Familiar Face and Voice Matching and Recognition in Children with Autism

Jill Boucher
University of Sheffield, U.K.

Vicky Lewis
The Open University, Milton Keynes, U.K.

Glyn Collis
University of Warwick, Coventry, U.K.

Relatively able children with autism were compared with age- and language-matched controls on assessments of (1) familiar voice–face identity matching, (2) familiar face recognition, and (3) familiar voice recognition. The faces and voices of individuals at the children’s schools were used as stimuli. The experimental group were impaired relative to the controls on all three tasks. Face recognition and voice recognition correlated significantly with voice–face identity matching, but not with each other, suggesting that the recognition impairments jointly cause the matching impairment. Neither chronological age nor verbal mental age were consistently related to the recognition and matching impairments.

Keywords: Autism, school children, face perception, voice perception, memory.

Abbreviations: APT-I: Action Picture Test Information Scale; CA: chronological age; MLD: moderate learning disability; SLI: specific language impairments

Introduction

Face and voice processing are the main routes by which the normal human being learns about other people. Human neonates respond preferentially to face-like stimuli (Goren, Sarty, & Wu, 1975; Maurer & Young, 1983) and to voices (Eisenberg, 1976), and infants recognise familiar faces and also familiar voices from a strikingly early age (Bushnell, Sai, & Mullin, 1989; DeCasper & Fifer, 1980). By the age of 4 years the normal child can recognise familiar faces and also voices at a near-adult level (Bartholomous, 1973; Ellis, 1990), although the range of those who count as familiar is necessarily limited. By contrast, the ability to recognise unfamiliar but recently seen faces and also voices develops more slowly and does not reach adult levels until about the age of 14 years (Mann, Diamond, & Carey, 1979).

As well as underpinning the ability to discriminate between people and to recognise familiar people, face and voice processing are also critically involved in nonverbal communication. Nonverbal communication is used largely to indicate emotional states and attitudes, and to co-ordinate conversation (Argyle, 1983). Infants communicate nonverbally, both actively and in their understanding of the nonverbal signals of others. For example, normal infants respond to the affective content of facial expressions and vocal tones by the age of 6 months (Walker, 1982; Walker-Andrews, 1988). It is difficult to demonstrate unambiguously that they employ these abilities in their everyday social interactions, but their early protoconversations with carers seem to be mediated by enhanced use of facial and vocal expression as well as by body movements (Kaye & Fogel, 1980; Stern, 1985) and presumably by the infants monitoring the adults’ direction of gaze (Butterworth, 1991).

Given the difficulties that people with autism have with personal relationships and with the perception and use of nonverbal communication, it is perhaps not surprising that abnormalities of face processing have been found in association with autism. It is known, for example, that when recognising highly familiar faces children with autism selectively attend to different facial features from those selectively attended to by children without autism (Langdell, 1978), and that inverting faces has less effect on autistic children’s recognition ability than it does on that of controls (Hobson, Ouston, & Lee, 1988a; Tantam, Monaghan, Nicholson, & Stirling, 1989). These latter findings suggest that faces are processed analytically by children with autism, rather than holistically, as is the case in normal children and adults.
Despite these abnormalities of encoding, tests of face identity matching have not always demonstrated impaired performance. In studies by Ozonoff, Pennington, and Rogers (1990) and by Boucher and Lewis (1992), autistic children’s performance on a test of matching to sample was unimpaired. In a study by Davies, Bishop, Manstead, and Tantam (1994), identity matching across changes in orientation and across changes in expression was impaired only in older and more able children with autism, relative to controls.

In addition to perceptual or structural encoding abnormalities, the ability of children with autism to recognize and to understand facial expression is impaired (Capps, Yirmiya, & Sigman, 1992; Fein, Lucci, Braverman, & Waterhouse, 1992), as is the ability to name facial expressions (Hobson, Ouston, & Lee, 1989). The ability to integrate visual and auditory aspects of facial speech (de Gelder, Vroomen, & van der Heide, 1991), and the ability to infer meaning from the direction of gaze (Baron-Cohen & Cross, 1992) are also impaired. However, the ability to judge the direction of looking is not impaired (Phillips, Baron-Cohen, & Rutter, 1992). The study by de Gelder et al. (1991) further shows an impairment of unfamiliar face recognition. Boucher and Lewis (1992) reported a similar finding.

In view of the abnormalities of face processing that have been shown to occur in autism, and in view of the similarities between the course of development of face and voice processing in normal children (Mann et al., 1979), it is surprising that very little experimental work has been carried out to assess voice processing abilities in autism. Three studies have been reported in which the ability to match facial expressions to vocal affect was assessed (Hobson, Ouston, & Lee, 1988b; Ozonoff et al., 1990; Van Lancker, Cornelius, & Kreiman, 1989). All of these studies found voice–face affect matching to be impaired in children with autism. Hobson et al. (1989) found that the ability to name vocal expression is impaired. Other studies of voice processing in autism have shown that high-functioning people with autism use communicative patterns of intonation less frequently than either individuals with Asperger’s syndrome or controls (Fine, Bartolucci, Ginsberg, & Szatmari, 1991); that autistic infants use idiosyncratic intonation patterns in their crying (Ricks, 1975); and that young children with autism differ from young normal children and from learning disabled children in not showing a preference for their mother’s voice as opposed to other speech stimuli (Klin, 1989, 1992). Klin suggests that abnormal processing of voices may be an important factor underlying the social abnormalities in autism, and such evidence as is available tends to support this suggestion.

