

The Awkward Moments Test: A Naturalistic Measure of Social Understanding in Autism

Lisa Heavey,^{1,4} Wendy Phillips,² Simon Baron-Cohen,³ and Michael Rutter¹

Details are given of a new advanced theory of mind task, developed to approximate the demands of real-life mentalizing in able individuals with autism. Excerpts of films showing characters in social situations were presented, with participants required to answer questions on characters' mental states and on control, nonsocial questions. When compared with control participants, adults with high-functioning autism and Asperger syndrome were most impaired in their ability to answer the questions requiring mind-reading ability. Although the present findings have implications for task modification, such naturalistic, dynamic stimuli are held to offer an important means of studying subtle difficulties in mentalistic understanding.

KEY WORDS: Theory of mind; mind-reading ability.

INTRODUCTION

It is well established that the capacity for understanding others' minds is impaired in autism (Baron-Cohen, 1995; Happé, 1994a). The inability of the majority of individuals with autism to appreciate others' thoughts and beliefs and to understand the relationship between mental states and behavior has been widely studied within the "theory of mind" deficit account of autism. Such difficulties in mentalizing have been shown in a range of experimental paradigms which include first-order false belief tasks (requiring the participant to infer a story character's mistaken belief about a situation; e.g., Baron-Cohen, Leslie, & Frith, 1985), second-order belief attribution tasks (requiring the participant to infer a story character's thoughts about another character's thoughts; Baron-Cohen, 1989a), the production of mental state terms in spontaneous speech (Tager-Flusberg, 1992) and tasks which require the dis-

inction between mental and physical entities, that is, understanding how physical objects differ from thoughts about objects (Baron-Cohen, 1989b). Through the inventive use of comparison groups and control conditions, studies of individuals with autism have elucidated the nature of the mentalizing deficit and its relative specificity to this clinical group (e.g. Baron-Cohen *et al.*, 1985; Leslie & Frith, 1988). Such a deficit in theory of mind (ToM) predicts a range of selective impairments at the behavioral level and thus accounts well for the social, communicative and imaginative difficulties characteristic of the disorder (Frith, 1989).

Many of the early pivotal studies in this research area focused on the task performance of mentally handicapped children with autism; with the methodology employed thus appropriate to the chronological age and intellectual ability of such a participant group. For example, in first- and second-order ToM tests dolls/figures were used to enact scenes (e.g., Baron-Cohen, 1989a; Baron-Cohen *et al.*, 1985); puppet characters have been used to explore deception (Sodian & Frith, 1992); short and simple stories were read aloud requiring the distinction between mental and physical entities (Baron-Cohen, 1989b). The finding that a minority of individuals with autism succeed on such tasks has led to the development of more advanced adult tests of theory of mind; aimed at avoiding possible ceiling effects and pro-

¹ MRC Child Psychiatry Unit, Institute of Psychiatry, London, England.

² South Warwickshire Combined Care NHS Health Service Trust,.

³ University of Cambridge, Cambridge, United Kingdom.

⁴ Address all correspondence to Lisa Heavey, MRC Child Psychiatry Unit, Institute of Psychiatry, De Crespigny Park, Denmark Hill, London, SE13 4AF; email: spjwljh@iop.kcl.ac.uk

viding more sensitive measures of mentalistic understanding. Indeed, tasks such as the Strange Stories Test (Happé, 1994b) and the Reading the Mind in the Eyes Test (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997) have revealed subtle mind-reading difficulties in high-functioning individuals with autism and Asperger syndrome, relative to IQ-matched controls. Autistic individuals who were able to pass standard first- and second-order ToM tests, showed impairments in providing mental state justifications for story characters' nonliteral utterances (Happé, 1994b) and in inferring a person's mental state from a photograph of their eye region (Baron-Cohen *et al.*, 1997). Such findings validate the search for milder manifestations of the basic cognitive impairment in mentalistic understanding.

Although only a limited number of advanced theory of mind tests are presently available, the reasons for developing and refining new measures are manifest. To understand the nature of the social deficit and its underlying cognitive mechanisms in able individuals with autism, further indices are needed; the study of "pure" autism in the absence of mental retardation providing a strong test of psychological theory. The finding that ToM deficits are not always apparent in such high-functioning individuals, particularly those with Asperger syndrome (Ozonoff, Rogers, & Pennington, 1991; Bowler, 1992), contrasts markedly with their observed difficulties in everyday social interaction. This underscores the importance of using appropriately sensitive tools to detect mild abnormalities. Advanced theory of mind tasks, with a "mature" and challenging content, are also more suitable for use with adults than the numerous measures developed for younger subject populations. The characterization of subtle mentalizing difficulties beyond childhood is vital if we are to understand the consequences of early ToM impairment and the sequelae of a developmental delay in acquiring mindreading skills (Baron-Cohen, 1989a).

A major impetus for developing such advanced cognitive measures stems from a growing interest in the broader phenotype of autism. Family and twin studies have revealed the strong genetic basis of the disorder (e.g., Bailey *et al.*, 1995; Bolton *et al.*, 1994) and shown the presence of a "lesser variant" among the close relatives of individuals with autism (e.g., see Bailey, Palferman, Heavey, & Le Couteur, 1998, for a review). The characterization of milder phenotypic expression is of theoretical interest; any specific cognitive deficit that plays a causal role in the disorder should also be present in a conceptually similar, but milder, form among the relatives of individuals with autism. Indeed, there is accumulating evidence of executive function

impairment among autism relatives (Hughes, Leboyer, & Bouvard, 1997; Hughes, Plumet, & Leboyer, in press; Ozonoff, Rogers, Farnham, & Pennington, 1993; Piven & Palmer, 1997) and preliminary findings of a fine-detailed processing style in autism parents (Baron-Cohen & Hammer, 1997; Happé, Briskman, & Frith, 1998). Findings are less consistent with regard to mentalizing ability in autism relatives (Baron-Cohen & Hammer, 1997; Ozonoff *et al.*, 1993), although these discrepant findings again highlight the importance of using mental age-appropriate measures. The accurate delineation of the broader phenotype has obvious utility within molecular genetic studies of autism; aiding the identification of affected relatives with the result of increased power for linkage studies.

