

Effects and Moderators of a Short Theory of Mind Intervention for Children with Autism Spectrum Disorder: A Randomized Controlled Trial

Sander Begeer, Patricia Howlin, Elske Hoddenbach, Cassandra Clauser, Ramon Lindauer, Pamela Clifford, Carolien Gevers, Frits Boer, and Hans M. Koot

Limited perspective taking or “Theory of Mind” (ToM) abilities are a core deficit of autism, and many interventions are aimed to improve ToM abilities. In this study, we investigated the effectiveness of a ToM treatment for children with autism spectrum disorders (ASD) and, for the first time, the moderating roles of social interaction style (SIS) and disruptive behavior (DB), to determine which children are most likely to respond to this intervention. The trial protocol is registered at www.trialregister.nl, trial number 2327 and published before the data collection was finished (www.trials-journal.com). Children with autism aged 7–12 years ($n = 97$) were randomized over a waitlist control or a treatment condition. Outcome measures included ToM and emotion understanding, parent and teacher questionnaires on children’s social skills, ToM-related social behavior, and autistic traits. Six-month follow-up parent reported data were collected for the treatment group. The treatment had a positive effect on ToM understanding, parent-reported ToM behavior, and autistic traits, but not on parent or teacher-reported social behavior. Passive SIS was associated with diminished treatment effects on autistic traits, but DB was unrelated to outcomes. The ToM intervention improved conceptual social understanding and ToM-related behavior of children with ASD. However, broader application of learned skills to other domains of functioning was limited. Individual differences with regard to treatment response are discussed. *Autism Res* 2015, 8: 738–748. © 2015 International Society for Autism Research, Wiley Periodicals, Inc.

Keywords: autism; treatment; randomized controlled trial; theory of mind; moderator

Introduction

Autism Spectrum Disorders (ASD) are lifelong neurodevelopmental disorders, defined by impairments in social interaction and communication, and restricted, repetitive behaviors [APA, 2013]. A core deficit of autism is limited perspective taking, or “Theory of Mind” (ToM) ability. ToM refers to the understanding of mental states, such as beliefs and desires, and how they influence behavior, and is generally regarded as a cornerstone of social competence [Wellman, Cross, & Watson, 2001]. Limitations in ToM impair functioning in almost all aspects of daily life [Peterson, Slaughter, & Paynter, 2007].

Various intervention approaches have been developed with a specific focus on ToM abilities [Fletcher-Watson, McConnell, Manola, & McConachie, 2014]. These include programs such as “Teaching ToM,” “Picture in the Head training,” or “Thought Bubble training,” all of which aim to teach children about mental representations of themselves and those around them. Despite

the popularity of these interventions, evidence supporting their effectiveness is limited. A consistent finding with regard to children with ASD is that they show improvements within the specific domain(s) taught, but fail to generalize these to other situations [Begeer et al., 2011; Hadwin, Baron-Cohen, Howlin, & Hill, 1996].

Several randomized controlled trials (RCTs) have highlighted the limitations of ToM interventions. For instance, while children with ASD (4–15-years old) showed improvements in their understanding of beliefs and emotions, the effects did not generalize to other domains, including imaginative play [Hadwin et al., 1996], teacher reported ToM skills [Fisher & Happe, 2005; Paynter & Peterson, 2013] or self-reported empathy or parent-reported social behavior [Begeer et al., 2011]. A computerized social skills program game did show treatment effects that generalized to parent- and teacher-reported social skills but there were no direct measures of children’s behavior [Beaumont & Sofronoff, 2008]. In short, ToM skills can be trained, but

From the VU University, Department of Developmental Psychology, Van der Boechorststraat 1, 1081BT Amsterdam, The Netherlands and EMGO Institute for Health and Care Research, Van der Boechorststraat 7, Amsterdam, The Netherlands (S.B., E.H., H.M.K.); Wei43, Weissenbruchstraat 43, 1058 KN Amsterdam (P.C., C.G.); De Bascule, Rijksweg 145 1115, AP Duivendrecht (E.H., C.C., R.L., F.B.); School of Psychology, Brennan MacCallum Building (A18) The University of Sydney NSW 2006 (S.B.); Institute of Psychiatry, King’s College, 16 De Crespigny Park London SE5 8AF (P.H.); Faculty of Health Sciences, C42 - Cumberland Campus, The University of Sydney Lidcombe NSW 2141, Australia (P.H.)

Received December 04, 2014; accepted for publication February 23, 2015

Address for correspondence and reprints: Sander Begeer, Developmental Psychology, VU University Amsterdam, Van der Boechorststraat 1, 1081BT Amsterdam, The Netherlands. E-mail: S.Begeer@vu.nl

Published online 6 April 2015 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/aur.1489

© 2015 International Society for Autism Research, Wiley Periodicals, Inc.

maintenance and generalization are limited [Fletcher-Watson et al., 2014].

The modest effects of these RCTs may indicate limited effects of the treatment, but could also be attributed to the way outcomes are assessed. Outcome measures, in particular parent or teacher questionnaires, are often designed to assess stable features or capacities in children and can be insensitive to change. For instance, the Children's Social Behavior Questionnaire [Begeer et al., 2011] includes questions about core autism features (e.g., "does your child fail to show empathy when another person is hurt or sad?") which may be unlikely to change after a short intervention. Instead, examining the *relative occurrence* of explicit treatment related behavior over a specified time period may improve the detection of behavioral change (e.g., "how often did you child fail to show empathy when another person was hurt or sad last week?"). In this study, we measured the frequency of ToM-related behavior during a designated time frame, using the ToM Behavior Checklist (ToMBC).