The main aim of the experiments reported here was, therefore, to add to existing evidence concerning voice processing in autism. We took as our starting point the question of whether familiar voice–face identity matching might be impaired in autistic children, in addition to the well-documented impairment in voice–face affect matching. We argued that if a test of voice–face identity matching showed no impairment in the children with autism, then this would suggest that voice–face affect matching impairments result from specific deficits in affect processing. This would be an important finding, since there is dispute in the literature concerning the extent to which the social deficits in autism involve the processing of affect, as opposed to information by which a person is identified.

We did not, however, expect autistic children’s voice–face identity matching to be unimpaired. We based this prediction on the following arguments. Evidence of impaired integration of vocal and facial speech (de Gelder et al., 1991) suggests that impaired voice–face affect matching results at least partly from a cross-modal processing deficit or from deficits in processing vocal and/or facial information in general, rather than from deficits in processing affective information in particular. There is no strong evidence indicating cross-modal processing deficits in autism (though see Martineau, Morris, Kay, & Levin, 1992). By contrast there is evidence, cited above, which suggests that children with autism have face and probably voice processing abnormalities that go beyond the processing of affect. In particular, our own work (Boucher & Lewis, 1992) and that of de Gelder et al. (1991) on unfamiliar face recognition suggests that matching familiar voices to familiar faces will be difficult for children with autism because of face recognition difficulties. Moreover, Klin’s (1992) evidence of abnormal attention to voices suggests that voice–face identity matching may also be impaired because of voice recognition difficulties. We therefore predicted not only that voice–face identity matching would be impaired in autism, but that follow-up experiments would demonstrate that this impairment results from a combination of deficits of familiar face and familiar voice recognition.

Our initial experiment assessing voice-face identity matching is reported below as Experiment 1.

**Experiment 1: A Test of Familiar Voice–Face Identity Matching**

**Method**

**Participants.** The assessment of familiar voice and face processing in children with special needs raises the difficulty that the voices and faces of famous people, such as are generally used as stimuli in this type of assessment, cannot be used since there is no certainty that all children will have had equal exposure to any set of “famous” people that might be used. In order to overcome this difficulty we selected participants from schools catering specifically for children with autism and for children with other types of developmental disorder associated with low language ability. This enabled us to use as stimuli the faces and voices of adults working at the schools, ensuring that children in the experimental and control groups had comparable exposure to the stimuli. Such schools are not easy to identify, and most such schools register only small numbers of children with autism, including a majority with little or no language. In order to assemble an experimental group of adequate size and language ability, we therefore selected participants from two schools, referred to below as School A and School B.

School A is a non-residential, city school attended by approximately 110 children aged between 3 and 12 years. The school caters for children with mild to moderate learning disability (MLD) sometimes associated with specific language impairments (SLI) and includes a small proportion of children with autism. School B is a non-residential city school catering for approximately 60 children aged between 4 and 12 years. The school caters for children who have either autism or SLI, a
proportion of whom have MLD in addition to their autism or SLI.

Two groups of participants were selected from these 2 schools, a group of 19 children with autism and a group of 20 control children with MLD/SLI without signs of autism. No child was included in the study who had sensory or motor difficulties, or who was unwilling to co-operate, or who failed to reach criterion on task-training procedures.

The children with autism had been diagnosed as autistic by independent psychiatrists or psychologists, and conformed to Wing’s (1988) criteria for the autistic spectrum. Thus all the children had impaired reciprocal social interaction and empathy, impaired conversational skills and nonverbal communication, and behavioural rigidity associated with lack of imagination and creativity. These criteria are comparable to the criteria given in DSM-IV (American Psychiatric Association, 1994).

Children in the MLD/SLI group all had impaired language. This was associated in some cases with overall low learning ability, in other cases with specific language impairment in absence of overall MLD, and in some other cases with a combination of MLD and SLI. Additional diagnostic information was not available for the control children attending School A. However, of the 12 control participants from School B, 8 had been diagnosed by speech and language therapists as having complex language learning difficulties including comprehension difficulties and phonological output difficulties. Two children were diagnosed as cases of oral dyspraxia. The remaining two control participants from School B were described as having poor language skills linked to moderate learning difficulties.

