Journal of Autism and Developmental Disorders, Vol. 25, No. 4, 1995

Teaching Theory of Mind: A New Approach to Social Skills Training for Individuals with Autism[1]

Sally Ozonoff[2] and Judith N. Miller
University of Utah

This study examined the effectiveness of a social skills training program for normal-IQ adolescents with autism. Five boys participated in the 4 1/2-month treatment condition; four boys matched on age, IQ, and severity of autism constituted the no-treatment control group. In addition to teaching specific interactional and conversational skills, the training program provided explicit and systematic instruction in the underlying social-cognitive principles necessary to infer the mental states of others (i.e., theory of mind). Pre- and post-intervention assessment demonstrated meaningful change in the treatment group's performance on several false belief tasks, but no improvement in the control sample. No changes, however, were demonstrated on general parent and teacher ratings of social competence for either group.

INTRODUCTION

Difficulty relating socially to other people is a hallmark of autism. This prominent and persistent symptom is critical to diagnosis of the condition, even in higher functioning individuals. Research has shown that social skills are related to long-term adjustment and prognosis for both autistic and nonautistic individuals (Matson & Swiezy, 1994; Parker & Asher, 1987;

[1]This project was partially supported by a University of Utah Undergraduate Research Opportunities Program grant to the authors. We gratefully acknowledge the assistance of Tom George and Joel Wendel, who served as social skills trainers and assisted in data collection. We also thank the subjects and their families for their long-term support of research on autism. Finally, the helpful comments of two anonymous reviewers on a draft of this manuscript are appreciated.
[2]Address all correspondence to Sally Ozonoff, Department of Psychology, 502 Behavioral Science Building, University of Utah, Salt Lake City, Utah 84112.

0162-3257/95/0800-0415$07.50/0

Schopler & Mesibov, 1983). Over the past 20 years, a variety of treatment approaches have been used in an attempt to remediate the social skills deficits of people with autism. Interventions have varied in the orientation of the approach (operant, cognitive-behavioral, social learning, developmental), the number of people involved (individual, dyadic, and group treatments), and the level of behavior targeted for intervention (molecular vs. molar approaches). While a thorough review of the history of social skills training with autistic people is beyond the scope of this paper, Matson and Swiezy (1994) provide a thoughtful summary and analysis of the topic.

In recent years, interest in group approaches to social skills training has increased. In the first investigation of group treatment modalities, Mesibov (1984) described a social skills program for verbal adolescents and adults with autism. The primary goals of the intervention were to increase interactional and communication skills, promote positive peer experiences, and enhance self-esteem. Techniques used included modeling, coaching, and role-play. Primary assessment of outcome was qualitative (e.g., participants', families', and staff members' impressions of change) and suggested strong satisfaction with the training model. Although quantitative data and formal statistical analyses were not presented, Mesibov stated that preliminary examination of outcome measures indicated empirical improvement as well.

Williams (1989) conducted a similar group social skills program with autistic children. A preliminary evaluation of this group also suggested improvement at post-treatment testing. A social behavior questionnaire was administered to a staff member at each participant's school before and after the 4-year intervention. All subjects demonstrated improvements as measured by this questionnaire.

The authors of both previous studies identified the need for more rigorous data collection and evaluation of treatment efficacy, as have others (La Greca, 1993). The present study, an investigation of the effectiveness of a group intervention for normal-IQ adolescent boys with autism, provided more extensive outcome measures and formal statistical analyses of effects. In addition, we employed a no-treatment control group design; without this, it is difficult to determine whether any improvements seen exceed those that would normally occur in the development of autistic children. We also used outcome ratings by evaluators blind to group membership and time of testing to more objectively evaluate the strength of our treatment. Finally, we introduced a new dimension to both our training program and our evaluation methods by formally measuring, teaching, and assessing change in one of the most consistently documented deficits of autism, theory of mind.

Theory of mind is defined as the ability to infer the mental states of others (e.g., their knowledge, intentions, beliefs, desires). In an important series of studies, Baron-Cohen and colleagues demonstrated that the ability of children with autism to attribute mental states to others was seriously deficient. Baron-Cohen, Leslie, and Frith (1985, 1986) found that 80% of autistic children were unable to correctly predict the beliefs of others, whereas most mentally retarded and normal controls of lower mental age were able to do so. Baron-Cohen (1989) found that 100% of mentally retarded autistic children failed a more difficult second-order belief attribution test. These basic findings have been replicated with several additional paradigms (Leekam & Perner, 1991; Leslie & Frith, 1988; Perner, Frith, Leslie, & Leekam, 1989), suggesting that theory of mind impairments are core, central features of autism.

