

Brief Report: Impaired Identification of Discrepancies Between Expressive Faces and Voices in Adults with Asperger's Syndrome

Kate O'Connor

Published online: 2 February 2007
© Springer Science+Business Media, LLC 2007

Abstract The aim of the present experiment was to examine the ability of adults with Asperger's syndrome and age-matched typically-developing controls to identify incongruent and congruent emotional information from the face and voice. In the first part of the experiment, participants determined whether simultaneously presented expressive faces and voices were the same or different. In the second part, participants identified expressive faces and voices in isolation. Results showed that relative to controls, adults with AS were less accurate at distinguishing between congruent and incongruent expressive faces and voices. Both groups obtained similar accuracy to expressive faces and voices presented in isolation. These findings may partially explain some of the difficulties individuals on the autistic spectrum have with social interaction.

Keywords Asperger's syndrome · Sensory integration · Expression

Although different individuals with the neurodevelopmental disorder Asperger's syndrome (AS) have described a desire to interact successfully with others they generally experience great difficulty (Attwood, 1998; Birch, 2003; Miller, 2003; Sainsbury, 2000). Unfortunately, their seemingly awkward approach to social interactions, unusual mannerisms and behaviour

often result in bullying, rejection and isolation (Birch, 2003; Jackson & Attwood, 2002; Miller, 2003).

An ability to accurately interpret and integrate information from the face and voice is an important skill for successful social interaction. Typically-developing individuals show a natural tendency to integrate simultaneously presented information in the face and voice during speech perception. This is observed in the McGurk effect (McGurk & McDonald, 1976), where dubbing an audible syllable onto a video of a face mouthing a different syllable classically results in subjects hearing a third syllable produced from the subconscious merging of visual and auditory information. Similarly, integration of facial and vocal information has been observed during the concurrent presentation of expressive faces and voices. For example, a sad face is generally perceived as less sad if presented simultaneously with a happy voice (de Gelder & Vroomen, 2000; de Gelder, Vroomen, de Jong, Masthoff, Tromenaars & Hodiamont, 2005). Likewise, a happy voice tends to be perceived as less happy in the presence of a sad face (de Gelder & Vroomen, 2000; de Gelder et al., 2005).

Two research groups have shown that typically-developing individuals are faster and more accurate at identifying simultaneously presented face and voice expressions when stimuli are emotionally congruent rather than emotionally incongruent (de Gelder & Vroomen, 2000; Dolan, Morris, & de Gelder, 2001). This may be because emotional information portrayed in the voice is often reflected on the face at the same time.

However, emotional expressions in the voice and face are not always congruent, especially in the context of more complex social situations. For example, a friend may (out of politeness) have a kind tone to their

K. O'Connor (✉)
Psychology Department, University of Auckland, Private
Bag, 92109 Auckland, New Zealand
e-mail: gamokatiao@yahoo.co.nz

voice when informing you that your speech was really interesting, despite having a slightly bored facial expression. Furthermore, an individual may sound unhappy, yet still smile in the company of others to try to conceal their sadness. It is important to be able to recognize these discrepancies in order to respond appropriately to the needs of another individual.

Although past research has not specifically examined processing of congruent and incongruent face and voice expressions in AS or Autistic Disorder, some research groups have examined processing of emotional auditory and visual information in autistics. For example, several studies have found adolescents and children with autism have difficulty selecting appropriate facial expressions to match emotional nonverbal vocalizations relative to typically-developing controls matched for mental and/or chronological age (Hobson, 1986; Hobson, Ouston, & Lee, 1988; Loveland, Tunali-Kotoski, Chen, Brelsford, & Ortegon, 1995). A more recent study (Hall, Szechtman, & Nahmias, 2003) found that relative to typically-developing adult controls, adults on the autistic spectrum were less accurate at matching expressive voices to facial expressions.

The purpose of the present experiment was to investigate processing of expressive faces and voices in Autism Spectrum Disorder further, by determining whether adults with AS have difficulty discriminating incongruent from congruent emotional faces and voices.