The two groups of children were equated for chronological age (CA), number of terms at their school, and for verbal mental age (VMA) using the Action Picture Test Information scale (APT-I) (Renfrew, 1972). The APT-I requires children to answer questions about what is happening in a series of pictures, and assesses the use of nonsubstantive as well as substantive vocabulary. Mean CA, VMA and number of terms in school for children attending School A and for children attending School B are shown in Table 1. Two-way ANOVAs (group by school) on each of these variables showed that there was no difference between the experimental and control groups and no interaction of the group factor with school. The mean CA of the children differed slightly but significantly between the two schools [$F(1,35) = 4.73, p < .05$]. Not surprisingly, given the CA difference, there was also a marginally significant difference between the two schools in the number of terms in school [$F(1,35) = 3.96, p = .055$]. However, the lack of significant interactions indicates that these differences were balanced between the experimental and control groups. Analyses of experimental data incorporating a schools factor failed to show any other differences between schools, confirming that there were no important differences between the children selected from the two schools.

**Materials.** For each school the following materials were prepared.

1. **Training stimuli.** Three black-and-white photographs of faces (10 cm × 12.5 cm, mounted on cards measuring 14 cm × 16 cm) were prepared. The photographs were of highly recognisable staff from the child’s own school (the Headteacher, the school secretary, and a male teacher). Individuals were photographed looking into the camera. Clothing was concealed by a dark cloth. However, accessories such as earrings and spectacles were not removed. In addition an audiotape was prepared, on which each of the three individuals whose faces appeared in the training set was recorded saying “Hello. Do you know who I am? Can you find my picture?” After a 3-second pause, this was repeated. Each recorded voice and its repeat appeared three times on the training tape, making a total of nine stimuli. The order of voices was randomised, with the requirement that each voice occurred once within the first three, once within the second three, and once within the final three voice stimuli.

2. **Test stimuli.** Twenty photographs of staff from the child’s own school (similar to the training photographs, but of different staff) were used as test stimuli. An audiotape was prepared on which each of the 20 individuals whose faces appeared as test stimuli was recorded as for the training stimuli.

**Procedure.** Children were seen individually in a quiet room in their school following at least one familiarisation session in which the tester spent time in the child’s classroom and also assessed the child’s language ability in the quiet room to be used for experimental testing. Training and testing took place in a single session. The training procedure was as follows. The tester and the child sat side by side at a table. The three training photographs were placed separately face up on a table in front of the child and the child was helped to name the people in the photographs and to say something about each of them. This was done to ensure that the children recognised the people in the photographs. The tester then drew the child’s attention to the tape recorder and said: “Listen to this voice. One of these
The first voice was then played, and the child was encouraged to point to the appropriate photograph. If the child hesitated or made an error, the second recording of the same voice was played, and the child was again asked to find the face matching the voice. If the child again failed to respond, or made an error, a correct response was elicited using prompts. The photograph corresponding to the first voice was then turned over. The second voice was then played, using the same procedure, and then the third voice. If the child made any errors or needed help, the three photographs were laid face-up again, and the process was repeated up to twice more, using the varied order recordings on the audiotape. When a child achieved three consecutively correct matches unaided the training materials were removed and testing began. Every child achieved this training criterion.

The test procedure was the same as the procedure used in training, except that all 20 test photographs were placed simultaneously face up in 4 rows of 5 in a predetermined order in front of the child. A record was made as to whether the child’s response was correct on the first voice presentation, or on the second, or whether errors occurred on both presentations, providing a score of 2, 1, or 0 respectively for each voice–face pair, making a maximum total score of 40. If the child identified the correct photograph it was turned over. If the child did not identify the correct photograph on both presentations, the tester pointed to the correct photograph, said “I think it was this person” and turned that photograph over. Thus the size of the array from which the child had to choose was progressively reduced after each trial. A proportion of correct responses were praised, and comments such as “You’re doing well”, “Here’s an easy one”, “Not many more to do” and so on were used to maintain the child’s motivation.

**Results**

Mean scores and standard deviations are shown for each group and each school in Table 2. Taking into account trial-to-trial reductions in the size of the array, and that the children were given up to two attempts and the possibility of scoring 0, 1, or 2 on each trial, the score expected by chance if a child had adopted a pure guessing strategy can be calculated as 9.793. The mean scores for both groups were significantly greater than this expected score [children with autism \( t(18) = 5.61, p < .001 \); controls \( t(19) = 9.10, p < .001 \). In fact, all four cell means in the group \( \times \) school layout were significantly greater than chance.

Because of differences in the stimuli used in the two schools, and slight differences in the ages of children at the two schools, the schools factor was incorporated in the analysis. A two-way analysis of variance (group by school) showed that the children with autism scored significantly lower than the control children on the test of voice–face identity matching \( F(1,35) = 10.3, p = .003 \). There was no effect of school \( F(1,35) = 2.04 \) and no interaction between group and school \( F(1,35) = .3 \).

Correlation analyses were carried out using regression/ANOVA procedures, focusing on age and language scores as continuous-variable predictors. Adding CA to the basic ANOVA (and the age \( \times \) group interaction) showed it to be a significant predictor of voice–face identity matching score \( F(1,33) = 15.0, p < .001 \). The interaction between age and group was not significant \( F(1,33) = 2.5, p > .12 \), indicating that the relationship between age and voice–face identity matching score was similar for the two groups. The magnitude of this relationship is described by a correlation coefficient of .63 (with group differences partialed out). This analysis also showed that the group difference remains highly significant when age is considered as a covariate.