Previous social skills training programs have not specifically targeted theory of mind abilities for intervention. We felt this would be important for a number of reasons. First, we surmised that children with autism might have different training needs from the nonautistic children for whom most social skills programs were developed. Examples of such approaches are *The Accepts Program* (Walker et al., 1983) and *The Prepare Curriculum* (Goldstein, 1988). The advantages of these programs include commercial availability, research development and validation, and empirical examination of effectiveness. These packaged programs, however, often rely upon basic social-cognitive skills that children with autism may lack. For example, an early lesson on negotiation in one of these manuals requires the child to first decide "if you and the other person disagree." Another lesson on giving compliments requires the child to "consider the wording to keep the other person from being embarrassed." Clearly, both of these exercises (and others) require inference of mental and emotional states, skills that research tells us are difficult for autistic individuals. These training programs assume certain fundamental abilities that an autistic client may not possess. Consequently, in addition to teaching overt social behaviors, our training program also included explicit and systematic instruction in perspective-taking and interpersonal problem-solving strategies.

A second reason we decided to teach theory of mind skills as part of our intervention was to examine the modifiability of this social-cognitive impairment. Would it even be possible to teach mental state concepts to individuals with autism (see also Baron-Cohen & Howlin, 1993)? Williams (1989) commented that, even after a 4-year social skills intervention, perspective-taking remained very difficult for most participants. Nonintervention studies of theory of mind have found it to be a persistent deficit that does not improve on its own as mental age advances (Holroyd & Baron-Cohen, 1993; Ozonoff & McEvoy, 1994). If a theory of mind can be taught,

what effects, if any, might it have on the broader range of autistic symptoms? Would learning this skill have relevance for other domains of social or cognitive functioning?

METHOD

Subjects

Nine male autistic adolescents were recruited through an announcement in the Autism Society of Utah newsletter. All met DSM-III-R diagnostic criteria for either Autistic Disorder or Pervasive Developmental Disorder Not Otherwise Specified. All were functioning in the nonretarded intellectual range, with Full Scale IQ scores above 70. Five adolescents participated in the treatment condition, while four made up the no-treatment control group. Group assignment was not random. Rather, placement was determined by practical constraints: Subjects whose schedules did not permit them to attend the weekly sessions at the appointed time were assigned to the control group. There were no differences between the treatment and control groups on age, IQ, severity of autism (as rated by the Childhood Autism Rating Scale; Schopler, Reichler, & Renner, 1988), or length of time between pre- and post-testing. See Table I for descriptive characteristics of the two groups.

All subjects lived at home. Two control subjects were attending regular classrooms, with no special education assistance. Two subjects (one in the

Table I. Descriptive Characteristics of the Sample[a]

	Treatment (n = 5)			Control (n = 4)		
	M	*SD*	Range	*M*	*SD*	Range
Age	13.8	0.24	13.5–14.0	13.6	2.4	11.3–16.2
VIQ	88.0	10.2	74–99	94.0	18.7	66–104
PIQ	97.6	16.1	80–113	95.7	12.3	84–112
FSIQ	91.8	12.2	76–106	94.2	14.9	73–107
CARS[b]	30.7	3.4	27.5–36.0	32.5	2.9	30.0–35.0
Time to post-testing (in months)	4.8	0.45	4–5	4.5	0.58	4–5

[a]No significant differences.
[b]Childhood Autism Rating Scale.

control group and one in the treatment group) were in part-time resource placements and mainstreamed for several classes. The other five subjects (four in the treatment group and one in the control group) were in self-contained classrooms (one Intellectually Handicapped, four Behavior Disorder).

Procedure

All nine subjects were administered a battery of theory of mind tests before the social skills training program began; the same measures were given again after the intervention was complete. Questionnaires rating the frequency of specific social behaviors were administered to each subject's parents and primary teacher at both time periods as well. The average time between pre- and post-testing was $4^1/_2$ months. All subjects in the treatment group were tested within 2 weeks of the end of the intervention.

Coding of responses on all theory of mind tasks was done by a member of the research team blind to subject identity, group membership, and time of testing.

Measures

Hughes and Sullivan (1988), in their critique of social skills outcome research, highlight the need for two levels of evaluation. Only designs that include both levels of measurement, they contend, are capable of determining functional relationships between target behaviors and outcomes. *Specifying measures* are those that determine whether a specific target behavior or skill has changed as a result of the treatment. A common type of specifying measure is a test of social cognition. In this study, we used false belief tasks developed for research use with autistic children. *Impact measures,* on the other hand, help determine whether the intervention influenced a socially valid outcome. Typical impact measures are peer, teacher, or parent ratings of sociometric status or adjustment. In this study, parent and teacher assessments of overall social competence were used.