In the study of subtle mind-reading deficits associated with autism and its broader phenotype, the use of *naturalistic* test formats and stimuli offer a potential means of increasing task sensitivity. Such measures, designed to approximate the demands of everyday social functioning, may succeed in tapping the problems experienced in processing mental states "on line"; thus revealing difficulties which remain undetected in most standard theory of mind tests. The present paper details one such task designed to provide a naturalistic measure of theory of mind competence. The methodology employed was relatively novel and innovative; social understanding was tested using dynamic, fast-paced stimuli, depicting real people in real-life contexts, requiring the processing of subtle, transient cues.

Adults with high-functioning autism/Asperger syndrome (autism group) and control participants were required to answer questions based on the social interaction of characters shown in various film clips. This was held to represent a measure of mental state understanding in that participants needed to infer actors' thoughts and feelings. Most of the individual films involved a character experiencing a socially uncomfortable and unpleasant moment, hence the title of "Awkward Moments Test." Questions on mental states were accompanied by control questions about the films which did not have a theory of mind/social component. Participants' understanding of film characters' intentions was also explored. Differences between the autism and control groups were predicted to be most apparent for the questions requiring mind-reading ability, reflecting the autism group's difficulty in processing mentalistic information. A shortened form of Happé's Strange Stories task was also administered to provide a separate index of the autism group's ToM competence. Convergent evidence of an impairment in attributing mental states in the autism group would help validate the Awkward

Moments Test as an advanced theory of mind measure (Baron-Cohen *et al.*, 1997).

Despite previous findings that suggest superior theory of mind ability in individuals with Asperger syndrome relative to those with high-functioning autism (Ozonoff *et al.*, 1991), the present experimental group was selected to include participants with either diagnosis. Although earlier studies may have failed to detect mentalizing difficulties among individuals with Asperger syndrome, Baron-Cohen *et al.* (1997) demonstrated how more advanced, naturalistic methods may be sensitive to the ToM difficulties of this particular diagnostic group. Thus, the inclusion of individuals with Asperger syndrome in the present study may not necessarily attenuate group differences on other similarly high-level measures. Also, as future applications of the Awkward Moments Test are planned to include the assessment of relatives with the broader autism phenotype, it was of interest to include a range of able individuals with autism spectrum diagnoses in the present study. Finally, although diagnostic boundaries remain to be firmly established between these disorders (Volkmar, Klin, & Cohen, 1997), individuals with Asperger syndrome, like those with high-functioning autism, exhibit marked social difficulties in everyday life; rendering them suitable for inclusion in a cognitive investigation of such problems.

METHOD

Participants

Sixteen individuals with high-functioning autism ($n = 14$) and Asperger syndrome ($n = 2$) were selected for participation in the present study. The gender ratio was 15:1 (M:F). All individuals had been reliably di-

agnosed according to accepted criteria (Rutter, 1978; DSM-III, DSM-IV, ICD-10). In nine cases, the Autism Diagnostic Interview (ADI; Le Couteur *et al.*, 1989) was used and for the remaining individuals, equivalent clinical interviews which predated the ADI had been administered. Participants were recruited through previous diagnostic assessment at the Maudsley Hospital (a tertiary referral center) and through participation in a Maudsley-based social skills training program for adults with autism and Asperger syndrome (some individuals having been referred from other diagnostic centers). The chronological age of the autism group ranged from 22 to 51 years ($M = 34.7$ years, $SD = 9.5$). A short-form version of the WAIS-R (Wechsler, 1981) was selected to assess intellectual function, comprising the vocabulary, comprehension, block design, and object assembly tests. Full-scale IQs were calculated using a regression equation (Crawford, Allan, & Jack, 1992). As participants were required to read material in the Awkward Moments Test, basic reading and reading comprehension ability was assessed using the Wechsler Objective Reading Dimensions (WORD). Mean WAIS-R scaled scores, full-scale IQ, and WORD standard scores are shown in Table I.

A comparison group of 15 men was selected to broadly match the autism group in terms of intellectual ability and reading competence. Individuals were recruited to the study through various sources including adult education centers, further education colleges, and job employment agencies. Those control participants whose IQ mapped the lower range in the autism group were contacted primarily through employment support schemes for individuals with mild learning difficulties. Care was taken to ensure that a diagnosis of Pervasive Developmental Disorder had never been suspected in these latter cases. The chronological age of the control

Table I. WAIS-R Scaled Scores, Full-Scale IQ, and WORD Standard Scores for Autism and Control Groups

	WAIS-R scaled scores				Full-scale IQ	WORD standard scores	
	Vocabulary	Comprehension	Block design	Object assembly		Basic reading	Reading comprehension
Autism ($n = 16$)							
<i>M</i>	7.88	7.19	9.38	7.88	89.63	103.38	83.81
<i>SD</i>	2.61	2.99	3.10	2.61	11.65	13.49	17.34
Range	5–13	2–14	4–14	3–11	67–110	66–114	58–112
Control ($n = 15$)							
<i>M</i>	8.67	9.40	10.07	8.07	95.53	95.87	98.33
<i>SD</i>	2.19	2.32	3.92	3.08	14.73	14.60	17.50
Range	5–12	7–15	4–19	2–12	74–115	70–114	76–118
<i>t</i> test (<i>p</i> value)	.37	.03	.85	.59	.22	.15	.03

group ranged from 22 to 45 years ($M = 30.7$, $SD = 8.1$), with the difference between groups in chronological age being nonsignificant, $F(1, 29) = 1.6$, ns. Mean WAIS-R scaled scores, full-scale IQ, and WORD standard scores are shown in Table I. Comparisons between groups for the WAIS-R vocabulary, block design, and object assembly scaled scores, full-scale IQ, and WORD basic reading standard scores were nonsignificant (see Table I). However, the WAIS-R comprehension scaled scores and WORD reading comprehension standard scores were significantly lower in the autism group (see Table I). This highlights the difficulty in individually matching such groups, given the “spiky” Wechsler intelligence test profiles of individuals with autism (e.g., Happé, 1994c) and the often reported discrepancy between their basic reading and reading comprehension ability (e.g., Frith & Snowling, 1983).