The nature of social interactive problems among individuals with ASD varies widely [Jones & Klin, 2009] and individual differences may moderate the effects of ToM interventions. Kanner's early writings [1943] focused on the social withdrawal of children with autism and indeed, many children with ASD do refrain from social engagement [Kanner, 1943]. However, others will actively seek out contact, albeit often in an odd manner, for instance by talking incessantly without monitoring others' interest in the topic [Beglinger & Smith, 2005]. Wing and Gould [1979] suggested three social subtypes of autism: active-but-odd, passive, and aloof. The active group of individuals does engage with others, but in an idiosyncratic, socially naive manner. Passive individuals may engage with others, but do not initiate the interaction. Aloof individuals remain indifferent even when others attempt to engage with them. Aloof behavior is most frequently observed in intellectually disabled children with ASD [Beglinger & Smith, 2005]. A recent study showed that social interaction style (SIS) can be reliably assessed in participants with ASD and average intelligence [Scheeren, Koot, & Begeer, 2012].

Children who present with the "classic" passive interaction style may benefit from interventions that target their motivation to interact with others, a common theme in many interventions aimed at preschool aged children with ASD, including joint attention and joint engagement programs [Kasari & Patterson, 2012]. Children with active-but-odd interaction styles do not lack the motivation to interact but fail in appropriately adapting their social interactions to the social context. For them, a fundamentally different type of intervention approach may be required. This study is the first to take interaction style into account when examining treatment effects.

In addition to interaction style, disruptive behavior (DB) has been postulated as a moderator of treatment effects in children with ASD. Up to one third of children and adolescents with ASD shows conduct problems, including disruptive or hyperactive behavior [Bears, Johnson, Handen, Smith, & Scahill, 2013; Kaat & Lecavalier, 2013]. Co morbid hyperactive and conduct problems have been found to reduce treatment effects [Antshel et al., 2011]. Similarly, children with attention deficit hyperactivity disorder (ADHD) (but no ASD) have been found to benefit very little from these types of treatments [Storebo, Gluud, Winkel, & Simonsen, 2012]. Parents of children with more behavioral difficulties may also face particular problems in encouraging the use of taught skills in the home setting.

Thus, besides the question of whether ToM interventions are *generally* successful, a further crucial question is "for whom are these interventions most effective?" Studies on moderators of treatment effects in autism are rare, and those that do exist tend not to focus on autism specific outcomes, but rather on issues such as the effects of medication on DB [e.g., Farmer et al., 2012].

In this RCT, we explored the impact of a shortened version of the ToM training program [Hoddenbach et al., 2012] for children with ASD and average intelligence quotient (IQ). While a variety of moderators can be studied, any treatment trial can only validly test one or two main hypotheses at a time [Kraemer, Wilson, Fairburn, & Agras, 2002]. We specifically focused on the moderating roles of SIS and DB because they represent important sources of variation in the style of real life social behavior of children with ASD [Kohls, Chevallier, Troiani, & Schultz, 2012; Kaat & Lecavalier, 2013]. SIS reflects the self-initiated inclination to seeking social interactions, which could be considered a prerequisite for using ToM. After all, a child's spontaneous orientation toward others will affect the need to use and develop ToM skills. DB qualifies the type of social interactions, in particular negative interactions, which may cause lower or different treatment effects.

Outcome measures focused on both social knowledge and social performance. As there is no single "gold-standard" measure of ToM ability, child measures included scores on two tasks of ToM and one of emotion understanding. Teacher and parent measures targeted their perceptions of children's performance on the domains of ToM and social skills. The maintenance of treatment effects 6-month post-treatment was evaluated using parent report measures.

Hypotheses

As the focus of the intervention was on ToM and emotional understanding, we expected to find the strongest effect on child measures specifically designed to target

these skills. We did not specify which of the three measures used would be likely to reflect most change, given the lack of existing data on optimal measures of ToM [Sprung, 2010]. With regard to parent and teacher reports, we expected improvements in general social skills, but we anticipated to find the strongest effects on the questionnaire targeting ToM-related behavior. Children with more DB were expected to benefit less from treatment as problems of this kind are likely to interfere with treatment gains. The moderating effect of SIS was explored, but given the lack of research in this area no specific hypotheses were proposed.

Method

Study design

The study was a RCT with an intervention and a waitlist control group. The Medical Ethics Committee of the VU University Medical Centre approved the project (project no. 2010/241). The protocol was specified before the start of the trial (www.trialregister.nl, trial number 2327) and published before the data collection was finished. Full details of the trial protocol are available at www.trialsjournal.com [Hoddenbach et al., 2012].

Participants

A total of 97 children (93% boys) aged between 7 and 12 years of age ($M = 9.6$; $SD = 1.2$) was recruited through referrals at the outpatient service of a treatment center, and met the eligibility criteria of: (1) an ASD according to the DSM-IV-TR [APA, 2001], based on multiple assessments by psychologists and psychiatrists working independently of this study, and (2) a verbal IQ score within the normal range or above (>70), based on the Peabody Picture Vocabulary Test – III-NL (PPVT) [Dunn & Dunn, 2004]. Before joining the study, parents gave informed consent. Figure 1 summarizes participant flow through the study.

