Development and Psychopathology **18** (2006), 591–617

Printed in the United States of America
DOI: 10.1017/S0954579406060305

REGULAR ARTICLE

Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia

OFER GOLAN AND SIMON BARON-COHEN
Autism Research Centre, University of Cambridge

Abstract

This study evaluated *Mind Reading*, an interactive systematic guide to emotions, for its effectiveness in teaching adults with Asperger syndrome (AS) and high-functioning autism (HFA) to recognize complex emotions in faces and voices. Experiment 1 tested a group of adults diagnosed with AS/HFA ($n = 19$) who used the software at home for 10–15 weeks. Participants were tested on recognition of faces and voices at three different levels of generalization. A matched control group of adults with AS/HFA ($n = 22$) were assessed without any intervention. In addition, a third group of general population controls ($n = 24$) was tested. Experiment 2 repeated the design of Experiment 1 with a group of adults with AS/HFA who used the software at home and met in a group with a tutor on a weekly basis. They were matched to a control group of adults with AS/HFA attending social skills training and to a general population control group ($n = 13$ for all three groups). In both experiments the intervention group improved significantly more than the control group on close, but not distant, generalization tasks. Verbal IQ had significant effects in Experiment 2. Using *Mind Reading* for a relatively short period of time allows users to learn to recognize a variety of complex emotions and mental states. However, additional methods are required to enhance generalization.

The ability to recognize and understand emotions and mental states develops in the first year of life, with infants able to discriminate different emotions from facial expressions and vocalizations (Fernald, 1992; Haviland & Lelwica, 1987). Development of emotion recognition skill continues through childhood and adulthood as part of "theory of mind" (Astington, Harris, & Olson, 1988), or what is also referred to as "mindreading" (Wellman, 1992) or "empathizing" (Baron-Cohen, 2002, 2003).

It is well established that emotion recognition and mental state recognition are core difficulties in people with autism spectrum conditions (ASCs; Baron-Cohen, 1995; Hobson, 1994). These underlie the social difficulties that are diagnostic. Such difficulties have

The first author was supported by the National Alliance for Autism Research, the Corob Charitable Trust, the Cambridge Overseas Trust and B'nai B'rith Leo Baeck scholarships. The second author was supported by the Shirley Foundation, the Medical Research Council, and the Three Guineas Trust. We are grateful to the following for helping with this study: our participants, Red Green and Blue Ltd, Jessica Kingsley Ltd, Pamela Yates for arranging and running the social skills groups, Andrea Macleod and Rob Whiskens (Autism West Midlands), Jill Howard (Autism London), Lynne Moxon (European Services for People with Autism), and Harriet Fisher and James Graham (The Interact Centre). Many thanks to Jacqueline Hill, Chris Ashwin, Sally Wheelwright, Matthew Belmonte, Yael Golan, Sarah Johnson, Emma Chapman, and Ilaria Minio Paluello. A version of this work was presented at the International Meeting For Autism Research (IMFAR), May 2004.

Address correspondence and reprint requests to: Ofer Golan, Autism Research Centre, University of Cambridge, Department of Psychiatry, Douglas House, 18b Trumpington Road, Cambridge CB2 2AH, UK; E-mail: og211@cam.ac.uk.

been found through cognitive, behavioral, and neuroimaging studies, and across different sensory modalities, both visual and auditory (Frith & Hill, 2004).

Because the human face is central in both the expression and communication of emotion, the majority of studies have focused on the face and tested recognition of six emotions (happiness, sadness, fear, anger, surprise, and disgust). These "basic emotions" are expressed and recognized universally (Ekman, 1993; Ekman & Friesen, 1971). Some studies reveal emotion recognition deficits among individuals with ASC, compared to typical or clinical control groups, using both static (Celani, Battacchi, & Arcidiacono, 1999; Deruelle, Rondan, Gepner, & Tardif, 2004; Macdonald et al., 1989) and dynamic stimuli (Hobson, 1986a, 1986b; Yirmiya, Sigman, Kasari, & Mundy, 1992). Other studies have found children and adults with high-functioning autism (HFA) or Asperger syndrome (AS) have no difficulties in recognizing these basic emotions from pictures (Adolphs, 2001; Grossman, Klin, Carter, & Volkmar, 2000) or films (Loveland et al., 1997), and that the deficit only becomes apparent when testing recognition of more "complex" emotions (such as embarrassment, insincerity, intimacy, etc.) in both adults and children with ASC (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Baron-Cohen, Wheelwright, & Jolliffe, 1997; Baron-Cohen, Wheelwright, Spong, Scahill, & Lawson, 2001; Golan, Baron-Cohen, & Hill, in press). These findings suggest recognition of basic emotions is relatively preserved among high-functioning individuals with ASC, and that they show greater difficulties recognizing more complex emotional and mental states.

Neuroimaging studies of emotion recognition from faces reveal that people with ASC show less activation in brain regions central to face processing, such as the fusiform gyrus (Critchley et al., 2000; Pierce, Muller, Ambrose, Allen, & Courchesne, 2001; Schultz et al., 2003). Behavioral studies show that children and adults with ASCs process faces differently compared to controls: participants with ASC tend to process faces in a feature-based approach, whereas controls process faces configurally (Hobson, Ouston, & Lee, 1988; Schultz et al., 2003; Teunisse & De Gelder, 1994; Young, 1998). There is also evidence of reduced activation in brain areas that play a major role in processing of emotion, such as the amygdala, when individuals with ASC process social–emotional information (Ashwin, Baron-Cohen, Wheelwright, O'Riordan, & Bullmore, 2005; Baron-Cohen et al., 1999; Critchley et al., 2000).

Emotion recognition from voices has been studied less frequently. Here, too, there are contradictory findings in relation to recognition of basic emotions (Boucher, Lewis, & Collis, 2000; Loveland, Tunali Kotoski, Chen, & Brelsford, 1995; Loveland et al., 1997). Regarding recognition of complex emotions from voices, several studies report a deficit in performance in high-functioning adults with ASC compared to controls (Golan et al., in press; Golan, Baron-Cohen, Hill, & Rutherford, 2005; Kleinman, Marciano, & Ault, 2001; Rutherford, Baron-Cohen, & Wheelwright, 2002). To date, there are no reported neuroimaging studies assessing emotion recognition in voices in ASC, perhaps because the sound of the scanner itself (the echo-planar system) makes auditory brain scanning difficult. One study used positron emission tomography to measure brain activity of participants with ASC and matched controls while listening to theory of mind stories. Both groups showed similar patterns of brain activity, but activation in the medial frontal area of the brain was less intensive and extensive in the AS group, compared to controls (Nieminen-von Wendt et al., 2003).

Studies assessing the ability of individuals with ASC to identify emotions and mental states from context have also shown deficits relative to the general population or to other clinical control groups (Baron-Cohen, Leslie, & Frith, 1986; Fein, Lucci, Braverman, & Waterhouse, 1992). For example, adolescents and adults with ASC have difficulties answering questions on the Strange Stories Test (Happe, 1994; Jolliffe & Baron-Cohen, 1999). This test assesses the ability to provide context-appropriate mental state explanations for nonliteral statements made by story characters (e.g., ironic or sarcastic statements).

When using this task in a neuroimaging study, reduced activation of the left medial prefrontal cortex was found in people with ASC compared to matched controls (Happe et al., 1996).

The integration of cross-modal emotional information from faces, voices, and context allows understanding and prediction of others' emotions and mental states (Baron-Cohen, 1995), and the more information available, the easier should be the recognition of these emotions or mental states. Studies assessing complex emotion and mental state recognition from ecologically rich social situations, containing multimodal sources of information, show a deficit in individuals with ASC, compared to controls (Golan, Baron-Cohen, Hill, & Golan, in press; Heavey, Phillips, Baron-Cohen, & Rutter, 2000; Klin, Jones, Schultz, Volkmar, & Cohen, 2002). These difficulties may be related to a failure to pick up the right emotional cues, and/or to a failure integrating them, explained by weak central coherence in the cognitive level (Frith, 1989), and underconnectivity between brain regions in the neurobiological level (Belmonte, Allen, et al., 2004; Belmonte, Cook, et al., 2004; Critchley et al., 2000). Although emotion recognition deficits in ASC are life-long, some high-functioning individuals develop compensatory strategies that allow them to recognize basic emotions. However, when recognition of more complex emotions and mental states is required, many find them hard to interpret.

In contrast to these difficulties, individuals with ASC show good and sometimes even superior skills in "systemizing" (Baron-Cohen, 2003). Systemizing is the drive to analyze or build systems, to understand and predict the behavior of nonagentive events in terms of underlying rules and regularities. Individuals with ASC are hyperattentive to detail and prefer predictable, rule-based environments, features that are intrinsic to systemizing. In addition, individuals with ASC are superior to controls on various tasks that involve searching for detail, analyzing and manipulating systems (Baron-Cohen, Richler, Bisarya, Gurunathan, & Wheelwright, 2003; Baron-Cohen, Wheelwright, Spong, et al., 2001; Jolliffe & Baron-Cohen, 1997; Lawson, Baron-Cohen, & Wheelwright, 2004; O'Riordan, Plaisted, Driver, & Baron-Cohen, 2001; Shah & Frith, 1993).

If high-functioning individuals with ASC possess good systemizing skills, it is possible they could use them to compensate for some of their empathizing difficulties. This might be hard to implement, because the socioemotional world is a context-related open system (Lawson, 2003), often unpredictable and difficult to conceptualize with strict rules. However, if provided with a *system* of emotions, it is plausible that systemizing skills could be harnessed to help individuals with ASC learn to recognize emotions.

Past attempts to teach emotion recognition to adults and children with ASC have either focused on the basic emotions (Hadwin, Baron-Cohen, Howlin, & Hill, 1996; Howlin, Baron-Cohen, & Hadwin, 1999), or have been part of social skills training courses, usually run in groups (Barry et al., 2003; Howlin & Yates, 1999; Rydin, Drake, & Bratt, 1999). These training programs typically do not focus specifically on systematically teaching emotion recognition, but instead address other issues, such as conversation, reducing socially inappropriate behavior, personal hygiene, and so forth. In such groups it is difficult to target the individual's specific pace of learning. Finally, such groups are socially demanding, and might therefore deter more socially anxious participants.

Other attempts to teach individuals with ASC social skills have used computer-based training (Bernard-Opitz, Sriram, & Nakhoda-Sapuan, 2001; Bolte et al., 2002; Hetzroni & Tannous, 2004; Rajendran & Mitchell, 2000; Silver & Oakes, 2001; Swettenham, 1996). The use of computer software for individuals with ASC has several advantages: first, individuals with ASC favor the computerized environment because it is predictable, consistent, and free from social demands, which they may find stressful. Second, users can work at their own pace and level of understanding. Third, lessons can be repeated over and over again, until mastery is achieved. Fourth, interest and motivation can be maintained through different and individually selected computerized rewards (Bishop, 2003; Moore, McGrath, & Thorpe, 2000; Parsons & Mitchell, 2002). Previous studies have found that the use of computers can help individuals with autism pass false belief tasks (Swettenham, 1996), recognize basic emotions from cartoons and still photographs (Bolte et al., 2002; Silver & Oakes, 2001), and solve problems in illustrated social situations (Bernard-Opitz et al., 2001). However, participants find it hard to generalize their knowledge from learnt material to related tasks.