Materials and Design

Awkward Moments Test

Film Clips. The Awkward Moments Test comprised eight individual films which ranged from 45 to 120 seconds in duration. Seven of the films were British television commercials and one film clip was taken from a British television series. Synopses of the eight films are given in the Appendix. Television commercials were selected as appropriate stimuli because of their high quality technical production, short duration, complete storyline, and prior knowledge of the characters was not required. The first of the eight films was used as a practice item with the remaining seven films constituting the test stimuli. The test films were set in a variety of situations, including a shop, cinema, garden, and domestic/home settings, both in Britain and international locations. The films featured characters of different ages, from teenagers to elderly individuals, and actors were depicted in various roles and relationships; such as friends, neighbors, lovers, and work colleagues. The test clips were related, however, in that they showed a character experiencing a socially awkward or unpleasant moment at some point during the film. The films were shown to participants on a 80486 notebook PC. They were captured from videotape to an animated video (AVI) format and played back using Windows MCI (multimedia interface).

Computer-Based Questions and Answers. Each film was followed by two questions displayed on the computer screen; the test question which pertained to the mental state of a central character and the control question which related either to a directly observable visual feature within the film or to information con-

tained within the dialog. The test questions all referred explicitly to a character’s feelings (e.g., “At the end of the film, how did the young man feel?”). To answer such questions accurately, participants were required to process more than direct visual cues such as facial expressions. They needed to appreciate (a) the character’s beliefs about a social situation (sometimes a false belief) and (b) the social significance of the character’s actions, that is, what the other characters in the film thought about the main character’s behavior. The control questions were unrelated to the social/mentalistic content of the film and thus served as general indices of attentional processing and memory for information. As the test questions explored participants’ mentalistic understanding, and the individuals with autism were predicted to experience the most difficulties on these particular items relative to the control participants, the test question always preceded the control question after each film in order to minimize the demands on memory for this question type. The full set of test and control questions are given in the Appendix. The test questions differed according to their time point of reference within the film. On four occasions, participants were asked how a character felt at the end of the film and the remaining test questions required the participant to report the character’s mental state at a specific moment occurring approximately half-way through the clip. Similarly, the control questions varied in terms of their reference to specific time points in each test film; five of the control questions referred to information in the first half of the film and two control questions related to dialog and events occurring towards the end of the film. Answers to the test and control questions were presented on screen in a multiple-choice format requiring the participant to select one response from four alternatives. Answers were in the form of single words, with the exception of three double/hyphenated words (two occurring as alternative responses to control questions and one constituting the correct response to a test question). Word frequency was controlled as closely as possible with the correct answer being comparable in word frequency with at least one of the alternatives within each group of four answers. For each test question, the four possible answers always included the correct mental state term (e.g., embarrassed, anxious), an incorrect/inadequate mental state term (e.g., confused, amazed) a simple emotional term (e.g., happy, afraid), and a physical state term (e.g., hot, thirsty). Mental states were held to require an appreciation of the thought processes and beliefs of the character in relation to the social context. The simple emotions were regarded as more “transparent” in being identifiable

from the characters' overt facial expressions and physical actions. To identify such emotions, an appreciation of the characters' thoughts in relation to external events was not necessary. The answers available for each control question comprised the correct response together with three alternatives varying in the degree to which they were semantically related to the target answer and the extent to which they were contextually appropriate to the film. The full set of correct answers to both sets of questions, together with the distracter words, are given in the Appendix.

Presentation Format of Films, Questions and Answers. After each film was shown without interruption, the test question was displayed on the screen, above a box marked "OK." When the question had been read, the participant was required to press a specific button on the computer keyboard, the OK key, which instructed the program to move on to the next screen display showing the four response alternatives. A cursor arrow which also appeared in the center of the screen equidistant from the four possible answers, was moved by the participant using a built-in mouse (directly next to the OK button) to one of the four alternative words and the answer was selected by pressing the OK key. The control question was next displayed on the screen until the participant indicated they have read the question by pressing the OK button. Selection of one of the four control answers was made by the participant moving the cursor arrow over to the selected response and again pressing the OK key. The program provided no feedback to the participant about the accuracy of their answers. The test films and questions could be presented in one of two orders: forward (film numbers 2 to 8) or reverse (numbers 8 to 2). In addition to registering the choice of answers, the Awkward Moments Test program recorded the time taken for participants to select an answer to each question. Responses were timed from when the four alternative answers first appeared on the screen to when the participant had pressed the OK key to make their selection. Time taken to read the questions was not recorded. The position of the correct response within each array of four answers was counterbalanced across all eight films (including the practice film) for each question type. Thus, the correct answer appeared twice in each of the four positions for both the test and control questions.

