J. Child Psychol. Psychiat. Vol. 42, No. 2, pp. 241–251, 2001
Cambridge University Press
© 2001 Association for Child Psychology and Psychiatry
Printed in Great Britain. All rights reserved
0021–9630/01 $15.00 + 0.00

The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism

Simon Baron-Cohen, Sally Wheelwright, Jacqueline Hill, Yogini Raste, and Ian Plumb

University of Cambridge, U.K.

In 1997 in this *Journal* we published the "Reading the Mind in the Eyes" Test, as a measure of adult "mentalising". Whilst that test succeeded in discriminating a group of adults with Asperger syndrome (AS) or high-functioning autism (HFA) from controls, it suffered from several psychometric problems. In this paper these limitations are rectified by revising the test. The Revised Eyes Test was administered to a group of adults with AS or HFA ($N = 15$) and again discriminated these from a large number of normal controls ($N = 239$) drawn from different samples. In both the clinical and control groups the Eyes Test was inversely correlated with the Autism Spectrum Quotient (the AQ), a measure of autistic traits in adults of normal intelligence. The Revised Eyes Test has improved power to detect subtle individual differences in social sensitivity.

Keywords: Theory of mind, Asperger's Disorder, autistic disorder, social cognition.

Abbreviations: AQ: Autism Spectrum Quotient; AS: Asperger syndrome; HFA: high-functioning autism; TS: Tourette's syndrome.

A challenge for psychology is to develop tests that are sensitive to subtle cognitive dysfunction. This is particularly important in the domain of social cognition. There is a wealth of basic-level social cognitive tests for use with young children (Flavell, Green, & Flavell, 1986; Flavell, Shipstead, & Croft, 1978; Wellman, 1990; Wimmer & Perner, 1983). However, there are few tests that can measure if an adult with normal intelligence may have a mild deficit in social understanding. Researchers in adult neuropsychology have developed some tests of face perception (Young, Hellawell, De Wal, & Johnson, 1996), but often in the case of patients with acquired brain damage the deficits are gross rather than subtle. In contrast, in neurodevelopmental conditions such as autism or Asperger syndrome, deficits in social cognition may persist across the lifespan. These may, however, be camouflaged as a result of learning compensatory strategies. Without a subtle and sensitive test, the investigator might erroneously conclude that the patient is "recovered" or "normal".

In our first efforts towards developing an adult test of social sensitivity, we described the "Reading the Mind in the Eyes" Test (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997). In this test, the participant is presented with a series of 25 photographs of the eye-region of the face of different actors and actresses, and is asked to choose which of two words best describes what the person in the photograph is thinking or feeling. This test was conceived of as a test of how well the participant can put themselves into the mind of the other person, and "tune in" to their mental state. For this reason, we described it as an "advanced theory of mind test". "Theory of mind" is shorthand for the ability to attribute mental states to oneself or another person (Premack & Woodruff, 1978), and this ability is the main way in which we make sense of or predict another person's behaviour. Theory of mind is also referred to as "mentalising" (Morton, Frith & Leslie, 1991), "mind reading" (Whiten, 1991), and "social intelligence" (Baron-Cohen, Jolliffe, et al., 1999), and overlaps with the term "empathy". Examples from the first version of the test are shown in Figs. 1 and 2.

A task analysis of the Eyes Test might include the following: The subject needs to have a mental state lexicon and know the semantics of these terms. The Eyes Test then involves mapping these terms to fragments of facial expressions of mental states—just the part of the face around the eyes. At a reportedly unconscious, rapid, and automatic level, subjects must match the eyes in each picture to examples of eye-region expressions stored in memory and seen in the context of particular mental states to arrive at a judgement of which word the eyes most closely match. Note that the Eyes Test is described as an advanced test of theory of mind but in fact only involves the first stage of attribution of theory of mind: attribution of the relevant mental state (e.g. *compassion*). It does not include the second stage: inferring the *content* of that mental state (e.g. compassion for her mother's loss). However, attribution of the *type* of mental state is nevertheless part of theory of mind, even if it is not all of it.

The results of this test showed that adult males in the general population scored a mean of 18.8 ($SD = 2.5$) whereas women scored slightly but significantly higher, with a mean of 21.8 ($SD = 1.8$). Adults with high-functioning autism (HFA) or Asperger Syndrome (AS) performed significantly worse than sex-matched normal

Requests for reprints to Simon Baron-Cohen, Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, U.K.
(E-mail: SB205@cus.cam.ac.uk).

Figure 1. An example of a (male) stimulus used: in the first version word choices were serious (correct) vs. playful. In the revised version the word choices were serious (correct), ashamed, alarmed, and bewildered.

Figure 2. A second (female) example from the Eyes Test: in the first version the word choice was reflective (correct) vs. unreflective. In the revised version the word choice was reflective (correct), aghast, irritated, and impatient.

controls, or adults with a Tourette's syndrome (TS) (a different psychiatric condition, and included as an additional control group). Thus, the adults with HFA or AS scored on average 16.3 out of 25 ($SD = 2.9$), whereas the adults with TS scored on average 20.4 out of 25 ($SD = 2.6$). Although this was only a 4-point difference, it was significant at the $p < .01$ level. The group with TS did not differ significantly on this test from the general population.