The treatment group included a similar number of girls ($n = 4$, 9%) compared with the control group ($n = 3$, 6%; ns.). The mean number of treatments the children previously received was 6.5 ($SD = 3.0$); 35% of children were on medication, and 87% were from two-parent families. Mothers (76%) and fathers (88%) were generally employed and most had a college degree or higher (mothers 63%, fathers 59%). No differences were found between the treatment and control groups on any of these demographic variables and scores on assessment measures did not differ (Table 1).

Procedure

Participants were recruited from an academic center for child and adolescent psychiatry in Amsterdam, The Netherlands, between April 2010 and March 2013.

Approximately 270 children annually are referred to the center's outpatient clinic for ASD. An independent researcher randomized the participating children to treatment or waiting list conditions using a digital random number generator. The randomization outcome was shared with the primary investigator, who informed parents about which arm of the trial they would be in. Baseline assessment took place before random allocation. Follow-up data, using parent reports on the social skills questionnaire (SSQ) and the ToMBC were collected 6 months after intervention ceased.

Intervention

The "Mini ToM intervention" is a manualized, weekly intervention for groups of five to six children, all aged within 3 years of each other. It involves eight sessions of approximately 1 hr. Sessions are supervised by certified therapists, licensed Counseling Psychologists, M.Sc. or Ph.D., registered with the Mental Health Council, who received specific training for this intervention. The training is delivered in a child psychiatric center. The program was based on a shortened version of a ToM-intervention [Begeer et al., 2011; Steerneman, Jackson, Pelzer, & Muris, 1996]. The intervention was shortened to be more cost effective, while the key elements of the treatment were preserved. All sessions followed the same structure: (1) discussing the homework assignment; (2) games and exercises related to the day's theme; (3) children summarizing the session to their parents; and (4) explanation of the following week's homework assignment. A detailed explanation of the treatment is freely available through the published trial protocol, available at www.trialsjournal.com [Hoddenbach et al., 2012].

Descriptive measures

Peabody picture vocabulary test – III-NL. The PPVT [Dunn & Dunn, 2004] is a receptive language and screening test for verbal comprehension. It is highly correlated with a more general measure of verbal IQ, the WISC-III verbal IQ [Hodapp & Gerken, 1999].

Primary outcome measures

Child-based measures. *ToM test* The ToM test [Muris et al., 1999] is a standardized interview for children aged 5–13 years. It measures ToM knowledge at three levels (Elementary, Intermediate, and Complex), with cognitive substages within each stage including perception and imitation, emotion recognition, elementary ToM, second-order belief understanding, and understanding of complex humor. Children listen to a hypothetical story and/or look at a picture and answer the corresponding question. The test contains 72 items, each scored 0 (incorrect) or 1 (correct). Higher scores indicate higher levels of ToM knowledge. Test-retest reliability is high (ICCs 0.80–0.99).

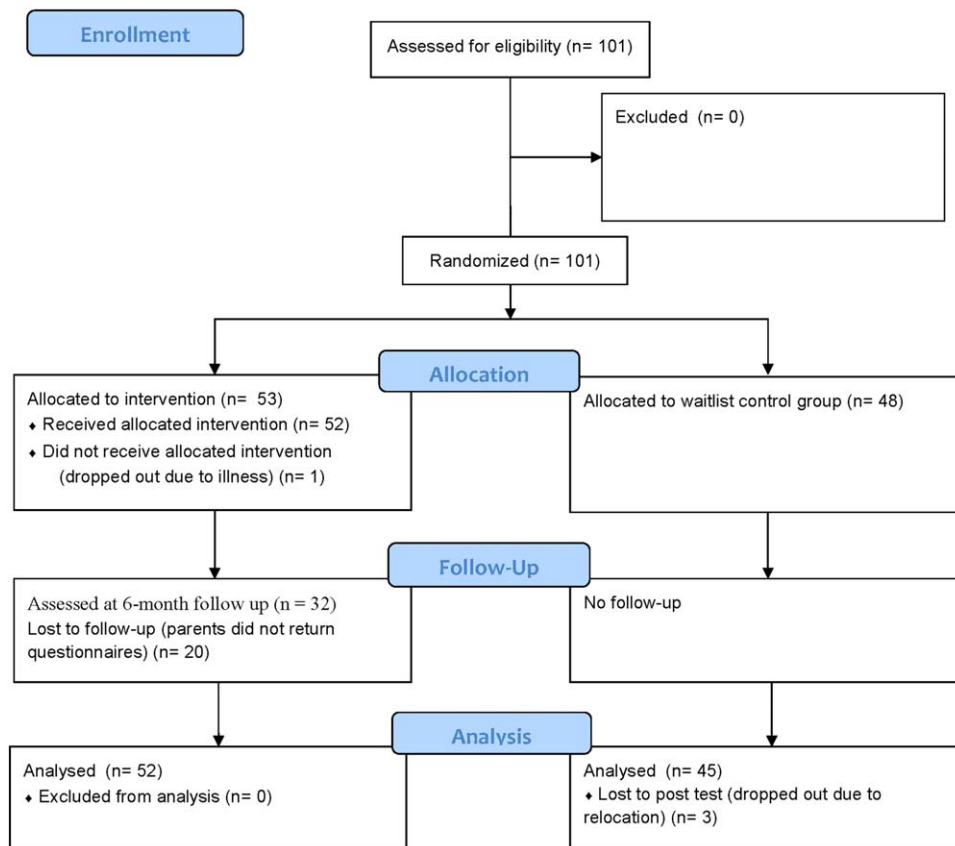


Figure 1. CONSORT 2010 flow diagram.