Questions on Characters' Intentions. Further to the computer-based mental state questions, participants' understanding of the film actors' intentions was explored in the form of a "mini-interview" after each film. A series of hierarchically structured questions was devised, with the questions for each film conforming to a general format. The experimenter first asked a nonspe-

cific probe ("What was the main character trying to do in this film?") and progressed to more specific directive prompts according to the quality of the participant's answer (e.g., "Was he trying to do anything more than this?"; "What did he want the other people to think about him?"). Thus, the questions proceeded from being neutral in containing no reference to mental states, to being explicit in their requirement for a mentalistic answer. For two films (6 and 7), the questions concerned the film director's intentions rather than those of the actors. These items required the participant to consider the viewer's expectations about the film (e.g., "What was the film director trying to do when he/she made this film? What is the viewer supposed to think is going to happen?"). A general scoring system was devised which could be applied across all of the different films. This system incorporated (a) the quality of the participant's answer to the experimenter's initial question; (b) the quality of their best answer, and (c) the number of prompts needed to obtain their best response. The quality of participants' initial and best responses was rated as follows: a score of 2 for a full, mental state-based answer; a score of 1 for an incomplete answer, and a score of 0 for a literal, physical state-based answer. The number of prompts needed for participants to obtain their best response was rated as follows: a score of 2 for responding correctly to the initial question, a score of 1 for requiring an additional prompt, a score of 0 when two or more prompts were needed. Thus, a maximum composite score of 6 would be awarded for a full answer in response to the experimenter's first question (i.e., "What was the main character trying to do in that film?"). As the films differed considerably in their content and storyline and therefore in terms of the characters' intentions, specific scoring criteria were derived from this general scoring system for each film. Participants' responses to the intention questions were rated by both the experimenter and a corater, blind to the identity and diagnosis of individuals. Agreement between raters for participants' total scores on the intention questions was examined (Intraclass correlation of .99), indicating a high level of interrater reliability.

Strange Stories Task

Happé's (1994b) original task was shortened to include eight stories: two control/physical passages and two stories relating to an understanding of irony, double bluff, and white lies. The physical stories constitute control items because they do not require an inference about characters' mental states. The remaining stories were considered to involve a theory of mind

component in requiring participants to infer a story character's intentions behind their indirect speech. From Happé's original findings, these specific stories were among the most difficult items for individuals with autism and thus were selected as appropriate for the present group of able adults. Each story together with the respective questions was printed on a separate A4 sheet and combined in a quasi-randomized order (so no two stories of the same type were presented in succession). In line with the current scoring procedure for this task (Happé, Winner & Brownell, 1998) participants' answers to the justification questions were scored with 2 for a full mental state answer, 1 for a partially correct response, and 0 for an incorrect answer.

Procedure

The Awkward Moments Test and the Strange Stories task were administered as part of a wider battery of cognitive tests, with all participants receiving the measures in the same order (Heavey, Phillips, & Rutter, 1999). The Awkward Moments Test was presented before the Strange Stories task. Participants were tested individually in a quiet room either at the Maudsley Hospital, at their education/employment center, or in their home.

Awkward Moments Test

The experimenter introduced the task by saying the following:

On this screen you will see some short pieces of film showing people talking to each other in social situations. Different people notice different things about social situations. This test is to see which things you notice. After each film, you will see a question on the screen about the feelings of one of the characters.

The format of presentation of the questions and the mode of response were then outlined to the participant, with the experimenter demonstrating the use of the mouse controls and the OK key. Instructions were given to read the questions carefully and to consider all response alternatives before making their selection. Participants were told that the first film was a practice item to familiarize them with the task. After viewing the film and answering the questions, the participants were given the choice of completing this item a second time.

Before the first of the test films was presented, participants were informed they would be asked additional questions at the end of each film. They were told:

People often have hidden personal reasons for what they do. After each of the films, I am going to ask you

about the story characters' private reasons for behaving in the way that they did.

The order of presentation of test films was counterbalanced so that equal numbers of participants in each group saw the films in forward (numbers 2 to 8) and reverse (numbers 8 to 2) order. After the first test film had been shown and the computer-based questions answered, the experimenter reiterated that the following (verbal) questions would concern the story character's personal reasons for their behavior in the film. After the intention questions, participants were asked whether they had seen the film before. Yes/no answers were recorded to provide an index of the familiarity of the film. Participants were then given the opportunity for a short pause before viewing the next test film.

Strange Stories Task

Participants were presented with the collection of eight stories and required to read each passage carefully before completing the respective questions in writing.

RESULTS

As the differences between groups on the WAIS-R comprehension subtest and the WORD reading comprehension measure were significant, these were specified as covariates within all of the following analyses. Also, in view of the diagnostic heterogeneity of the autism group and previous findings of superior theory of mind performance of individuals with Asperger syndrome relative to high-functioning autism, analyses were performed in which the two participants with Asperger syndrome were excluded. As the pattern of findings was consistent with those obtained for the full autism group, the following analyses include the two individuals with Asperger syndrome.

Awkward Moments Test

Number of Correct Answers

The total number of questions answered correctly by participants across all seven test films was analyzed using ANCOVA with one between-subjects factor of group (autism vs. controls) and one within-subjects factor of question type (test vs. control). This revealed a significant main effect of group, $F(1, 27) = 18.30, p < .001$, with the control group showing superior performance to the autism group ($M = 5.97$ and 4.54 , respectively), a significant main effect of question type, $F(1, 29) = 90.47, p < .001$, with the control questions being easier than test questions ($M = 6.29$ and 4.17 , respectively),

and a significant interaction between these two factors, $F(1, 29) = 10.17, p < .01$. Group means are given in Table II. Planned simple effects analyses revealed that the difference between groups for the test questions, $F(1, 27) = 24.60, p < .001$, partial $\eta^2 = .50$, was greater than the difference between groups for the control questions, $F(1, 27) = 11.93, p < .01$, partial $\eta^2 = .32$.

Response Times

ANCOVA of response times revealed a significant main effect of question type, $F(1, 29) = 35.71, p < .001$, with both groups responding more slowly on the test than control questions ($M = 11.91$ and 6.87 seconds, respectively), no significant main effect of group, $F(1, 27) = 2.02, ns$, and no significant interaction, $F(1, 29) = 0.04, ns$. Group mean response times are given in Table II.

Intention Question Scores

Participants' total scores for the intention questions were entered into a one-way ANCOVA. This revealed a significant main effect of group, $F(1, 27) = 34.01, p < .001$, with the control group showing better performance than the autism group on this measure ($M = 22.60$ and 8.31 , respectively).