The computer-based interventions above used drawings or photographs for training, rather than more lifelike stimuli. This might have made generalization harder than if more ecologically valid stimuli were used. In addition, the programs teaching emotion-recognition focused on basic emotions, and only on facial expressions. No reported program to date has systematically trained complex emotion recognition in both visual and auditory channels, with lifelike faces and voices. The study reported here evaluates such a program in two formats: individually based, and via a group intervention. The question tested in both was "Can the good systemizing skills that individuals with ASC possess be used to teach them to improve their recognition of complex emotions?" To test this question, emotions and mental states had to be presented as a *system* for learners to analyze and study.

Mind Reading: A Systematic Guide to Emotions

Mind Reading (Baron-Cohen, Golan, Wheelwright, & Hill, 2004) is an interactive guide to emotions and mental states. It is based on a taxonomic system of 412 emotions and mental states, grouped into 24 emotion groups, and six developmental levels (from age 4 to adulthood). The emotions and mental states are organized systematically, according to the emotion groups and developmental levels. Each emotion group is introduced and demonstrated by a short video clip giving some clues for later analysis of the emotions in this group. Each emotion is defined and demonstrated in six silent films of faces, six voice recordings, and six written examples of situations that evoke this emotion. The resulting library of emotional "assets" (video clips, audio clips, or brief stories) comprise $412 \times 18 = 7{,}416$ units of emotion information to learn to recognize or understand. This is therefore a rich and systematically organized set of educational material. The face videos and voice recordings comprise actors of both genders, various ages, and ethnicities, to facilitate generalization. Faces and voices are presented separately for each emotion (i.e., silent face films and faceless voice recordings) to encourage analysis of the emotion in each modality. All face video clips and voice recordings were validated by a panel of 10 independent judges, and were included in *Mind Reading* if at least 8 judges agreed the emotional label given described the face/voice ($p < .05$, binomial test).

This emotion database is accessed using three applications: (a) *The emotion library* allows users to browse freely through the different emotions and emotion groups, play the faces, voices and scenarios giving examples of the emotions, read stories, add their own notes, and compare different emotional expressions in the face and the voice using a scrapbook. (b) *The learning center* uses lessons, quizzes, and several reward collections to teach about emotions in a more structured and directive way. In addition to teaching about the 24 emotion groups, it also includes lessons and quizzes about the 20 and 100 most commonly used emotions, as well as a "build your own lesson/quiz" option. The various reward collections were chosen for their potential appeal to users with ASC and were arranged systematically (e.g., pictures and information about space elements, clips of birds arranged by families, different types of trains to collect, etc.). A reward is given when a quiz question is answered correctly on the first attempt. (c) *The game zone* comprises five educational games, allowing users to enjoy a game while studying about emotions. The software was created for the use of children and adults of various levels of functioning. Vocal and animated helpers give instructions on every screen. Figure 1 shows screen shots from the software (www.jkp.com/mindreading).

The study reported here includes two experiments. Experiment 1 tested for any improvement in adults with AS/HFA in emotion

Figure 1. Screenshots from *Mind Reading: The Interactive Guide to Emotions*, by S. Baron-Cohen, O. Golan, S. Wheelwright, & J. J. Hill, 2004, London: Jessica Kingsley Limited. Copyright 2003 by the University of Cambridge. Reprinted with permission. [A color version of this figure can be viewed online at www.journals.cambridge.org]

recognition skills following *independent* use of the software, and the extent to which these users can generalize their acquired knowledge. Experiment 2 tested the same in a group of participants who, in addition to using the software individually, attended a weekly *group* session led by a *tutor*. Using group discussion, role play, worksheets, and analysis of emotions in newspapers and television programs, this group aimed to consolidate participants' computer-acquired knowledge to improve generalization. Both interventions took place over a period of 10–15 weeks, to assure a meaningful period for training, recognizing that a longer duration might lead to individuals dropping out. This allowed us to test whether the educational software (used in both experiments) has an effect independent of any human tutor or social group effects, or whether addition of a human tutor and social group enhances any effects. In both experiments, participants were tested before and after the intervention. A no-computer intervention control group of adults with AS/HFA was matched to the intervention group in both experiments. This AS/HFA control group was also tested before and after a similar period of time. In Experiment 1 the AS/HFA control group had no intervention, thereby simply controlling for the passage of time. In Experiment 2 the AS/HFA control group undertook a 10–15 week social skills course, but without using *Mind Reading*, thereby controlling for the specific use of this educational software. Finally, a typical control group from the general population was matched to the intervention groups in both experiments. This group was only tested once, to obtain baseline measures.

For both experiments we predicted the following:

1. the performance of participants with AS/HFA would be lower than typical controls on all emotion recognition tasks at Time 1;
2. AS/HFA groups using the software would perform better on all emotion recognition tasks at Time 2, compared to Time 1, across generalization levels; and
3. participants using the software would improve more than AS/HFA controls on all tasks.

In addition, the association of software usage time with improvement on the different tasks was examined for the intervention groups.

Experiment 1

This experiment compared the effect of using *Mind Reading* at home alone to that of taking the assessment twice with no intervention, at three levels of generalization. The need for a no-intervention AS/HFA group was to assess whether any improvement was related to the intervention or merely to taking the tasks twice.

Method

Design. A controlled trial design with multiple repeated measures was used for this experiment. Two groups, the home intervention and the AS/HFA control, were assessed twice; at Time 1 and, 10–15 weeks after, at Time 2. In each assessment, generalization was assessed using stimuli in two modalities (visual and auditory) and three generalization levels:

1. *Close generalization:* this was tested using faces and voices that were included in the computer intervention but were presented using different software, and with more challenging multiple choice answers.
2. *Feature-based distant generalization:* using faces and voices that were not included in the training, we tested the ability to transfer emotion recognition skills separately in faces and voices.
3. *Holistic distant generalization:* using scenes from feature films, this level comprised holistic socioemotional stimuli, including faces, voices, body language, and context. Participants were then asked to identify the emotion of one of the characters in the scene, thus testing the ability to transfer acquired knowledge to complex holistic stimuli. This level was only assessed at Time 2.

In addition, the two groups were compared to the typical control group at Time 1.

Participants. Three groups took part in this experiment; one AS/HFA intervention group, one AS/HFA control group, and one typical control group. Participants in the clinical groups had all been diagnosed with AS/HFA in specialist centers using established criteria (American Psychiatric Association, 1994; World Health Organization, 1994). Participants filled in the Autism Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001), to assess their self-reported level of autistic traits. Eighty percent of the participants scored above a cutoff point of 32, which matches the percentage originally reported by Baron-Cohen et al. (2001). Eighty-eight percent of the participants scored above 26, which has recently been suggested in two separate studies as a more sensitive cutoff point for the AQ (Kurita, Koyama, & Osada, 2005; Woodbury-Smith, Robinson, Wheelwright, & Baron-Cohen, 2005). Participants were recruited from several sources, including a volunteer database, a local clinic for adults with AS/HFA, and an advert in the National Autistic Society magazine *Communication*. Participants were accepted to the study only if they had not participated in any related intervention during the last 3 months and had no plans for engaging in another intervention while the study was ongoing. Participants were randomly allocated into two groups:

Software home users. Nineteen participants (14 males, 5 females) were asked to use the software (provided free of charge) at home by themselves for 2 hr/week over a period of 10 weeks, a total of 20 hr. Participants were included in the study if they completed a minimum of 10 hr of work with the software. If they did not complete this minimum, participants were given an extension of up to 4 weeks to do more work with the software. Out of 24 participants originally recruited to this group, 3 withdrew during the 10-week period and 2 others were excluded at the end, as they failed to reach the 10-hr minimum. No specific pattern was found for these participants: they varied in their age range (21–43), education (3 had carried on studying beyond compulsory education, 2 had not), and employment status (2 were unemployed, 3 were employed). Their IQ, AQ, and Time 1 assessment task scores ranged within 1 *SD* of their group means. All of them related dropping out/not completing their work to being too busy and not getting to do the required amount of work. This finding will be discussed further.

AS/HFA control group. Twenty-two participants (17 males, 5 females) attended the assessment meetings with a 10–15-week period between them, during which they did not take part in any intervention related to emotion recognition.

Typical control group. Twenty-eight participants were recruited for this group from a local employment agency. Participants reported no psychiatric history and no occurrence of ASC in their families. After screening for autistic spectrum conditions using the AQ (Baron-Cohen, Wheelwright, Skinner, et al., 2001), four participants were excluded for scoring above the cutoff of 26. The remaining 24 participants (19 males, 5 females) attended one assessment meeting.

All participants were given the Wechsler Abbreviated Scale of Intelligence (WASI), comprising the vocabulary, similarities, block design, and matrix reasoning tests. The WASI produces verbal, performance, and full-scale IQ scores, with correlations of .88, .84, and .92 with the full Wechsler scales (Wechsler, 1999). All participants scored above 70 on both verbal and performance scales.

All participants completed the AQ (Baron-Cohen, Wheelwright, Skinner, et al., 2001). One-way analysis of variance (ANOVA) of AQ scores was significant, $F\ (2, 62) = 81.01$, $p < .001$. Tukey post hoc comparisons revealed that the two clinical groups scored significantly higher on the AQ compared to the typical control group. Participants were asked to report any psychiatric comorbid diagnoses. Five participants in each AS/HFA group had another psychiatric diagnosis, such as depression or attention deficit/hyperactivity disorder. No difference was found between the AS/HFA groups in the proportion of participants with psychiatric comorbidity, $\chi^2\ (1) = .07$, *ns* (see Table 1).

The three groups were matched on age, verbal and performance IQ, handedness, and gender. They spanned an equivalent range of

Table 1. *Means (standard deviations) and ranges of age, IQ, and AQ scores for the intervention and control groups of Experiment 1*

	Software Home Users ($N = 19$)	AS/HFA Controls ($N = 22$)	Typical Controls ($N = 24$)	F (2, 62)
Age	30.5 (10.3)	30.9 (11.2)	25.3 (9.1)	2.14
	17.5–48	17.5–52	17.5–51	
Verbal IQ	108.3 (13.3)	109.7 (10.0)	115.8 (13.7)	2.31
	80–127	93–129	86–138	
Performance IQ	112.0 (12.6)	115.3 (12.3)	112.5 (8.9)	0.53
	94–134	97–140	92–129	
AQ	37.2 (8.4)	38.2 (7.5)	14.0 (5.9)	81.02*
	20–47	16–49	6–26	
				$\chi^2(2)$
Females (%)	26.3	22.7	20.8	0.18
Left handed (%)	21.1	9.1	8.3	1.91
Employed (%)	47.4	40.9	45.8	0.20
A levels or above (%)[a]	68.4	63.6	58.3	0.47
Comorbid diagnoses (%)	26.3	22.7	NA	0.07

[a]A levels are the first component of noncompulsory education in Britain.
*$p < .001$. All other test results are not significant ($p > .05$).

employment and educational levels. As shown in Table 1, no significant differences were found between the groups for age, F (2, 62) = 2.14, *ns*, verbal IQ, F (2, 62) = 2.31, *ns*, performance IQ, F (2, 62) = 0.53, *ns*, gender, χ^2 (2) = 0.18, *ns*, handedness, χ^2 (2) = 1.91, *ns*, education, χ^2 (2) = 0.47, *ns*, and employment, χ^2 (2) = 0.20, *ns*. In addition, no difference was found between the two AS/HFA groups in the time between the two assessment meetings, t (27.7) = 1.57, *ns*.

