability in the control sample and fitted a linear regression line on top of these data using the coefficients from the final HLM model. For any value of verbal ability, we were able to predict rated score. We then examined data from the AS group to determine how many of the rated means differed from the prediction, using one-sample *t*-tests. Figure. 2 shows that half of the participants with AS fell significantly below the predicted level in their rated understanding ( $p < .02$  in all cases). Most of the remaining participants were not different from the predicted level of performance and one was even significantly above the predicted level ( $p = .02$ ). In other words, many of the participants in the AS group underperformed in their rated understanding of figurative speech, but this was not inevitable and neither was it purely incidental to measured verbal ability. We used the same statistical strategy in the ensuing analyses.

For the Sarcasm scenario, the intraclass correlation coefficient ( $\rho$ ) showed that 36% of the variance in rated score was between participants. This established, the fixed effect of age was included as a Level 2 predictor in the next model, and the associated coefficient was significant ( $t(22) = -4.65, p = <.001$ ). However, neither the inclusion of raw verbal score, nor BADS profile score, nor group significantly increased the level of prediction. A model which included only age as a predictor accounted for approximately 13% of the between-participant variance. In other words, rated score varied with age, but not with measured verbal ability, and neither did it differ between groups. Curiously, the direction of the relation was for younger participants to gain higher ratings for understanding sarcasm than older participants.

Turning attention to participants' responses to inappropriate requests, those rated as responding inappropriately in one scenario tended also to be rated as responding inappropriately in the other:

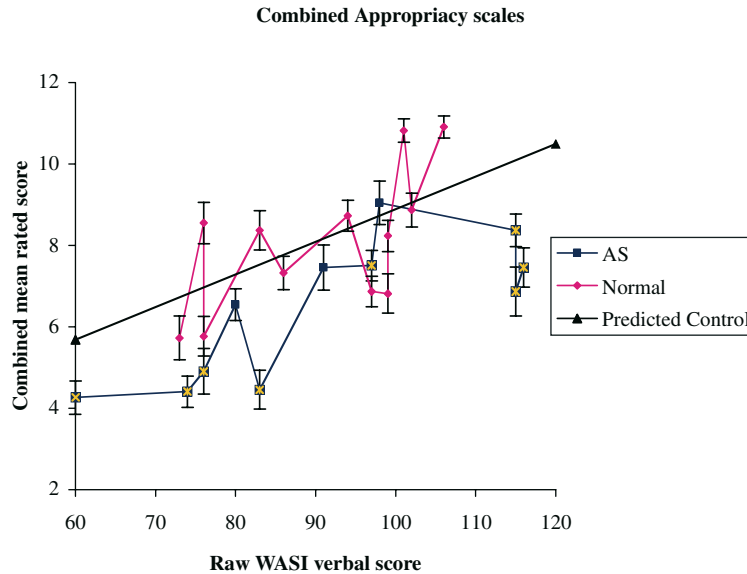


**Fig. 2.** Rated understanding of figure of speech for the AS and control groups, plotted against verbal score. The 'predicted' control line is derived from the relation between rated understanding and verbal score in the control group. Asterisks superimposed on the points on the graph indicate significant differences from the predicted control.

$r = 0.56$ ,  $p = .001$ . In the interest of brevity, the data from the Appropriacy scenarios were combined (see fig. 3) and it was found that the intraclass correlation coefficient was 41%. The fixed effect factor for age was included as a Level 2 predictor and it significantly contributed to the prediction of rated score ( $t(21) = 2.678$ ,  $p = 0.01$ ). The older participants were rated as being better at responding to an inappropriate request. Group was then added to the model which significantly reduced the deviance statistic:  $\chi^2(1) = 5.94$ ,  $p = 0.01$ . Then raw verbal score was included as a predictor and its coefficient was significant ( $t(19) = 4.672$ ,  $p < .001$ ), but the coefficient for age was no longer significant ( $t(19) = 1.826$ ,  $p = .08$ ). This implies that age and verbal ability account for some shared variance. BADS profile score was added to the model including group and raw verbal score, but its coefficient was nonsignificant. Therefore, the best model to fit the data included both raw verbal score and group, accounting for 62% of the variance. These relations are depicted in Fig. 3, which shows that 8 out of the 11 participants with AS were rated significantly below prediction based on a regression line derived from the control group data (one-sample  $t$ -tests,  $p \leq .006$  in all cases). These patterns of results closely corresponded with those that we gained from analysing data from the two appropriacy scenarios independently.