Familiarity of Films and Test Performance

To examine whether the familiarity of the film related to the participants' accuracy in answering the test (theory of mind) questions, chi-square analyses, using Fisher's Exact Test, were performed for both groups, for each of the seven test films. In view of the number of analyses performed, α was adjusted to 1%. All of the calculations were nonsignificant, which suggested no consistent association between familiarity and accuracy for the theory of mind questions in either group.

Theory of Mind Performance and IQ

Although the two groups were selected to be of comparable intellectual ability, the IQs of the autism group spanned a marginally lower intellectual range than that of the control group (full-scale IQs of 67–110 and 74–115, respectively). Thus, to further deal with intellectual differences between groups, analyses were performed in which the two lowest functioning autism participants and two most intellectually able controls were excluded. The modified range in full-scale IQs of the two groups was 77–110 (autism group) and 74–110 (control group). For the number of correct responses to the computerized Awkward Moments Test questions, with Wechsler comprehension test scaled scores and WORD reading comprehension standard scores specified as covariates, the main effect of group, $F(1, 23) = 17.21, p < .001$; the main effect of question type, $F(1, 25) = 124.16, p < .001$, and the group by question type interaction, $F(1, 25) = 13.86, p < .01$, were significant. The group difference for the total scores on the intention questions was significant, $F(1, 23) = 34.35, p < .001$, as was the difference between groups for the Strange Stories task, $F(1, 22) = 10.52, p < .01$.

The relationship between participants' performance on the specific theory of mind measures (i.e., total number of test questions answered correctly, response times for the test questions, total scores for the intention questions) and IQ was further examined using correlational analyses (see Table III). These showed a relationship between intellectual ability and theory of mind performance in the control group but not in the autism group. In view of previous studies noting the relationship between verbal ability and theory of mind competence in autism (e.g., Happé, 1995; Sparrevohn & Howie, 1995), the nonsignificant correlations between performance on the theory of mind measures and

Table II. Performance on Awkward Moments Test by Autism and Control Groups

	No. of correct responses (max = 7)		Response time (sec)		Total scores for intention questions (max = 42)
	Test questions	Control questions	Test questions	Control questions	
Autism ($n = 16$)					
<i>M</i>	3.13	5.94	13.61	8.40	8.31
<i>SD</i>	1.20	0.77	6.61	3.07	3.74
Range	1–5	5–7	4.92–27.93	4.41–16.07	2–14
Control ($n = 15$)					
<i>M</i>	5.27	6.67	10.09	5.23	22.60
<i>SD</i>	1.22	0.49	7.51	3.31	7.41
Range	3–7	6–7	4.43–34.33	2.19–15.22	9–37

Table III. Correlations Between Awkward Moments Test theory of Mind Measures and IQ Measures

	WAIS-R scaled scores				
	Vocabulary	Comprehension	Block design	Object assembly	Full-scale IQ
Autism (<i>n</i> = 16)					
No. of correct test responses	-.23	.09	.08	.20	.05
Response times for test questions	-.13	-.08	.28	.44	.18
Intention scores	.24	.42	-.09	.27	.29
Control (<i>n</i> = 15)					
No. of correct test responses	.70 ^b	.61 ^a	.34	.60 ^a	.63 ^a
Response times for test questions	-.29	-.06	-.27	-.26	-.27
Intention scores	.78 ^b	.61 ^a	.58 ^a	.57 ^a	.74 ^b

^a *p* < .05.^b *p* < .01.

the WAIS-R verbal subtests in the autism group are of particular interest. These findings should be treated with caution, however, given the small samples and the number of correlations calculated.

Correlations between Awkward Moment Test performance (theory of mind measures) and IQ (Wechsler individual subtests and full-scale IQ) were compared between the two groups (Cramer, 1994). The correlations between vocabulary and number of correct test responses differed significantly between groups ($z = 2.75, p < .05$) as did the correlations between vocabulary and total intention scores ($z = 2.00, p < .05$). All other intergroup comparisons of correlations were nonsignificant.

Strange Stories Task

Participants total scores for the test stories requiring mental state understanding (double bluff, irony, white lie) were entered into a one-way ANCOVA, with WAIS-R comprehension scaled scores and WORD reading comprehension standard scores specified as covariates. This revealed a significant difference between groups, $F(1, 29) = 9.52, p = .005$, with the autism group showing impaired performance relative to controls (M scores = 6.20, $SD = 2.37$, and 9.53, $SD = 1.73$, respectively). The two groups did not differ significantly in their ability to answer the physical, control stories ($Z = -0.98, ns$).

The relationship between participants' performance on the two TOM tests was examined. For the autism group, the correlations between performance on the Strange Stories task and the Awkward Moments Test were as follows: number of test questions answered correctly ($r = .48, ns$), response times for test questions ($r = .03, ns$) and total score on the intention questions ($r = .16, ns$). Within the control group, the following corre-

lations between Awkward Moments Test performance and Strange Stories task performance were obtained: number of test questions answered correctly ($r = .60, p < .05$), response time for test questions ($r = -.132, ns$) and total score for the intention questions ($r = .46, ns$).

DISCUSSION

The results support the general experimental prediction: the theory of mind measures within the Awkward Moments Test revealed group differences. The autism group answered fewer of the mental state-based questions correctly and were deficient in their ability to explain film characters' intentions. These findings were consistent with the results of the Strange Stories task: the autism group was impaired in providing mental state explanations for story characters' nonliteral speech when compared to the control group. This pattern of findings provides some support for the utility of the methodology employed within the Awkward Moments Test. By approximating the demands of the real social world, naturalistic stimuli in the form of film excerpts may represent an appropriate format for revealing subtle mind-reading difficulties.