Specifying Measures: Theory of Mind Tasks[3]

M&Ms False Belief Task (Perner et al., 1989). In this measure of first-order perspective-taking, subjects were shown a box of M&Ms and

[3]Further detail on the theory of mind tasks and their scoring can be obtained in the primary sources cited in the text. All measures were administered in the manner specified in the original references, using the standard number of trials, control questions, and prompts.

asked what it contained. After a response was given, the box was opened to reveal that it actually held a pencil. Subjects were then asked to predict what another child, who had never seen the box, would think it contained. A pass was scored if the subject responded, "M&Ms" (or "candy").

Second-Order Belief Attribution Task (Baron-Cohen, 1989). This task required a more advanced perspective-taking skill, that of predicting what one person thinks another person thinks. It employed a model of a town constructed from railroad miniatures. Subjects were told a story about two children, John and Mary, playing in a park (see Baron-Cohen, 1989, for story narrative). After seeing the story enacted, subjects were asked to predict Mary's beliefs about John's whereabouts (the Belief Question). A correct answer required subjects to recursively reason about mental states (i.e., what does Mary think John thinks?), and from this, predict where she would believe he had gone. The Belief Question was scored pass or fail. In the Justification Question, subjects were asked to explain why Mary held the belief she did. Responses were scored according to the number of mental state attributions made by the subject: 0 (i.e., no mental state attributions were made), 1 (i.e., mental states were attributed to only one character), and 2 (i.e., the subject accounted for the mental states of both characters). For further detail on the scoring procedure, see Baron-Cohen (1989). In a previous investigation (Ozonoff & McEvoy, 1994), both autistic and nonautistic subjects occasionally received credit on the Justification Question for using mental state terms *with incorrect content* (e.g., the response "she *thinks* he is at the church" would be given a score of 1, using Baron-Cohen's 1989 scoring system, for referring to Mary's mental state, although the content of the answer is incorrect). Thus, in the present study, we added an additional code that assessed the accuracy of the subject's response to the Justification Question (coded in a pass/fail manner). The maximum score possible on the Second-Order Belief Attribution Task was 4 (1 for the Belief Question, 2 for the Justification Question, and 1 for the accuracy code). This scoring procedure, with the additional accuracy code, was used for the two advanced theory of mind tasks described below as well. The decision to add the accuracy code appeared warranted, as fully 11% of the Justification responses given in this study used mental state terminology with inaccurate content.

Overcoat Story (Bowler, 1992). This task also measured second-order belief attribution abilities, but was more difficult than the previous one because the story was read aloud and no visual cues were provided. It has been argued that high-functioning autistic people may possess the *ability* to reason about others' mental states; their problems may instead

occur in the *application* of that knowledge under demanding conditions (Bowler, 1992; Rutter & Bailey, 1993) (e.g., when not given sufficient time, when no visual cues are provided, or in social situations with multiple competing stimuli). Thus, this measure tested second-order theory of mind abilities under the more demanding and naturalistic condition of *listening* to the scenario. As in the previous task, subjects were asked to predict what one person thinks another person thinks (Belief Question) and then explain why (Justification Question). An accuracy code was again used, making four the maximum score possible on the task. Scoring was performed as described by Bowler (1992).

Prisoner Story (Happe, 1994). In this advanced theory of mind measure, subjects read a short story about a prisoner of war who is being interrogated about the location of his army's tanks. The story states that the interrogators assume the prisoner will lie. In an effort to save his comrades, the prisoner instead tells the truth. Subjects were asked to predict where the interrogators would look for the tanks (Belief Question) and why (Justification Question). An accuracy code was also used, as above. Happe (1994) has suggested that this task requires a "third-order" theory of mind (i.e., "he *knows* they *think* he will *lie*").

Impact Measures: The Social Skills Rating System

The Social Skills Rating System (SSRS; Gresham & Elliott, 1990) provides a broad, multirater assessment of social behavior. Parents and teachers estimate the frequency with which a variety of specific social skills are displayed. For example, items include demonstrating interest in others, initiating conversations, inviting peers to the home, joining group activities, and waiting turns in games. The SSRS includes national norms defined by age, sex, and handicap status, and demonstrates good reliability and validity (Gresham & Elliott, 1990).

Separate parent and teacher versions of the SSRS were administered to obtain information about social competence in the home and classroom settings. Ratings were done at both the pre- and post-treatment assessments to examine change in the perceived frequency of specific social skills. Although it was not possible to keep parents blind to group assignment, teachers filled out the SSRS unaware of whether the child had participated in the treatment group or not. Raw scores obtained on the SSRS were converted to standard scores ($M = 100$, $SD = 15$) on the basis of age and gender. Handicapped norms were not used.