Thus, we had succeeded in developing a test of social sensitivity or mind-reading that was able to reveal subtle mind-reading difficulties in adults with HFA or AS. This had been predicted on the basis of more basic mind-reading deficits in younger children with autism (Baron-Cohen, 1995). This was also of interest because it demonstrated that normal adults could judge mental states from even minimal cues (expressions around the eyes alone). Having established that the ability to "read the mind in the eyes" was testable, we considered in what ways the test could be improved.

Problems with the Original Version of the Test

(1) The first version of the task involved a forced choice between only two response options (the two words presented), so chance performance on each trial is $p = .5$. Across the test as a whole one would therefore need to score 17 or above out of 25 to be significantly above chance (Binomial Test). This meant that the range of scores in which the test can reveal individual differences whilst still being above chance is only 9 points (17–25). This is too narrow. Ideally, a test such as this would have a wider range, in order to be able to identify individual differences with greater power.

(2) When the first version of the test was given to parents of children with AS, they too scored below the general population level (Baron-Cohen & Hammer, 1997). This had been predicted on the basis that they might have the "broader phenotype" (Bailey et al., 1995), since one or both of such parents might be carrying the genes for autism. However, parents scored at a similar level to people with HFA or AS (fathers scoring on average 17.3 out of 25 ($SD = 1.6$), and mothers scoring a mean of 18.9 ($SD = 2.1$), even though they did not have the condition themselves. This highlights that the test has too narrow a range of scores to be able to distinguish between someone with the "lesser variant"/"broader phenotype" (e.g., in a first-degree relative of someone with autism), and someone with the condition itself.

(3) The narrow range of scores that are significantly above chance on the first test can lead to a score in the normal range being close to the ceiling of the test. Ceiling effects are obviously undesirable because one loses power to detect individual differences.

There are two simple modifications we can make to the test to remedy these three limitations: increase the number of items in the test, and increase the number of response options on each trial. In the revised version of the test reported in this paper, we have made both of these modifications: the total number of items (photographs) is increased from 25 to 36, and the number of response options (forced-choice words) is increased from 2 to 4 per trial. This means that chance is $p = .25$ per trial, and that

one only needs to score 13 or above, out of 36, to be performing significantly above chance (Binomial Test). In effect, this provides a bigger window of 24 points (from 13–36) in which to be able to reveal individual differences in ability on this test. It also decreases the risk of normal performance approaching the ceiling of the test.

(4) The first version of the test included both basic and complex mental states, and so contained some items that were too easy, and which therefore risked producing ceiling effects. Basic emotions are happy, sad, angry, afraid, and disgust. They are basic because they are recognised universally; because they can be recognised purely as emotions, without the need to attribute a belief to the person; and because they are recognised even by very young normally developing children (Ekman & Friesen, 1971, Harris, 1991; Walker, 1982). Complex mental states in contrast involve attribution of a belief or intention—a cognitive mental state—to the person. In the revised version of the test we limited the items to complex mental states so as to make the task that much more challenging, and in this way increasing the likelihood of obtaining a greater range of performance in a random sample of adults.

(5) In the original version, there were some items that could be solved simply by checking the gaze direction of the face. The words for such items were "noticing" or "ignoring", etc., (mental states linked to perception), such that gaze-direction might be all that a participant needed to attend to in order to arrive at the correct answer. This could be too easy a clue for someone with a subtle mind-reading difficulty. These are therefore excluded in the revised version of the test.

(6) The original version had more female faces than male faces, and it was unclear if this may have biased the test in some way. In the revised version of the test, this was carefully controlled by having an equal number of male and female faces in the photographs. The advantage of this was that it allowed a control condition—judging gender from the eyes—to be closely matched to the experimental condition—judging mental states from the eyes.

(7) In the original version of the test the target word and its foil were always semantic opposites (e.g., concerned vs. unconcerned, or sympathetic vs. unsympathetic), again making the test too easy. The test essentially was asking the participant to distinguish chalk from cheese, or black from white—in this case, asking them to distinguish between mental states of opposite emotional valence (positive vs. negative). In the revised version of the test we have again increased the level of difficulty by ensuring that as far as possible the three foil words have the same emotional valence as the target word. For example, if the target word was "serious", the foil words might be "ashamed", "alarmed", and "bewildered". This effectively means that a person has to distinguish the correct target word from three close imposters, on each trial. As such, we are testing the ability to distinguish shades of gray, or different types of cheese, as it were, so as to add to the challenging nature of the test, thereby maximising the possibility of revealing subtle individual differences. Figures 1 and 2 show two examples of pictures taken from the original test but with the new choice of four words with each.

(8) Finally, given that the Eyes Test involves mapping a word to a picture, it is unclear if comprehension problems with the words themselves might have contributed to an individual's score. This is particularly a concern with a group of patients with HFA in whom there will have been language delay. In the revised version of this test, we rectified this problem by including a glossary of all the mental state terms, which subjects were encouraged to consult in any case where they were unsure of a word.

The study below reports data from the revised version of this test, and had several additional aims. (1) To test a group of adults with AS or HFA on the revised version of the test. This was in order to check if the deficit in this group of patients that had been found on the original version (Baron-Cohen, Jolliffe, et al., 1997) and related tests (Baron-Cohen, Wheelwright, & Jolliffe, 1997) could be replicated. (2) To test if in a sample of normal adults, an inverse correlation would be found between performance on the Eyes Test (Revised) and the Autism Spectrum Quotient (AQ) (Baron-Cohen & Wheelwright, in press). The AQ measures the degree to which any individual (adult) of normal IQ possesses traits related to the autistic "spectrum" (Wing, 1988). The AQ is a self-report questionnaire. Scores range from 0–50, and the higher the score, the more autistic traits a person possesses. (3) To test if the sex difference (female superiority) found on the first version of the test (Baron-Cohen, Jolliffe, et al., 1997) replicated.