To gain further insight into the basis of the ratings of inappropriacy in the AS group, we coded participants on whether they offered to lend money (any amount, not just £100), the full £100, and if they disclosed a full home address (house number, road name). These classifications were then point-biserially correlated with the mean rated appropriacy scores for the AS group. There was a significant correlation of  $r_{pb} = 0.69$  ( $p = .01$ ) when the criterion was offering to lend (any) money. The correlation was  $r_{pb} = 0.57$  ( $p = .05$ ), when the criterion was offering to lend the full amount; it was  $r_{pb} = 0.72$  ( $p = .01$ ) when the criterion was disclosing the home address. Apparently, raters were heavily influenced by whether or not participants offered to lend money and disclose their home address, though not all the variance in rated appropriacy is explained by these criteria. Perhaps raters were also influenced by other aspects of the interaction and were rating not only what was said, but how it was said.

To investigate the link between the social and linguistic aspects of communication in AS, a correlation was conducted between the scores for Appropriacy and Figure of Speech:  $r = 0.65$  ( $p = 0.03$ ). This correlation was independent of age but not independent of raw verbal score or BADS profile score. That is, when raw verbal score and BADS profile score were entered as partial correlates (in



**Fig. 3.** Ratings of responses for combined Appropriacy scenarios for the AS and control groups, plotted against verbal score. The 'predicted' control line is derived from the relation between rated understanding and verbal score in the control group. Asterisks superimposed on the points on the graph indicate significant differences from the predicted control.

separate analyses) the correlation was no longer significant.

### Comparisons between AS and the TS Group

If individuals with AS have problems understanding nonliteral speech and with inappropriate requests, independent of executive ability, then there should be a significant difference at the group level, with the TS group rated as having better understanding or responding more appropriately (despite the difference between the three experimental groups approaching significance on BADS profile score, this is taken into account statistically on an individual level in HLM). To address this question, we conducted a series of analyses analogous to those presented above.

For the Figure of speech scenario the intraclass correlation coefficient was 28%. This established, the fixed effect of BADS profile score was included as a Level 2 predictor and the associated coefficient was significant ( $t(22) = 4.37, p < .001$ ). Neither the inclusion of age nor group significantly increased the level of prediction. A model which included only the BADS profile score as a predictor accounted for approximately 39% of the between-participant vari-

ance, while a model which used only group accounted for about 20% of the variance. The model using both Level 2 predictors accounted for approximately 48% of the variance. In other words, any difference in rated score between groups could largely be accounted for by differences in measured executive ability.

A pattern that was somewhat less clear emerged for the Sarcasm scenario. The intraclass correlation coefficient was 29% and Level 2 models showed that the use of BADS profile score as a coefficient did not significantly predict rated score ( $t(22) = 1.97, p = .06$ ) and neither was there an increase in the level of prediction when group membership was added to the model ( $t(22) = 1.15, p = 0.26$ ), nor age ( $t(22) = -0.55, p = .59$ ). In other words, rated score did not differ between groups, did not vary with measured executive ability and did not vary with age.

When the Appropriacy scenarios were combined, the analysis showed that the intraclass correlation coefficient was 38%. This established, the fixed effect of BADS profile score was included as a Level 2 predictor in the next model, and its associated coefficient significantly improved the level of prediction ( $t(18) = 3.99, p = 0.001$ ). However, neither the inclusion of age nor group significantly improved the level of prediction any further. Hence, the final

model, which included only the BADS profile score as a predictor, accounted for approximately 48% of the between-participant variance. In other words, only executive ability predicted the ratings participants received for the appropriateness of their responses when asked to lend money and disclose their home address. These patterns of results closely corresponded with those gained from analysing data from the two appropriacy scenarios independently.

## DISCUSSION

Reputedly, it is notoriously difficult to coax individuals with autism to participate in role-play. In the current study, however, individuals with AS took to the computer-mediated role-play and adopted their character as introduced in the prologue. The dialogues generated by the participants included details that were appropriate for their character and were not merely a literal rendition of their own details. This in itself is surprising when considering results from studies which suggest individuals with autism have difficulty with generativity (Turner, 1999).

Generally, the data convey the heterogeneity of performance across individuals with autism. It was not the case that individuals with AS uniformly underperformed on the various measures. Some performed well and some performed not so well. This gives a preliminary impression that there is not something essentially autistic that leads individuals with AS to underperform in understanding a figure of speech or in responding to an inappropriate request. An interaction between autism, age and certain levels of verbal or executive ability might predict understanding of nonliteral language and responding suitably to inappropriate requests. We consider this possibility below.