Methods

Participants

Eighteen adults with AS and 18 age- and gender-matched typically-developing controls were tested. Controls were selected from the Auckland community, while adults with AS were recruited from the Auckland Autistic Association. Adults with AS had been diagnosed by a registered medical professional experienced with autistic spectrum disorders according to DSM-IV (APA, 1994) criteria. In addition, AS subjects met both the DSM-IV and Gillberg and Gillberg (1989) criteria for AS when reassessed in the laboratory. Four adults with AS were on medication for the treatment of depression and/or anxiety. IQ was not tested, since individuals with AS often show an uneven profile of abilities on IQ tests which does not reflect their academic attainment (Attwood, 1998). However, both groups had received a similar number of years of education and were not behind their peers in terms of

academic achievement. Participants with an existing neurological condition (epilepsy, head injury, significant sensorimotor impairment, schizophrenia or dyslexia) were excluded and all participants had normal or corrected-to-normal visual acuity.

Control and AS subjects had a mean age of 25.2 years (SD = 6.5 years, range = 19–47 years) and 26.9 years (SD = 7.8 years, range = 19–50 years), respectively, each group consisting of 16 males and 2 females. All procedures were approved by the University of Auckland Human Subjects Ethics Committee and written consent was obtained from each subject prior to participation.

Stimuli

Visual Stimuli

Colour photographs of unfamiliar actors expressing the basic expressions happy, sad and angry were selected from QuickTime files in the Mind Reading Emotions Library (Baron-Cohen, Golan, Wheelwright, & Hill, 2003). All photographs were cropped to remove the ears, shoulders and part of the hair so that the face was the central focus. The final size of each photograph was 7.1 × 5.3 cm. Greyscale images were obtained using Adobe Photoshop software so that skin colour and tone did not detract from emotional expression. Each expression was rated by 48 typically-developing adults and expressions with a high degree of accuracy (>85%) were selected for the present task (Fig. 1). The final stimulus set contained photographs of eight adults (four males and four females) expressing each of the three emotions (a total of 24 stimuli).

Auditory Stimuli

Six female and eight male actors were instructed to pronounce a semantically neutral sentence (“I want to



Fig. 1 Examples of angry, sad and happy facial expressions. Each expression was presented by eight different people, four male and four female. Adapted from Baron-Cohen et al. (2003); © University of Cambridge, United Kingdom. Used with permission

go to the other movies”) in a happy, sad and angry voice. These sentences were recorded on a computer connected to an external microphone using Adobe Audition software. Each sentence was rated by 14 typically-developing adults and the four most accurate sentences from male and female actors were chosen for each expression (all recognized with 80% or greater accuracy).

Visual and auditory stimuli were presented using E-Prime (version 1.0 Beta 5) software run on a lap-top computer (14-inch screen).

Procedure

Participants were seated 70 cm from the computer screen in a quiet room at The University of Auckland Department of Psychology. In the first half of the experiment (Part 1), subjects were exposed to simultaneously-presented face and voice expressions (Fig. 2). Subjects were instructed to press the ‘Sm’ keyboard button if the face and voice expressions were the same and the ‘Di’ button if they were different (corresponding to letters z and m on the keyboard, respectively). Auditory and visual stimuli were presented for a mean duration of 2.8 s. Once a response was made, a blank-screen was presented for 1000 ms before the next trial began.

Expressive faces and voices were paired to produce 24 same (congruent) and 24 different (incongruent) combinations. Male voices were only matched with male faces and likewise for females. Trials were presented twice at random over two blocks, each block consisting of 48 trials, balanced with respect to gender and condition (congruent, incongruent). Blocks were separated by a 5-min rest interval. Prior to the experiment, all subjects were required to participate in a brief practice session to ensure they understood the task.

The second half of the experiment (Part 2) was comprised of two blocks separated by a 5-min rest

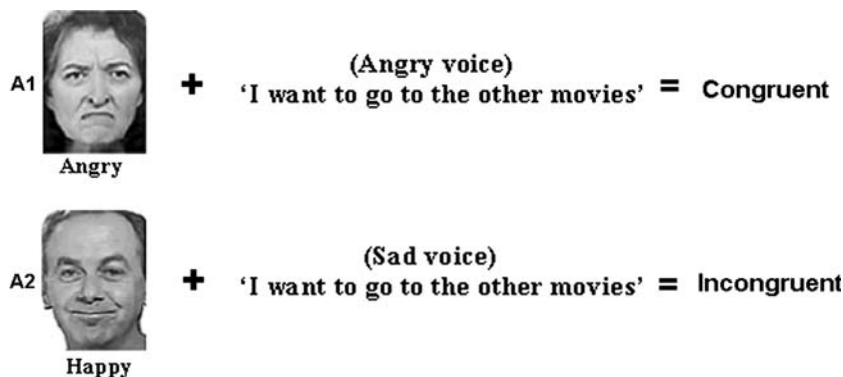
interval. Facial expressions were presented in one block, while expressive voices were presented in the other. The purpose of this was to confirm whether any potential between group differences in Part 1 could be attributed to discrimination between congruent and incongruent stimuli or identification of expressive faces and/or voices presented in isolation.