Method

Subjects

Table 1 shows the four groups of subjects tested.

Group 1 comprised adults with AS or HFA ($N = 15$, all male). They were recruited via adverts in the U.K. National Autistic Society magazine, or equivalent support groups. They had all been diagnosed in specialist centres using established criteria (American Psychiatric Association, 1994; World Health Organisation, 1994). They spanned an equivalent range of socioeconomic classes and educational levels as seen in Group 2. They were all given the short WAIS-R (Wechsler, 1939) comprising the Block Design, Vocabulary, Similarities, and Picture Completion, and all scored in the normal range (mean $= 115$, $SD = 16.1$).

Group 2 comprised normal adults ($N = 122$) drawn from adult community and education classes in Exeter, or from public library users in Cambridge. They had a broad mix of daytime occupations ranging from unemployment through manual and clerical workers, to professionals. They also had a broad mix of educational level, some having no education beyond secondary school, others having either occupationally related training, or college degrees. Data on age was available for $N =$ 88 of these.

Group 3 comprised normal adult students ($N = 103$, 53 male, 50 female) all studying for undergraduate degrees in Cambridge University (71 in science, 32 in other subjects). Since this university has very stringent entrance requirements (typically three grade As at Advanced Level [school leaving] examination), this group is not representative of the general population and they can be assumed to have high IQ.

Group 4 comprised randomly selected individuals in the general population ($N = 14$) who were IQ matched with Group 1 (mean $= 116$, $SD = 6.4$). Groups 1 and 4 did not differ significantly or IQ, or on age. See Table 1.

Procedure

Subjects in all four groups were tested on the revised adult Eyes Test, as described earlier. This was individually administered in a quiet room in Cambridge or Exeter. Subjects in the AS/HFA group were also asked to judge the gender of each

Table 1
Subject Characteristics

		Chronological age		IQ	
	N	Mean	*SD*	Mean	*SD*
Group 1					
AS/HFA adults	15	29.7	14.5	115	16.1
Group 2					
General population controls[a]	88	46.5	16.9	—	—
Group 3					
Students	103	20.8	0.8	—	—
Group 4					
IQ-matched controls	14	28.0	9.0	116	6.4

[a] $N = 122$ for Eyes Test.

Table 2
Percentage of Subjects in Groups 2 and 3 Combined, Who Chose Each Word on Each Item

Item	Target	Foil 1	Foil 2	Foil 3
1	31.6	1.8	26.2	40.4
2	53.1	4.0	5.8	37.1
3	78.7	4.9	12.0	4.4
4	82.1	5.4	4.9	7.6
5	84.9	4.0	2.2	8.9
6	79.6	1.3	8.0	11.1
7	79.9	7.6	10.3	2.2
8	79.5	3.6	13.8	3.1
9	72.9	6.7	14.7	5.8
10	74.7	12.9	8.9	3.6
11	83.6	4.9	8.9	2.7
12	48.4	34.7	13.3	3.6
13	68.4	20.4	8.4	2.7
14	73.8	3.1	12.0	11.1
15	85.8	6.7	5.3	2.2
16	72.9	7.1	4.0	16.0
17	86.7	6.2	5.3	1.8
18	76.0	1.8	13.3	8.9
19	79.6	9.3	4.0	7.1
20	63.4	18.8	16.1	1.8
21	68.3	10.3	4.5	17.0
22	64.4	10.2	17.3	8.0
23	88.0	5.3	6.7	0.0
24	77.3	12.4	8.9	1.3
25	84.9	1.3	3.6	10.2
26	80.9	0.4	4.0	14.7
27	75.6	8.0	4.0	12.4
28	64.9	5.8	21.8	7.6
29	72.9	2.7	4.9	19.6
30	64.4	1.8	21.8	12.0
31	65.8	4.9	22.2	7.1
32	71.9	16.5	0.9	10.7
33	90.2	2.2	4.4	3.1
34	52.0	16.4	11.6	20.0
35	60.4	10.2	23.6	5.8
36	65.8	6.7	23.1	4.4
37	79.1	0.9	16.4	3.6
38	73.3	10.7	8.9	7.1
39	81.3	0.9	2.2	15.6
40	60.0	3.1	26.7	10.2

person in each photo, as a control task, given anticipated impairments on mental state recognition. Normal adults were found to be at ceiling on the gender recognition task during piloting so, to save time, were not required to do this task. In addition, subjects in Groups 1, 3, and 4 completed the AQ (Baron-Cohen & Wheelwright, in press). Finally, subjects were asked at the outset to read through the glossary (see Appendix B) and indicate any word meanings they were unsure of. They were then encouraged to read these particular meanings and were told that they could return to this glossary at any point during the testing.

Eyes Test Development

Target words and foils were generated by the first two authors and were then piloted on groups of eight judges (four male, four female). The criterion adopted was that at least five out of eight judges agreed that the target word was the most suitable description for each stimulus and that no more than two judges picked any single foil. Items that failed to meet this criterion had new target words, foils, or both generated and were then re-piloted with successive groups of judges until the criterion was met for all items.