Adding the APT-I language scores into the basic analysis as a predictor variable (and the language \( \times \) group interaction, omitting subjects with missing APT-I data—see Table 1) showed a significant relationship between this measure and voice–face identity matching score \( F(1,31) = 12.8, p = .001 \). A significant interaction between APT-I score and group \( F(1,31) = 5.9, p = .021 \) showed that the relationship was different for the two groups: the regression line relating voice–face identity matching to the APT-I score was much less steep for the group of children with autism than for the controls (see Fig. 1, and the correlations were weaker (for the children with autism \( r = .37, p = .12 \); for the control group \( r = .69, p < .002 \)).

Over the whole sample \( (N = 39) \), the correlation between APT-I scores and CA, partiailing out group differences, was \( r = .45 (p = .006) \). Nonetheless, with both these predictors simultaneously included in the analysis (plus the group \( \times \) APT-I interaction), both remained significant or very nearly so, indicating that they are independent predictors of voice–face identity matching \( [CA—F(1,32) = 8.9, p = .005; APT-I—F(1,32) = 3.9, p = .055] \).

**Discussion**

The results of Experiment 1 show that children with autism are significantly impaired in matching familiar voices to faces, relative to a mixed group of children without autism but of low language ability. Covariance analyses indicated that this group difference was not simply a consequence of any differences between the groups in age or language ability (see Table 1). There was no effect of school and no interaction between groups and schools, indicating that the different stimuli used to test the children in the two schools, and the slight difference in the ages of the children in the two schools, had no differential effects on voice–face identity matching ability.

There were significant (and independent) relationships between voice–face identity matching and both age and verbal ability in the control group. Voice–face matching was significantly related to age, and weakly related to
Our first prediction of impaired voice–face identity matching ability in children with autism was therefore supported. As suggested in the Introduction, this finding, and previous findings of impaired voice–face affect matching, might be explained in terms of a deficit in cross-modal processing. However, there is very little evidence to support such a hypothesis. Alternatively, the matching deficits might be explained in terms of impaired processing of social stimuli, and we hypothesised that impaired identity matching results from a combination of impairments of familiar face and familiar voice recognition. Experiment 2, reported next, was designed to assess familiar face recognition.

**Experiment 2: A Test of Familiar Face Recognition**

The ability of individuals with autism to recognise familiar faces (i.e. faces that an individual has been exposed to over time) has not previously been assessed. Langdell (1978) reported a study that was concerned with the way in which familiar face recognition is achieved, rather than with recognition ability itself. In Langdell’s study, a small set of highly familiar faces (of classmates) was used as stimuli, and no recognition impairment was reported in the children with autism. However, it seems likely that ceiling effects occurred in this study. The fact that two studies (Boucher & Lewis, 1992; de Gelder et al., 1991) have reported impaired recognition of unfamiliar faces suggests that recognition of familiar faces will be impaired in autism, and this was our predicted finding in Experiment 2.

**Method**

**Participants.** The same children who took part in Experiment 1 were assessed in Experiment 2 (see Table 1).

**Materials.** For each school the following materials were prepared:

1. **Training stimuli.** Black-and-white photographs of similar size and mountings to those used in Experiment 1 were used. Six photographs of the exteriors of both schools taken from various viewpoints made up a set of 12 training “buildings”. These were arranged in a predetermined order so that photographs of familiar buildings (the child’s own school) and unfamiliar buildings (the other school) were evenly distributed throughout the set, with no more than three photographs of either type occurring consecutively. The set began and ended with a highly recognisable photograph of the child’s own school. In addition, the 3 photographs of faces of highly recognisable staff at each school, which were used for training in Experiment 1, made up a set of 6 faces that was duplicated, making a complete set of 12 training “faces”. The 12 faces in the training set were arranged in a predetermined order as described above for the photographs of buildings.

2. **Test stimuli.** Twenty photographs of faces of adults working at each school (excluding those used in training) and the 12 photographs of buildings used in training, made up a set of 52 test stimuli. These were arranged in a predetermined order as described above. The set began and ended with a highly recognisable photograph of the child’s own school.

Two rectangular “post boxes” approximately 46 cm high were made, with a slot in the front of each large enough for the mounted photographs to be posted through. An unmounted, additional copy of the photograph of the child’s own school used as the first stimulus in the test set was prepared. This was used to identify one of the post boxes, as described in the Procedure.