Furthermore, another aspect of nonliteral language which may need to be considered is the developmental perspective. In Happé's (1994) and Jolliffe and Baron-Cohen's (1999) studies, nonliteral language comprehension was considered to be an index of theory of mind. However, no acknowledgement was given to developmental factors, despite the ages of participants with autism ranging from 8.9 to 45.1 years in Happé's (1994) study and 18–49 years in Jolliffe and Baron-Cohen's (1999) replication (see Winner, Rosentiel, & Gardner, 1976, for a study investigating the development of metaphoric understanding and Dermorest, Meyer, Phelps, Gardner, &

Winner, 1984, for study looking at false remarks, including sarcasm). In our study, participants' age was taken in account as a variable in the analysis. It revealed that age was not a factor in understanding figure of speech, but it was in understanding sarcasm (and responding to an inappropriate request).

Rated understanding of sarcasm is quite different from rated understanding of figurative language in this respect. Evidently, it would not be wise to treat these two measures as tapping into the same underlying competence. While the main challenge posed by figurative language might be to understand what is meant rather than what is said, sarcasm might pose the challenge of understanding how an utterance alludes to an implicitly or explicitly stated expectation that has not come to fruition (Gibbs, 1986, 1994; Jorgensen, Miller, & Sperber, 1984). Furthermore, in sarcasm, the speaker's statement is out of line with his/her behaviour and intonation, and these two clues may be crucial in accurately assessing the purpose of the statement (Demorest *et al.*, 1984). This may be why it is difficult to show understanding of sarcasm without those cues. Nevertheless, our results suggest that some of the between-participant variance is accounted for by age, and understanding may even be weaker in the older participants we tested.

Turning to the Appropriacy scenarios, it seems that raters were influenced by whether or not the participant offered to lend money or disclosed their home address. The correlation between ratings and these categorical data were strong but imperfect, suggesting that other aspects of communication were also influencing the raters, perhaps concerning the manner in which participants responded to the inappropriate request. Participants' rated responses to inappropriate requests were predicted by their age, and measures of executive and verbal ability. The participants who were most likely to be rated as responding appropriately had higher scores in verbal and executive ability and they tended to be older.

Interestingly, after taking into account measured executive ability, it was not possible to discriminate between individuals with AS and TS in their rated responses to figures of speech and inappropriate requests. This contrasts with the comparison between the AS and Control groups, for whom executive ability did not predict rated score for any of the scenarios. In short, there is nothing mysteriously autistic in the performance on these tasks by individuals with AS. However, since we have no way of knowing how verbal ability might be involved in differentiating between the AS and TS groups we

cannot be certain if executive ability is the only factor that explains differences between these groups. Nevertheless, the results are somewhat surprising in relation to Happé's (1994) conclusion. She found a link between understanding figurative language and being able to diagnose higher-order beliefs. Her account seems to suggest that a uniquely autistic deficit in mentalising can explain autistic difficulties with figurative language.

The relation between executive and language ability has been closely investigated by Bishop and Norbury (2005a, 2005b) who found that nonautistic children with pragmatic language impairment (PLI) were just as impaired as children with autism on a test of response inhibition and had problems generating novel ideas ('generativity', see Turner, 1999). Additionally, Liss *et al.* (2001) found that differences in performance on the Wisconsin Card Sorting Task between language impaired and autistic children were no longer significant when verbal IQ was covaried.

The results of this study suggest, then, that not all of our participants with AS had problems with figurative language and inappropriate requests. Furthermore, the correlation between rated understanding of figurative language and rated appropriacy suggests a link between the linguistic and social aspects of understanding others' minds. Moreover, any such problems seemed to be secondary to aspects of verbal ability and executive functioning. This finding might pose a problem for the specificity claim of the theory of mind hypothesis of autism (Baron-Cohen, 1995), which states that autism involves a domain-specific deficit. The results of the current study, rather, are consistent with theories that propose that impairments in mentalising are subordinate to more general cognitive or executive impairments (Frye, Zelazo, & Palfai, 1995; Riggs, Peterson, Robinson, & Mitchell, 1998). Neurological evidence also points to interlinking in mentalising and executive function, given that both kinds of ability are located in the frontal lobes (Stuss, Gallup, & Alexander, 2001). On the other hand, Fine, Lumsden and Blair (2001) report a single case of a man with early left amygdala damage who performed well on a battery of executive function tasks, including the BADS, but failed a number of theory of mind tests, including Happé's Strange Stories.