A note of caution is necessary, however, in interpreting these preliminary findings. Although differences between groups were most apparent on the theory of mind measures, the autism group was poorer than the control group at answering the control, nonsocial questions about the films. This group difference was largely attributable to the autism participants' difficulty with two specific control items (Films 5 and 7), both of which referred to background visual information extraneous to the plot. This suggestion of decreased vigilance in the

autism group may reflect more generalized difficulties in attentional processing (e.g., Courchesne *et al.*, 1994), which may in part underlie their problems on the theory of mind test questions. Certainly, because of their length, pace, and complexity, the films do demand close attention and monitoring. However, the autism groups' success on the remaining control questions, which were more closely related to the storyline of the films, does not support an explanation of their impaired task performance in terms of general attentional factors. Thus, the Awkward Moments Test should still be regarded as tapping a more specific facet of cognitive processing; that is, the ability to infer characters' thoughts and feelings. It is possible that the difference between groups for the control questions reflects the focus of attentional resources on the more social, interpersonal elements of the film for the individuals with autism, particularly as the task was emphasized as being a measure of social understanding. For the controls, the processing of social information may prove less effortful, with additional cognitive resources available for dealing with more incidental details.

In the context of other psychological theories of autism, it is necessary to consider further the degree to which the Awkward Moments Test represents a "pure" theory of mind task. Appreciation of the films' content requires the retention of information from various scenes, the selection of relevant social cues and the integration of information across scenes in order to derive meaning; requirements that imply the involvement of executive function (e.g., Hughes, Russell, & Robbins, 1994) and central coherence (e.g., Happé, 1994a). However, this problem of "contamination" is inherent in using such naturalistic measures; real-life social understanding makes demands on executive processes and contextual/global processing. The solution to refining and strengthening film clip measures as theory of mind tests may lie not in the revision of test stimuli but in the modification of control questions. This requires the formulation of inferential questions involving the retention and integration of information within the film but lack the social/mentalistic content of the test questions. Thus, continued difficulties in answering the test compared to the control questions would be directly attributable to the theory of mind component rather than the executive and central coherence demands of the task.

A number of further experimental findings from the Awkward Moments Test merit discussion. Although the autism group showed impaired performance on both the Awkward Moments Test and the Strange Stories task, the association between participants' performance on the two measures was not significant. This may, in

part, reflect differences between the tasks in terms of presentation formats and response demands. Also, in the absence of true mentalistic understanding in the autism group, this lack of association between tasks may derive from the application of different alternative processing strategies in solving the relevant questions. For the present study, which employs only two tasks and a relatively small sample, such interpretations remain speculative. However, the development of an increasing number of advanced ToM tasks offers the opportunity to further compare both autism and control participants' performance across various measures, with the observed relationships between tasks having implications for our conceptualization of theory of mind as a multifaceted construct.

Regarding response times obtained on the Awkward Moments Test, the two groups did not differ in their speed of response to the computer-based questions after each film, although there was a trend for the autism group to be slower across both types of question. As the task instructions did not refer to the timing of responses or the need to respond as quickly as possible, little can be made of these findings. Nevertheless, there is no evidence of impulsivity of response or a speed-error trade-off in the autism group. Although not a particularly informative measure within the present study, response time may prove a useful index of mentalizing ability in future studies, if accompanied by appropriate task instructions. For example, in family studies of autism, mildly affected individuals may be at ceiling in terms of their accuracy in identifying characters' mental states, yet their underlying difficulties in making such inferences may be revealed by slower response times (e.g., Bowler, 1992).

The most discriminative measure of performance on the Awkward Moments Test proved to be the intention questions; the autism group were impaired in providing explanations of characters' intentions and motives within the films. Intentional understanding was explored by requiring participants to verbally generate their answers; a response format that in itself may place demands on the pragmatic/theory of mind ability and verbal skills of the participant (Happé, 1994c). However, it is worth noting that a perfect answer could be as little as two words in length (e.g., the character was trying to "show off"); participants were not required to give lengthy and complicated descriptions. Indeed, some individuals in the autism group provided quite complex, inventive, and eloquent explanations of characters' behavior which unfortunately lacked any relevant mention of the characters' true motives.

It is of interest that theory of mind performance and IQ (both full-scale and individual subtest performance)

were found to be related in the control group but not in the autism group. Thus, control participants' test performance seemed to be mediated, at least in part, by intellectual ability, whereas for the autism group, such intellectual factors did not appear to confer an advantage in answering the mental state-based questions. However, on closer inspection of intergroup comparisons of correlation coefficients, the key difference appeared to be on a measure of verbal ability: Wechsler vocabulary scores. The absence of an association between this verbal measure and Awkward Moments Test scores in the autism group contrasts with previous reports of a relationship between verbal ability and performance on standard first and second-order theory of mind tests in autism (e.g., Happé, 1995), although Baron-Cohen *et al.* (1997) reported no significant correlation between IQ and performance on the Eyes Test with able autistic adults. It is worth considering whether the discrepancy between the present findings and those of most previous studies may relate to the IQ measures selected. Certainly, several previous studies in which such a relationship has been found employed the full Wechsler intelligence test battery (Bauminger & Kasari, 1999; Yirmiya, Solomonica-Levi, Shulman, & Pilowsky, 1996) rather than the short-form version. However, in the present study the IQ–ToM relationship is examined not only in relation to full-scale IQ but also in terms of individual subtest performance. This is of interest, specifically for verbal subtests, given that some studies have employed vocabulary tests as the primary measure of verbal IQ and observed a significant verbal IQ–ToM relationship (Happé, 1995; Sparrevohn & Howie, 1995). Clearly, differences between various studies in terms of task format, response requirements, characteristics of participant groups, and measures of verbal ability need to be examined to further elucidate the nature of this relationship.

To conclude, the Awkward Moments Test represents a potentially useful methodology for exploring the subtle mentalizing difficulties associated with autism. Although the present findings have implications for the modification of task design, such a general approach involving naturalistic stimuli and varied, graded response measures may prove especially informative in helping to characterize and dimensionalize the broader autism phenotype.