ToM-advanced test The ToM-advanced test [Scheeren, de Rosnay, Koot, & Begeer, 2013] measures five advanced forms of ToM: understanding second-order false belief, emotional display rules, violation of social rules, double bluff, and sarcasm. The children listen to a story and answer questions about mental states. Each question is scored on a three-point Likert scale (0 = incorrect, 1 = correct but not complete, 2 = correct). Higher scores indicate a better advanced ToM understanding. The inter-rater reliability is good to very good (Kappa's 0.57–1.00).

Levels of emotional awareness scale for children The levels of emotional awareness scale for children (LEAS-C) [Bajgar, Ciarrochi, Lane, & Deane, 2005] is a questionnaire to assess children's emotional awareness and contains 12 scenario's describing hypothetical social situations.

Children are asked how they would feel in the described situation. The use of complex emotions (e.g., guilt or embarrassment) and double perspective (highlighting own and other person's feelings) are coded. They can attribute these emotions to themselves (one point), the other (two points), or to both (three points). Higher scores indicate more awareness of emotions. Internal consistency is moderate (alpha 0.64–0.71) and convergent validity acceptable.

Parent and teacher measures. *Social skills questionnaire* The SSQ is a widely used questionnaire designed to assess parents' (SSQ-P) and teachers' (SSQ-T) perception of the child's social skills. It contains 30 items, rated 0 (not true), 1 (sometimes true), or 2 (mostly true). A

Table 1. Baseline Demographic and Clinical Characteristics of the ToM Treatment and the Waitlist Control Groups

Group	ToM treatment (<i>n</i> = 45) Mean (SD) [range]	Waitlist control (<i>n</i> = 52) Mean (SD) [range]	<i>t</i> -value
Child age (y)	9.7 (1.3) [7.5–12.5]	9.5 (1.2) [7.5–12.2]	−0.78 ^{ns}
Verbal ability (PPVT standard score)	107.6 (13.6) [79–133]	105.1 (12.5) [82–126]	−0.84 ^{ns}
Social interaction style			
Active-but-odd scale	29.5 (4.8) [0–78]	28.5 (4.8) [0–78]	−1.04 ^{ns}
Passive scale	24.6 (3.6) [0–78]	25.5 (3.3) [0–78]	1.33 ^{ns}
Disruptive behavior			
DBD attention deficit	13.3 (5.4) [0–27]	12.8 (5.8) [0–27]	−0.42 ^{ns}
DBD hyperactivity	11.8 (5.2) [0–27]	10.6 (6.0) [0–27]	−1.00 ^{ns}
DBD ODD	7.8 (4.7) [0–24]	8.6 (5.7) [0–24]	0.77 ^{ns}
DBD CD	1.8.7 (2.7) [0–48]	2.1 (2.5) [0–48]	0.70 ^{ns}
Social Responsiveness Scale			
SRS total score	84.0 (22.2) [0–260]	84.0 (19.5) [0–260]	0.29 ^{ns}
SRS social awareness	11.7 (3.2) [0–32]	12.2 (3.1) [0–32]	0.82 ^{ns}
SRS social cognition	16.5 (5.3) [0–48]	16.2 (4.5) [0–48]	−0.27 ^{ns}
SRS social communication	28.1 (8.3) [0–88]	28.4 (7.5) [0–88]	0.17 ^{ns}
SRS social motivation	12.5 (5.3) [0–44]	13.3 (3.7) [0–44]	0.87 ^{ns}
SRS autistic mannerisms	14.0 (5.0) [0–48]	14.0 (5.0) [0–48]	−0.07 ^{ns}

Note. DBD, disruptive behavior disorders rating scale; ODD, oppositional defiant disorder; CD, conduct disorder.
^{ns} not significant.

total score based on summing the numerical ratings for each item with higher scores reflecting greater social skills (range is 0–60). Internal consistency for the SSQ is good (Guttman split-half reliability of 0.90) [Spence, Donovan, & Brechman-Toussaint, 1999].

ToM behavior checklist The ToMBC is a parent questionnaire designed to assess parental observations of specific ToM-related behaviors of their child. Parents were asked to specify the frequency over the past week (0-never to 5-very often) of eight domains of behavior (understood a joke, comforted somebody, asked about someone's feelings, figured out his/her story was not interesting to others, apologized, paid close attention to somebody's story, spontaneously complimented someone, asked an interested question). These behaviors were chosen based on information provided by parents during meetings set up to evaluate the previous version of the intervention, described in Begeer et al. [2011]. Higher scores indicated a higher frequency of ToM-related behaviors. Internal reliability of the ToMBC in this sample was 0.81 [Begeer et al., 2011].

Secondary outcome measure

Social responsiveness scale. The social responsiveness scale (SRS) [Constantino & Gruber, 2007] is a parent questionnaire designed to assess autistic features. It comprises 65 items covering five subscales: social awareness, social cognition, social communication, social motivation, and autistic manners. Each item is rated from 0 (never true) to 3 (almost always true). Higher scores indicate more autistic features. Good reliability and validity have been reported [Bolte, Poustka, & Constantino, 2008]. As the SRS focuses specifically on social

skills, it was also used as an additional outcome measure following the intervention. In this sample, nine children did not meet the cutoff for an autism diagnosis [i.e., *T* score ≥ 60 ; cf. Bolte et al., 2008]. However, excluding these children (whose *T* scores ranged from 34 to 59) from the analysis did not change any of our findings.