Social Skills Training Program

The social skills group met weekly for $4^1/_2$ months (14 sessions with holiday breaks). Each meeting was 90 minutes long and was organized and run by one primary leader with three additional staff helping. Sessions began with a snack period in which the participants could mingle with each other and the staff, practicing conversational skills. This was followed by a group discussion of the day's topic. The particular social skill being taught was broken down into simple, concrete components that could be explained easily to group members. The importance of the skill was discussed (e.g., people will like you better if you show interest in what they are saying). Trainers then modeled the skill through a role-play while the group members watched. This was followed by several videotaped role-plays by participants, accompanied by coaching from trainers. The videotapes were then reviewed and each child was given reinforcement and constructive feedback on his performance. Finally, the session ended with a game (e.g., Bingo, Wheel of Fortune).

The training program was divided into two major units, each 7 sessions long, separated by a 3-week break for the winter holidays. The first unit addressed basic interactional and conversational skills, including how to begin, maintain, and appropriately end conversations, how to choose topics that interest others, how to read, interpret, and appropriately express nonverbal signals and emotional expressions, how to negotiate and share, how to listen, how to give compliments, and how to express interest in others. While these skills were the major focus of the first training unit, they continued to be addressed throughout the remainder of the social skills program.

The second module focused on teaching perspective-taking and theory of mind skills. We first illustrated this idea physically, by having group members lead a blindfolded trainer through a maze. Children were taught how to take the blindfolded person's physical perspective, providing a good description of obstacles and possible routes, without assuming the blindfolded person could see what the child could see.

After demonstrating how physical and visual perspectives may differ, we then addressed how cognitive points of view could also differ, namely, how one person could know something that another person does not know. Borrowing from an approach suggested by Baron-Cohen and Howlin (1993), we emphasized that *perception influences knowledge,* that what one sees or hears determines what one knows (i.e., "a person will know x only if s/he saw or heard about x," Baron-Cohen & Howlin, 1993, p. 473). The participants acted out several role-plays that illustrated these principles. The role-plays were similar in form to standard first-order false belief tasks (e.g., Baron-Cohen et al., 1985, 1986). However, importantly, the specific

content of the role-plays differed from that of the outcome measures described above. In one scenario, for example, child A and child B hide a toy together, then child B changes the hiding place outside of child A's view. The group members were asked to predict where child A would look for the toy. We were surprised when, after two sessions on this topic, the boys seemed to have mastered first-order perspective-taking. Accordingly, more advanced theory of mind skills were targeted next.

For three additional sessions, we worked on second-order perspective taking, that is, predicting what one person thinks another person thinks. The role-plays mirrored the format of standard second-order false belief tasks (e.g., Baron-Cohen, 1989), but again, the specific content was different from the scenarios employed in the research measures. For example, in one role-play, children A, B, and C decide to go bowling together later that evening. After child A leaves to change clothes before bowling, B and C revise the plans, deciding instead to see a movie. B and C then part ways as well. On the way home, B stops by A's house to tell him of the new plan. They agree to meet at the movie theatre and A heads there. C, after changing his own clothes, stops by A's house and finds he is not there. The group members were asked to predict where C would think A had gone. A number of similar scenarios were acted out and discussed over the three sessions. While the boys found these quite a bit more challenging and frustrating than the first-order role-plays, they eventually seemed to appreciate the underlying heuristic. At the end of the training, most were able to articulate that since C had not *seen* B speak with A, he could not *know* that A *knew* of the revised plans.

As Mesibov (1992) has pointed out, there are several reasons why individuals with autism experience social difficulties *beyond* knowledge or skill deficits. For example, many children with autism, even those who are high-functioning, have few friends and, thus, few opportunities to socialize with others of similar age. In addition, many have been teased or rejected by peers. Therefore, in addition to teaching social skills to our participants, another goal of our program was to demonstrate to them that social interactions can be enjoyable and desirable.

We did this both within sessions, by ending each meeting with games or other pleasurable activities, and between meetings, by alternating didactic clinic sessions with community outings to places chosen by the group members themselves (e.g., video arcades, restaurants, malls). While fun, these field trips and games were also used, somewhat more covertly, to address social skills. For example, on an outing to a pizza parlor, participants were given social goals to work on (e.g., starting two conversations with people seated near them) and provided with feedback on their behavior after the session. Similarly, games were used to address the particu-

lar social skill targeted at that meeting; for example, clues in a game of Wheel of Fortune might be "a nonverbal signal that shows you are listening" or "what people do when they are frustrated."

Over the course of the intervention, we also held several parties, in which adults and children from outside the group were invited. These parties, very popular with the group members, served as an additional opportunity to practice social skills, negotiation, and planning. The boys helped organize the food, music, entertainment, and decorations for the parties. Emphasis was placed on taking the perspective of the guests and inferring what they would enjoy, not what the group members preferred. The boys were encouraged to act as hosts, introducing the guests to each other, offering food, and making appropriate conversation. While these goals were not always met, the boys became more skilled and self-confident over the course of the intervention. The highlight of the last party was an activity decided upon by the group members: They set up a video cassette recorder and showed favorite role-plays to party guests, proudly pointing out improvements in their social skills during the $4^1/_2$-month program.