The data from Groups 2 and 3 did not differ from each other, so the results were combined, creating a sample of $N = 225$. Table 2 shows the results of an item analysis on this combined group. New criteria were applied to these data: at least 50% of subjects had to select the target word and no more than 25% could select any one of the foils. These criteria were arbitrarily selected but with the aim of checking that a clear majority of the normal controls selected the target word and that this was selected at least twice as often as any foil. Items 1, 2, 12 and 40 failed to meet these criteria and were therefore dropped. Subsequent analyses were carried out using the 36 items. Thus target words were established on the basis of consensus from a large population, since there is no objective method for identifying the underlying mental state from an expression. The complete list of target mental state words (in italic) and their foils are shown in Appendix A. The glossary of mental state terms is shown in Appendix B.

Predictions

Based on the previous studies we predicted that:

(1) The AS/HFA group would score significantly lower on the mental state judgements on the Eyes Test, but be unimpaired on the gender control judgements.
(2) The AS/HFA group would score significantly higher in the AQ.
(3) Females in the "normal" groups (2 and 3) would score higher than males on the Eyes Test.
(4) Males in the "normal" group (3) would score higher than females on the AQ.
(5) Scores on the AQ and the Eyes Test would be inversely correlated.

Results

Subjects in the four groups did not differ in the number of words in the glossary that they were unsure of, and in all subjects, the number of words checked never exceeded

Table 3
Performance on the Revised Eyes Test and AQ

		Eyes Test		AQ	
	N	Mean	*SD*	Mean	*SD*
Group 1					
AS/HFA adults					
All	15	21.9	6.6	34.4[a]	6.0
Group 2					
General population controls					
All	122	26.2	3.6	—	—
Males	55	26.0	4.2	—	—
Females	67	26.4	3.2	—	—
Group 3					
Students					
All	103	28.0	3.5	18.3[b]	6.6
Males	53	27.3	3.7	19.5[c]	6.7
Females	50	28.6	3.2	16.6[d]	6.1
Group 4					
IQ matched controls					
All	14	30.9	3.0	18.9	2.9

[a] $N = 14$, due to 1 unreturned AQ.
[b] $N = 79$, due to 24 unreturned AQs.
[c] $N = 47$, due to 6 unreturned AQs.
[d] $N = 32$, due to 18 unreturned AQs.

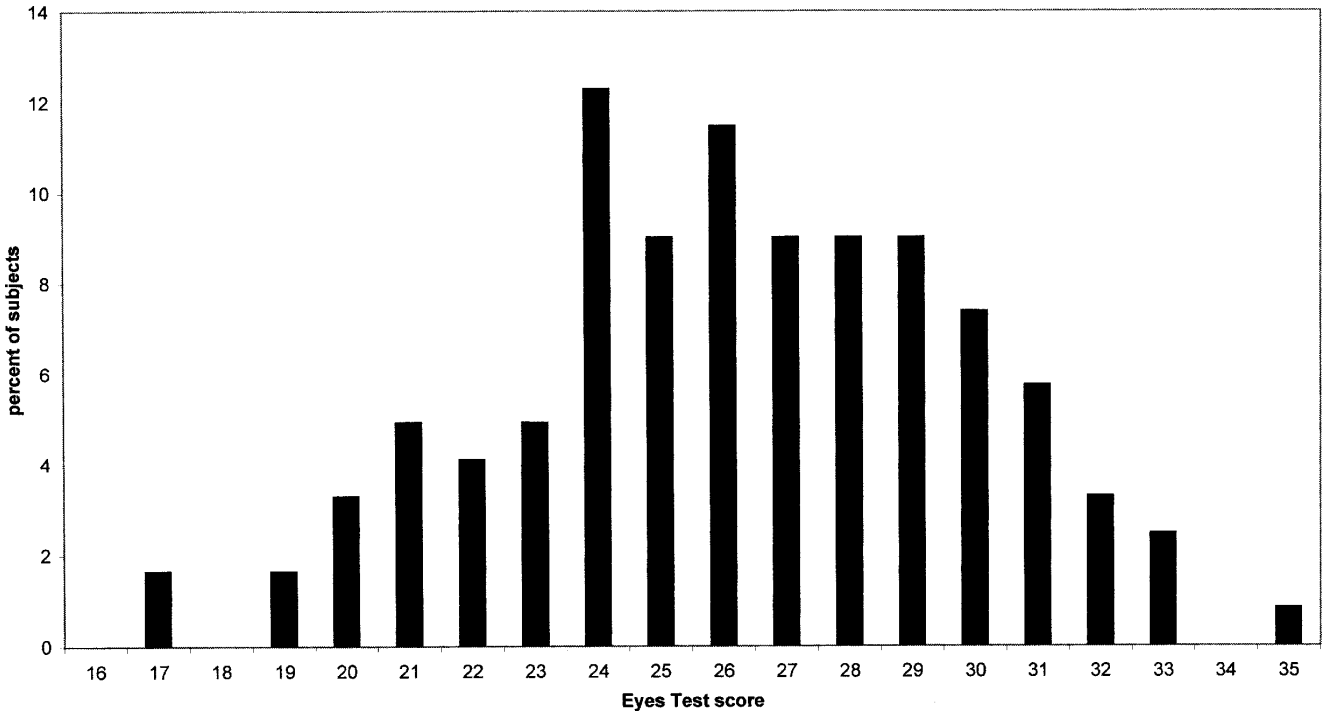

Figure 3. Distribution of Eyes Test scores in Groups 2 and 3.