Measures of moderating variables

Wing subgroups questionnaire. The Wing subgroups questionnaire (WSQ) [O'Brien, 1996] is a parent questionnaire to determine the social subtypes of autism described by Lorna Wing [Wing & Gould, 1979]. We used just two subtypes (active-but-odd vs. passive), to distinguish between an active and a passive interaction style. We did not focus on aloof behavior, as this type of behavior, which represents a more extreme form of passiveness, is less prevalent among children with ASD of average IQ. The WSQ contains 13 descriptions of active-but-odd and passive behavior. Parents evaluated how often the described behavior occurred for their children on a seven-point Likert scale (0 (never) to 6 (always)). Higher scores indicate a higher prevalence of the SIS under concern. The internal consistency of the WSQ is moderate to good [Scheeren et al., 2012].

Disruptive behavior disorders rating scale. The disruptive behavior disorders (DBD) [Pelham, Gnagy, Greenslade, & Milich, 1992] parent questionnaire assesses DB in children aged 6–16-years old, including 42 items related to attention deficits (9 items), hyperactivity/impulsivity (9 items), Oppositional-Defiant

Table 2. Child Measures and Parent/Teacher Reports at Pretest and Post-Test for the ToM Treatment and Waitlist Control Groups

Outcomes [range]	ToM treatment Mean (SD)		Waitlist control Mean (SD)		$F_{(1, 96)}^a$	η_p^2
	Pretest	Post-test	Pretest	Post-test		
Child measures						
Theory of Mind total score [0–72]	51.68 (6.37)	58.15 (5.18)	50.55 (6.32)	54.54 (5.94)	6.25**	0.06
Advanced ToM test [0–10]	3.14 (1.71)	3.89 (1.69)	3.65 (1.81)	4.24 (1.87)	0.20	0.00
LEAS-C total [0–48]	34.34 (7.66)	37.80 (7.43)	32.02 (7.92)	33.15 (7.38)	2.68	0.03
LEAS-C mixed emotions [0–36]	1.80 (2.63)	2.32 (3.75)	0.82 (1.61)	1.13 (2.22)	0.13	0.00
LEAS-C complex emotions [0–36]	2.24 (4.07)	3.61 (4.01)	2.04 (3.31)	2.57 (4.36)	0.88	0.01
Parent reports						
SSQ [0–60]	41.71 (7.16)	42.33 (8.01)	38.75 (6.69)	37.98 (6.69)	2.94	0.03
ToMBC [0–40]	21.81 (4.67)	25.61 (4.77)	21.09 (4.97)	22.06 (4.80)	5.00*	0.05
SRS total [0–260]	82.81 (22.19)	77.94 (22.65)	84.04 (19.52)	86.29 (19.41)	6.85*	0.07
SRS social awareness [0–32]	11.68 (3.06)	11.71 (3.22)	12.19 (3.06)	12.73 (2.92)	0.78	0.01
SRS social cognition [0–48]	16.51 (5.29)	14.75 (4.65)	16.24 (4.48)	17.19 (4.55)	14.92**	0.14
SRS social communication [0–88]	28.09 (8.32)	26.39 (8.31)	28.36 (7.54)	28.35 (7.89)	1.96	0.02
SRS social motivation [0–44]	12.49 (5.25)	11.65 (4.85)	13.28 (3.67)	13.29 (4.08)	1.28	0.01
SRS autistic mannerisms [0–48]	14.04 (5.01)	13.43 (5.14)	13.96 (5.72)	14.74 (4.98)	3.78	0.04
Teacher reports						
SSQ [0–60]	40.22 (8.60)	43.99 (6.77)	37.36 (9.34)	40.22 (9.18)	0.36	0.00

Note.

^a Post-treatment analyses evaluated the effect of treatment condition while accounting for baseline group differences.

LEAS-C, levels of emotional awareness scale for children; SSQ, social skills questionnaire; ToMBC, ToM behavior checklist; SRS, social responsiveness scale.

* $P < 0.05$; ** $P < 0.01$.

Disorder (8 items), and Conduct Disorder (16 items). Parents evaluate how well each item fits their child from 0 (not at all) to 3 (a lot). Higher scores reflect greater disorder. Adequate psychometric properties of the DBD have been reported in both community samples [Pelham et al., 1992] and ASD samples [Demurie, Roeyers, Baeyens, & Sonuga-Barke, 2011; Geurts, Vertie, Oosterlaan, Roeyers, & Sergeant, 2004; Scheeren et al., 2012].

Statistical analyses

Pretest differences in demographic and clinical characteristics were investigated using Chi-square tests and analyses of variance. To test intervention effects, repeated measures analyses of variance were performed to compare group differences in changes on child and informant measures. To test moderating effects of SIS and DB, Group (0 = waitlist control, 1 = treatment) \times SIS and Group \times DB effects on change in conceptual and practical skills were tested using multiple regression analyses.

Moderating effects of SIS and DB on the ToM intervention were analyzed for those outcomes that showed a significant treatment effect. All moderator variables were centered by subtracting the means, and interaction terms with group were computed before entering them into a regression model. Outcome measures were

analyzed using their pretest values in the first step of the regression model.

Results

Child-based measures

ToM test. Compared with the control group, the treatment group showed significantly more improvement in their total score on the ToM test from pretest to post-test, $F_{(1, 95)} = 6.95$, $P = 0.01$, $\eta_p^2 = 0.07$ (See Table 2). Subscale analysis indicated that group differences occurred only at the Intermediate level which includes first-order and false belief reasoning tasks $F_{(1, 95)} = 14.01$, $P = 0.001$, $\eta_p^2 = 0.13$. No group differences were found for the easier Elementary level which includes perception and imitation tasks, $F_{(1, 95)} = 1.62$, $P = 0.21$, $\eta_p^2 = 0.02$, or the more Complex level, which includes second-order belief reasoning and understanding of humor, $F_{(1, 95)} = 0.81$, $P = .37$, $\eta_p^2 = 0.01$.