Instruments.

The Cambridge Mindreading (CAM) Face–Voice Battery. This battery (Golan et al., in press) tests complex emotion and mental state recognition from short silent clips of faces and from voice recordings. Fifty faces and 50 voices, taken from *Mind Reading*, test recognition of 20 different complex emotions and mental states (e.g., intimate, insincere). The face task comprises silent clips of adult actors, both male and female, of different ethnicities, expressing the emotions in the face. An example showing one frame from one of the clips is shown in Figure 2. The voice task comprises recordings of short sentences expressing various emotional intonations. In both tasks four adjectives were presented after each stimulus is played and participants were asked which adjective best describes how the person feels. Both tasks were run on a computer, using the experimental software DMDX (Forster & Forster, 2003). A handout of definitions of all the adjectives used in the battery was available for the participants at the beginning and through the assessment. There was no time limit for answering. The battery provides an overall facial and an overall vocal emotion recognition score (max = 50 for each of them), as well as individual scores for each of the 20 emotions assessed (pass/fail, i.e., recognized above chance or not) and an overall number of the emotions correctly recognized (max = 20). Individuals with AS/HFA have been found to score significantly lower than controls on all three scores of the battery (Golan et al., in press). Test–retest correlations, calculated for the AS/HFA control group in the current study were found to be $r = .94$ for the face task and $r = .81$ for the voice task, $p < .001$ for both. This battery was used for the assessment of close generalization in both faces and voices.

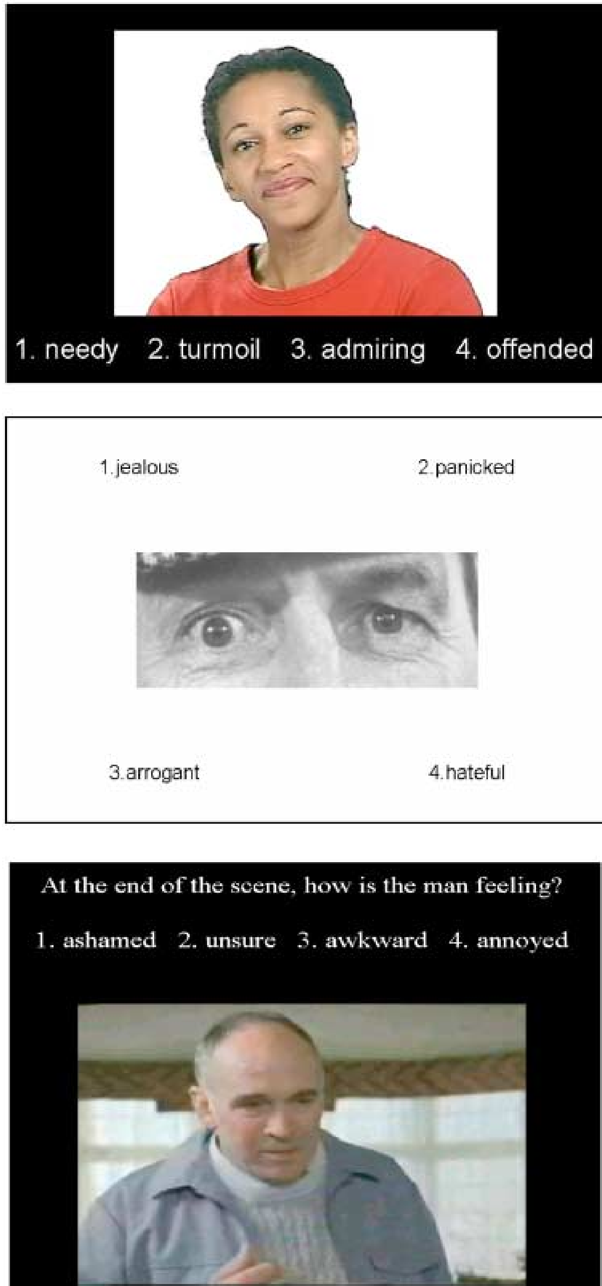


Figure 2. Examples of visual tasks from the three generalization levels: (top) close generalization adapted from the CAM Face task from Golan et al. (in press), (middle) feature-based distant generalization adapted from Reading the Mind in the Eyes Test from Baron-Cohen et al. (2001), and (bottom) holistic distant generalization adapted from Reading the Mind in Films task (Golan & Baron-Cohen 2006). Screenshot from *Lost for Words*, by D. Longden, 1999, London: Yorkshire Television. Copyright 1999 ITN. Reprinted with permission. [A color version of this figure can be viewed online at www.journals.cambridge.org]

Reading the Mind in the Eyes task (revised, adult version). The task (Baron-Cohen, Wheelwright, Hill, et al., 2001) has 36 items, in which participants are presented with a photograph of the eyes region of the face and must choose one of four adjectives or phrases to describe the mental state of the person pictured (see Figure 2 for an example). A definition handout is provided at the beginning of the task. In the present study, the pictures and adjectives were presented on the computer screen (using DMDX software), to avoid possible difficulties due to communication with a human examiner (Ozonoff, 1995a). Items were presented in a random order. There was no time limit for answering. Test–retest correlation, calculated for the AS/HFA control group in the current study was found to be $r = .86$ ($p < .001$). This task was used as a facial feature based generalization task.

Reading the Mind in the Voice task (revised). We used a revised version (Golan et al., 2005) of the original task (Rutherford et al., 2002). In the original task, 40 segments of speech, taken from BBC drama series, were played on audio tape to the participants, who were asked to choose one out of two possible answers, describing the speaker's mental state in each item. The task was revised as follows: each of the test items was allocated two more foils, taken from the same level: one level above or one level below the correct answer (based on the emotion taxonomy in *Mind Reading*). Foils were selected to match the content of the verbalizations but not the intonation, thus making the task harder to answer. This avoided ceiling effects to which the original version of the test was prone. Seven items were removed because the authors found the correct answer inappropriate to the verbalization. Eight more items were excluded after validation by a sample of 15 typically developing adults. The final task included 25 items with four possible answers for each of them. The test items were "cleaned" as far as possible and played on a computer (using DMDX software) in random order, preceded by an instruction slide and two practice items. Participants were given a definition handout before the beginning of the task. There was no time limit for answering. Test–retest correlation, calculated for the AS/HFA control group in the current study was $r = .80$ ($p < .001$). This task was used as a vocal feature based generalization task.

Reading the Mind in Film task. This task (Golan, Baron-Cohen, Hill, et al., in press) comprises 22 short social scenes taken from feature films. Each scene includes visual, vocal, and some contextual information. Scenes are presented on the computer screen (using DMDX software). Participants are presented with four adjectives and are asked to choose the one that best describes the way a target character feels at the end of the scene. Foils were selected to match some aspects of the scene (e.g., content of speech) but not all of them (e.g., facial expression, intonation, etc.), thus making the task harder to answer. A handout of definitions of all the adjectives used in the task was available for the participants at the beginning and through the assessment. There was no time limit for answering. Task score is the number of correctly recognized emotions (max = 22). Participants with AS/HFA were found to perform significantly worse on this task, comparing to matched controls. Figure 2 presents a screenshot from one of the task items. This task was used as a holistic distant generalization task and was only taken at Time 2.

Procedure

Participants were individually tested at the Autism Research Center in Cambridge. The first author and three trained assistants helped the participants through the assessments. The assistants were blind as to which group the participants belonged. Participants in the intervention group were asked to help in the evaluation of a new piece of software. It was explained they would need to commit to using *Mind Reading* for 2 hr/week over a period of 10 weeks, and to be assessed before and after this training period. Participants of both control groups were asked to take part in an emotion recognition study, helping to validate new tasks. For this reason, participants in the AS/HFA control group were asked to come for two assessments, separated by a 10- to 15-week period. Participants' written consent was obtained. All participants were told they were free to withdraw from the study at any time.

In the first assessment background information was collected, followed by the emotion recognition tasks. Participants were seated in front of IBM compatible computers with 15-in. monitors and were given headphones for the voice tasks. The emotion recognition tasks were presented to them in random order (with the exception of the Film task). The CAM battery included two breaks in each task due to its length. Participants were also allowed a break to freshen up between tasks. In between the tasks, two subtests of the WASI were administered. After the assessment was completed, *Mind Reading* was introduced to the participants of the intervention group in detail, including a presentation of the emotion taxonomy, the different areas, and a demonstration of a systematic analysis of an emotion, comparing different faces and voices to identify the unique facial/intonation features of this emotion. Participants were encouraged to analyze the facial and vocal stimuli systematically. They were asked to use the emotions library and learning center as they please, but not to use the game zone for more than a third of the usage time. They were also told *Mind Reading* logs their work, and they were asked to bring the log file to the second assessment meeting, for usage time verification. The whole assessment meeting took about 3 hr. During the time between the two assessments participants of the intervention group were approached by telephone at least once, to check they were still committed to the study and to working with the software.

In the second assessment participants took the same emotion recognition tasks, including the Film task and the other two WASI subtests. Task administration order was changed. The log files of participants in the intervention group were checked to verify they had used the software for the required amount of time. Average usage time was 17.5 hr (SD = 6.7, range = 10–36). These participants were then asked for their feedback about the program, their experience with it, and comments about its usefulness. Participants of the AS/HFA control group were asked about any possible effects of the first assessment meeting on their interest in emotions during the period between the assessments. All participants were then debriefed about the aims and design of the study and were rewarded with a complimentary copy of *Mind Reading* (or were allowed to keep the copy they used, with no charge). This assessment meeting took about 3 hr.

Results

The performance of the three groups on the emotion recognition tasks at Time 1 was explored first. Five one-way analyses of variance were conducted, testing group differences on the emotion recognition tasks used at Time 1. Using Holm's sequential rejective Bonferroni procedure (Holm, 1979; see also Zhang, Quan, Ng, & Stepanavage, 1997), significant differences were found between the groups on the CAM face task, F (2, 62) = 13.82, $p <$ .001, voice task, F (2, 60) = 11.53, $p < .001$, the number of emotional concepts recognized, F (2, 60) = 12.77, $p < .001$, the Reading the Mind in the Eyes task, F (2, 62) = 6.10, $p <$ .01, and the Reading the Mind in the Voice-R task, F (2, 62) = 4.92, $p < .02$. Preplanned comparisons with Bonferroni corrections revealed no significant differences between the two clinical groups on any of the task scores, and significantly higher scores of the typical control group on all tasks, comparing to the two AS/HFA groups. These findings support hypothesis 1. Table 2 shows the means and standard deviations of the groups' emotion recognition scores at Time 1.

Next, five multivariate analyses of covariance (MANCOVA) with repeated measures were conducted, to examine the differences between the intervention and AS/HFA control group on the various tasks at Time 1 and Time 2. Age, verbal, and performance IQ were used as covariates. Using Holm's sequential rejective Bonferroni procedure, significant Time × Group interactions were found on CAM faces, F_{Wilks} (1, 35) = 11.82, $p < .002$, CAM voices, F_{Wilks} (1, 34) = 7.51, $p < .01$, CAM number of concepts recognized, F_{Wilks} (1, 33) = 8.38, $p < .01$, but not for Reading the Mind in the Eyes, F_{Wilks} (1, 36) = 1.46, *ns*, or Reading the Mind in the Voice-R, F_{Wilks} (1, 36) = 0.47, *ns*.