Apart from the possibility that some individuals with AS responded to figurative language and inappropriate requests in a suitable way, we also found that some individuals without autism, notably some of those with TS, did not always perform ideally. It

seems that individual differences in rated scores were attributable largely to executive ability, suggesting that executive ability may be a more important factor in predicting rated performance than clinical diagnosis. Hence, impaired performance on these tasks may not be unique to autism.

The procedure of using computer-mediated role-play to investigate understanding of nonliteral language comprehension is arguably more useful than a 'paper and pencil' test, both in principle and practically. It is a useful task in principle because participants demonstrate their working understanding in a simulated conversation, which might have more in common with the level of functioning that is involved in real-life performance, in contrast to reflecting on a vignette whose relevance to real life is uncertain. Indeed, Green, Gilchrist, Burton and Cox (2000) found that despite good abstract understanding of social relationships, adolescents with AS showed a profound lack of social ability in every day life.

Computer-mediated role-play is also a useful tool on practical grounds because the scripts can be blind rated by multiple raters on the degree of understanding. Hence, the procedure is sensitive to subtleties in levels of understanding and is not confined to categorising participants as either passing or failing. The limitations of dichotomous 'all-or-nothing' tests has been criticized by Klin (2000). He states that because success on theory of mind tasks depends on verbal ability, there is too much discrepancy between theory of mind task performance and everyday social functioning in autism. This is because social ability lies along a continuum, relating to degree of competence rather than passing or failing (Klin, 2000). Our gradation of competence via use of blind raters and rating scales also allows statistical analysis at the level of the individual participant, thereby demonstrating the heterogeneity of performance within a group of individuals with autism. There was a striking level of consistency between raters, as indicated by high estimates of the average intercept reliability for all the models. These high levels of reliability indicate good agreement in what raters thought of as evidence for showing understanding sarcasm, figure of speech and appropriate behaviour.

Apart from being a useful tool for investigating understanding of nonliteral language, computer-mediated role-play is also valuable for investigating aspects of socialisation, such as responses to inappropriate requests. The measure seems reliable, given that if a participant is rated as inappropriate in

lending money, they are also likely to be rated as inappropriate in disclosing a home address. Impairment in socialisation is perhaps the most important of the autistic triad in the light of Gillman, Carter, Volkmar and Sparrow (2000) reporting that 48% of the variance in diagnosis is accounted for by impairment in socialisation on its own.

Channon, Charman, Heap, Crawford and Rios (2001) state that although clinical reports show that individuals with AS have difficulties in social situations, these are seldom documented systematically. In the current study raters were essentially asked to make judgments about the appropriacy of specific responses to a social situation. The consistency of the raters, coupled with the strong correlation between Appropriacy scenarios, suggests that there is consensus about social norms; it is widely recognised as inappropriate to lend money on first meeting someone and to disclose your home address.

Two studies have attempted to investigate understanding of social behaviour in autism (Channon *et al.*, 2001; Loveland, Pearson, Tunali-Kotoski, Ortegon, & Cullen Gibbs, 2001), and both required participants to make reflective judgements about others' interactions or to supply solutions to problems arising in social contexts. Loveland *et al.* note that high-functioning individuals with autism sometimes show social understanding when asked to reflect on social behaviour and yet show inappropriate behaviour in everyday life. They suggest that either the measures used under test conditions are too coarse to detect subtle group differences, or that individuals with autism may know about social appropriateness, but may not act upon their knowledge. Computer-mediated role-play might offer a more direct and valuable method of investigation because it allows us to observe how individuals apply their knowledge of social appropriateness, rather than merely reflect on what is appropriate or inappropriate.

## ACKNOWLEDGEMENTS

We thank the students, parents, teachers and staff of Leicestershire Autism Outreach Team, Leicestershire and Derbyshire Autistic Support Groups, The Nottingham Blue Coat School, Brookfield Community School, Burleigh Community College, Holbrook Centre for Autism, Mapplewell Hall School, Shepshed High School, Whitegates

Further Education Unit and the patients and staff at The Queen Elizabeth Psychiatric Hospital, Birmingham. Special thanks goes to everyone who helped with the blind rating, and to Eamon Ferguson for invaluable statistical help. This research was completed while the first author held an Economic and Social Research Council studentship and then an ESRC Fellowship at The University of Nottingham. Bubble Dialogue was developed by the Language Development and Hypermedia Research Group, The University of Ulster at Coleraine under the direction of Harry McMahon who generously allowed us to use it.