**Procedure.** Children were seen individually in a quiet room in their own school. The training procedure was as follows. The two boxes were placed a few centimetres apart on a table with the slots facing forward, and equidistant from the child. The child sat or stood at the table (according to their ability to reach the slots easily), and the tester sat beside them. The child was shown the unmounted photograph of their own school and asked to identify it. Identification was reinforced by discussing the photo, naming the child’s school or using such phrases as “Your school”. The child was then invited to decide which box was to belong to their school, and to stick the photograph onto that box using Blutak. A note was made as to the position of the
Children with autism

The child's preferred label for the name, or the tester named, the person in the first picture (the face up in a pile in front of the child. The child was invited to 'rubbish' box. 'No', respectively, and the description of the S-box as the already mentioned included the frequent use of 'Yes' and choice of labels. Labels used by the children other than phrases to verbalise labels for each box, and reinforcing the child's training face photograph and the S

The second training picture was of the other school and the child was invited to post the second training picture into the S+ box. This procedure continued for the set of 12 training photographs of buildings was then placed face up in a pile in front of the child and equidistant from the two boxes. The child was asked to look at the first picture, which was of their own school and was a duplicate of the one on the front of the S+ box. The fact that the two photographs were the same, and that both were of the child's own school was pointed out, and the child was invited to post the first training picture into the S+ box. The second training picture was of the other school and the child was invited to post the second training picture into the S− box. This procedure continued for all 12 photographs of buildings, the tester encouraging the child to verbalise labels for each box, and reinforcing the child's choice of labels. Labels used by the children other than phrases already mentioned included the frequent use of "Yes" and "No", respectively, and the description of the S-box as the "rubbish" box.

The set of 12 training photographs of faces was then placed face up in a pile in front of the child. The child was invited to name, or the tester named, the person in the first picture (the Headteacher at their school). The child's preferred label for the S+ box was utilised to establish the link between the first training face photograph and the S+ box, and the tester ensured that the child posted the photograph in the S+ box. The second training face photograph was of someone at the other school, and the tester ensured that child posted the photograph in the S− box. Help was given on subsequent items only if needed to ensure correct responding. Training terminated when the child made four consecutive correct responses without help. All children included in the study met this criterion. The relative simplicity of the recognition task used in training ensured that this constituted response training rather than an assessment of face recognition ability.

The test procedure was similar to the procedure used in training except that errors were not corrected. However, if a child posted four photographs in succession into one of the boxes, the tester reminded the child that some photographs might be appropriately placed in the other box. Praise and encouragement were given as in Experiment 1.

Following the test the photographs were collected from each box and the numbers of correct and incorrect (false positive and false negative) responses were noted.

### Results

All the children posted all the pictures of buildings correctly, and these pictures were excluded from the analysis. The mean number of correctly posted photographs of faces, and standard deviations, are shown for each group and each school in Table 3. There was a guessing rate of 50% correct, i.e. 20 items out of 40, so it is clear that all the children in both groups performed well above chance. In fact seven children in the control group and three children in the autism group performed at ceiling.

Because of differences in the stimuli used to test the children at the two schools, the schools factor was incorporated in the analysis. A two-way analysis of variance (group by school) confirmed that the children with autism correctly posted fewer faces than the controls $[F(1,35) = 4.09, p = .05]$. There was no significant difference between the two schools $[F(1,35) = .98]$ and no significant interaction $[F(1,35) = .21]$. When CA and APT-I scores were included as additional predictors, neither was significant [for age, $F(1,33) = 1.3$; for APT-I, $F(1,31) = 0.3$] and there was no evidence that these predictors interacted with group. However, when the score on the voice–face identity matching task (Experiment 1) was included, this was a significant predictor $[F(1,33) = 5.2, p = .028]$ even with the main effect of group partialed out. The magnitude of this relationship is described by a partial correlation coefficient of .41. There was no evidence that this relationship differed between the groups since the interaction of group and voice–face identity matching score was not significant $[F(1,33) = 1.5, p > .2]$. This analysis also showed that, with voice–face identity matching scores considered as a covariate, the main effect of group was considerably weakened $[F(1,33) = 2.14, p = .15]$.

The distributions of face-recognition scores are shown in Fig. 2. The scores for the control children were bunched at or just below ceiling, whereas the modal score for the children with autism is somewhat lower. A separate-variances $t$-test confirmed that the difference between the means from the two groups was significant even when allowance was made for unequal variances $[t(31) = 2.261, p = .03]$. The results of Experiment 2 show that children with autism are significantly impaired relative to children of comparable language ability at recognising the faces of

### Discussion

The results of Experiment 2 show that children with autism are significantly impaired relative to children of comparable language ability at recognising the faces of
people to whom they have been exposed over a period of time. A significant impairment was demonstrated for the children with autism despite the fact that seven of the children in the control group were attached, and one might speculate that the impairment in the children with autism might have been more marked had ceiling effects been avoided. However, this was not a task that any of the children found particularly difficult, suggesting that impaired recognition of familiar faces is not a cause of autism, although it forms part of the pattern of social impairments.

The fact that all children performed at ceiling on the recognition of buildings provides firm evidence that the relatively poorer performance of the children with autism was caused neither by lack of motivation nor by any misunderstanding of the procedural aspects of the task. The autistic children’s ability to recognize buildings does not provide a control for the precise cognitive requirements of the task we used to assess face recognition, since classifying together different views of a particular school is not strictly analogous to classifying together photographs of different people working at a particular school. However, the fact that all the children performed well above chance indicates that they understood the cognitive requirements of the face recognition task. It might be suggested that the face recognition impairment in the children with autism results from a generalised visual processing deficit, and that their better performance on the recognition of buildings does not exclude this interpretation because of the ceiling effects that occurred on buildings recognition. This possibility cannot be ruled out. However, it seems unlikely in view of the weight of evidence suggesting that face processing involves dedicated neural systems.