Table 2. *Experiment 1 means (standard deviations) of the three groups on all tasks at Time 1 and Time 2*

	Software Home Users[a]		AS/HFA Controls		Typical Controls
	Time 1	Time 2	Time 1	Time 2	
CAM face task (max score = 50)	31.3 (8.8)	37.5 (7.8)	32.5 (8.4)	34.8 (8.2)	42.0 (5.2)
CAM voice task (max score = 50)	33.8 (6.6)	38.9 (6.2)	35.2 (7.4)	36.6 (7.9)	42.1 (4.2)
CAM no. concepts recognized (max score = 20)	9.8 (5.2)	13.6 (4.8)	10.5 (5.2)	11.3 (5.4)	16.1 (3.0)
Reading the Mind in the Eyes (max score = 36)	23.1 (6.7)	23.8 (4.7)	23.9 (6.7)	23.0 (7.3)	28.5 (3.1)
Reading the Mind in the Voice (max score = 25)	16.1 (2.9)	16.7 (3.9)	16.1 (3.9)	17.4 (3.5)	18.6 (2.4)
Reading the Mind in Films (max score = 22)		11.8 (3.8)		12.8 (3.4)	15.5 (2.4)

[a]Two participants of this group did not complete the CAM voice task at Time 1 and one participant did not complete the CAM faces at Time 2. These participants were excluded from the analysis of the three CAM scores. Other than that, groups' sizes are identical to Table 1.

No main effects were found significant, but the effect of the covariate verbal IQ was significant in relation to CAM voice scores, F (1, 34) = 5.11, $p < .05$.

Simple main effects analyses for the three CAM scores with Bonferroni corrections revealed the intervention group improved significantly from Time 1 to Time 2 on all three scores: for faces, t (17) = 5.37, $p < .001$; for voices, t (16) = 5.24, $p < .001$; for concepts recognized: t (15) = 3.96, $p < .005$. The AS/HFA control group scores did not change significantly from Time 1 to Time 2 on the CAM voices, t (21) = 1.43, *ns*, and the number of concepts recognized, t (21) = 1.25, *ns*, but did so on the CAM faces task, t (21) = 3.51, $p < .005$. However, when a t test was conducted on Time 2–Time 1 score differences, the improvement of the intervention group was significantly greater than that of the AS/HFA control group, t (38) = 3.38, $p < .005$. Table 2 and Figure 3 show the mean scores of all tasks at Time 1 and Time 2.

Then, we computed score differences for the 20 CAM concepts (Time 2 score minus Time 1 score). A multivariate ANOVA (MANOVA) was conducted for the 20 difference scores, with group as the independent variable. The MANOVA did not yield a significant group difference beyond emotional concept, F_{Wilks} (20, 19) = 1.36, *ns*, but yielded significant individual between group effects for the concepts: grave, F (1, 38) = 5.81, $p < .05$, uneasy, F (1, 38) = 5.2, $p < .05$, lured, F (1, 38) = 6.98, $p < .05$, intimate, F (1, 38) = 5.70, $p < .05$, insincere, F (1, 38) = 6.79, $p < .05$, and nostalgic, F (1, 38) = 18.48, $p < .001$. These findings suggest that the intervention group's recognition of these emotional concepts improved more than it did in the AS/HFA control group. However, care should be taken when interpreting these findings, due to the lack of an overall effect for the MANOVA, and the multiple tests conducted under it.

Next, an analysis of the Film task scores was conducted, to test for holistic distant generalization. This task was only used at Time 2, so differences could only be measured between groups. Therefore, a one way ANOVA was conducted on the three groups' Film task scores and was significant, F (2, 62) = 7.68, $p < .01$. Preplanned contrasts with Bonferroni correction revealed no difference between the intervention group (M = 11.8, SD = 3.8) and the AS/HFA control group (M = 12.8, SD = 3.4) for this task, t (62) = 1.07, *ns*. However, the typical control group (M = 15.5, SD = 2.4) scored significantly higher than both groups, t (62) = 3.81, $p < .001$. These results suggest the intervention group did not perform better than the AS/HFA control group on this level of generalization.

Last, a correlation analysis was conducted between the time participants had used the software and the improvement scores of each task. Nonparametric analysis was used, due to the small group size. Significant Spearman rho correlations were found for software usage time with improvement on the CAM voice task ($r_{\text{Spearman}} = .59$, $p < .01$), and with Reading the Mind in Film task scores ($r_{\text{Spearman}} = .45$, $p < .05$). No other correlations were significant.

Discussion

In this experiment, the use of *Mind Reading* at home was compared to no intervention. Results showed that following 10–20 hr of using the software over a period of 10–15 weeks, users significantly improved in their ability to recognize complex emotions and mental states from both faces and voices, compared to their performance before the intervention, and relative to the control group. This finding is interesting, considering the short usage time and the large number of emotions included in the software, and because participants were not asked to study these particular emotions.

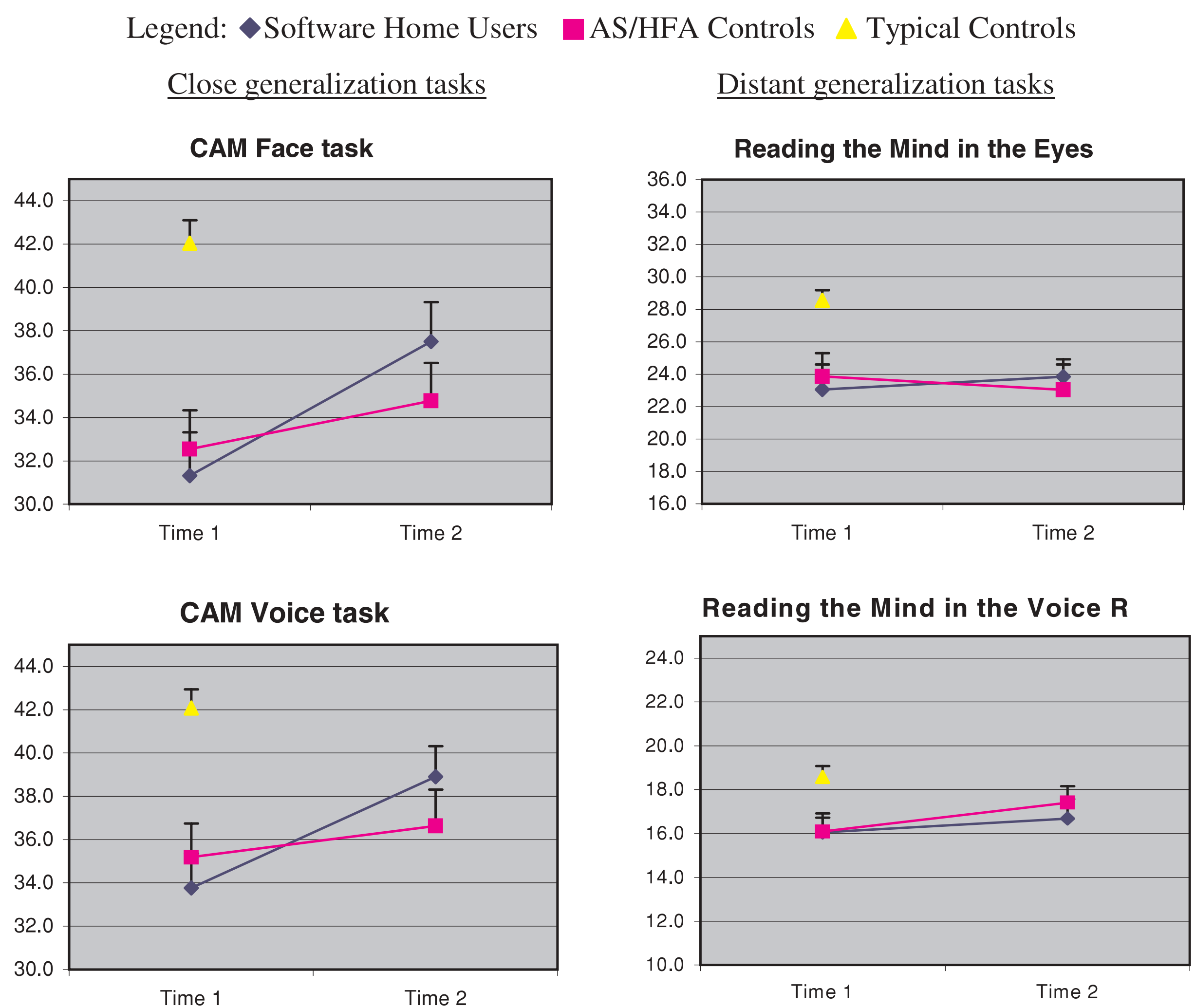
Time 1 results confirmed the marked difficulties individuals with ASC have in recognizing emotions from faces, voices, and eyes, and from social situations including the above with context and body language. These findings replicate results of previous studies (Baron-Cohen et al., 1997; Baron-Cohen, Wheelwright, Hill, et al., 2001; Golan, Baron-Cohen, & Hill, in press; Golan, Baron-Cohen, Hill, et al., in press; Heavey et al., 2000; Kleinman et al., 2001; Rutherford et al., 2002). Because no differences were found between the clinical groups at Time 1, any difference at Time 2 can be attributed to the intervention.

The intervention group improved significantly on close generalization measures, including faces, voices, and the emotions individuals with AS/HFA had particular difficulties with (Golan et al., in press). Improvement in the ability of individuals with ASC to recognize mental states such as *intimate, insincere*, or *grave* might have a positive effect on their confidence, willingness, and functioning in interpersonal situations. This, together with participants' reports of greater attention to faces and emotions, and improved eye contact, suggests that the analysis of emotions using *Mind Reading* allows people with ASC to improve emotion recognition skills from both faces and voices.

Improvement following the intervention was limited to close generalization tasks, that is, to faces and voices taken from *Mind Reading*. No difference was found between the intervention and AS/HFA control group on either feature based or holistic tasks of distant generalization. Similar findings of poor generalization have been found in studies teaching theory of mind, emotion recognition, and social skills to individuals with ASC. These findings will be discussed in more detail later. However, software usage time was positively correlated with film task scores, suggesting that the more participants used the software, the higher they scored on the distant generalization task. It is possible that a longer period of usage would have led to improved generalization among software users.

Interestingly, the control group significantly improved on the CAM face task between Time 1 and Time 2, despite having no formal intervention. This could be the result of taking the same task for a second time after a relatively brief interval. In addition, when interviewed at the end of the second assessment, many participants in this group reported greater interest in emotions following the assessment at Time 1. Therefore, it is possible that the assessment itself served as a limited short-term intervention, arousing participants' awareness of the importance of faces and emotions. This new awareness was not sufficient to cause an improvement on voices, but did allow for improvement on faces, which might suggest this domain is more easily changed through intervention. However, the improvement in this group's CAM face task scores was significantly smaller than that of the software home users' group.