## REFERENCES

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed. text revised). Washington, DC.
- Baddeley, A., & Wilson, B. A. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and Cognition*, 7, 212–230.
- Baker, M.J., Koegel, R.L., & Kogel, L.K. (1998). Increasing the social behaviour of young children with autism using their obsessive behaviours. *Journal of the Association for Persons with Severe Handicaps*, 23(4), 300–308.
- Baron-Cohen, S. (1995). *Mindblindness*. Cambridge, MA: MIT Press.
- Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or High-Functioning Autism. *Journal of Autism and Developmental Disorders*, 29(5), 407–418.
- Bishop, D. V. M., & Norbury, C. F. (2005a). Executive functions in children with communication impairments, in relation to autistic symptomatology - 1: Generativity. *Autism*, 9(1), 7–27.
- Bishop, D. V. M., & Norbury, C. F. (2005b). Executive functions in children with communication impairments, in relation to autistic symptomatology - 2: Response inhibition. *Autism*, 9(1), 29–43.
- Bornstein, R. (1990). Neuropsychological performance in children with Tourette syndrome. *Psychiatry Research*, 33, 73–81.
- Bornstein, R. (1991). Neuropsychological correlates of obsessive characteristics in Tourette syndrome. *Journal of Neuropsychiatry*, 3, 157–163.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models for social and behavioral research: Application and data analysis models*. Newbury Park, CA: Sage.
- Bryk, A. S., & Raudenbush, S. W. (2002). *Hierarchical linear models for social and behavioral research: Application and data analysis models* (2nd ed.). Newbury Park, CA: Sage.
- Bryk, A. S., Raudenbush, S. W., & Cogden, R. (2001). *Hierarchical linear modeling with HLM 5 programs* (version 5.04) [computer program manual]. Chicago, Illinois: Scientific Software International.
- Channon, S., Charman, T., Heap, J., Crawford, S., & Rios, P. (2001). Real-life type problem solving in Asperger's syndrome. *Journal of Autism and Developmental Disorders*, 31(5), 461–469.
- Davidson, G. (1996). *Chambers guide to grammar and usage*. St Ives: Chambers.
- Demorest, A., Meyer, C., Phelps, E., Gardener, H., & Winner, E. (1984). Words speak louder than actions: Understanding