two. Table 3 shows the means and standard deviations on the Revised Eyes Task for each of the four groups, and the results of the AQ for Groups 1, 3, and 4 only. A one-way ANOVA comparing the four groups on the Revised Eyes Task revealed that there was a significant main effect of group, $F(3, 250) = 17.87$, $p = .0001$. Further examination of this result using Scheffé's tests indicated that, as predicted, Group 1 performed significantly worse than the other three groups, who did not differ from each other. This is shown graphically in Fig. 3. Sex differences were examined in Groups 2 and 3, using an ANOVA of Group × Sex. The sex difference approached significance, $F(1, 224) = 3.38$, $p = .067$, with females scoring higher than males, whilst the interaction was insignificant, $F(1, 224) = 0.79$, $p = .376$. Separate group item analyses are shown in Table 4. All subjects with AS/HFA scored 33 or above out of 36 on the gender recognition control task. There were no within-group differences in Group 3 (students) according to subject studied, $F(1, 99) = 1.39$, $p = .24$.

On the AQ, as expected, Group 1 scored significantly higher than Groups 3 and 4: one-way ANOVA of group, $F(2, 103) = 23.4$, $p = .00001$; Scheffé's tests indicated Group 1 scored significantly higher at the .05 level than Groups 3 and 4, for which there was no difference. The predicted sex difference on the AQ (males scoring higher

Table 4
Item Analysis of the Eyes Test (*Including Only the 36 Retained Items*), *Showing the Percentage of Each Group Passing Each Item*

Item	Group 1 AS/HFA adults ($N = 15$)	Group 2 General population controls ($N = 122$)	Group 3 Students ($N = 103$)	Group 4 IQ-matched controls ($N = 14$)
1	60.0	85.2	70.9	100.0
2	73.3	78.7	85.4	100.0
3	66.7	86.1	83.5	100.0
4	33.3	73.0	87.4	78.6
5	66.7	77.0	82.5	85.7
6	86.7	80.3	77.7	85.7
7	46.7	68.0	78.6	92.9
8	60.0	67.2	83.5	78.6
9	80.0	77.0	91.3	85.7
10	66.7	73.0	63.1	85.7
11	46.7	68.0	80.6	92.9
12	53.3	87.7	83.5	92.9
13	60.0	69.7	76.7	78.6
14	73.3	80.3	94.2	100.0
15	66.7	69.7	83.5	85.7
16	80.0	77.0	82.5	78.6
17	53.3	65.6	60.2	85.7
18	46.7	58.2	79.6	71.4
19	66.7	69.7	58.3	71.4
20	86.7	88.5	87.4	92.9
21	53.3	73.8	81.6	85.7
22	60.0	79.5	91.3	85.7
23	53.3	77.9	84.5	85.7
24	80.0	73.8	77.7	92.9
25	60.0	71.3	57.3	71.4
26	53.3	65.6	81.6	100.0
27	46.7	65.6	63.1	64.3
28	40.0	66.4	65.0	57.1
29	80.0	77.9	64.1	92.9
30	53.3	91.0	89.3	85.7
31	53.3	51.6	52.4	85.7
32	46.7	50.0	72.8	78.6
33	66.7	58.2	74.8	85.7
34	60.0	77.0	81.6	85.7
35	46.7	65.6	82.5	92.9
36	66.7	76.2	87.4	92.9

than females) in Group 3 was also found ($t = 1.97$, $p = .03$ for one-tailed significance). Finally, the correlation between AQ score, IQ score, and Eyes Test score was computed. Combining the groups, there was no correlation between the Eyes Test and IQ ($r = .09$, $p = .6$) or between the AQ and IQ ($r = .05$, $p = .77$). The AQ and Eyes Test were, as expected, inversely correlated ($r = -.53$, $p = .004$). This was true for all three groups where both measures were used. In the student group, the Eyes Test was inversely correlated with the social skills category ($r = .27$, $p = .015$) and the communication category ($r = .25$, $p = .027$).

Discussion

This study reports normative data on the Revised Eyes Test for adults. The modifications were designed to render this test a more sensitive measure of adult social intelligence. As was hoped, the modifications from the original version led to normal performance being significantly below ceiling. This is important if the test is to do more than discriminate extreme performance and instead detect meaningful individual differences. This study replicated the earlier finding that adults with AS or HFA are significantly impaired on such tests, whereas they are not impaired on the gender recognition control test (Baron-Cohen, Jolliffe, et al., 1997; Baron-Cohen, Wheelwright, et al., 1997). This therefore validates it as a useful test with which to identify subtle impairments in social intelligence in otherwise normally intelligent adults.

In a series of single case studies we have also found that this test distinguishes very high-functioning adults with AS/HFA from controls (Baron-Cohen, Wheelwright, Stone, & Rutherford, 1999). The Revised Eyes Test may be relevant to clinical groups beyond those on the autistic spectrum (e.g., brain-damaged patients following amygdalectomy or prefrontal cortical lesions). The test has recently been used with these groups (Stone, Baron-Cohen, & Knight, 1999; Stone, Baron-Cohen, Young, & Calder, 1998). We have recently developed a child version of this test, reported separately (Baron-Cohen, Wheelwright, Spong, Scahill, & Lawson, in press). The adult Eyes Test has been used during fMRI, revealing amygdala activity in the normal (but not in the autistic) brain (Baron-Cohen, Ring, et al., 1999).