Participants were required to identify happy, angry and sad expressions in each block by pressing the Ha, An and Sa labelled computer keys (corresponding to letters a, g and l on the keyboard). Each block consisted of the three expressions made by all eight actors, resulting in a total of 24 trials per block. The mean duration of auditory stimuli was 2.6 s, while visual stimuli were presented for 3.0 s. Once a response was made, a blank-screen was presented for 1000 ms before the next trial began. Trials were presented at random within each block and blocks were randomized across groups so that half the control and AS subjects were administered the facial expression block first and the other half the expressive voice block. As this task was relatively straight-forward, participants were not required to take part in a practice session.

Data analysis

Part 1: Accuracy scores were converted to percentages and analysed using a 2-way analysis of variance (ANOVA) with the factors group (control, AS) and condition (congruent, incongruent). A signal detection theory (SDT) type analysis was also employed to measure the discrimination ability or sensitivity (d') of each subject to discriminate between congruent and incongruent trials (Green & Swets, 1966; Macmillian & Creelman, 1991). The analysis of d' over accuracy values is advantageous as d' removes any response bias an individual may have. The d' values for each subject were calculated using the percentage of ‘hits’ (correct identification of incongruent trials) and ‘false alarms’ (incorrect identification of congruent trials), thus

Fig. 2 Example of a congruent (A) and incongruent (B) trial. Adapted from Baron-Cohen et al. (2003); © University of Cambridge, United Kingdom. Used with permission



representing the ability of control and AS subjects to discriminate incongruent from congruent trials. Low d' values indicate difficulty discriminating incongruent from congruent expressions in the face and voice, while high d' values indicate superior ability during this discrimination. Mean d' values were calculated for each group and examined using an unpaired t -test (2-tailed).

Part 2: The percentage of trials angry, sad and happy expressions were correctly identified in the facial expression and expressive voice blocks were calculated for each subject. Data was analysed using a repeated measures ANOVA with the factors group (control, AS) and expression (angry, happy, sad) for each block.

Results

Part 1: The mean accuracy for congruent trials was 87.1% (SD = 5.6) in control and 82.4% (SD = 11.9) in AS subjects, while mean accuracy for incongruent trials was 85.9% (SD = 7.1) and 73.6% (SD = 11.3), respectively (Fig. 3). ANOVA revealed a significant effect for group ($F(1, 34) = 14.58, P = 0.001$), with higher accuracy scores in control relative to AS subjects. A main effect of condition was also observed ($F(1, 34) = 5.13, P = 0.03$), with both groups obtaining higher accuracy to congruent relative to incongruent trials. The interaction between group and condition did not reach significance ($F(1, 34) = 3.10, P = 0.11$).

The mean d' value was 2.34 (SD = 0.53) for control and 1.70 (SD = 0.59) for AS subjects. An unpaired t -test found the difference between these values to be significant ($t(34) = 3.42, P = 0.002$). This indicates that in comparison to controls, adults with AS had more

difficulty discriminating between congruent and incongruent trials (i.e., determining whether expressive faces and voices were the same or different).

Part 2: The mean accuracy for happy, angry and sad facial expressions was 98.8% (SD = 3.2), 97.2% (SD = 4.6) and 90.6% (SD = 8.0) in control and 97.2% (SD = 5.8), 94.4% (SD = 8.6) and 87.2% (SD = 11.3) in AS subjects, respectively (Fig. 4A). ANOVA revealed a significant effect for expression ($F(2, 68) = 16.43, P < 0.001$), with higher accuracy to angry and happy relative to sad faces ($t(34) = 3.86, P = 0.001$ and $t(34) = 5.04, P < 0.001$, respectively). Similar performance was obtained for angry and happy expressions. The group effect and the group by facial expression interaction were not significant ($F(1, 34) = 2.81, P = 0.10$ and $F(2, 68) = 0.13, P = 0.88$, respectively).