APPENDIX

Details of Films

Synopsis of plot; test and control questions and answers; model answers for intention questions

Practice Film (1)

An attractive woman knocks on the door of her new neighbor's apartment. A man opens the door and she asks to borrow some coffee. He invites her inside and they exchange pleasantries.

There is subtle interplay between characters indicating mutual attraction.

Test Question. At the end of the film, how did the man feel?

Answer. Interested (Distracters: bored, happy, thirsty)

Control Question. When the man opened the door, what did the woman ask for?

Answer. Coffee (Distracters: cat food, milk, coffee, sugar)

Film 2

The film is set in the British embassy in Moscow. The main character is a government agent who, in front of an audience of colleagues, accuses the British ambassador of criminal dealings. However, evidence is presented which proves that the ambassador is innocent and the main character realizes he has made a mistake.

Test Question. At the end of the film, how did the main character feel?

Answer. Horrified (Distracters: hot, unhappy, amazed)

Control Question. Who was hanging upside down?

Answer. Ambassador (Distracters: secretary, cowboy, policeman)

Model Answer for Intention Question. The main character is trying to impress everyone with his skilful handling of the situation.

Film 3

The film is set in a high-rise apartment. A man knocks on the door of his new neighbor's apartment. The door is opened by an attractive woman. He asks to borrow some tea. She invites him inside and introduces her pet dog. After she has left the room, the man throws a ball for the dog to fetch but accidentally throws the ball out of the window. The dog jumps out of the window to chase the ball. The film ends by showing that the dog is unhurt.

Test Question. When the dog jumped out of the window, how did the man feel?

Answer. Embarrassed (Distracters: thirsty, surprised, sad)

Control Question. When the woman answered the door, what did the man ask for?

Answer. Tea (Distracters: lawnmower, beer, telephone)

Model Answer for Intention Question. The man was attracted to the woman and wanted to get to know her.

Film 4

Two neighbors are talking over the garden fence. The older man draws attention to strange marks in the younger man's garden and says that they have been made by aliens. The young man knows that the old man made the patterns in his garden and is trying to play a trick on him. The young man calls the old man's bluff by threatening to call the police to investigate the aliens.

Test Question. At the end of the film, how did the young man feel?

Answer. Smug (Distracters: confused, restless, afraid)

Control Question. When the young man said he was making a phonecall, who did he say he would phone?

Answer. Police (Distracters: mother, gardener, ambulance)

Model Answer to Intention Question. The young man was calling the older man's bluff, he was playing a trick on him.

Film 5

A young man wakes up in his parents' house. He had a party the night before and must tidy the house before his parents return from their holiday. He notices a scratch on a table and arranges for a craftsman to repair this. Just when he thinks that he has cleared away all evidence of the party, he sees that someone has defaced a portrait on the wall.

Test Question. When they boy saw the picture on the wall, how did he feel?

Answer. Shocked (Distracters: sleepy, disappointed, unhappy)

Control Question. At the beginning of the film, how many people were in the house?

Answer. Two (Distracters: none, five, seven)

Model Answer to Intention Question. The boy was trying to hide from his parents the fact that he had a party.

Film 6

A boyfriend and girlfriend sat together in a cinema. The boyfriend leaves the room to visit the bathroom. He returns to sit down in the cinema and puts his arm around the person sitting next to him. He turns to see that he has his arm around another man, rather than

his girlfriend. He has sat in the wrong seat. The film playing at the cinema has a soundtrack which creates a feeling of suspense.

Test Question. When the boyfriend saw who he was sitting next to, how did he feel?

Answer. Embarrassed (Distracters: afraid, lost, puzzled)

Control. What kind of film was showing at the cinema?

Answer. Horror (Distracters: comedy, musical, western)

Model Answer to Intention Question. The director was trying to make the viewer think that something bad was going to happen in the cinema.

Film 7

A man and a woman are talking about their gardener, noting that he is getting too old to be doing such heavy work. They call him over to speak to him. They present him with a new lawnmower.

Test Question. When the woman called the gardener over, how did he feel?

Answer. Anxious (Distracters: happy, weary, interested)

Control Question. When the gardener arrived at the house, what was he pushing?

Answer. Bicycle (Distracters: broom, grassroller, trolley)

Model Answer to Intention Question. The director was trying to make the viewer think that the gardener would lose his job.

Film 8

The setting is a North African country. The main character is in a small shop, trying to barter with the shopkeeper to buy a rug. He realizes that he is speaking to the shopkeeper in the wrong language. He leaves the shop and is pleased with his purchase and his handling of the situation but he does not realize that the rug has caught fire.

Test Question. At the end of the film, how did the older man feel?

Answer. Self-satisfied (Distracters: pleased, worried, hot)

Control Question. At the end of the film, what was the older man carrying?

Answer. Rug (Distracters: vase, beach-ball, towel)

Model Answer for Intention Question. The older man was trying to impress the other people with his knowledge of the language and local customs.

