Journal of Autism and Developmental Disorders, Vol. 32, No. 3, June 2002 (© 2002)

Reading the Mind in the Voice: A Study with Normal Adults and Adults with Asperger Syndrome and High Functioning Autism

M. D. Rutherford,[1,2] Simon Baron-Cohen,[1] and Sally Wheelwright[1]

People with high functioning autism (HFA) and Asperger syndrome (AS) have deficits in theory of mind (ToM). Traditional ToM tasks are not sensitive enough to measure ToM deficits in adults, so more subtle ToM tests are needed. One adult level test, the Reading the Mind in the Eyes test has shown that AS and HFA subjects have measurable deficits in the ability to make ToM inferences. Here we introduce a test that extends the above task into the auditory domain and that can be used with adults with IQ Scores in the normal range. We report the use of the test with an adult sample of people with AS/HFA and with two adult control groups. Results suggest that individuals with AS/HFA have difficulty extracting mental state information from vocalizations. These results are consistent with previous results suggesting that people with HFA and AS have difficulties drawing ToM inferences.

KEY WORDS: Autism; theory of mind; Asperger syndrome; auditory.

INTRODUCTION

People with autism have a selective theory of mind (ToM) deficit. They have difficulty in inferring the mental states of others, as measured by "false belief" tasks (Baron-Cohen, Leslie, & Frith, 1985) and a variety of other tasks (Baron-Cohen, 1995; Happé, 1994; Perner, Leslie, Leekam, & Frith, 1989). People with the related condition of Asperger syndrome (AS) also have a ToM deficit (Baron-Cohen *et al.*, 1997; Happé *et al.*, 1996). AS is a condition that is similar to autism, sharing social deficits, and most authors consider AS part of the autistic continuum (Wing, 1988).

Traditional ToM tests, which are designed for young children, are not subtle enough to detect deficits in adults of normal intelligence. Adults with AS may pass traditional ToM tests, but passing may not indicate normal ToM function. Perhaps adults with AS solve these tasks using mental processes other than typical ToM processes, having developed compensatory techniques. As a group, these adults tend to pass both first-order false belief tasks and more subtle second-order false belief tasks (Bowler, 1992; Ozonoff, Pennington, & Rogers, 1991; Ozonoff, Rogers, & Pennington, 1991), so it has been difficult to accurately measure the social deficits of this group. More subtle tests of mindreading abilities are needed.

There are a few adult-level social cognitive tests. One early adult-level test revealing ToM deficits in autism is the Strange Situation test (Happé, 1994), in which subjects provide an explanation for an ambiguous action in a short story. The Faux Pas task, in which subjects identify a faux pas in a short story, has been used to measure ToM in stroke and head trauma patients and in AS populations (Baron-Cohen, O'Riordan, Stone, Jones, & Plaisted, 1999; Stone, Baron-Cohen, & Knight, 1998). Another approach has been to make

[1] Departments of Experimental Psychology and Psychiatry, Autism Research Centre, University of Cambridge, Downing Street, Cambridge, CB2 3EB, U.K.

[2] Correspondence should be addressed to M. D. Rutherford, Department of Psychology, 2155 S. Race St., Denver, Colorado 80208; e-mail: mrutherf@du.edu

0162-3257/02/0600-0189/0 © 2002 Plenum Publishing Corporation

false belief tasks more sensitive by creating multiple embeddings (Kinderman, Dunbar, & Bentall, 1998; Rutherford, 2001) or by recording reaction time (Rutherford, 2001). Cartoons in which the humor relies on the appreciation of a false belief have been used to measure ToM sensitivity in patients with schizophrenia (Corcoran, Cahill, & Frith, 1997), and acquired brain damage (Happé, Brownell, & Winner, 1998) and in normal adults (Rutherford, 2001).

The adult ToM task most relevant to the current study is the Reading the Mind in the Eyes task in its original (Baron-Cohen *et al.*, 1997), revised (Baron-Cohen *et al.*, 2001), and child (Baron-Cohen *et al.*, in press) version. In each of a series of trials, subjects are presented with a photograph of the eye region of the face and must choose one of two adjectives or phrases to describe the mental state of the person pictured. Subjects are asked, "Which word best describes what the person in the photo is thinking or feeling?" In control trials, subjects are asked to judge the person's sex. On both the adult and child tests, AS/HFA groups show significant deficits, despite normal intelligence. The measure presented in the current paper extends this test by expanding it into the auditory domain.