In the present study, among the general population controls and student group, there was a trend towards a sex difference (female superiority) ($p = .07$). This echoes the sex difference found with the previous version of this test. One possible reason why the sex difference did not reach significance with the new test is that if the effect size is relatively small, the chance of detecting a sex difference would be low. There was no significant correlation between IQ and the Eyes Test, suggesting this is independent of general (nonsocial) intelligence.

Performance on the Revised Eyes Test was inversely correlated with performance on the Autism Spectrum Quotient (AQ), suggesting that both measure degrees of autistic traits across the notional spectrum (Wing, 1988). The AQ is not diagnostic but may serve as a useful instrument for quantifying the extent of an individual's "caseness" in terms of AS/HFA, measuring personality traits. The present results confirm our earlier finding that adults with HFA or AS score significantly higher on the AQ than do general population controls.

A criticism of the Revised Eyes Test might be that, even with the new modifications, the stimuli are static, whereas the real world never is. Future studies might usefully employ dynamic stimuli of eye expressions. Static stimuli, however, make the test quick and easy to use, since it can be administered as a "pencil and paper" test. In our laboratory, we are also experimenting with computer-presentation of the Eyes stimuli so as to record response time in subjects' judgements of the most appropriate mental state term to match each picture. Such speed of processing approaches may be a fruitful way to explore individual differences on this task. However, it is clear that even a nonautomated format is sufficient to reveal group differences. The Eyes test stands as an example of how experimental methods can be applied to the social domain.

Acknowledgements—SBC and SW were supported by the MRC and the McDonnell Pew Foundation during the period of this work. IP and YR submitted the normative data, which they collected, as part of their final year project for the degree of BSc in Psychology, University of Cambridge. JH collected the normative data as part of her MSc in Psychological Research Methods at the University of Exeter.

References

American Psychiatric Association (1994). *DSM-IV Diagnostic and statistical manual of mental disorders* (4th ed). Washington, DC: Author.

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*. Boston: 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., 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., Ring, H., Wheelwright, S., Bullmore, E., Brammer, M., Simmons, A., & Williams, S. (1999). Social intelligence in the normal and autistic brain: An fMRI study. *European Journal of Neuroscience*, *11*, 1891–1898.

Baron-Cohen, S., & Wheelwright, S. (in press). The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*.

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a "language of the eyes"? Evidence from normal adults and adults with autism or Asperger syndrome. *Visual Cognition*, *4*, 311–331.

Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V., & Lawson, J. (in press). Are intuitive physics and intuitive psychology independent? A test with children with Asperger syndrome. *Journal of Developmental and Learning Disorders*.

Baron-Cohen, S., Wheelwright, S., Stone, V., & Rutherford, M. (1999). A mathematician, a physicist, and a computer scientist with Asperger syndrome: Performance on folk psychology and folk physics test. *Neurocase*, *5*, 475–483.

Ekman, P., & Friesen, W. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*, 124–129.

Flavell, J. H., Green, E. R. & Flavell, E. R. (1986). Development of knowledge about the appearance-reality distinction. *Society for Research in Child Development*, *51*.

Flavell, J., Shipstead, S., & Croft, K. (1978). Young children's knowledge about visual perception: Hiding objects from others. *Child Development*, *49*, 1208–1211.

Morton, J., Frith, U., & Leslie, A. (1991). The cognitive basis of a biological disorder: Autism. *Trends in Neurosciences*, *14*, 434–438.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a "theory of mind"? *Behaviour and Brain Sciences*, *4*, 515–526.

Stone, V., Baron-Cohen, S., & Knight, K. (1999). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, *10*, 640–656.

Stone, V., Baron-Cohen, S., Young, A., & Calder, A. (1998). *Patients with amygdalectomy show impairments in theory of mind*. Cambridge: University of Cambridge.

Walker, A. S. (1982). Intermodal perception of expressive behaviours by human infants. *Journal of Experimental Child Psychology*, *33*, 514–535.

Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore, MD: Williams & Wilkins.

Wellman, H. (1990). *Children's theories of mind*. Cambridge, MA: MIT Press.

Whiten, A. (1991). *Natural theories of mind*. Oxford: Basil Blackwell.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128.

Wing, L. (1988). The autistic continuum. In L. Wing (Ed.), *Aspects of autism: Biological research*. London: Gaskell/Royal College of Psychiatrists.

World Health Organisation. (1994). *International classification of diseases* (10th ed.). Geneva: Author.

Young, A., Hellawell, D., De Wal, C., & Johnson, M. (1996). Facial expression processing after amygdalectomy. *Neuropsychologia*, *34*, 31–39.

Manuscript accepted 30 June 2000

Appendix A

List of Target Mental State Terms for Each Item (in Italic) and Their Distractors