REFERENCES

- American Psychiatric Association. (1980). *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed.). Washington, DC: Author.
- American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: Author.
- Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., & Rutter, M. (1995). Autism as a strongly genetic disorder: Evidence from a British twin study. *Psychological Medicine*, 25, 63–77.
- Bailey, A., Palferman, S., Heavey, L., & Le Couteur, A. (1998). Autism: the phenotype in relatives. *Journal of Autism and Developmental Disorders*, 28, 369–392.
- Baron-Cohen, S. (1989a). The autistic child's theory of mind: A case of specific developmental delay? *Journal of Child Psychology and Psychiatry*, 30, 285–298.
- Baron-Cohen, S. (1989b). Are autistic children behaviorists? An examination of their mental-physical and appearance-reality distinctions. *Journal of Autism and Developmental Disorders*, 19, 579–600.
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and the theory of mind*. Cambridge, MA: MIT Press/Bradford Books.
- Baron-Cohen, S., & Hammer, J. (1997). Parents of children with Asperger syndrome: What is the cognitive phenotype? *Journal of Cognitive Neuroscience*, 9, 548–554.
- 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 of Asperger Syndrome. *Journal of Child Psychology and Psychiatry*, 38, 813–822.
- Bauminger, N., & Kasari, C. (1999). Brief report: Theory of mind in high-functioning children with autism. *Journal of Autism and Developmental Disorders*, 29, 1, 81–86.
- Bolton, P., MacDonald, H., Pickles, A., Rios, P., Goode, S., Crowson, M., Bailey, A., & Rutter, M. (1994). A case-control family history study of autism. *Journal of Child Psychology and Psychiatry*, 35, 877–900.
- Bowler, D. (1992). "Theory of mind" in Asperger's syndrome. *Journal of Child Psychology and Psychiatry*, 33, 877–893.
- Courchesne, E., Townsend, J., Akshoomoff, N.A., Saitoh, O., Yeung-Courchesne, R., Lincoln, A.J., James, H.E., Haas, R.H., Schreibman, L., & Lau, L. (1994). Impairment in shifting attention in autistic and cerebellar patients. *Behavioral Neuroscience*, 108, 848–865.
- Cramer, D. (1994). *Introductory statistics for social research: Step-by-step calculations and computer techniques using SPSS*. Routledge.
- Crawford, J., Allan, K., & Jack, A. (1992). Short forms of the UK WAIS-R: Regression equations and their predictive validity in a general population sample. *British Journal of Clinical Psychology*, 31, 191–202.
- Frith, U. (1989). *Autism: Explaining the enigma*. Oxford, UK: Basil Blackwell.
- Frith, U., & Snowling, M. (1983). Reading for meaning and reading for sound in autistic and dyslexic children. *Journal of Developmental Psychology*, 1, 329–342.
- Happé, F. (1994a). *Autism: An introduction to psychological theory*. UCL Press.
- Happé, F. (1994b). 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. (1994c). Wechsler IQ profile and theory of mind in autism: a research note. *Journal of Child Psychology and Psychiatry*, 35(8):1461–1471.
- Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, 66, 843–855.
- Happé, F., Briskman, J., & Frith, U. (1998). Exploring the cognitive phenotype of autism: Weak central coherence in parents and siblings of children with autism. Paper presented at BPS Developmental Section Conference, Lancaster.
- Happé, F., Winner, E., & Brownell, H. (1998). The getting of wisdom: Theory of mind in old age. *Developmental Psychology*, 34, 358–362.
- Heavey, L., Phillips, W., & Rutter, M. (1999). Theory of mind, executive function and central coherence in high-functioning autism. Manuscript in preparation.
- Hughes, C., Russell, J., & Robbins, T. (1994). Evidence for executive dysfunction in autism. *Neuropsychologia*, 32, 477–492.
- Hughes, C., Leboyer, M., & Bouvard, M. (1997). Executive function in parents of children with autism. *Psychological Medicine*, 27, 209–220.
- Hughes, C., Plumet, M-H, & Leboyer, M. (1999). Towards a cognitive phenotype for autism: siblings show strong span and weak executive function. *Journal of Child Psychology and Psychiatry*, 40(5), 705–718.
- Le Couteur, A., Rutter, M., Lord, C., Rios, P., Robertson, S., Holdgrafer, M., & McLennan, J. (1989). Autism Diagnostic Interview: A standardized investigator-based instrument. *Journal of Autism and Developmental Disorders*, 19, 363–387.
- Leslie, A., & Frith, U. (1988). Autistic children's understanding of seeing, knowing and believing. *British Journal of Developmental Psychology*, 6, 315–324.
- Ozonoff, S., Rogers, S. J., & Pennington, B. F. (1991). Asperger's syndrome: Evidence of an empirical distinction from high-functioning autism. *Journal of Child Psychology and Psychiatry*, 32, 1107–1122.
- Ozonoff, S., Rogers, S., Farnham, J., & Pennington, B. (1993). Can standard measures identify subclinical markers of autism? *Journal of Autism and Developmental Disorders*, 23, 429–441.
- Piven, J., & Palmer, P. (1997). Cognitive deficits in parents from multiple-incidence autism families. *Journal of Child Psychology and Psychiatry*, 38, 1011–1022.
- Rutter, M. (1978). Diagnosis and definition. In M. Rutter & E. Schopler (Eds), *Autism: A reappraisal of concepts and treatment*. New York Plenum Press.
- Sodian, B., & Frith, U. (1992). Deception and Sabotage in autistic, retarded and normal children. *Journal of Child Psychology and Psychiatry*, 33(3) 591–605.
- Sparrevoorn, R., & Howie, P. (1995). Theory of mind in children with autistic disorder: Evidence of developmental progression and the role of verbal ability. *Journal of Child Psychology and Psychiatry*, 36, 249–263.
- Tager-Flusberg, H. (1992). Autistic children's talk about psychological states: deficits in the early acquisition of a theory of mind. *Child Development*, 63, 161–172.
- Volkmar, F. R., Klin, A., & Cohen, D. (1997). Diagnosis and classification of autism and related conditions: consensus and issues. In D. J. Cohen & F. R. Volkmar (Eds), *Handbook of autism and pervasive developmental disorders* (2nd ed.). Wiley.
- Yirmiya, N., Solomonica-Levi, D., Shulman, C., & Pilowsky, T. (1996). Theory of mind abilities in individuals with autism, Down Syndrome, and mental retardation of unknown etiology: The role of age and intelligence. *Journal of Child Psychology and Psychiatry*, 37, 1003–1014.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scales-Revised*. The Psychology Corporation: New York.
- World Health Organization. (1992). *The ICD-10 classification of mental and behavioral disorders: Clinical descriptions and diagnostic guidelines*. Geneva: Author.