In the Reading the Mind in the Voice test (The Voice Test), subjects hear a brief sample of dialogue on audiotape and then choose between two adjectives to best describe the mental state of the speaker. In control trials, subjects choose between two age ranges to best describe the speaker's age. The Voice Test requires a subject to use cues in vocalizations to infer someone's mental state. Of interest is whether the performance of the HFA/AS group on experimental trials, relative to performance on the control trials, distinguishes them from two neurologically normal adult groups.

Past research suggests that children and adolescents with autism may have deficits in perceiving mood or emotion based on vocal cues. They have shown deficits when asked to match vocal segments to videos of faces (Loveland *et al.*, 1995), to match vocal segments to photographs of faces (Hobson *et al.*, 1989), and nonverbal vocalizations to line drawings of body postures (Hobson, 1986a) or to line drawings of facial expressions (Hobson, 1986b). One study, however, failed to show a deficit in naming nonverbal vocalized emotions, relative to control task performance (Hobson *et al.*, 1989), and another study showed no effect for autism diagnosis on ability to identify emotions in video segments that included spoken words (Loveland *et al.*, 1997). The current study is intended as a test of whether an adult AS/HFA group has difficulty extracting mental state information from vocalizations.

METHOD

Participants

The experimental participants were 19 adults with HFA or AS (17 men and 2 women). They had all been referred to specialist centers and were diagnosed in these centers using established APA (1994) and *ICD-10* (1994) diagnostic criteria (the Autism Diagnostic Interview-Revised [Lord, Rutter, & LeCouteur, 1994] and the Autism Diagnostic Observation Schedule-Generic [Lord *et al.*, 1989]). In addition, all were screened for the current study using the Autism Spectrum Quotient (AQ) (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) and scored above the cutoff of 32 (maximum = 50). The mean age of the HFA/AS group was 29 years (median, 23 years; range, 16–59 years), and the group had an average WAIS verbal IQ of 101, (see Table I for details). These participants did not have any other psychologic diagnoses.

The participants in the first control group consisted of 78 adults (38 men and 40 women) recruited on the University of Cambridge campus, in residential colleges and departmental buildings. The participants in the second control group consisted of 20 neurologically normal adults (17 men and 3 women) who were neither university graduates nor university students. They were recruited via postings around the university and newspaper advertisements (see Table I for age and IQ information). Participants in all three groups were paid for their time.

Materials

The answer sheet had two types of side-by-side comparison. For experimental trials, the answer sheet had two adjectives to describe the speaker's mental attitude or emotion. For control trials, the answer sheet had two options to represent the person's age (e.g., "over 42" or "under 42"). These two types alternated on the answer sheet.

The audio stimuli were composed of segments of dialogue taken from audiocassette tapes of dramatic performances. Each speech segment was either a sentence or a phrase. Each speech segment lasted for approximately 2 seconds with a 3-second pause between speech segments, during which participants marked their choice. Table II shows each speech segment and the target and foil adjective for each.

Table I. Chronological Age and WAIS-R IQ

		WAIS		
	Chronological age	Verbal IQ	Performance IQ	Overall IQ
Autism/Asperger syndrome (n = 19)				
M	29	101	116	107.93
SD	14.52	10.5	17.2	12.69
Range	16–59	79–117	80–143	87–133
Typically developing adults (n = 20)				
M	36	103	98	101
SD	10.68	11.165	9.3	7.34
Range	18–53	85–120	80–112	91–116

Construction of the Task

First, 50 segments of dialogue were recorded from dramatic audio books. Two possible mentalistic adjectives per segment were proposed by the authors. Next, four independent judges were asked to judge between the two options as they listened to the audiotape. During this phase of construction, judges could pause the audiotape between segments and could repeat segments. Those items on which the judges were unanimous were kept.

Finally, a new set of four independent judges took the test in real time, with a 3-second pause between segments. Items were included in the test if at least three of these judges agreed with the previous set of judges. Ten items were excluded, leaving 40 segments of speech.

The control task was developed following a similar procedure. The same recording of speech segments was used, and a forced-choice age judgment was made. Again, four independent judges were given the questionnaire and asked to judge between the two options as they listened to the audiotape. If all four judges agreed on the response, then the item was kept. Finally, four new judges took the test in real time, and items were only included in the test if at least three of these judges agreed with the previous set of judges.

Procedure

Participants were tested individually in their homes or in a quiet room at the university. Immediately before the test, participants were shown the answer sheets and asked whether there were any words that they did not understand. Participants were given the opportunity to indicate which, if any, words they did not know. (No participant asked for words to be clarified.)