RESULTS

Effect Size and Clinical Significance

When samples are small, many writers have suggested that the magnitude or size of effects be reported, in addition to or instead of statistical *p* values, to permit more appropriate interpretation of results (Baer & Ahern, 1993; Cohen, 1992; Paquin, 1983; Rosnow & Rosenthal, 1988). Particularly in treatment outcome studies with small samples, reliance upon traditional significance tests of mean differences may underestimate the magnitude of treatment effects and potentially lead to an inappropriate acceptance of the null hypothesis (Kazdin, 1980).[4]

In the present study, the low prevalence of high-functioning autism, the nature and intensity of the therapeutic intervention, and the no-treatment control group design made a large sample prohibitive. Thus, in this paper, we use both traditional significance tests and several alternative indices of change to evaluate the clinical significance of our intervention.

[4]As Kazdin (1980) has written, "statistical significance is a function of many different features of an experiment, only *one* of which is whether there is a relation between the independent and dependent variables" (p. 359). Statistical significance is also, of course, directly proportional to the size of the sample. Just as highly significant *p* values do not necessarily reflect important effects when the number of subjects is large, neither do nonsignificant *p* values always indicate trivial or nonmeaningful effects in a small sample.

The primary index used is *effect size,* that is, the absolute magnitude of change or the degree to which the treatment group improves relative to control group subjects (Jacobson, Follette, & Revenstorf, 1984; Rosnow & Rosenthal, 1988). Several other indices are also used to evaluate the efficacy of our treatment, including the proportion of clients who improve and the generalizability of improvement (Kazdin, 1980).

Performance on Theory of Mind Measures

Scores on the four theory of mind measures were summed to form an overall performance composite (maximum score possible = 13). Repeated measures analyses of variance (ANOVAs) were then used to examine performance on the theory of mind composite as a function of both group and time. There was no main effect of group, but a marginally significant main effect of time, $F(1, 7) = 12.84, p < .11$, and a marginally significant interaction of group and time, $F(1, 7) = 12.84, p < .11$. Planned contrasts of the cell means indicated that there were no differences between the two groups at pre-treatment testing; however, the treatment group's perform-

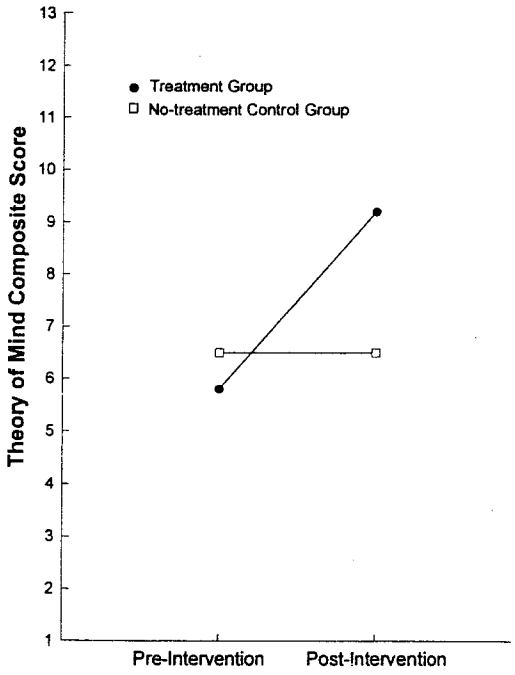

Fig. 1. Group mean performance on the theory of mind composite at pre- and post-treatment testing.

ance had improved at post-treatment assessment, $t(4) = -2.43$, $p < .08$, while that of the control group had not (Figure 1).

Two effect size indices were also calculated. The effect size of the group difference on the theory of mind composite at posttreatment testing was 0.64, classified as a medium to large effect (Cohen, 1992). The effect size of the group difference in change scores (i.e., post-treatment minus pre-treatment composite) was 1.6, considered a very large effect.[5] Thus, while our small sample made statistical significance at the .05 p level difficult to obtain, the magnitude of the differences suggests that our intervention was effective in improving performance on false belief tests.

Proportion of Subjects Improving

Table II contains the theory of mind composite scores of individual subjects in the two groups at the two time periods. Eighty percent of the treatment group (4 of 5 subjects) demonstrated improvement on the theory of mind composite relative to their pre-treatment performance, while only 25% (1 of 4 subjects) in the no-treatment control group did.