The mean accuracy for happy, angry and sad voice expressions was 81.1% (SD = 15.3), 84.4% (SD = 10.9) and 88.8% (SD = 9.0) in control and 77.2% (SD = 18.4), 83.9% (SD = 14.6) and 86.1% (SD = 12.9) in AS subjects, respectively (Fig. 4B). ANOVA revealed a significant effect for expression ($F(2, 68) = 4.98, P = 0.01$), with higher accuracy to sad relative to happy voices ($t(34) = 2.77, P = 0.03$). All other comparisons were not significant. There was no significant effect of group ($F(1, 34) = 0.48, P = 0.49$) or interaction between group and voice expression ($F(2, 68) = 0.21, P = 0.82$).

These results show that adults with AS did not differ significantly from controls when identifying happy, sad and angry expressions from the face or voice presented in isolation.

Discussion

The present study shows that in comparison to typically-developing controls, adults with AS are less accurate at discriminating incongruent from congruent expressive faces and voices. This result was not due to problems identifying happy, sad or angry faces and voices, as both groups obtained similar accuracy to expressive faces and voices presented in isolation. These findings suggest that adults with AS may have difficulty integrating face and voice expressions, consistent with past studies in children and adults on the autistic spectrum (Hall et al., 2003; Hobson, 1986; Hobson et al., 1988; Loveland et al., 1995).

Successful social interaction requires the ability to simultaneously integrate information from multiple sensory modalities. Distinguishing between congruent and incongruent emotional information in the face and

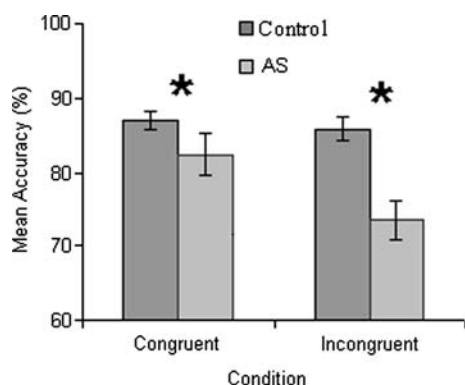
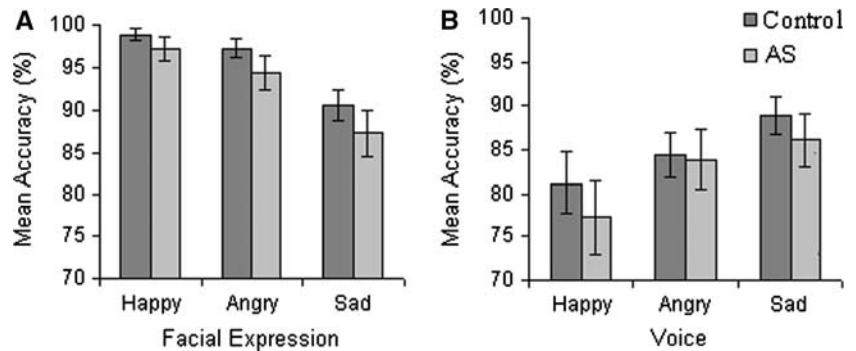


Fig. 3 Mean accuracy (percentage correct) for congruent and incongruent face and voice expressions in control and AS adults. Significant differences between groups were observed for both congruent and incongruent trials (indicated by asterisks). Error bars represent standard errors

Fig. 4 Mean accuracy (percentage correct) for control and AS adults on the (A) Emotional Voice Recognition and (B) Facial Expression Recognition blocks. There were no significant differences between groups for any condition. Error bars represent standard errors



voice is particularly important when the expression in an individual's voice is not necessarily the same as the expression on their face. For example, the hypothetical speaker discussed in the introduction would continue to deliver boring talks if they were unable to identify incongruent information in the face and voice of the listener. The difficulty adults with AS had discriminating congruent from incongruent face and voice expressions in the present task may explain some of the problems they experience with social interaction in daily life.

Several evoked potential studies have found individuals on the autistic spectrum to exhibit delayed latencies of the N170 face-sensitive component to faces relative to typically-developing controls (McPartland, Dawson, Webb, Panagiotides, & Carver, 2004; O'Connor, Hamm, & Kirk, 2005). Any delay in processing facial information during social interactions would likely disrupt contingencies necessary for associative learning as suggested by McPartland et al. (2004). For example, delayed processing of facial information could lead to inaccurate association of facial expression with expressive information in the voice. Consequently, this may explain why adults with AS were less accurate at distinguishing congruent from incongruent face and voice expressions in this experiment.