Unlike our previous findings (Golan et al., in press), verbal IQ was found to have a significant effect on CAM voice task scores. The


Legend: Software Home Users AS/HFA Controls Typical Controls
Close generalization tasks
Distant generalization tasks
CAM Face task
44.0
42.0
40.0
38.0
36.0
34.0
32.0
30.0
Time 1
Time 2
Reading the Mind in the Eyes
36.0
34.0
32.0
30.0
28.0
26.0
24.0
22.0
20.0
18.0
16.0
Time 1
Time 2
CAM Voice task
44.0
42.0
40.0
38.0
36.0
34.0
32.0
30.0
Time 1
Time 2
Reading the Mind in the Voice R
24.0
22.0
20.0
18.0
16.0
14.0
12.0
10.0
Time 1
Time 2

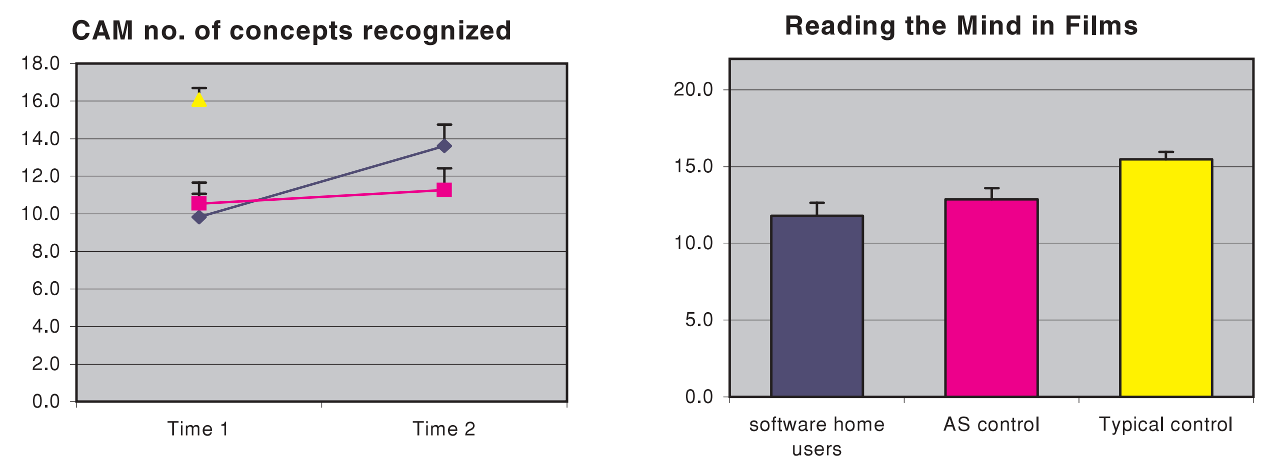


Figure 3. Mean scores (with standard error bars) of the three groups on two levels of generalization for Experiment 1. [A color version of this figure can be viewed online at www.journals.cambridge.org]

inclusion of verbal content in the voice task segments of speech (rather than just intonation) might account for this effect and for the lack of it in the face task. However, no such effect was found for Reading the Mind in the Voice (Revised), which also includes verbal content in its items. Experiment 2 was in part intended to help determine how central the role is that verbal IQ plays in the ability to recognize emotions.

Experiment 2

This experiment compared the effect of using *Mind Reading* with a weekly support of a tutor in group meetings to that of taking part in a 10-week social skills training course without using the software. Once again, improvement of emotion recognition skills in the three levels of generalization was the target outcome measure. This experiment allowed us to examine any extra value of using the software to that of attending group training for social skills that had less if any systematic method for teaching emotion recognition.

Method

Design and instruments. The design and the instruments used in this experiment were identical to those of Experiment 1.

Participants. Three groups took part in this experiment: two AS/HFA intervention groups (who were independent from those in Experiment 1) and one typical (general population) control group. Participants in the clinical groups had all been diagnosed with AS/HFA in specialist centers using established criteria (American Psychiatric Association, 1994; World Health Organisation, 1994). They were recruited via two support organizations and two colleges for individuals with ASC, where group meetings were also held. Because they were recruited via organizations that had volunteered to help with the study, participants were not randomly allocated to the groups, but instead were assigned by their recruiting organization.

Participants filled in the AQ (Baron-Cohen, Wheelwright, Skinner, et al., 2001) to assess their self-reported level of autistic traits. Forty percent of them scored above the cutoff of 32, a result that was unexpectedly low given earlier studies using the AQ (Woodbury-Smith et al., 2005), so a parent version of the AQ was sent to the parents or tutors of those who scored below the cutoff to check if this reflected underreporting of symptoms by this group or if diagnosis was in question. All but three parents/tutors returned the questionnaires. Using these reports, 70% of the participants scored above the cutoff of 32 (Baron-Cohen, Wheelwright, Skinner, et al., 2001) and all participants scored above the more sensitive cutoff of 26 (Kurita et al., 2005; Woodbury-Smith et al., 2005), an indicator that diagnosis was reliable. Participants were accepted to the study if they had not participated in any related intervention during the last 3 months and had no plans for engaging in another intervention while the study was ongoing. There were three groups in this experiment.

Software and tutor group. Thirteen participants (12 males, 1 female) were asked to use *Mind Reading* alone for 2 hr/week, over a 10-week period. In addition, these participants attended 10 weekly sessions in small groups of up to six members. A tutor worked with each group following a protocol that included analysis of features in different facial and vocal expressions of emotions, examples of situations from participants' everyday life, and the emotions they evoke, analysis of emotions in pictures from newspapers and scenes from feature films and TV programs. Tutors were free to choose the materials they wanted to analyze in the group, and were asked to relate lessons to emotion groups in *Mind Reading*, to help associate the software with group activities and with everyday life. Each of the tutors was given a copy of *Mind Reading* and was asked to become familiar with it before the course. Two of the tutors were support workers and one was a teacher, all experienced in working with adults with ASC. Three such groups were run: two in support centers for individuals with ASC and one in a college for adolescents and young adults with AS/HFA.

As in Experiment 1, participants were included if they completed a minimum of 10 hr using *Mind Reading*. Out of 18 participants who originally started the course, 3 withdrew during the course, and 2 were excluded after failing to complete the 10 hr of minimum use. Two groups were given extra time, so that participants could complete the minimum usage time. As in Experiment 1, no pattern was found for the participants who withdrew or did not complete the program: their age range varied between 19 and 46; 2 of them had continued studying beyond compulsory education, whereas 3 did not; 1 was unemployed, 3 were students, and 1 was employed. Their IQ levels and Time 1 assessment task scores ranged within 1 *SD* from their group average. The reasons for their not having completed the program included not being able to commit to working with the software for 2 hr/week due to their studies or work. One participant had left the group after falling in love with another participant and being rebuked.

Social skills course group. Thirteen participants (10 males, 3 females) took part in 10 sessions of social skills training. Two courses, with 9 participants in each, were facilitated by a clinical psychologist who specializes in social skill training for individuals with ASC, together with a staff member of the institution where the training took place. The facilitators also recruited the participants. The two groups followed the same curriculum, which included themes such as conversation rules, emotional expressions and body language, job interviews, and friendship. The facilitators used a variety of techniques, such as stand-up teaching, group discussion, role play, and picture analysis. The facilitators were blind to the curriculum of the other research group and to the software. In addition, the experimenters were blind to the group's curriculum until after the Time 2 assessment. Out of the 18 participants who originally attended the course, 2 did not complete the course and 3 others were excluded because their IQ scores fell below 70.

Typical (general population) control group. Thirteen participants (10 males, 3 females) of the typical control group described in Experiment 1 were matched to the two AS/HFA groups in this experiment.

A one-way ANOVA conducted on groups' AQ scores was significant for both self-report, F (2, 36) = 16.98, $p < .001$, and parent/tutor report, F (2, 36) = 69.34, $p < .001$. Tukey post hoc comparisons revealed that the two clinical groups scored significantly higher on the AQ compared to the typical control group, and did not differ from each other. Participants were also asked to report any comorbid diagnosis. Five participants in the software and tutor group and four participants in the social skills group had comorbid diagnoses such as depression, learning difficulties or OCD. The two AS/HFA groups did not differ on the proportion of participants with comorbid diagnoses, χ^2 (1) = 0.17, *ns* (see Table 3).

The three groups were matched on age, verbal and performance IQ, handedness, and gender. They spanned an equivalent range of socioeconomic classes and educational levels. No significant differences were found between the groups for age, F (2, 36) = 0.7, *ns*, verbal IQ, F (2, 36) = 2.36, *ns*, performance IQ, F (2, 36) = 2.70, *ns*, gender, χ^2 (2) = 1.39, *ns*, handedness, χ^2 (2) = 2.44, *ns*, education, χ^2 (2) = 0.22, *ns*, and employment, χ^2 (2) = 0.21, *ns*. In addition, no difference was found between the two AS/HFA groups on the time between the two assessment meetings, t (17.3) = 1.61, *ns*. Table 3 presents the groups' age and IQ data.

Procedure

Participants were tested at the local support centers and colleges for individuals with ASC. Participants were tested in groups of three by the first author and one of three trained assistants. The assistants were blind to which group participants belonged.

Participants of the software and tutor group were asked to help in the evaluation of a new intervention program. They were asked to commit to using *Mind Reading* for 2 hr/week over a period of 10 weeks and to attend all 10 group sessions. *Mind Reading* was briefly introduced to the participants of this group at the first assessment. The group tutors then introduced it in more detail at the first group meeting.

Table 3. *Means (standard deviations) and ranges of age, IQ, and AQ scores for the intervention and control groups of Experiment 2*

	Software and Tutor Group ($N = 13$)	Social Skills Group ($N = 13$)	Typical Controls ($N = 13$)	F (2, 36)
Age	25.5 (9.3) 17–50	24.4 (6.4) 17–42	25.5 (9.6) 17–51	0.70
Verbal IQ	105.7 (16.1) 76–123	96.5 (15.5) 78–126	109.2 (14.4) 86–128	2.37
Performance IQ	103.9 (19.8) 71–129	95.5 (6.0) 87–105	106.3 (6.0) 92–117	2.72
AQ (self-report)	25.1 (7.1) 16–37	29.7 (8.1) 17–45	13.9 (5.7) 8–26	16.98*
AQ (with parent/tutor report[a])	33.7 (4.4) 27–42	34.2 (4.9) 29–45	13.9 (5.7) 8–26	69.34*
				χ^2 (2)
Females (%)	7.7	23.1	23.1	1.39
Left handed (%)	15.4	30.8	7.7	2.44
Employed (%)	38.5	38.5	46.2	0.21
A levels or above (%)[b]	38.5	30.8	38.5	0.22
Comorbid diagnoses (%)	38.5	30.8	NA	0.17

[a]Parent/tutor filled AQ replaced self-report if the score was lower than the cutoff.
[b]A levels are the first component of noncompulsory education in Britain.
*$p < .001$. All other test results are not significant ($p > .05$).

Participants of the social skills group were told the study was to evaluate how social skills groups teach people to recognize emotions. They were asked to take part in the assessments at the beginning and the end of the course (which was free of charge). All participants were told they were free to withdraw from the study at any time. In the first assessment some background information was obtained, followed by administration of the emotion recognition tasks. Participants were seated in front of IBM compatible laptop computers with 15-in. monitors, and were given headphones for the voice tasks. The testing procedure was similar to that of Experiment 1. The assessment meetings took about 3 hr. During the intervention time between the two assessments, participants' use of the software was monitored by the tutors. The tutors were also in charge of collecting the log files made by the software to verify usage time. Average usage time of the software in the tutor and software group was 14.9 hr ($SD = 3.1$, range = 10–23). In the second assessment, participants of both intervention groups were asked for their detailed feedback about the program. They were then debriefed about the aims and design of the study and were rewarded with a complimentary copy of *Mind Reading* (or were allowed to keep the copy they used with no charge). This assessment meeting also took about 3 hr.