Advanced ToM and emotion understanding. No group differences were found in pretest to post-test gains made in advanced ToM understanding, $F_{(1, 95)} = 0.05$, $P = 0.82$, $\eta_p^2 = 0.00$, or measures of emotion understanding, including scores on the LEAS-C total, $F_{(1, 95)} = 3.46$, $P = 0.07$, $\eta_p^2 = 0.04$, LEAS-C mixed emotions, $F_{(1, 95)} = 0.51$, $P = 0.48$, $\eta_p^2 = 0.00$, and the LEAS-

Table 3. Regression Analysis of Predictors of Treatment Effects

Predictor	Post-test scores on								
	ToM task			ToMBC			SRS		
	β	R^2 change	Total R^2	β	R^2 change	Total R^2	β	R^2 change	Total R^2
Social interaction style (SIS)									
Pretest score	0.63	0.43***		0.61	0.44***		0.65	0.62***	
Group	0.26	0.07***		0.24	0.05**		-0.17	0.03**	
WSQ-active-but-odd	0.18	0.00		-0.11	0.00		0.17	0.00	
WSQ passive	-0.08	0.00		-0.02	0.00		0.00	0.01	
Group*WSQ active-but-odd	-0.19	0.02		0.05	0.00		-0.03	0.00	
Group*WSQ passive	0.03	0.00		-0.01	0.00		0.22	0.02*	
			0.53***			0.50***			0.69***
Disruptive behavior (DB)									
Pretest score	0.62	0.43***		0.55	0.44***		0.76	0.62***	
Group	0.27	0.07***		0.24	0.05**		-0.17*	0.03**	
DBD attention deficit	0.01	0.01		0.02	0.02		-0.01	0.02	
DBD hyperactivity	0.06	0.00		0.00	0.00		0.02	0.00	
DBD ODD	-0.08	0.01		-0.10	0.02		0.04	0.22	
DBD CD	0.07	0.00		-0.07	0.00		0.01	0.12	
Group* DBD attention deficit	-0.09	0.01		-0.08	0.00		0.02	0.00	
Group*DBD hyperactivity	-0.07	0.00		0.15	0.00		-0.03	0.14	
Group* DBD ODD	-0.05	0.00		-0.11	0.00		-0.02	0.01	
Group* DBD CD	-0.05	0.00		-0.15	0.00		0.02	0.02	
			0.53***			0.55***			0.65***

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

WSQ, Wing subgroups questionnaire; SRS, social responsiveness scale; DBD, disruptive behavior disorders rating scale; ODD, oppositional ant disorder; CD, conduct disorder.

From the VU University, Department of Developmental Psychology and EMGO Institute for Health and Care Research, Amsterdam, The Netherlands (S.B., E.H., H.M.K.); Wei43, Amsterdam (P.C., C.G.); De Bascule, Amsterdam (E.H., C.C., R.L., F.B.); School of Psychology, University of Sydney (S.B.); Institute of Psychiatry, King's College, London (P.H.); Faculty of Health Sciences, University of Sydney (P.H.)

C complex emotions, $F_{(1, 95)} = 0.84$, $P = 0.36$, $\eta_p^2 = 0.01$.

Parent and teacher measures

Social skills questionnaire. No group differences were found in pretest to post-test gains on the SSQ for either parent rated, $F_{(1, 95)} = 2.94$, $P = 0.09$, $\eta_p^2 = 0.03$, or teacher rated forms, $F_{(1, 69)} = 1.34$, $P = 0.25$, $\eta_p^2 = 0.02$.

ToM behavior checklist. Significantly higher pretest to post-test gains were reported on the ToMBC in the treatment compared with the waitlist control condition, $F_{(1, 95)} = 4.58$, $P = 0.04$, $\eta_p^2 = 0.05$.

Social responsiveness scale. The treatment group showed a significant reduction of autistic features from pretest to post-test on the SRS total score compared with the control group, $F_{(1, 95)} = 6.85$, $P = 0.01$, $\eta_p^2 = 0.07$. The difference in reduction on the SRS scores in the treatment versus the control groups was primarily driven by the strong reductions on the SRS subscale

social cognition, $F_{(1, 95)} = 14.95$, $P = 0.00$, $\eta_p^2 = 0.14$ (see Table 3).

Follow-up measures

The return rate for the 6-month follow-up measure was low ($n = 30$, 31%). No change was found on scores on the parent reported SSQ scores from post-test ($M = 42.6$, $SD = 8.4$) to 6-month follow-up ($M = 44.8$, $SD = 5.7$), $F_{(1, 28)} = 3.48$, $P = 0.07$, $\eta_p^2 = 0.11$, nor on the ToMBC scores from post-test ($M = 26.2$, $SD = 4.0$) to 6-months follow-up ($M = 26.0$, $SD = 3.1$), $F_{(1, 28)} = 0.11$, $P = 0.74$, $\eta_p^2 = 0.00$.