Scores on the face recognition task (Experiment 2) were highly correlated with scores on the voice–face identity matching task (Experiment 1), despite the fact that ceiling effects on the face recognition task caused some reduction in the variance on this task. However, face recognition and voice–face identity matching were strikingly different in their relationships with CA and language scores. Explanations of these findings are explored in the General Discussion.

Experiment 3: A Test of Familiar Voice Recognition

There is no evidence concerning autistic people’s ability to recognize either unfamiliar (recently experienced) or familiar voices. However, the findings of Fine et al. (1991), Ricks (1975), and Klin (1989, 1992), referred to above, suggest that abnormalities of voice perception and encoding may be present, and this indirectly suggests that voice recognition may be impaired in autism. This was our prediction.

Experiment 3 was originally designed to assess familiar voice recognition using as stimuli the voices of the 20 adults from each school whose voices and faces were used as stimuli in Experiments 1 and 2. The experiment as originally designed was, however, initially carried out with the children attending School A, and marked floor effects occurred, four out of seven children with autism and four out of nine children in the control group performing at chance. For children at School B we therefore reduced the number of S+ stimuli to 15, omitting the voices of those 5 adults least often correctly matched by the children in the voice–face identity matching task. We also introduced a further modification designed to reduce the likelihood of floor effects. This modification consisted of a pretest procedure designed to establish for the child the pool of S+ voices that they would be required to identify in the main experiment.

Because of these changes in experimental method, Experiment 3 was carried out with children from School B only.

**Method**

**Participants.** The same children from School B who took part in Experiments 1 and 2 also took part in Experiment 3, except that one child in the original control group had left the school before Experiment 3 was carried out, and the linguistically most able child in the control group was therefore excluded in order to keep the experimental and control groups equated for language ability. Participant details for the remaining 12 children with autism and 8 controls are shown in Table 4.

**Materials**

1. **Pretest stimuli.** Photographs of the faces of the 15 adults from School B whose voices were to be used as S+ stimuli in this experiment were used in the preliminary procedure. The photographs were drawn from the test set used in Experiment 1.
2. **Training stimuli.** The voices of the three members of staff at School B and the three members of staff at School A that had been used as training stimuli in Experiment 1 were also used as training stimuli in the present experiment. For this experiment each voice was recorded saying: “Hello. Do you know who I am? Am I at your school?” After a 3-second pause, this was repeated. The six voices were recorded in random order, with a highly familiar voice (that of the Headteacher) occurring first on the tape. Each of the six recorded voices appeared three times on the training tape. Each voice was included once in the first six stimuli, once in the next six stimuli, and once in the final six stimuli. The order of voices was varied in each set of six stimuli.

3. **Test stimuli.** An audiotape was prepared of the voices of each of 15 staff at School B (S+ stimuli) and the voices of 15 staff at School A (S− stimuli). Each voice was recorded as for the training stimuli. S+ and S− stimuli were evenly distributed throughout the set, with no more than three voices of staff from either school occurring consecutively. The first voice on the tape was an S+ stimulus.

The two post boxes which were used in Experiment 2 were also used in Experiment 3.

**Procedure.** Prior to testing, photographs of the faces of the 15 familiar staff whose voices would be heard were presented to the child one at a time and the child was asked to name them. If the child could not name the individual, the tester provided a name and an identifying comment, such as “He’s the teacher in Class 2, isn’t he?” or “She’s the secretary. She’s in the office and answers the telephone when it rings.”

For training and for the test, the child sat at a table with the two boxes within easy reach on the table and equidistant from the child. The S+ box, with the photograph of the child’s own school, was positioned to the child’s left or right as selected by the child in Experiment 2. The tester sat beside the child. The training procedure was as follows. Children were reminded that
in a previous game they had posted photographs of their school and photographs of people at their school into the box with the picture of their school on it, and that other photographs had gone into “the rubbish box”. The child was then shown the tape recorder and told that “Someone is going to say something to you. Listen carefully.”. Both recordings of the first (very familiar) voice were then played, and the child was encouraged to respond to the recorded question “Am I at your school?” by touching the S+ box and/or by replying “Yes/My school” etc. Both recordings of the second (unfamiliar) voice were then played and the child was encouraged to respond by touching the S− box and/or by replying “No/Rubbish box” etc. This procedure was continued using the training tape until the child achieved three consecutive correct responses unaided.

The test procedure was the same as that used in training except that errors were not corrected. However, if a child indicated the same answer four times in succession, the tester reminded the child that some of the voices might appropriately belong in the other box. Praise and encouragement were given as in the other experiments. A record was kept as to whether the child’s response was correct or incorrect.