- deliberately false remarks. *Child Development*, 55, 1527–1534.
- Evans, J. J., Chua, S. E., McKenna, P. J., & Wilson, B. A. (1997). Assessment of the dysexecutive syndrome in schizophrenia. *Psychological Medicine*, 27, 635–646.
- Fine, C., Lumsden, J., & Blair, R. J. R. (2001). Dissociation between 'theory of mind' and executive functions in a patient with early left amygdala damage. *Brain*, 124, 287–298.
- Frith, U. (1989). *Autism: Explaining the enigma*. Oxford: Blackwell.
- Frith, U. (2003). *Autism: Explaining the enigma* (2nd ed.). Oxford: Blackwell.
- Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, 10, 483–527.
- Gibbs, R. W. (1986). On the psycholinguistics of sarcasm. *Journal of Experimental Psychology: General*, 115, 3–15.
- Gibbs, R. W. (1994). *The poetics of mind. Figurative thought, language, and understanding*. Cambridge: Cambridge University Press.
- Gillman, J. E., Carter, A. S., Volkmar, F. R., & Sparrow, S. S. (2000). Towards a developmental operational definition of autism. *Journal of Autism and Developmental Disorders*, 30(4), 269–278.
- Gray, B., Creighton, N., McMahon, M., & Cunningham, D. (1991). Getting started with Bubble Dialogue, Language Development and HyperMedia Research Group Internal Report, University of Ulster at Coleraine, Ulster.
- Green, J., Gilchrist, A., Burton, D., & Cox, A. (2000). Social and psychiatric functioning in adolescents with Asperger syndrome compared with conduct disorder. *Journal of Autism and Developmental Disorders*, 30(4), 279–293.
- Hadwin, J., Baron-Cohen, S., Howlin, P., & Hill, K. (1996). Can we teach children with autism to understand emotions, belief, or pretence?. *Development and Psychopathology*, 8, 345–365.
- Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped and normal children and adults. *Journal of Autism and Developmental Disorders*, 24, 129–154.
- Happé, F. G. E. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, 66, 843–855.
- Jolliffe, T., & Baron-Cohen, S. (1999). The Strange Stories test: A replication of with high-functioning adults with autism or Asperger syndrome. *Journal of Autism and Developmental Disorders*, 29(5), 395–406.
- Jorgensen, J., Miller, G., & Sperber, D. (1984). A test of the mention theory of irony. *Journal of Experimental Psychology: General*, 10, 1–9.
- Klin, A. (2000). Attributing social meaning to ambiguous visual stimuli in higher-functioning autism and Asperger syndrome: The social attribution task. *Journal of Child Psychology and Psychiatry*, 41, 831–846.
- Kreuz, R. J., & Glucksberg, S. (1989). How to Be Sarcastic: The Echoic Reminder Theory of Verbal Irony. *Journal of Experimental Psychology: General*, 118(4), 373–386.
- Liss, M., Fein, D., Allen, D., Dunn, M., Feinstein, C., Morris, R., & Waterhouse Rapin, L. I. (2001). Executive functioning in high-functioning children with autism. *Journal of Child Psychology and Psychiatry*, 42(2), 261–270.
- Loveland, K. A., Pearson, D. A., Tunali-Kotoski, B., Ortegón, J., & Cullen Gibbs, M. (2001). Judgements of Social Appropriateness by Children and Adolescents with Autism. *Journal of Autism and Developmental Disorders*, 31(4), 367–376.
- Mitchell, P., Saltmarsh, R., & Russell, H. (1997). Overly literal interpretations of speech in autism: Apprehending the mind behind the message. *Journal of Child Psychology and Psychiatry*, 38, 685–692.
- Norris, G., & Tate, R. L. (2000). The Behavioural Assessment of the Dysexecutive Syndrome (BADS): Ecological, Concurrent and Construct validity. *Neuropsychological Rehabilitation*, 10(1), 33–45.
- O'Neill, M., & Jones, R. S. P. (1997). Sensory-perceptual abnormalities in Autism: A case for more research?. *Journal of Autism and Developmental Disorders*, 27(3), 283–293.
- Parsons, S., & Mitchell, P. (2002). The potential of virtual reality in social skills training for people with autistic spectrum disorders. *Journal of Intellectual Disability Research*, 46(5), 430–443.
- Prior, M., Eisenmajer, R., Leekam, S., Wing, L., Gould, G., Ong, B., & Dove, D. (1998). Are there subgroups within the autistic spectrum? A cluster analysis of a group of children with autistic spectrum disorders. *Journal of Child Psychology and Psychiatry*, 39(6), 893–902.
- Rajendran, G., & Mitchell, P. (2000). Computer mediated interaction in Asperger's syndrome: The Bubble Dialogue program. *Computers and Education*, 35, 187–207.
- Rickards, H. (1995). Tics and fits. The current status of Gilles de la Tourette syndrome and its relationship with epilepsy. *Seizure*, 4, 259–266.
- Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality?. *Cognitive Development*, 13, 73–90.
- Roberston, M., Trimble, M. R., & Lees, (1988). The psychopathology of the Gilles de la Tourette syndrome: A phenomenological analysis. *British Journal of Psychiatry*, 152, 383–390.
- Steiger, H., Gauvin, L., Jabalpurwala, S., Séguin, J. R., Stotland, S. (1999). Hypersensitivity to social interactions in bulimic syndromes: Relationship to binge eating. *Journal of Consulting and Clinical Psychology*, 67(5), 765–775.
- Stuss, D. T., Gallup, G. G., & Alexander, M. P. (2001). The frontal lobes are necessary for 'theory of mind'. *Brain*, 124, 279–286.
- Swettenham, J. S. (1996). Can Children with autism be taught to understand false belief using computers?. *Journal of Child Psychology and Psychiatry*, 37, 157–165.
- Turner, M. A. (1999). Generating novel ideas: Fluency performance in high-functioning and learning disabled individuals with autism. *Journal of Child Psychology and Psychiatry*, 40, 189–201.
- Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. The Psychological Corporation.
- Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H. C., & Evans, J. J. (1996). *The Behavioural Assessment of the Dysexecutive Syndrome*. Thames Valley Test Company: Flempton, Bury St. Edmunds.
- Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of Autism and Developmental Disorders*, 9, 11–29.
- Winner, E., Rosenstiel, A. K., & Gardner, H. (1976). The development of metaphoric understanding. *Developmental Psychology*, 12(4), 289–297.
- Woolfe, T., Want, S. C., & Siegal, M. (2002). Signposts to development: Theory of mind in deaf children. *Child Development*, 73, 768–778.