Moderators of treatment effect

Social interaction style. All regression models included the pretest score in the first step and Group in the second step, followed by WSQ-active-but-odd, WSQ-passive, and the interaction terms of Group*WSQ-active-but-odd and Group*WSQ-passive. No moderation effects were found of SIS (WSQ) on the ToM test performance or the ToMBC parent questionnaire. However, with regard to the scores on the SRS, a significant Group*WSQ-passive interaction effect predicted post-

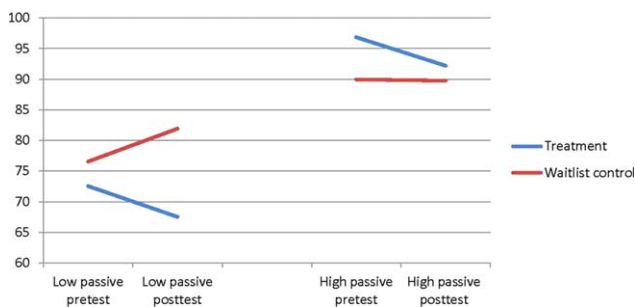


Figure 2. Pretest and post-test scores on the SRS for the Low passive and the High passive participants in the waitlist control and the treatment groups.

test improvements, unstandardized $b = 1.84$ SE = 0.74, $t = 2.47$, $P = 0.01$, $\Delta R^2 = 0.02$, showing that passive interaction styles moderated treatment effects on the SRS. Figure 2 depicts the improvement on the SRS in the waitlist control and the treatment condition for the Low passive and High passive participants separately, using median split. The High passive children showed more autistic features overall (i.e., higher SRS scores), but both Low and High passive children seem to benefit from the treatment, as their SRS scores drop at a similar rate. However, the Low passive children in the treatment condition *improved more* (i.e., scored lower on the SRS at post-test vs. pretest) *relative* to the waitlist control condition, where a slight increase in SRS scores was seen. In the High passive group, the pretest to post-test gains in treatment versus waitlist control group were less pronounced. This can be interpreted as indicating that a Low passive interaction style is associated with a worsening in SRS scores in the absence of an intervention, but with an improvement in SRS when treated. Active-but-odd interaction style was not associated with treatment response.

Disruptive behavior. Similar regression analyses were conducted on the DBD. We entered Group, followed by four DBD subscales (attention, hyperactivity, ODD, and CD) and the interaction effects of these subscales with Group. None of the DBD subscales was associated with different levels of gains on the SRS in the treatment or the waitlist control group.

Discussion

This study examined the effectiveness of a ToM intervention for children with ASD. Besides determining *whether* this intervention was successful, we also explored *for whom* the intervention was most effective. As expected, the treatment had a positive effect on child-based measures, including ToM understanding, although no effects were found on children's ability to

reason about emotions or more advanced ToM problems. Parent reports on the ToMBC indicated a small, positive effect of the treatment and these scores were maintained at the 6 months post-test. Scores on the SRS, which is a more general measure of social responsiveness also improved following intervention contrary to predictions, however, the parental report SSQ, one of our primary outcome measures, did not show a treatment effect and no further changes were noted at the 6 months post-test. This suggests that the intervention has some effects on basic ToM abilities, although more complex aspects of ToM understanding and more general social skills remain unaffected, in line with previous studies [see for a recent overview: Fletcher-Watson et al., 2014].

No moderating effects were found for the presence of DB (attention deficits, hyperactivity, oppositional, and conduct problems). This indicated that DB in children with ASD did not interfere with this treatment. These findings are in contrast to Antshel et al. [2011], who showed clear failure to improve social skills in children with autism and clinical diagnosis of comorbid ADHD. The key difference with our study is that we measured DB on a continuous measure and may have targeted children with milder problems. Indeed, the percentage of children scoring in the clinical range on the parent reported DBD was low: 14% attention deficit, 10% hyperactivity, 4% ODD, and 2% CD. In combination with the close supervision during the sessions, these low rates of problem behavior may have minimized tendencies for DB.

Children's SIS (socially passive vs. socially active) did not affect ToM tasks or parent reports of ToM-related skills. However, a moderating effect of interaction style was found on SRS scores. Children with a low passive interaction style tended to show a worsening in SRS scores if they did not receive the intervention, but if they received the intervention, their SRS scores improved. It could be that Low passive children are more motivated to seek out social interactions, but fail to do this in an appropriate or competent manner. This may cause increased autism scores in the absence of treatment. While this finding needs to be further explored, if the finding were to be replicated this could indicate that we should offer more support to Low passive children because otherwise they are likely to experience an increase in autism symptoms. These suggestions are in line with recent findings on increased treatment effects in younger children who showed higher levels of sociability at baseline [Kaale, Fagerland, Martinsen, & Smith, 2014]. Importantly, the presence of an active-but-odd interaction style was *not* related to a higher treatment effect, suggesting that children may need to attain a threshold level of social engagement to benefit from the treatment, but beyond that point a

more socially active predisposition is no further advantage.

Intervention trials with very young children [Green et al., 2010] suggest that treatment effects are likely to be most evident on proximal measures that assess behaviors that are the specific target of treatment. In this study, treatment effects were found for a general social responsiveness measure (SRS) but not on the SSQ, which also assesses social behavior. Differences between the SRS and the SSQ likely play a role here. The SRS has many more items and covers a broad range of behaviors whereas the SSQ focuses more specifically on social reciprocity and adherence to social rules (e.g., Apologizes when does something wrong; does kind things for others; shares things; follows the rules in games, etc.) which may be more difficult to change. In addition, the social cognition subscale of the SRS seemed to be the domain that improved most following intervention, suggesting that outcome measures most closely linked to the skill taught in the intervention show the highest improvement. It is difficult to conclude from these findings what the most appropriate outcome measure would be for this study although a combination of specific and more general outcomes should always be considered. The main benefits of the ToM treatment are improved social understanding, albeit with little generalization to wider social behavior.