Good initial performance at pre-treatment testing reduces the amount of change possible at post-treatment assessment, however. Since two con-

Table II. Individual Subjects' Scores on the Theory of Mind Composite (Maximum = 13)

Subject	Pretreatment	Posttreatment	Change
	Treatment group		
1	6	11	+5
2	8	8	0
3	8	10	+2
4	5	13	+8
5	2	4	+2
	Control group		
1	10	9	-1
2	11	11	0
3	4	2	-2
4	1	4	+3

[5]Effect sizes for differences between independent means greater than 0.5 are considered medium effects, while those above 0.8 indicate large effects (Cohen, 1992).

Table III. Percentage of Variables Not Passed at Pretreatment Testing But Improved at Posttreatment

Subject	Treatment group	Control group
1	67	0
2	75	50
3	60	0
4	100	44
5	25	—
Mean	65.4	23.5

trol subjects (1 and 2) performed relatively well at initial testing and thus had less room for improvement, it was necessary to calculate an alternate index of change that would minimize ceiling effects. The number of variables showing improvement at post-treatment testing was divided by the number of tests not passed at pre-treatment. By individualizing the denominator of this ratio for each subject, penalties for initial ceiling performance were largely eliminated.

As can be seen in Table III the treatment group demonstrated improvement on 65.4% of tests not initially passed, while the average improvement of the control group was only 23.5%. The effect size of this group difference in mean improvement was 1.54, again a very large effect that is likely meaningful. The only subject passing all false belief tests at follow-up, who might be said to possess and reliably employ a theory of mind, was in the treatment group; this subject (4) passed less than half of the tasks at pre-treatment testing.

Correlations

Correlations between the independent variables and the post-treatment theory of mind composite were calculated to examine the influence of subject variables on treatment outcome (Table IV). Despite the high magnitude of many of the relationships, none attained statistical significance at the .05 level, due to the small sample. The effect size of most correlations, however, was medium to large, indicating clinical significance.[6]

[6]The effect size of a correlation coefficient is equal to the absolute value of *r* (Ozer, 1985); *r*s above .3 are considered medium in size, while those above .5 are considered large effects (Cohen, 1992).

Table IV. Pearson Product–Moment Correlations Between Independent Variables and the Theory of Mind Composite at Posttreatment Testing

Group	*n*	VMA[a]	Age	VIQ	PIQ	CARS
Treatment	5	.47[b]	.67[c]	.38[b]	–.07	–.79[c]
Control	4	.79[c]	.92[c]	.44[b]	–.44[b]	.00

[a]Verbal mental age = (VIQ × Age)/100.
[b]Medium effect.
[c]Large effect.

Generalization of Change

Parent and teacher SSRS scores were assessed with repeated measures ANOVAs as a function of group and time. There were no significant main or interaction effects for either parent or teacher ratings. The groups differed at neither initial testing nor posttreatment follow-up. Significant change with time was not apparent in either group. As can be seen in Table V, the magnitude of the group and time differences was very small and not always in the predicted direction. In addition, correlations between posttreatment performance on the theory of mind composite and posttreatment SSRS scores were negative and moderate in magnitude (–.2 to –.6 range), indicating that subjects scoring high on theory of mind measures were rated low by both parents and teachers on general social skills. Thus, the SSRS null results do not appear simply secondary to low statistical

Table V. Performance on the SSRS (in Standard Scores, $M = 100$, $SD = 15$)[a]

	Pretreatment		Posttreatment	
	M	SD	*M*	SD
Parent ratings				
Treatment group	74.4	10.9	71.8	10.9
Control group	78.0	11.9	79.8	16.3
Teacher ratings				
Treatment group	83.6	13.0	84.8	8.7
Control group	86.0	19.1	92.7	14.6

[a]No significant differences.

power but seem to reflect a genuine lack of treatment effects on this measure.

DISCUSSION

This study found that a social skills training program providing systematic instruction in theory of mind principles was able to substantially improve performance on several false belief tasks. Although small sample size limited the ability to demonstrate statistically significant differences between the treatment and control groups, the magnitude of the group differences and the effect size of the changes strongly suggest that the intervention had a meaningful impact. We believe that reliance on traditional statistical tests would have led to an inappropriate acceptance of the null hypothesis and an erroneous conclusion that treatment cannot modify theory of mind abilities.

In contrast to the improvement seen in the treatment group, the mean performance of the no-treatment control group did not change over the $4^1/_2$-month follow-up period. Consistent with previous longitudinal research (Holroyd & Baron-Cohen, 1993; Ozonoff & McEvoy, 1994), theory of mind abilities in this group remained static and did not improve without intervention.