Results

For differences in group performance at Time 1, five one-way ANOVAs were conducted on the emotion recognition task scores, using Holm's sequential rejective Bonferroni procedure. Significant differences were found between the groups for the CAM face task, F (2, 36) = 9.76, $p < .001$, voice task, F (2, 36) = 5.64, $p < .01$, and the number of emotional concepts recognized, F (2, 36) = 7.77, $p < .005$; as well as for Reading the Mind in the Eyes, F (2, 36) = 6.75, $p < .005$, and Reading the Mind in the Voice-R, F (2, 36) = 4.99, $p < .02$. Preplanned comparisons with Bonferroni corrections revealed no significant differences

Table 4. *Experiment 2 means (standard deviations) of the three groups on all tasks at Time 1 and Time 2*

	Software and Tutor		Social Skills		Typical Controls
	Time 1	Time 2	Time 1	Time 2	
CAM face task (max score = 50)	32.3 (8.1)	36.2 (8.9)	26.8 (9.7)	29.3 (9.5)	40.8 (6.0)
CAM voice task (max score = 50)	33.2 (9.1)	38.9 (7.6)	31.1 (9.1)	31.8 (10.9)	41.1 (5.2)
CAM no. concepts recognized (max score = 20)	10.2 (4.9)	13.5 (5.2)	7.7 (5.8)	8.5 (6.3)	15.1 (3.7)
Reading the Mind in the Eyes (max score = 36)	21.6 (6.3)	23.8 (4.2)	21.5 (5.6)	19.2 (6.8)	28.2 (3.6)
Reading the Mind in the Voice (max score = 25)	15.1 (2.8)	16.2 (3.5)	13.9 (4.5)	14.7 (4.6)	18.1 (2.8)
Reading the Mind in Films (max score = 22)		11.9 (3.7)		10.5 (3.2)	14.8 (2.6)

Note: $N = 13$ for every group.

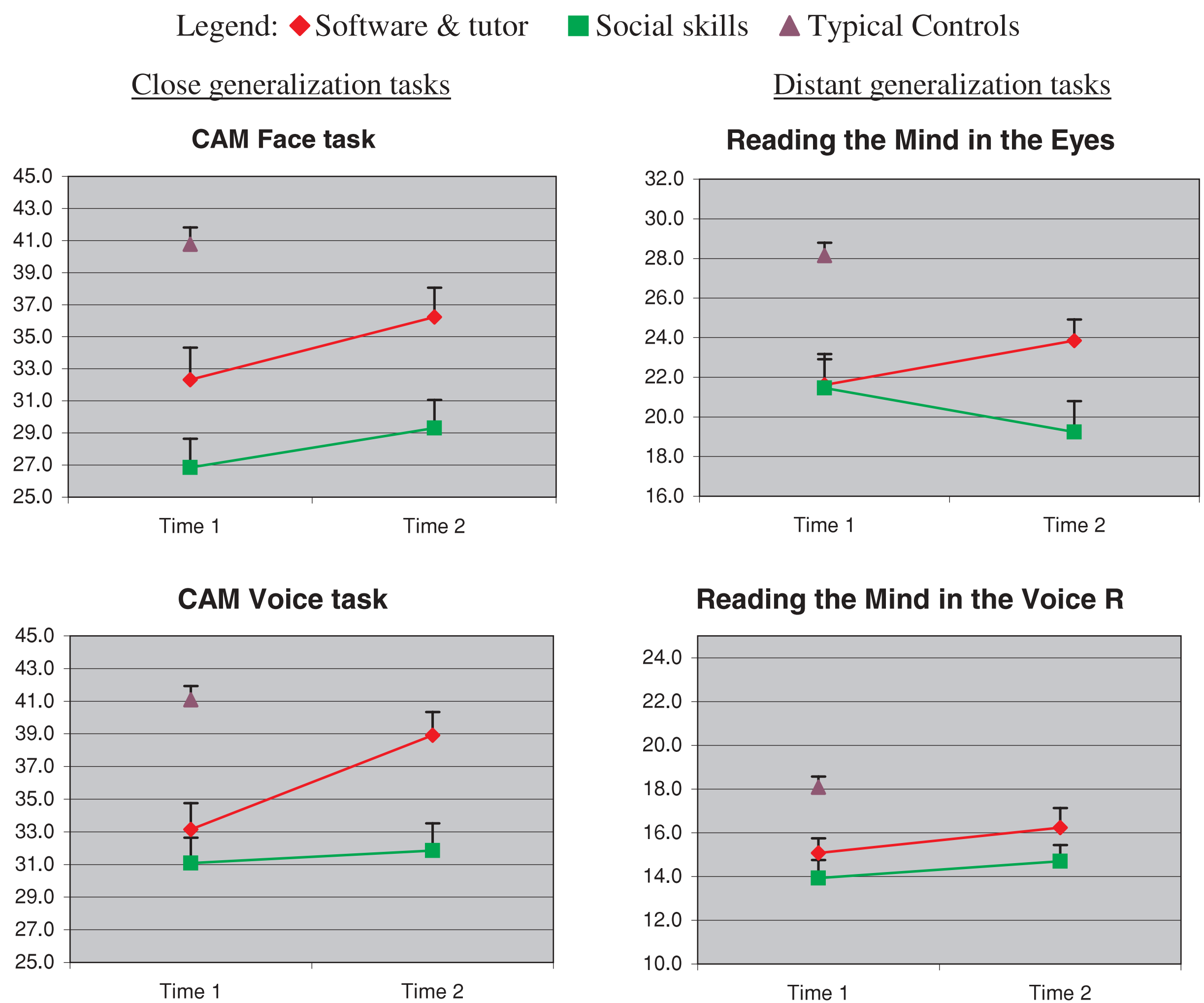
between the two clinical groups on any of the task scores, and significantly higher scores by the typical control group on all tasks, compared to the two AS/HFA groups. These findings support hypothesis 1. Means and standard deviations of the groups' emotion recognition scores at Time 1 appear in Table 4.

Five MANCOVAs with repeated measures were conducted to test for group differences on the various tasks at Time 1 and Time 2. Verbal IQ was entered as a covariate. Using Holm's sequential rejective Bonferroni procedure, Time × Group interactions were found for CAM voices, F_{Wilks} (1, 23) = 6.5, $p < .012$, CAM number of concepts recognized, F_{Wilks} (1, 23) = 6.04, $p < .016$, and Reading the Mind in the Eyes, F_{Wilks} (1, 23) = 8.4, $p < .01$, but not for CAM faces, F_{Wilks} (1, 23) = 0.23, *ns*, or Reading the Mind in the Voice-R, F_{Wilks} (1, 23) = 0.11, *ns*. Verbal IQ had a significant effect as a covariate on all tasks, beyond time, F (1, 23) = 17.89 for CAM faces, 19.4 for CAM voices, 17.05 for CAM number of concepts, 11.2 for Reading the Mind in the Eyes, 10.0 for Reading the Mind in the Voice-R, $p < .01$ for all.

Simple main effects analyses with Bonferroni correction revealed the software and tutor group improved significantly from Time 1 to Time 2 on CAM voice task, t (12) = 4.65, $p < .01$, and CAM number of concepts recognized, t (12) = 5.2, $p < .001$, whereas the social skills group did not on either CAM voices, t (12) = 0.56, *ns*, or CAM number of concepts, t (12) = 1.72, *ns*. Simple main effect analysis of the Time × Group interaction for Reading the Mind in the Eyes with Bonferroni correction revealed no significant time effects for either the software and tutor group, t (12) = 2.01, *ns*, or the social skills group, t (12) = −2.5, *ns*. Because the strong effect of verbal IQ might have overshadowed any main effect or interaction in the CAM faces scores, simple main effect analyses were conducted for the two groups despite the lack of significant interaction. Paired t tests for CAM face scores at Time 1 and Time 2 with Bonferroni correction showed a significant improvement in the software and tutor group, t (12) = 4.2, $p < .005$, but not in the social skills group, t (12) = 2.0, *ns*. Mean scores of all tasks at Time 1 and Time 2 can be found in Table 4 and Figure 4.

The MANOVA conducted for the 20 CAM concept difference scores failed to reach significance, F_{Wilks} (20, 5) = 1.16, *ns*. This may have been due to the small number of participants in each group and the small range of difference scores for each emotional concept. Significant effects were found for two concepts only: vibrant, F (1, 24) = 4.88, $p < .05$, and mortified, F (1, 24) = 10.04, $p < .01$.

The ANOVA conducted for the film task scores, testing for holistic distant generalization at Time 2 only, was significant, F (2, 36) =


Legend: Software & tutor Social skills Typical Controls
Close generalization tasks
Distant generalization tasks
CAM Face task
Reading the Mind in the Eyes
CAM Voice task
Reading the Mind in the Voice R
Time 1
Time 2

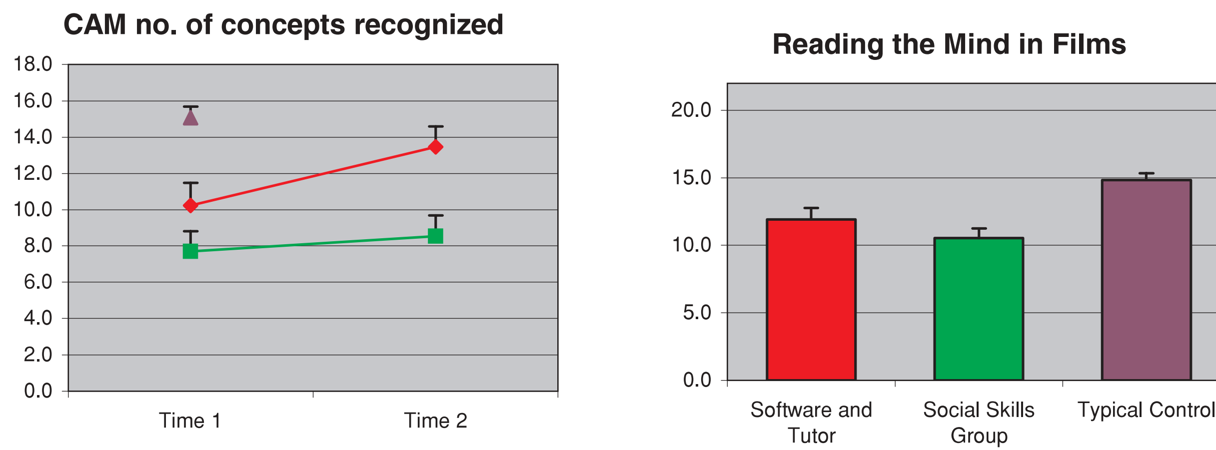


Figure 4. Mean scores (with standard error bars) of the three groups on two levels of generalization for Experiment 2. [A color version of this figure can be viewed online at www.journals.cambridge.org]

6.15, $p < .01$. However, preplanned contrasts with Bonferroni correction revealed no difference between the software and tutor group ($M = 11.9$, $SD = 3.7$) and the social skills group ($M = 10.5$, $SD = 3.2$) for this task, t (36) = 1.1, *ns*, although the typical control group ($M = 14.8$, $SD = 2.6$) scored significantly higher than both of the other groups, t (36) = 3.33, $p < .01$. These results suggest that as seen in Experiment 1, the AS/HFA group using *Mind Reading* did not perform better than the AS/HFA group who did not use it, at this level of generalization.