PI[a]	jealous	*panicked*	arrogant	hateful
1	*playful*	comforting	irritated	bored
2	terrified	*upset*	arrogant	annoyed
3	joking	flustered	*desire*	convinced
4	joking	*insisting*	amused	relaxed
5	irritated	sarcastic	*worried*	friendly
6	aghast	*fantasizing*	impatient	alarmed
7	apologetic	friendly	*uneasy*	dispirited
8	*despondent*	relieved	shy	excited
9	annoyed	hostile	horrified	*preoccupied*
10	*cautious*	insisting	bored	aghast
11	terrified	amused	*regretful*	flirtatious
12	indifferent	embarrassed	*sceptical*	dispirited
13	decisive	*anticipating*	threatening	shy
14	irritated	disappointed	depressed	*accusing*
15	*contemplative*	flustered	encouraging	amused
16	irritated	*thoughtful*	encouraging	sympathetic
17	*doubtful*	affectionate	playful	aghast
18	*decisive*	amused	aghast	bored
19	arrogant	grateful	sarcastic	*tentative*
20	dominant	*friendly*	guilty	horrified
21	embarrassed	*fantasizing*	confused	panicked
22	*preoccupied*	grateful	insisting	imploring
23	contented	apologetic	*defiant*	curious
24	*pensive*	irritated	excited	hostile
25	panicked	incredulous	despondent	*interested*
26	alarmed	shy	*hostile*	anxious
27	joking	*cautious*	arrogant	reassuring
28	*interested*	joking	affectionate	contented
29	impatient	aghast	irritated	*reflective*
30	grateful	*flirtatious*	hostile	disappointed
31	ashamed	*confident*	joking	dispirited
32	*serious*	ashamed	bewildered	alarmed
33	embarrassed	guilty	fantasizing	*concerned*
34	aghast	baffled	*distrustful*	terrified
35	puzzled	*nervous*	insisting	contemplative
36	ashamed	nervous	*suspicious*	indecisive

[a]PI: practice item.

Appendix B

Glossary for Adult Eyes Test

ACCUSING	blaming The policeman was *accusing* the man of stealing a wallet.
AFFECTIONATE	showing fondness toward someone Most mothers are *affectionate* to their babies by giving them lots of kisses and cuddles.
AGHAST	horrified, astonished, alarmed Jane was *aghast* when she discovered her house had been burgled.
ALARMED	fearful, worried, filled with anxiety Claire was *alarmed* when she thought she was being followed home.
AMUSED	finding something funny I was *amused* by a funny joke someone told me.
ANNOYED	irritated, displeased Jack was *annoyed* when he found out he had missed the last bus home.
ANTICIPATING	expecting At the start of the football match, the fans were *anticipating* a quick goal.
ANXIOUS	worried, tense, uneasy The student was feeling *anxious* before taking her final exams.
APOLOGETIC	feeling sorry The waiter was very *apologetic* when he spilt soup all over the customer.
ARROGANT	conceited, self-important, having a big opinion of oneself The *arrogant* man thought he knew more about politics than everyone else in the room.
ASHAMED	overcome with shame or guilt The boy felt *ashamed* when his mother discovered him stealing money from her purse.

Appendix B (cont.)

ASSERTIVE	confident, dominant, sure of oneself The *assertive* woman demanded that the shop give her a refund.
BAFFLED	confused, puzzled, dumfounded The detectives were completely *baffled* by the murder case.
BEWILDERED	utterly confused, puzzled, dazed The child was *bewildered* when visiting the big city for the first time.
CAUTIOUS	careful, wary Sarah was always a bit *cautious* when talking to someone she did not know.
COMFORTING	consoling, compassionate The nurse was *comforting* the wounded soldier.
CONCERNED	worried, troubled The doctor was *concerned* when his patient took a turn for the worse.
CONFIDENT	self-assured, believing in oneself The tennis player was feeling very *confident* about winning his match.
CONFUSED	puzzled, perplexed Lizzie was so *confused* by the directions given to her, she got lost.
CONTEMPLATIVE	reflective, thoughtful, considering John was in a *contemplative* mood on the eve of his 60th birthday.
CONTENTED	satisfied After a nice walk and a good meal, David felt very *contented.*
CONVINCED	certain, absolutely positive Richard was *convinced* he had come to the right decision.
CURIOUS	inquisitive, inquiring, prying Louise was *curious* about the strange-shaped parcel.
DECIDING	making your mind up The man was *deciding* who to vote for in the election.
DECISIVE	already made your mind up Jane looked very *decisive* as she walked into the polling station.
DEFIANT	insolent, bold, don't care what anyone else thinks The animal protester remained *defiant* even after being sent to prison.
DEPRESSED	miserable George was *depressed* when he didn't receive any birthday cards.
DESIRE	passion, lust, longing for Kate had a strong *desire* for chocolate.
DESPONDENT	gloomy, despairing, without hope Gary was *despondent* when he did not get the job he wanted.
DISAPPOINTED	displeased, disgruntled Manchester United fans were *disappointed* not to win the Championship.
DISPIRITED	glum, miserable, low Adam was *dispirited* when he failed his exams.
DISTRUSTFUL	suspicious, doubtful, wary The old woman was *distrustful* of the stranger at her door.
DOMINANT	commanding, bossy The sergeant major looked *dominant* as he inspected the new recruits.
DOUBTFUL	dubious, suspicious, not really believing Mary was *doubtful* that her son was telling the truth.
DUBIOUS	doubtful, suspicious Peter was *dubious* when offered a surprisingly cheap television in a pub.
EAGER	keen On Christmas morning, the children were *eager* to open their presents.
EARNEST	having a serious intention Harry was very *earnest* about his religious beliefs.
EMBARRASSED	ashamed After forgetting a colleague's name, Jenny felt very *embarrassed.*
ENCOURAGING	hopeful, heartening, supporting All the parents were *encouraging* their children in the school sports day.
ENTERTAINED	absorbed and amused or pleased by something I was very *entertained* by the magician.
ENTHUSIASTIC	very eager, keen Susan felt very *enthusiastic* about her new fitness plan.

Appendix B (cont.)