**Results**

Three children with autism failed to meet the training criterion and were excluded from the study. A further two children with autism reached criterion on training but performed at chance level on the first 15 (out of 30) voices. These two children showed some distress and unwillingness to continue and testing was terminated for ethical reasons. We therefore had complete results for 7 of the original group of 12 children with autism and for 8 control subjects. Details of the reduced group of 7 children with autism are shown in Table 4.

The mean number of voices correctly responded to by the children with autism in the reduced group was 17.43 (SD 2.1) out of a maximum of 30. A t-test showed that the autistic children’s scores were significantly above chance $[t(6) = 2.99, p < .024]$. The mean number of voices correctly responded to by the controls was 19.88 (SD 2.0), which was also well above chance $[t(7) = 6.79, p < .001]$. The difference between the means was significant $[F(1,13) = 5.1, p = .04]$. Across all 15 subjects, the voice recognition scores were significantly correlated with their voice–face identity matching scores from Experiment 1 ($r = .57, p < .03$). With voice–face identity matching scores as a covariate the main effect of group disappeared $[F(1,12) = 2.5, p = .14]$. There was no evidence that the relationship between voice–face identity matching and voice recognition scores differed between the two groups. Voice recognition scores did not correlate with the face recognition scores from Experiment 2, nor with CA or APT-I scores.

**Discussion**

The results of the test of voice recognition at School B show that children with autism are significantly impaired at recognising familiar voices relative to children with low language ability associated with SLI, MLD, or combinations of SLI with MLD. The significance level is not high ($p = .04$). However, the groups were small, and had we been able to include full scores for the 2 children with autism who reached the training criterion (i.e. who understood the task) but who performed at chance on the first 15 items, the difference between the 2 groups would have been greater. It is also possible that the three children with autism who failed to reach the training criterion did so because they have no ability to recognise voices, rather than because they failed to understand the task. The task used in Experiment 3 was based closely on that used in Experiment 2, which all the children had learned without difficulty. The simplest explanation of the three children’s training failure is, therefore, that they could not recognise even the three highly familiar and repeated voices used in training.

It might be suggested that the low scores of most of the children in this experiment (including several of the control group) reflect nonspecific task difficulty, rather than specific voice recognition difficulty. However, this is unlikely to be the case since voice recognition correlated significantly with performance on voice–face identity matching (Experiment 1), as was logically to be expected. This would not have occurred had low scores in Experiment 3 resulted from nonspecific difficulties, unless the same nonspecific difficulties impaired the performance of the children with autism in both Experiment 1 and Experiment 3. If that was the case then it seems odd that these nonspecific difficulties failed to produce a significant correlation between scores on Experiments 2 and 3.
General Discussion

The three experiments reported here add to existing evidence of impaired processing of social stimuli by children with autism. Experiment 1 shows that children with autism are impaired in matching the voices and faces of people to whom they have had as much exposure as controls. Experiment 2 demonstrated that recognition of familiar faces is impaired relative to controls, although neither group found this task difficult. Experiment 3 demonstrated impaired recognition of familiar voices.

The result of Experiment 1 can be explained in terms of the results of Experiments 2 and 3. This interpretation of the results of Experiment 1 is supported by the demonstration that face recognition and voice recognition both correlated significantly with voice–face identity matching, but not with each other. It is also supported by the observation that when voice–face identity matching scores were used as a covariate, the difference between the groups’ face recognition scores was considerably weakened, and the difference between their voice recognition scores disappeared. Alternative explanations of the finding in Experiment 1 in terms of impaired cross-modal processing or task complexity are therefore unlikely.

Regression analyses of data from the three experiments using CA and APT-I scores as covariates show that neither age nor expressive language ability are strongly related to either face recognition or voice recognition. Age and language ability were independently related to voice–face identity matching ability in both groups of children (though the relationship between language and matching ability was weak in the children with autism), but the group difference in matching ability remained highly significant when the effects of both age and language ability were partialled out. These results imply that the significant relationships between age and matching ability, and between language and matching ability, derive from relationships between these measures and some ability, or abilities, required for the matching task but not for the recognition tasks. Such abilities might include cross-modal processing or visual scanning ability. Language ability and cross-modal processing might well be related. However, it is not clear as to why CA should correlate significantly with either cross-modal processing or visual scanning: none of the children tested were developmentally normal, and one would have expected abilities such as these to correlate with developmental age rather than with CA. The question as to why CA and matching ability should have correlated so clearly in both groups of children is an intriguing one. However, it is not one we will pursue here, except to point out that the children with autism were somewhat, but not significantly, older than the controls, and would therefore have been slightly advantaged over the controls in whatever aspect of the task was age-related. This strengthens the finding of impaired matching ability in the children with autism.