Nonparametric correlation analysis conducted for software usage time with task improvement scores in the software and tutor group revealed significant correlations of software usage time with improvement on the CAM voice task ($r_{\text{Spearman}} = .60$, $p < .05$), the number of CAM concepts correctly recognized ($r_{\text{Spearman}} = .53$, $p < .05$) and with film task scores ($r_{\text{Spearman}} = .50$, $p < .05$). No other correlations were found significant.

Discussion

This experiment compared the use of *Mind Reading* at home with weekly support of a tutor in group sessions to participating in a social skills course without the use of the software. As in Experiment 1, the group using the software improved significantly more than the control group on two measures of close generalization: recognition of voices and number of emotional concepts recognized, both taken from *Mind Reading*. The close generalization face task showed that the software and tutor group improved following training, whereas no improvement was found for the social skills group. These findings confirm the value of home use of *Mind Reading* compared to the effect of group meetings.

As in Experiment 1, software users failed to improve more than controls on feature-based distant generalization tasks and did not perform better than controls on the holistic distant generalization level: the film task. Software usage time in the software and tutor group was correlated with greater improvement on CAM voices task and CAM's number of emotions correctly recognized, as well as with higher scores on the holistic film task. This suggests that longer use of the software leads to improved generalization, although caution should be used when interpreting these results, due to the small group size.

The most striking effect in this experiment was that of Verbal IQ. Unlike Experiment 1, its effect was significant on all tasks beyond time. Yet, despite this, the software and tutor group improved significantly more than the social skills group on the recognition of two emotional concepts. The difference between the two experiments in the potency of verbal IQ could be related to the IQ difference between the groups in the two experiments. Although the groups in Experiment 1 had significantly above average intelligence, the two groups in Experiment 2 had average (and many lower than average) IQ, were recruited as groups via colleges and support centers, and had tutors take responsibility for their work. As a result of these differences, the groups of Experiment 2 might have found the assessments more challenging and the words in the tasks more difficult. Although definition handouts were offered, it is possible that those who have higher verbal IQ coped better with the tasks, as they were less distracted and stressed by the need to use the handouts. Future studies could include nonverbal tasks in the evaluation, to improve the validity of assessment in lower functioning groups.

Another limitation of this experiment lies in the different number of hours the two groups have received. The social skills group had no input at home, to match the homework of the Software and tutor group. This difference in the number of intervention hours, as well as having other subjects (except for emotion recognition) discussed in the social skills group, might have accounted for some of the improvement difference. Future studies should try to match the control intervention more carefully in both time and curriculum.

Last, group size limits the power of findings in this experiment. This is related to practical reasons of keeping the social skills group and tutor and software groups small, to optimize learning conditions, as well as to drop out rate. Future studies should evaluate the use of *Mind Reading* in group settings with more participants to further validate findings of this experiment.

General Discussion

This study evaluated a new intervention program for systematically teaching emotion recognition in faces and voices to adults with ASC. We investigated the effectiveness of individual independent use of *Mind Reading*, and the effectiveness of using it in conjunction with a tutor and group support. Results of both experiments revealed that the use of *Mind Reading* led adults with ASC to improve significantly in their emotion recognition skills. This improvement was achieved over a relatively short period of time, on a variety of complex and socially important emotions and mental states, in both faces and voices. The improvement was also independent of (or additional to) tutor and group support.

However, improvement following the use of the software in both experiments was limited to different presentation and variations on taught stimuli. Participants found it hard to generalize their knowledge to other tasks of emotion recognition from voices and eyes, and did not perform better than controls on a task involving integration of facial, vocal, and contextual cues. Similar generalization problems have been reported in previous studies attempting to teach theory of mind, emotion recognition, and social skills (Bolte et al., 2002; Hadwin et al., 1996; Swettenham, 1996, 2000).

Generalization difficulties have been reported to be characteristic of ASC (Rimland, 1965). A focus on small details at the expense of being able to see the larger picture (Frith, 1989), abstraction difficulties, and insistence on sameness may make generalization a challenge for individuals with ASC. Various models try to explain the generalization difficulties: adherence to rule-based categorization while failing to use prototype-based categorizations (Klinger & Dawson, 1995); an inability to recognize the similarities between stimuli (Plaisted, 2001); or the inability to deal with open systems of the social world, instead focusing on closed system, rule-based atomic physical phenomena (Klin, Jones, Schultz, & Volkmar, 2003; Lawson, 2003).

Our own view is that reduced generalization is not so much a reflection of a deficit as a reflection of the strong drive to systemize in people with ASC (Baron-Cohen, 2002). Good systemizing requires that one pays attention to the small details between variables in case these are important to understanding the workings of the system. As such, a good systemizer resists grouping variables together until there is reliable evidence that these variables have no functional differences. To group them together risks losing key information. With this view, generalization by individuals without ASC is a sign of reduced systemizing, and reduced generalization by individuals with ASC is a sign of their talent at systemizing, as revealed on analytical tests (Baron-Cohen, Wheelwright, Spong, et al., 2001, Folk Physics Test; Jolliffe & Baron-Cohen, 1997, Embedded Figures Test). Because the socioemotional world is difficult if not impossible to systemize, a systemizing strategy for learning about emotions may lead to limited generalization.

Mind Reading was used as the intervention for adults with ASC to exploit their good systemizing skills. As a systematic guide to emotions, *Mind Reading* trained recognition in faces and voices separately. Our results show that this was successful. Paradoxically, this way of teaching encouraged an atomized learning style leading to improvement, and made it harder to generalize to holistic material. Hence, we recommend that *Mind Reading* should be viewed as a first step in a training program. The next steps would need to deal with the systematic introduction of context and integration of different socioemotional cues into one (flexible) picture. Each step will need to be explicitly connected to the previous ones, with the main features pointed out, to ease generalization (Ozonoff, 1995b).

One side effect of the study was the relatively high dropout rate of participants in the software using groups (21% in Experiment 1, 28% in Experiment 2). The reasons individuals gave for withdrawing were usually related to their inability to find the time to use the software at home for 2 hr/week. Assuming that this is not an extensive amount of time to ask for, and bearing in mind that these participants had joined the study willingly, while knowing the work requirements, this finding could relate to difficulties individuals with AS/HFA have prioritizing, planning ahead and adhering to goals. Such difficulties could be explained by the executive dysfunction theory of autism (Ozonoff, 1995b; Russell, 1997), and suggest that even high-functioning adults with ASC may benefit from support of this kind. Future studies will need to evaluate the effect of possible executive dysfunction on the ability of adults with ASC to independently benefit from computer-based interventions.

In addition, as some of the participants pointed out, joining the study gave them an opportunity for desirable social contact with the experimenters and their group. Whereas this reason may have been secondary to their wish to acquire socioemotional skills, for some it may have been the main reason to join the study. It is therefore possible that participants who withdrew from the study felt the social gain was not strong enough to justify spending so much time using the software on their own. This again strengthens the need to use *Mind Reading* in conjunction with group activities, to boost both generalization and motivation.

The use of computer-based tasks in the evaluation of learning and generalization in this study has its limitations. Although such tasks allow for controlled and structured assessment of emotion recognition skills, testing different modalities separately, they are quite different to real life experience. Hence, the relevance of improvement among the software users in this study to real life functioning should be considered with care. Indeed, some participants commented they found recognizing emotions and mental states on the computer easier than doing this in real social situations, which requires cross-modal information processing and an immediate reaction in real time. This brings up the question of the appropriate evaluation of emotion recognition abilities. It also raises the question of the relevance of emotion recognition training to actual social functioning in adults with ASC. Most computerized emotion recognition and social skills programs were created for children with ASC. Whereas one was made for adolescents (Parsons, Mitchell, & Leonard, 2005), we are not aware of any other program made to address the needs of adults with ASC. More research into such interventions and their association with real life social functioning is required.

In both experiments, significant correlations were found between usage time and improvement scores of some tasks. Of particular interest was the positive correlation with the holistic film task, suggesting the use of the software might be associated with distant generalization measures. Research studies suggest there is both a developmental delay and long lasting difficulties in mental state recognition by individuals with ASC (Baron-Cohen, 1995; Frith & Hill, 2004). The present results suggest that learning aspects of empathizing skills (such as the emotion recognition component) is possible by people with ASC even into adulthood and that improvement in such areas is achieved when the intervention harnesses their systemizing strengths. It may be the case that even greater improvement would be achieved if intervention was started at a developmentally earlier time point. We are currently investigating the effectiveness of this method with children with ASC (Golan & Baron-Cohen, 2006).

The results of this study, as well as the participants' reports of looking more at faces and engaging in more eye contact following the use of *Mind Reading*, call for neuroimaging studies to examine possible changes in the functioning of brain areas (e.g., in the amygdala, fusiform gyrus, or prefrontal cortex), and gaze tracking studies to examine subtle behavioral changes following the systematic study of emotions using *Mind Reading*. Such studies would throw light on whether the observed cognitive changes reported here are arising from changes in those neural regions that are typically recruited by the nonautistic brain, or if they are due to compensatory strategies by other neural regions. We conclude that the present experiments indicate that complex emotion–recognition can be improved over a relatively brief training period (10–15 weeks), when systematic methods such as *Mind Reading* are employed. Additional systematic methods (possibly over a longer time period) are required to improve generalization.

References

Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology*, *11*, 231–239.

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

Ashwin, C., Baron-Cohen, S., Wheelwright, S., O'Riordan, M., & Bullmore, E. T. (2005). *Differential activation of the social brain during emotion processing in adults with autism*. Manuscript submitted for publication.

Astington, J. W., Harris, P. L., & Olson, D. R. (1988). *Developing theories of mind*. Cambridge: Cambridge University Press.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Boston: MIT Press/Bradford Books.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences*, *6*, 248–254.

Baron-Cohen, S. (2003). *The essential difference: Men, women and the extreme male brain*. London: Penguin.

Baron-Cohen, S., Golan, O., Wheelwright, S., & Hill, J. J. (2004). *Mind reading: The interactive guide to emotions*. London: Jessica Kingsley Limited (www.jkp.com).

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and Intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology*, *4*, 113–125.

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., & Wheelwright, S. (2003). The Systemising Quotient (SQ): An investigation of adults with Asperger Syndrome or High Functioning Autism and normal sex differences. *Philosophical Transactions of the Royal Society, Series B*, *358*, 361–374.

Baron-Cohen, S., Ring, H. A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A., et al. (1999). Social intelligence in the normal and autistic brain: An fMRI study. *European Journal of Neuroscience*, *11*, 1891–1898.

Baron-Cohen, S., Wheelwright, S., Hill, J. J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test—Revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*, 241–251.

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

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

Barry, T. D., Klinger, L. G., Lee, J. M., Palardy, N., Gilmore, T., & Bodin, S. D. (2003). Examining the effectiveness of an outpatient clinic-based social skills group for high-functioning children with autism. *Journal of Autism and Developmental Disorders*, *33*, 685–701.

Belmonte, M. K., Allen, G., Beckel-Mitchener, A., Boulanger, L. M., Carper, R. A., & Webb, S. J. (2004). Autism and abnormal development of brain connectivity. *Journal of Neuroscience*, *24*, 9228–9231.

Belmonte, M. K., Cook, E. H., Anderson, G. M., Rubenstein, J. L., Greenough, W. T., Beckel-Mitchener, A., et al. (2004). Autism as a disorder of neural information processing: Directions for research and targets for therapy. *Molecular Psychiatry*, *9*, 646–663. [Unabridged edition at http://www.cureautismnow.org/media/3915.pdf]

Bernard-Opitz, V., Sriram, N., & Nakhoda-Sapuan, S. (2001). Enhancing social problem solving in children with autism and normal children through computer-assisted instruction. *Journal of Autism and Developmental Disorders*, *31*, 377–398.