Answer sheets alternated between a ToM judgment and an age judgment. The trials were ordered such that participants would hear and make a judgment on each speech segment and then start again from the beginning, making age judgments for those segments on which they had made mentalistic judgments before, and vice versa. The audiotape played the entire series twice, and participants circled one of the two choices. According to experimenter observation, 3 seconds was sufficient time for every subject; no one fell behind. Participants were allowed to adjust the volume as necessary before and during the test. The test was administered in conjunction with a battery of other tasks. The Voice Test took about 11 minutes, and the entire battery of tests took about 90 minutes.

RESULTS

The primary prediction was that HFA/AS participants would perform worse on experimental trials than would control participants (compared with each group's performance on control trials). An analysis of variance revealed that there was a significant interaction between group and trial type [$F(1, 114) = 3.75$, $p = .027$] (see Table III). A planned linear contrast of the difference scores (differences between the mentalistic and control trials) showed that the AS/HFA group was different from the two control groups on the ToM trials [$t(114) = 2.779$, $p = .006$, two-tailed].[1]

[1] A one-way ANOVA on just the difference scores was also significant. [$F(116) = 3.860$, $p = .02$].

Table II. Target Mental State Terms, Foil Terms, and Percent Correct for Each Group

Target (correct)	Foil (incorrect)	Spoken phrase	AS/HFA	College control	Noncollege control
1. Earnest	Alarmed	"No, honestly I do."	81	89	75
2. Friendly	Grateful	"Collie said you were up here."	74	99	100
3. Confused	Angry	"Your brother? I don't remember you ever speaking of a brother."	84	94	100
4. Suspicious	Intrigued	"Where did you get them?"	79	99	100
5. Worried	Insulted	"Please! We must go."	79	95	90
6. Concerned	Pleased	"Does he know she found them?"	95	99	100
7. Hateful	Irritated	"I think your message is quite clear, Mr. Lathimer."	53	78	55
8. Apologetic	Hurried	"I'm afraid he's gone out, sir."	74	92	85
9. Pleading	Horrified	"I swear I have."	89	85	95
10. Perplexed	Accusatory	"What on earth do you mean?"	84	90	85
11. Nervous	Assured	"There's . . . uh . . . there's something I want to ask you."	68	84	90
12. Irritated	Surprised	"Keep the damn thing!"	100	99	100
13. Surprised	Sarcastic	"This is quite fun, Debra."	89	99	90
14. Joyous	Scared	"What a pair!"	74	94	100
15. Embarrassed	Furious	"I haven't, doctor, but I will."	89	92	100
16. Terrified	Angry	"Oh, my god!"	74	87	85
17. Enraged	Scared	"Why should I? Why should any of us?"	84	89	95
18. Disappointed	Apologetic	"But I had hoped . . ."	89	91	90
19. Reassuring	Amused	"It's bound to take a little time."	100	96	100
20. Sincere	Impatient	"I really am most grateful."	79	91	100
21. Melancholy	Resolved	"I've no idea what she thought of me."	79	94	90
22. Suspicious	Scared	"You . . . you suspect someone?"	95	99	100
23. Resentful	Hesitant	"It was taken for granted."	89	94	90
24. Concerned	Relaxed	"So, where're you off to now?"	100	92	95
25. Sincere	Menacing	"I won't harm him, I promise you."	63	95	95
26. Tentative	Impatient	"You may find this part rather strange."	100	100	100
27. Derogatory	Angry	"I think she was trying to make some sort of . . . gesture."	84	84	85
28. Stern	Curious	"But I rather think that we have a few things to discuss."	84	92	100
29. Sarcastic	Indifferent	"You seem to have done very well."	79	92	80
30. Defensive	Joking	"Yeah, well, I know nothing about that."	100	99	95
31. Regretful	Apologetic	"I can't help wondering if I was wrong about her."	68	94	85
32. Insulted	Disappointed	"What sort of people do you think we are?"	89	96	95
33. Resigned	Irritated	"Life must go on, Mr. Wilson."	100	95	80
34. Reflective	Worried	"We weren't likely to forget."	63	81	75
35. Frightened	Contemplative	"What does all this mean?"	47	81	55
36. Desperate	Flirtatious	"Will you come away with me, for a week . . .?"	74	84	90
37. Hopeful	Irritated	"Katherine, perhaps you'd come to help."	95	87	95
38. Annoyed	Apologetic	"Yes, of course, Vector dear . . . I'll just . . ."	11	90	75
39. Relieved	Understanding	"I don't want a half-hour like that again."	100	92	95
40. Urgent	Exhausted	"To the letter!"	84	94	95