Moreover, given the large amount of information available during social interactions the superior ability of many individuals on the autistic spectrum to attend to detail (Happé & Frith, 2006; Mottron, Dawson, Soulières, Hubert, & Burack, 2006; Plaisted, O'Riordan, & Baron-Cohen, 1998) would make social information particularly difficult to process. The potential distraction of details that most typically-developing individuals do not even notice or filter out could potentially result in reduced integration of information needed to interpret social situations such as face and voice expressions. Consequently, this could result in reduced expertise at recognizing similarities and differences between expressions in the face and voice,

thus offering an alternative explanation for the findings in Part 1 of this study.

In summary, these results suggest that adults with AS are less accurate at discriminating incongruent from congruent expressive faces and voices relative to typically-developing subjects. Increased attention to detail or delayed processing of facial information may result in less experience at integrating auditory and visual information, thus contributing to the present findings. It is possible that impaired integration of face and voice expressions may underlie some of the difficulties individuals with AS have interacting with others.

Acknowledgments This research was supported by funding from the Ministry of Research, Science and Technology, New Zealand. I would like to thank the adults with AS and control subjects who participated in this study. Also thank you to Joss O'Connor for commenting on an earlier version of this manuscript and Dr Jeff Hamm for assistance with the signal detection analysis.

References

- Attwood, T. (1998). *Asperger's syndrome: A guide for parents and professionals* (1st ed.). London: Jessica-Kingsley Publishers Ltd.
- APA. (1994). *DSM-IV diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.
- Baron-Cohen, S., Golan, O., Wheelwright, S., & Hill, J. (2003). *Mind reading emotions library*. London: Jessica-Kingsley Publishers Ltd (www.jkp.com/mindreading).
- Birch, J. (2003). *Congratulations! It's Asperger's syndrome*. London: Jessica-Kingsley Publishers.
- de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and eye. *Cognition and Emotion*, 14, 289–311.
- de Gelder, B., Vroomen, J., de Jong, S., Masthoff, E., Trome-naars, F., & Hodiamont, O. (2005). Multisensory integration of emotional faces and voices in schizophrenics. *Schizophrenia Research*, 72, 195–203.
- Dolan, R., Morris, J., & de Gelder, B. (2001). Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences*, 98(17), 10006–10010.

- Gillberg, C., & Gillberg, I. (1989). Asperger syndrome - some epidemiological considerations: A research note. *Journal of Child Psychology and Psychiatry*, 30, 631–638.
- Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Hall, G., Szechtman, H., & Nahmias, C. (2003). Enhanced salience and emotion recognition in autism: A PET study. *American Journal of Psychiatry*, 160(8), 1439–1441.
- Happe, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 36(1), 5–25.
- Hobson, R. (1986). The autistic child's appraisal of expressions of emotion. *Journal of Child Psychology and Psychiatry*, 27(3), 321–342.
- Hobson, R., Ouston, J., & Lee, A. (1988). Emotion recognition in autism: Coordinating faces and voices. *Psychological Medicine*, 18, 911–923.
- Jackson, L., & Attwood, T. (2002). *Freeks, geeks and Asperger's syndrome*. London: Jessica Kingsley.
- Loveland, K., Tunali-Kotoski, B., Chen, R., Brelsford, K., & Ortegon, J. (1995). Intermodal perception of affect by persons with Autism or Down syndrome. *Development and Psychopathology*, 7, 409–418.
- MacMillan, N., & Creelman, C. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
- McGurk, H., & McDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748.
- McPartland, J., Dawson, G., Webb, S., Panagiotides, H., & Carver, L. (2004). Event-related brain potentials reveal anomalies in temporal processing of faces in autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, 45(7), 1235–1245.
- Miller, J. (2003). *Women from another planet?: Our lives in the Universe of autism*. Milan: Dancing Mind Books.
- Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, 36(1), 27–43.
- O'Connor, K., Hamm, J., & Kirk, I. (2005). The neurophysiological correlates of face processing in adults and children with Asperger's syndrome. *Brain and Cognition*, 59(1), 82–95.
- Plaisted, K., O'Riordan, M., & Baron-Cohen, S. (1998). Enhanced discrimination of novel, highly similar stimuli by adults with autism during a perceptual learning task. *Journal of Child Psychology and Psychiatry*, 39, 765–775.
- Sainsbury, C. (2000). *Martian in the Playground*. London: The Book Factory.