A critically important question is whether the treatment program truly provided the theory of mind that autistic individuals seem to lack or whether it simply taught subjects how to solve false belief problems. That is, did we teach the ability or the task? In fact, we believe the latter possibility is more likely. Although performance on paper-and-pencil theory of mind measures improved, it was our impression that the ability to translate these principles to everyday conversations and interactions remained limited. As noted by others, there seemed to be a chasm between *possession* of mental state inference skills and regular *performance* of these skills in real-world situations (Bowler, 1992; Happe, 1994; Rutter & Bailey, 1993). Apparently, we were more successful in teaching our subjects to "hack out" rules and strategies to infer the mental states of others than in teaching them how to apply these skills under demanding "on-line" social conditions.

Despite these caveats, we still believe it is important to demonstrate that strategies for solving false belief tasks can be taught. This study found that impairment on mental state inference tasks is not necessarily an inevitable consequence of autism; with training, performance can improve to age-appropriate "normal" levels in many cases. Although it is generally recognized that intervention is effective in treating other core symptoms of autism (e.g., communication problems, stereotypic behaviors), this is the

first study to demonstrate that some aspects of theory of mind impairment can also be helped by treatment.

To promote generalization of skills taught in the training program to the real world, we provided many opportunities, modalities, and situations for practice and employed multiple trainers to teach skills. Our most important effort at promoting skill generalization, however, was to teach underlying problem-solving principles and cognitive mediational strategies that could be applied in multiple situations. For example, by teaching the principle "perception influences knowledge" (i.e., that what one sees and hears influences what one knows; Baron-Cohen & Howlin, 1993), we hoped that our group members would be able to predict the beliefs of others in a variety of situations outside the clinic.

The change seen in theory of mind performance after treatment, however, did not extend to more general ratings of social competence made by important figures in the subjects' lives. In fact, posttreatment correlations between parent and teacher SSRS scores and false belief performance were negative. If the intervention did not alter outside raters' impressions of the children, the effectiveness of the treatment might appear questionable.

On the other hand, there may be little reason to expect that theory of mind performance would predict SSRS ratings. There is little direct overlap between the constructs assessed by the two measures. For example, no questions on the SSRS specifically measure perspective-taking. Other research has found, in fact, that correlations between "specifying" measures, such as social cognition tests, and naturalistic "impact" measures are generally low and nonsignificant (Berler, Gross, & Drabman, 1982; Gresham, 1983; Hughes & Sullivan, 1988; Malik & Furman, 1993). Therefore, failure to demonstrate changes on the SSRS in the treatment group may not reflect failure of generalization, but may indicate only that the SSRS is an insensitive measure of theory of mind.

It may also be the case that the social behaviors measured by the SSRS are difficult to alter in the course of a $4^1/_2$-month intervention. For example, several SSRS items measure behavior that may take considerable time to acquire (e.g., "is self-confident in social situations," "is liked by others").

Finally, it is possible that more improvement in SSRS scores would have been evident had the measure been administered upon completion of the first 7-week unit of the training program, rather than after the full $4^1/_2$-month intervention. The SSRS appears to measure basic interactional and conversational skills, which were the focus of the first training module. The ability to measure changes in these skills may have been attenuated by performing the posttreatment assessment 10 weeks after completion of the most directly relevant component of the training. We chose, however, to admin-

ister the SSRS at the end of the full program because we were interested in whether understanding others' perspectives and gaining some insight into theory of mind processes would alter general social skills ratings. Since we considered the training provided in the two modules to be interdependent, the SSRS was included as an impact measure for the full training program, rather than as a specifying measure for the first unit alone.

The present study improves on those preceding it by using both specifying and impact assessment measures, blind ratings of outcome, and a no-treatment control group. There is clearly room for improvement in future studies, however. Especially important is the development of more naturalistic, socially valid measures of outcome and new methods to assess generalization. For example, analog measures that evaluate the ability to infer mental states in everyday situations would be very helpful. Work is currently underway in our laboratory to examine the extent of change on such measures. Assessing the impact of false belief training on pragmatic communication could also prove informative.

Another important charge for future investigators is the training and assessment of a wider range of mental state inferences. In the present study, we examined the ability to understand the mental states of knowledge, belief, deception, and intention. It is important to explore the modifiability of other mental state inferences, including desire, pretense, and misunderstanding.

Future research projects should also include an assessment of the durability of treatment effects. We tested our subjects within 2 weeks of the end of treatment. It is important to examine whether improvements on false belief task performance maintain over an extended follow-up period.

In this day of uncertain insurance coverage and third-party payments, professionals must hold themselves accountable for their treatments as never before. We must document the effectiveness of our interventions, examine which techniques are most important to facilitating improvement, assess which subject characteristics are most predictive of change, and examine how long the treatment must last for maximal benefit. The present study is a first step toward answering these difficult questions.

Finally, we hope that this work begins to bridge the gap between basic and applied research that exists in our field. While there has been some success in recent years in identifying central, core impairments of autism, few of these findings have yet been translated into clinical interventions. As many parents of research subjects wonder, how can our empirical investigations of underlying causal mechanisms help their children? This study is a first attempt to address that question. We hope it encourages many more.