Because the majority of participants in the AS/HFA group were male, we compared the men of this group with the men of the two control groups. An analysis of variance revealed a significant interaction between group and trial type when just men were included [$F(1, 69) = 4.791$, $p = .01$] (Table III). Again, a planned linear contrast on the difference scores showed that the AS/HFA group was different from the two control groups on the ToM trials [$t(69) = 3.12$, $p = .003$, two-tailed.][2]

There was no significant difference in performance on the voice trials or significant interaction between sex

[2] A one-way ANOVA on just the difference scores was also significant. [$F(71) = 4.923$, $p = .01$].

Table III. Number of Items Correct on Experimental and Control Trials (of 40)

	Mentalistic judgment			Control judgment			
Condition	*M*	*SD*	Range	*M*	*SD*	Range	Mean difference
Asperger syndrome	32.53	4.59	20–38	30.42	4.13	22–37	2.11
College control	36.81	2.16	31–40	32.29	2.71	26–37	4.51
Noncollege control	36.3	2.31	31–40	31.0	3.36	25–36	5.3

and trial type within the college control group. In this group, men on average scored 36.76 correct on experimental trials (31.82 on control trials) compared with women who scored an average of 36.85 on experimental trials (32.75 on control trials). The other two groups had too few women to reliably test for sex differences.

Finally, one might ask whether there was a correlation between verbal IQ and performance on the voice trials, consistent with the literature. In fact, there was no significant correlation between verbal IQ and performance on the voice task for either the AS/HFA group or the noncollege control group.

DISCUSSION

There are three major conclusions suggested by this study. First, the Voice Test distinguishes a group of people with HFA and AS from a control group. Second, these data are in agreement with past research that suggests a specific ToM deficit in people with AS and autism. Third, these results suggest that the ToM deficit is amodal.

The Voice Test may help to discriminate between HFA/AS and control participants. Note, however, that because there is overlap in the ranges of the two groups, the test in its current form is not diagnostic on its own. It may be used in a battery of diagnostic tests, or it may be distilled until there is a definite criterion point.

The deficit in social perception is likely not unique to autism. The Voice Test may prove useful for testing head trauma or stroke patients with subtle ToM deficits or for measuring ToM deficits in people with diseases of the amygdala (Adolphs *et al.*, 1998; Stone, 1999) or orbitofrontal cortex (Stone, 1999; Stone *et al.*, 1998). The test may also be used to characterize the broader phenotype of autism. It has been suggested that parents and grandparents of probands may show some subtle, hard-to-measure social deficits (Bailey *et al.*, 1995; Baron-Cohen and Hammer, 1997).

Care should be used when assessing individuals with brain trauma, however, because other cognitive skills are required to complete the task. The Voice Test would be affected by deficits in attention, audition, or language, for example. The test in its current form would not be useful for all neurologic cases, although the control condition was designed in part to control for possible deficits in these areas. As mentioned earlier, the participants in this study had no comorbid diagnoses.

The results of this study are in agreement with previous work that suggests that people with HFA/AS have a specific deficit making social inferences. The Voice Test is novel because it was designed to measure ToM deficits in the auditory domain. This difference in performance is not easily explained by differences in vocabulary, because HFA/AS participants had verbal IQs matched with the noncollege control group and they were asked to review the words before the tests.

It may be possible to improve the Voice Test by adding more items and including only the most diagnostic items. Item analysis showed that the groups performed significantly differently on several individual items. Therefore, it may be possible to develop an extremely sensitive and effective voice test by distilling the test.

The Voice Test might be improved in other ways as well. First, one could increase the sensitivity of the measure by increasing the number of alternatives available on the answer sheet (paralleling a recent improvement in the Reading the Mind in the Eyes test [Baron-Cohen *et al.*, 2001]). Second, one could increase the sensitivity of the test by measuring reaction time as well as accuracy. Third, concerns about different verbal ability might be ameliorated by ensuring that the test words are common words, requiring only a lower vocabulary level. Fourth, participants could perform a forced choice sentence completion task to ensure minimum vocabulary levels.

ACKNOWLEDGMENTS

This study was supported by the Fulbright Commission and NIH postdoctoral training grant 2 T32 MH15780-20 (M.D.R.) and by the MRC and the McDonnell-James Foundation (S.B.C. and S.W.). Participant payment was funded by a grant from Trinity College Cambridge to Valerie Stone. We are grateful to Valerie Stone for useful discussions of this work.