FANTASIZING	daydreaming Emma was *fantasizing* about being a film star.
FASCINATED	captivated, really interested At the seaside, the children were *fascinated* by the creatures in the rock pools.
FEARFUL	terrified, worried In the dark streets, the women felt *fearful*.
FLIRTATIOUS	brazen, saucy, teasing, playful Connie was accused of being *flirtatious* when she winked at a stranger at a party.
FLUSTERED	confused, nervous and upset Sarah felt a bit *flustered* when she realised how late she was for the meeting and that she had forgotten an important document.
FRIENDLY	sociable, amiable The *friendly* girl showed the tourists the way to the town centre.
GRATEFUL	thankful Kelly was very *grateful* for the kindness shown by the stranger.
GUILTY	feeling sorry for doing something wrong Charlie felt *guilty* about having an affair.
HATEFUL	showing intense dislike The two sisters were *hateful* to each other and always fighting.
HOPEFUL	optimistic Larry was *hopeful* that the post would bring good news.
HORRIFIED	terrified, appalled The man was *horrified* to discover that his new wife was already married.
HOSTILE	unfriendly The two neighbours were *hostile* towards each other because of an argument about loud music.
IMPATIENT	restless, wanting something to happen soon Jane grew increasingly *impatient* as she waited for her friend who was already 20 minutes late.
IMPLORING	begging, pleading Nicola looked *imploring* as she tried to persuade her dad to lend her the car.
INCREDULOUS	not believing Simon was *incredulous* when he heard that he had won the lottery.
INDECISIVE	unsure, hesitant, unable to make your mind up Tammy was so *indecisive* that she couldn't even decide what to have for lunch.
INDIFFERENT	disinterested, unresponsive, don't care Terry was completely *indifferent* as to whether they went to the cinema or the pub.
INSISTING	demanding, persisting, maintaining After a work outing, Frank was *insisting* he paid the bill for everyone.
INSULTING	rude, offensive The football crowd was *insulting* the referee after he gave a penalty.
INTERESTED	inquiring, curious After seeing Jurassic Park, Huge grew very *interested* in dinosaurs.
INTRIGUED	very curious, very interested A mystery phone call *intrigued* Zoe.
IRRITATED	exasperated, annoyed Frances was *irritated* by all the junk mail she received.
JEALOUS	envious Tony was *jealous* of all the taller, better-looking boys in his class.
JOKING	being funny, playful Gary was always *joking* with his friends.
NERVOUS	apprehensive, tense, worried Just before her job interview, Alice felt very *nervous*.
OFFENDED	insulted, wounded, having hurt feelings When someone made a joke about her weight, Martha felt very *offended*.
PANICKED	distraught, feeling of terror or anxiety On waking to find the house on fire, the whole family were *panicked*.
PENSIVE	thinking about something slightly worrying Susie looked *pensive* on the way to meeting her boyfriend's parents for the first time.
PERPLEXED	bewildered, puzzled, confused Frank was *perplexed* by the disappearance of his garden gnomes.
PLAYFUL	full of high spirits and fun Neil was feeling *playful* at his birthday party.

Appendix B (cont.)

PREOCCUPIED	absorbed, engrossed in one's own thoughts Worrying about her mother's illness made Debbie *preoccupied* at work
PUZZLED	perplexed, bewildered, confused After doing the crossword for an hour, June was still *puzzled* by one clue.
REASSURING	supporting, encouraging, giving someone confidence Andy tried to look *reassuring* as he told his wife that her new dress did suit her.
REFLECTIVE	contemplative, thoughtful George was in a *reflective* mood as he thought about what he'd done with his life.
REGRETFUL	sorry Lee was always *regretful* that he had never travelled when he was younger.
RELAXED	taking it easy, calm, carefree On holiday, Pam felt happy and *relaxed.*
RELIEVED	freed from worry or anxiety At the restaurant, Ray was *relieved* to find he had not forgotten his wallet.
RESENTFUL	bitter, hostile The businessman felt very *resentful* towards his younger colleague who had been promoted above him.
SARCASTIC	cynical, mocking, scornful The comedian made a *sarcastic* comment when someone came into the theatre late.
SATISFIED	content, fulfilled Steve felt very *satisfied* after he had got his new flat just how he wanted it.
SCEPTICAL	doubtful, suspicious, mistrusting Patrick looked *sceptical* as someone read out his horoscope to him.
SERIOUS	solemn, grave The bank manager looked *serious* as he refused Nigel an overdraft.
STERN	severe, strict, firm The teacher looked very *stern* as he told the class off.
SUSPICIOUS	disbelieving, suspecting, doubting After Sam had lost his wallet for the second time at work, he grew *suspicious* of one of his colleagues.
SYMPATHETIC	kind, compassionate The nurse looked *sympathetic* as she told the patient the bad news.
TENTATIVE	hesitant, uncertain, cautious Andrew felt a bit *tentative* as he went into the room full of strangers.
TERRIFIED	alarmed, fearful The boy was *terrified* when he thought he saw a ghost.
THOUGHTFUL	thinking about something Phil looked *thoughtful* as he sat waiting for the girlfriend he was about to finish with.
THREATENING	menacing, intimidating The large, drunk man was acting in a very *threatening* way.
UNEASY	unsettled, apprehensive, troubled Karen felt slightly *uneasy* about accepting a lift from the man she had only met that day.
UPSET	agitated, worried, uneasy The man was very *upset* when his mother died.
WORRIED	anxious, fretful, troubled When her cat went missing, the girl was very *worried.*