REFERENCES

Baer, L., & Ahern, D. K. (1993). Statistical problems with small sample size. *American Journal of Psychiatry, 150,* 356.

Baron-Cohen, S. (1989). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry, 30,* 285-297.

Baron-Cohen, S., & Howlin, P. (1993). The theory of mind deficit in autism: Some questions for teaching and diagnosis. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from autism* (pp. 466-480). New York: Oxford University Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind?" *Cognition, 21,* 37-46.

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

Berler, E. S., Gross, A. M., & Drabman, R. S. (1982). Social skills training with children: Proceed with caution. *Journal of Applied Behavior Analysis, 15,* 41-53.

Bowler, D. (1992). Theory of mind in Asperger's Syndrome. *Journal of Child Psychology and Psychiatry, 33,* 877-893.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

Goldstein, A. P. (1988). *The Prepare Curriculum: Teaching prosocial competencies.* Champaign, IL: Research Press.

Gresham, F. M. (1983). Social validity in the assessment of children's social skills: Establishing standards for social competency. *Journal of Psychoeducational Assessment, 1,* 299-307.

Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system.* Circle Pines, MN: American Guidance Service.

Happe, F. G. E. (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.

Holroyd, S., & Baron-Cohen, S. (1993). Brief report: How far can people with autism go in developing a theory of mind? *Journal of Autism and Developmental Disorders, 23,* 379-385.

Hughes, J. N., & Sullivan, K. A. (1988). Outcome assessment in social skills training with children. *Journal of School Psychology, 26,* 167-183.

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15,* 336-352.

Kazdin, A. E. (1980). *Research design in clinical psychology.* New York: Harper & Row.

La Greca, A. M. (1993). Social skills training with children: Where do we go from here? *Journal of Clinical Child Psychology, 22,* 288-298.

Leekam, S., & Perner, J. (1991). Does the autistic child have a theory of representation? *Cognition, 40,* 203-218.

Leslie, A. M., & Frith, U. (1988). Autistic children's understanding of seeing, knowing and believing. *British Journal of Developmental Psychology, 6,* 315-324.

Malik, N. M., & Furman, W. (1993). Problems in children's peer relations: What can the clinician do? *Journal of Child Psychology and Psychiatry, 34,* 1303-1326.

Matson, J. L., & Swiezy, N. (1994). Social skills training with autistic children. In J. L. Matson (Ed.), *Autism in children and adults: Etiology, assessment and intervention* (pp. 241-260). Pacific Grove, CA: Brooks/Cole.

Mesibov, G. B. (1984). Social skills training with verbal autistic adolescents and adults: A program model. *Journal of Autism and Developmental Disorders, 14,* 395-404.

Mesibov, G. B. (1992). Treatment issues with high-functioning adolescents and adults with autism. In E. Schopler & G.B. Mesibov (Eds.), *High-functioning individuals with autism* (pp. 143-155). New York: Plenum Press.

Ozer, D. J. (1985). Correlation and the coefficient of determination. *Psychological Bulletin, 97,* 307-315.

Ozonoff, S., & McEvoy, R. E. (1994). A longitudinal study of executive function and theory of mind development in autism. *Development and Psychopathology, 6,* 415-431.

Paquin, M. J. R. (1983). Beyond significant but meaningless results in psychotherapy research. *Psychotherapy: Theory, Research and Practice, 20,* 38-40.

Parker, J. G., & Asher, S. R. (1987). Peer relations and later personal adjustment: Are low-accepted children at risk? *Psychological Bulletin, 102,* 357-389.

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

Rosnow, R. L., & Rosenthal, R. (1988). Focused tests of significance and effect size estimation in counseling psychology. *Journal of Counseling Psychology, 35,* 203-208.

Rutter, M., & Bailey, A. (1993). Thinking and relationships: Mind and brain (some reflections on theory of mind and autism). In S. Baron-Cohen, H. Tager-Flusberg & D. Cohen (Eds.), *Understanding other minds: Perspectives from autism* (pp. 481-504). New York: Oxford University Press.

Schopler, E., & Mesibov, G. B. (Eds.). (1983). *Autism in adolescents and adults.* New York: Plenum Press.

Schopler, E., Reichler, R. J., & Renner, B. R. (1988). *The Childhood Autism Rating Scale (CARS).* Los Angeles: Western Psychological Services.

Walker, H. M., McConnell, S., Holmes, D., Todis, B., Walker, B., & Golden, N. (1983). *The Walker Social Skills Curriculum: The Accepts Program.* Austin, TX: Pro-Ed.

Williams, T. I. (1989). A social skills group for autistic children. *Journal of Autism and Developmental Disorders, 19,* 143-155.