Bishop, J. (2003). The Internet for educating individuals with social impairments. *Journal of Computer Assisted Learning*, *19*, 546–556.

Bolte, S., Feineis-Matthews, S., Leber, S., Dierks, T., Hubl, D., & Poustka, F. (2002). The development and evaluation of a computer-based program to test and to teach the recognition of facial affect. *International Journal of Circumpolar Health*, *61*(Suppl. 2), 61–68.

Boucher, J., Lewis, V., & Collis, G. M. (2000). Voice processing abilities in children with autism, children with specific language impairments, and young typically developing children. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*, 847–857.

Celani, G., Battacchi, M. W., & Arcidiacono, L. (1999). The understanding of the emotional meaning of facial expressions in people with autism. *Journal of Autism and Developmental Disorders*, *29*, 57–66.

Critchley, H. D., Daly, E. M., Bullmore, E. T., Williams, S. C., Van Amelsvoort, T., Robertson, D. M., et al. (2000). The functional neuroanatomy of social behaviour: Changes in cerebral blood flow when people with autistic disorder process facial expressions. *Brain*, *123*, 2203–2212.

Deruelle, C., Rondan, C., Gepner, B., & Tardif, C. (2004). Spatial frequency and face processing in children with autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*, *34*, 199–210.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, *48*, 384–392.

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

Fein, D., Lucci, D., Braverman, M., & Waterhouse, L. (1992). Comprehension of affect in context in children with pervasive developmental disorders. *Journal of Child Psychology and Psychiatry*, *33*, 1157–1167.

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391–428). London: Oxford University Press.

Forster, K. L., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers*, *35*, 116–124.

Frith, U. (1989). *Autism: Explaining the enigma*. Oxford: Blackwell.

Frith, U., & Hill, E. (2004). *Autism: Mind and brain*. Oxford: Oxford University Press.

Golan, O., & Baron-Cohen, S. (2006). *Teaching children with Asperger syndrome and high functioning autism to recognize emotions using interactive multimedia*. Unpublished manuscript.

Golan, O., Baron-Cohen, S., & Hill, J. (in press). The Cambridge Mindreading (CAM) Face–Voice Battery: Testing complex emotion recognition in adults with and without Asperger Syndrome. *Journal of Autism and Developmental Disorders*.

Golan, O., Baron-Cohen, S., Hill, J. J., & Golan, Y. (in press). "Reading the Mind in Films"—Testing recognition of complex emotions and mental states in adults with and without autism spectrum conditions. *Social Neuroscience*.

Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. D. (2005). *The "Reading the Mind in the Voice Test—Revised": A study of complex emotion recognition in adults with and without autism spectrum conditions*. Manuscript submitted for publication.

Grossman, J. B., Klin, A., Carter, A. S., & Volkmar, F. R. (2000). Verbal bias in recognition of facial emotions in children with Asperger syndrome. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*, 369–379.

Hadwin, J., Baron-Cohen, S., Howlin, P., & Hill, K. (1996). Can we teach children with autism to understand emotions, belief, or pretence? *Development and Psychopathology*, *8*, 345–365.

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

Happe, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., et al. (1996). "Theory of mind" in the brain. Evidence from a PET scan study of Asperger syndrome. *Neuroreport*, *8*, 197–201.

Haviland, J. M., & Lelwica, M. (1987). The induced affect response: 10-week-old infants' responses to three emotion expressions. *Developmental Psychology*, *23*, 97–104.

Heavey, L., Phillips, W., Baron-Cohen, S., & Rutter, M. (2000). The Awkward Moments Test: A naturalistic measure of social understanding in autism. *Journal of Autism and Developmental Disorders*, *30*, 225–236.

Hetzroni, O. E., & Tannous, J. (2004). Effects of a computer-based intervention program on the communicative functions of children with autism. *Journal of Autism and Developmental Disorders*, *34*, 95–113.

Hobson, P. (1994). Understanding persons: The role of affect. In S. Baron-Cohen, H. Tager Flusberg, & D. Cohen (Eds.), *Understanding other minds* (pp. 204–227). London: Oxford University Press.

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

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

Hobson, R. P., Ouston, J., & Lee, A. (1988). What's in a face? The case of autism. *British Journal of Psychology*, *79*, 441–453.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Howlin, P., Baron-Cohen, S., & Hadwin, J. (1999). *Teaching children with autism to mind-read: A practical guide for teachers and parents*. Chichester: Wiley.

Howlin, P., & Yates, P. (1999). The potential effectiveness of social skills groups for adults with autism. *Autism*, *3*, 299–307.

Jolliffe, T., & Baron-Cohen, S. (1997). Are people with autism and Asperger syndrome faster than normal on the Embedded Figures Test? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *38*, 527–534.

Jolliffe, T., & Baron-Cohen, S. (1999). The Strange Stories Test: A replication with high-functioning adults with autism or Asperger Syndrome. *Journal of Autism and Developmental Disorders*, *29*, 395–406.

Kleinman, J., Marciano, P. L., & Ault, R. L. (2001). Advanced theory of mind in high-functioning adults with autism. *Journal of Autism and Developmental Disorders*, *31*, 29–36.

Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: Lessons from autism. *Philosophical Transactions of the Royal Society*, *358*, 345–360.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, *59*, 809–816.

Klinger, L. G., & Dawson, G. (1995). A fresh look at categorization abilities in persons with autism. In E. Schopler & G. Mesibov (Eds.), *Learning and cognition in autism* (pp. 119–136). New York: Plenum Press.

Kurita, H., Koyama, T., & Osada, H. (2005). Autism-spectrum quotient—Japanese version and its short forms for screening normally intelligent persons with pervasive developmental disorders. *Psychiatry and Clinical Neuroscience*, *59*, 490–496.

Lawson, J. (2003). Depth accessibility difficulties: An Alternative conceptualisation of autism spectrum conditions. *Journal for the Theory of Social Behaviour*, *33*, 189–202.

Lawson, J., Baron-Cohen, S., & Wheelwright, S. (2004). Empathising and systemising in adults with and without Asperger Syndrome. *Journal of Autism and Developmental Disorders*, *34*, 301–310.

Loveland, K. A., Tunali Kotoski, B., Chen, R., & Brelsford, K. A. (1995). Intermodal perception of affect in persons with autism or Down syndrome. *Development and Psychopathology*, *7*, 409–418.

Loveland, K. A., Tunali Kotoski, B., Chen, Y. R., Ortegon, J., Pearson, D. A., Brelsford, K. A., et al. (1997). Emotion recognition in autism: Verbal and nonverbal information. *Development and Psychopathology*, *9*, 579–593.

Macdonald, H., Rutter, M., Howlin, P., Rios, P., Le Conteur, A., Evered, C., et al. (1989). Recognition and expression of emotional cues by autistic and normal adults. *Journal of Child Psychology and Psychiatry*, *30*, 865–877.

Moore, D., McGrath, P., & Thorpe, J. (2000). Computer-aided learning for people with autism—A framework for research and development. *Innovations in Education and Training International*, *37*, 218–228.

Nieminen-von Wendt, T., Metsahonkala, L., Kulomaki, T., Alto, S., Autti, T., Vanhala, R., et al. (2003). Changes in cerebral blood flow in Asperger syndrome during theory of mind tasks presented by the auditory route. *European Child and Adolescent Psychiatry*, *12*, 178–189.

O'Riordan, M. A., Plaisted, K. C., Driver, J., & Baron-Cohen, S. (2001). Superior visual search in autism.

Journal of Experimental Psychology: Human Perception and Performance, *27*, 719–730.

Ozonoff, S. (1995a). Reliability and validity of the Wisconsin Card Sorting Test in studies of autism. *Neuropsychology*, *9*, 491–500.

Ozonoff, S. (1995b). Executive functions in autism. In E. Schopler & B. Mesibov (Eds.), *Learning and cognition in autism* (pp. 199–219). New York: Plenum Press.

Parsons, S., & Mitchell, P. (2002). The potential of virtual reality in social skills training for people with autistic spectrum disorders. *Journal of Intellectual Disabilities Research*, *46*, 430–443.

Parsons, S., Mitchell, P., & Leonard, A. (2005). Do adolescents with autistic spectrum disorders adhere to social conventions in virtual environments? *Autism*, *9*, 95–117.

Pierce, K., Muller, R. A., Ambrose, J., Allen, G., & Courchesne, E. (2001). Face processing occurs outside the fusiform "face area" in autism: Evidence from functional MRI. *Brain*, *124*, 2059–2073.

Plaisted, K. C. (2001). Reduced generalization in autism: An alternative to weak central coherence. In J. A. Burack (Ed.), *The development of autism: Perspectives from theory and research* (pp. 149–169). Mahwah, NJ: Erlbaum.

Rajendran, G., & Mitchell, P. (2000). Computer mediated interaction in Asperger's syndrome: The Bubble Dialogue program. *Computers and Education*, *35*, 189–207.

Rimland, B. (1965). *Infantile autism: The syndrome and its implications for a neural theory of behavior*. London: Methuen.

Russell, J. (1997). *Autism as an executive disorder*. Oxford: Oxford University Press.

Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal Autism and Developmental Disorders*, *32*, 189–194.

Rydin, O. T., Drake, J., & Bratt, A. (1999). The effects of training on emotion recognition skills for adults with an intellectual disability. *Journal of Applied Research in Intellectual Disabilities*, *12*, 253–262.

Schultz, R. T., Grelotti, D. J., Klin, A., Kleinman, J., Van der Gaag, C., Marois, R., et al. (2003). The role of the fusiform face area in social cognition: Implications for the pathobiology of autism. *Philosophical Transactions of the Royal Society London B: Biological Sciences*, *358*, 415–427.

Shah, A., & Frith, U. (1993). Why do autistic individuals show superior performance on the block design task? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *34*, 1351–1364.

Silver, M., & Oakes, P. (2001). Evaluation of a new computer intervention to teach people with autism or Asperger syndrome to recognize and predict emotions in others. *Autism*, *5*, 299–316.

Swettenham, J. (1996). Can children with autism be taught to understand false belief using computers? *Journal of Child Psychology and Psychiatry*, *37*, 157–165.

Swettenham, J. (2000). Teaching theory of mind to individuals with autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed., pp. 442–456). Oxford: Oxford University Press.

Teunisse, J. P., & De Gelder, B. (1994). Do autistics have a generalized face processing deficit? *International Journal of Neuroscience*, *77*, 1–10.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX: The Psychological Corporation.

Wellman, H. M. (1992). *The child's theory of mind*. Cambridge, MA: MIT Press.

Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening adults for Asperger syndrome using the AQ: A preliminary study of its diagnostic validity in clinical practice. *Journal of Autism and Developmental Disorders*, *35*, 331–335.

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

Yirmiya, N., Sigman, M. D., Kasari, C., & Mundy, P. (1992). Empathy and cognition in high-functioning children with autism. *Child Development*, *63*, 150–160.

Young, A. W. (1998). *Face and mind*. Oxford: Oxford University Press.

Zhang, J., Quan, H., Ng, J., & Stepanavage, M. E. (1997). Some statistical methods for multiple endpoints in clinical trials. *Controlled Clinical Trials*, *18*, 204–221.