Lack of correlation between language ability and familiar face recognition is not particularly surprising in view of the fact that face processing involves predominantly visual skills. However, our finding is at odds with others, who have shown that verbal ability is related to the ability to process facial expression (Braverman, Fein, Lucci, & Waterhouse, 1989; Ozonoff et al., 1990). It was, in fact, because of Braverman et al.’s and Ozonoff et al.’s findings that we equated our experimental and control groups for verbal rather than for nonverbal ability. So why might language ability relate to the comprehension of facial expression, but not to face recognition? It seems likely that the explanation lies in the fact that affect processing shares a communicative function with language, whereas face recognition serves an identifying rather than a communicative function. The surprising lack of correlation between language ability and familiar voice recognition in our experiments might also be explained in this way.

It might be suggested that the main findings of impaired voice–face identity matching and face and voice recognition in the children with autism can be explained in terms of nonspecific methodological factors. However, there are several reasons for arguing that this was not the case. In the first place, there is no reason to think that children with autism should find the task demands harder than other children with low language ability. In fact the children with autism may, as a group, have been slightly more able nonverbally than the control group, since the control group included some children with generalised learning difficulties whereas in autism low language ability is not usually associated with low nonverbal ability. In addition, it has already been pointed out that the children with autism would have been slightly advantaged by age, which correlated with performance in the matching task. In the second place, the suggestion that children with autism are unmotivated to co-operate with testing is generally false. In the face recognition task all the children correctly identified all the buildings, demonstrating satisfactory motivation as well as comprehension of the task demands. If the children with autism were well motivated to co-operate in one of the experiments there is no reason why they should have been unco-operative in the other two experiments. In the third place, task training and practice were used rather than verbal instruction, and the materials used in training were simple and repetitive, so that the children could concentrate on learning what they had to do without being distracted by difficult material. Also, children were not included in an experiment if they did not reach the training criterion. Finally, and most importantly, logically predictable patterns of correlation occurred between performance across the different tasks: individual children’s performance on voice–face identity matching correlated significantly with their performance on face recognition and on voice recognition; but voice recognition did not correlate with face recognition. These logical patterns of correlation would not have occurred if nonspecific task demands had significantly influenced the autistic children’s performance.

A second possible methodological criticism concerns our language matching procedure. This breaks down into three issues. The first issue concerns whether the somewhat lower language ability of the children with autism on the APT-I might explain the main results. This seems very unlikely. In the first place the difference in verbal ability was not significant. In the second place APT-I scores did not correlate with performance except in Experiment 1 (and here only weakly in the children with
autism), and in Experiment 1 the group differences in performance remained when the effects of language were partialled out. The second issue concerning our language matching procedure relates to the use of a test of expressive language, the APT-I. We chose a test of expressive language, and specifically the APT-I, because it requires children to interpret what is happening in pictures and to describe what is happening using generative language. We believe that the APT-I is for both these reasons a test that probes real-life language ability better than many other quick screening tests, and that this is likely to make it a relatively difficult test for children with autism (cf. Tsai & Beisler, 1984). This should bias against finding predicted impairments in the children with autism, relative to children of comparable language ability, and is in general a methodologically sound strategy. The third issue concerning language matching procedures is the inclusion of two boys in the control group who could not be tested on the APT-I and whose scores on the Test of Reception of Grammar (Bishop, 1982) were at or above the average expressive language ability of the remainder of the group (see Table 1). It might be suggested that the inclusion of these two boys influenced the main findings. Inspection of individual scores shows that this was not the case, since both boys scored at or below the average for the control group in all three experiments. Nevertheless, the inclusion of some children with SLI in the control group does raise the question as to whether comprehension ability in the control group as a whole might have been superior to comprehension in the autistic group, and, if so, whether this might explain our results on the tests involving the recognition of voices, in which the between-group differences were strongest. This issue is currently being explored in further work.

The question as to why children with autism have impaired face and voice processing also needs to be followed up. The most obvious hypothesis is that autistic children do not attend to the faces and voices of other people to the same extent as normally sociable children. This explanation is consistent with Klin’s findings (1989, 1992) on attention to voices, and with the social withdrawal and reduced eye contact that are typical of children with autism. However, further hypotheses might be proposed. For example, it might be the case that children with autism have perceptual processing difficulties for both voices and faces, which impairs encoding and subsequent recognition. There is certainly evidence of abnormal encoding of faces (Davies et al., 1994; Langdell, 1978), although no data are yet available on autistic children’s perceptual processing of voices. A third possibility is that children with autism know less about people in general, and about specific individuals in particular, because of specific sociocognitive impairments. These hypotheses raise interesting questions concerning what we mean when we talk about “familiar” faces or voices. In this paper we have specifically used the term “familiar” in the restricted sense of “having had exposure to” (or, to be even more accurate: “having had opportunities for exposure to”). The autistic children tested in our experiments had, in fact, had rather more opportunities than controls for exposure to the adults whose voices and faces were used as stimuli given that they had been in school for longer than the controls (see Table 1). However, the results suggest that they either did not, or could not, utilise these opportunities to achieve normal familiarity.

Acknowledgements—We would like to thank children and teachers from the Rowan School, Sheffield, and from the Uffcolme School, Birmingham, for their co-operation with this research.

References


Accepted manuscript received 18 March 1997