REFERENCES

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature, 393,* 470–474.

Bailey, T., 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.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind.* Cambridge, MA: MIT Press.

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., & Hammer, J. (1997). Parents of children with Asperger syndrome: What is the cognitive phenotype? *Journal of Cognitive Neuroscience, 9,* 548–554.

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

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,* 407–418.

Baron-Cohen, S., Wheelwright, S., & Hill, J. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42,* 241–252.

Baron-Cohen, S., & Wheelwright, S. (In press). The Autism-Spectrum Quotient: A new instrument for screening for Asperger syndrome and high-functioning autism in adults of normal intelligence. *Journal of Autism and Developmental Disorders.*

Baron-Cohen, S., Wheelwright, S., Scahill, V., Spong, A., & Lawson, J., (2001). Are intuitive physics and intuitive psychology independent? A test with children with Asperger syndrome. *Journal of Developmental and Learning Disorder, 5,* 47–78.

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

Bowler, D. M. (1992). Theory of mind in Asperger syndrome. *Journal of Child Psychology and Psychiatry, 33,* 877–893.

Corcoran, R., Cahill, C., & Frith, C. D. (1997). The appreciation of visual jokes in people with schizophrenia: A study of 'mentalizing' ability. *Schizophrenia Research, 24,* 319–327.

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thought and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism & Developmental Disorders, 24*(n2), 129–154.

Happé F., Brownell, H., & Winner, E. (1999). *Acquired 'theory of mind' impairments following stroke, 70,* 211–240.

Happé, F., Ehlers, S., Fletcher, P., Frith, U., *et al.* (1996). 'Theory of mind' in the brain: Evidence from a PET scan study of Asperger syndrome. *NeuroReport: An International Journal for the Rapid Communication of Research in Neuroscience, 8,* 197–201.

Hobson, R. P. (1986a). The autistic child's appraisal of expressions of emotion. *Journal of Child Psychology and Psychiatry, 27,* 321–342.

Hobson, R. P. (1986b). The autistic child's appraisal of expressions of emotion: A further study. *Journal of Child Psychology and Psychiatry, 27,* 671–680.

Hobson, R. P., Ouston, J., & Lee, A. (1989). Emotion recognition in autism: Coordinating faces and voices. *Psychological Medicine, 18,* 911–923.

Kinderman, P., Dunbar, R. I. M., & Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology, 89,* 191–204.

Lord, C., Rutter, M., Goode, S., & Heemsbergen, J. (1989). Autism Diagnostic Observation Schedule: A standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders, 19,* 185–212.

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

Loveland, K., Tunali-Kotoski, B., Chen, R., Brelsford, K., & Ortegon, J. (1995). Intermodal perception of affect by persons with Autism or Down syndrome. *Development and Psychopathology, 7,* 409–418.

Loveland, K., Tunali-Kotoski, B., Chen, R., Brelsford, K., & Ortegon, J. (1997). Emotion recognition in autism: Verbal and nonverbal information. *Development and Psychopathology, 9,* 579–593.

Ozonoff, S., Pennington, B., & Rogers, S. (1991). Executive function deficits in high-functioning autistic children: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry, 32,* 1081–1106.

Ozonoff, S., Rogers, S., & Pennington, B. (1991). Asperger's syndrome: Evidence of an empirical distinction from high-functioning autism. *Journal of Child Psychology and Psychiatry, 32,* 1107–1122.

Perner, J., Leslie, A. M., Leekam, S. R., & Frith, U. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development, 60,* 689–700.

Rutherford, M. D. (April 2001). An experimentally-induced dissociation of the theory of mind mechanism. Presented at the Bi-Annual Meetings of the Society for Research in Child Development, Minneapolis, MN.

Stone, V. E. (1999). The role of the frontal lobes and the amygdala in theory of mind. In S. Baron-Cohen, D. Cohen, & H. Tager-Flusgerg (Eds.), *Understanding other minds: Perspectives from autism and developmental cognitive neuroscience,* Oxford: Oxford University Press.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10,* 640–656.

Wing, L. (1988). The continuum of autistic characteristics. In E. Schopler & G. Mesibov (Eds.) *Diagnosis and assessment in autism* (pp. 91–110). New York: Plenum Press.

World Health Organization (1994). *The ICD-10 classification of mental and behavioral disorders: Diagnostic criteria for research.* Geneva: World Health Organization.