This study has several strengths. It includes an RCT design with analyses of follow-up effects, based on multiple informants. The protocol was specified before the start of the trial (www.trialregister.nl, trial number 2327) and published before the data collection was finished [Hoddenbach et al., 2012]. The large sample allowed detailed analysis of the child characteristics that were related to outcome, and the re-evaluation of the social subtype classification of the late Lorna Wing provides an important domain for future research.

The study is limited by a number of factors. These include the absence of detailed diagnostic instruments such as the Autism Diagnostic Observation Scale (ADOS) [Lord et al., 2000] and the Autism Diagnostic Interview-Revised (ADI-R) [Lord, Rutter, & Lecouteur, 1994], due to limited resources. Although all participants had extensive clinical assessment and the SRS [Constantino & Gruber, 2007] validated the diagnosis in the majority (92%) of participants, the inclusion of ADOS and ADI-R scores might have resulted in a sample with higher autism severity. However, as omitting participants that did not meet SRS cutoff criteria for autism had no effect on the results, we do not expect that a more stringent approach to inclusion, above the clinical diagnosis and SRS score, would make fundamental changes to the current outcomes. The inclusion of more than one primary child outcome measure is also a weakness, but in the absence of existing data on the rel-

ative sensitivity of any single ToM assessment, and as indicated in the trial protocol (www.trialregister.nl, trial number 2327), we decided to use a combination of assessments to tap different aspects of ToM. In addition, no social validity measures were used, and follow-up measures were limited, as only parent information was obtained and the return rate at follow-up (31%) was poor.

In conclusion, as highlighted in a number of recent reviews [Georgiades, Szatmari, & Boyle, 2013; Gwynette, 2013], the heterogeneity of children with ASD requires more focused interventions, tailored to the specific needs of each individual. This exploration of the impact of social style on outcome represents one more step in identifying variables that are important to consider in designing appropriately focused treatments.

Acknowledgments

The study is funded by Fonds Psychische Gezondheid, (project number 2009 6442). We wish to thank all the children, parents, and clinicians that were involved in the study. We also want to thank the graduate students for their assistance in the data collection: Halima Azdad, Tessa Glasbergen, Bianca van Ligten, Suzanne Haddad, Rehanna Mohammad, Anouk Schilder, Cerisa Krist and Heleen van der Velde.

References

- Antshel, K.M., Polacek, C., McMahon, M., Dygert, K., Spenceley, L., Dygert, L., et al. (2011). Comorbid ADHD and anxiety affect social skills group intervention treatment efficacy in children with autism spectrum disorders. *Journal of Developmental and Behavioral Pediatrics*, 32, 439–446.
- APA. (2001). *Diagnostic and statistical manual of mental disorders (DSM-IV)*. Washington: APA.
- APA. (2013). *The Diagnostic and statistical manual of mental disorders (5th ed.) (Vol. 5)*. Washington, DC: APA.
- Bajgar, J., Ciarrochi, J., Lane, R., & Deane, F.P. (2005). Development of the levels of emotional awareness scale for children (LEAS-C). *British Journal of Developmental Psychology*, 23, 569–586.
- Bearss, K., Johnson, C., Handen, B., Smith, T., & Scahill, L. (2013). A pilot study of parent training in young children with autism spectrum disorders and disruptive behavior. *Journal of Autism and Developmental Disorders*, 43, 829–840.
- Baumont, R.B., & Sofronoff, K. (2008). A multi-component social skills intervention for children with Asperger syndrome: The junior detective training program. *Journal of Child Psychology and Psychiatry*, 49, 743–753.
- Begeer, S., Gevers, C., Clifford, P., Verhoeve, M., Kat, K., Hoddenbach, E., et al. (2011). Theory of mind training in children with autism: A randomized controlled trial. *Journal of Autism and Developmental Disorders*, 41, 997–1006.
- Beglinger, L., & Smith, T. (2005). Concurrent validity of social subtype and IQ after early intensive behavioral intervention

- in children with autism: A preliminary investigation. *Journal of Autism and Developmental Disorders*, 35, 295–303.
- Bolte, S., Poustka, F., & Constantino, J.N. (2008). Assessing autistic traits: Cross-cultural validation of the social responsiveness scale (SRS). *Autism Research*, 1, 354–363.
- Constantino, J.N., & Gruber, C.P. (2007). *Social responsiveness scale*. Los Angeles: Western Psychological Services.
- Demurie, E., Roeyers, H., Baeyens, D., & Sonuga-Barke, E.J.S. (2011). Common alterations in sensitivity to type but not amount of reward in ADHD and autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, 52, 1164–1173.
- Dunn, L.M., & Dunn, L.M. (2004). *The peabody picture vocabulary test (PPVT)-III-NL*. Amsterdam: Harcourt Test Publishers.
- Farmer, C., Lecavalier, L., Yu, S., Arnold, L.E., McDougle, C.J., Scahill, L., et al. (2012). Predictors and moderators of parent training efficacy in a sample of children with autism spectrum disorders and serious behavioral problems. *Journal of Autism and Developmental Disorders*, 42, 1037–1044.
- Fisher, N., & Happe, F. (2005). A training study of theory of mind and executive function in children with autistic spectrum disorders. *Journal of Autism and Developmental Disorders*, 35, 757–771.
- Fletcher-Watson, S., McConnell, F., Manola, E., & McConachie, H. (2014). Interventions based on the theory of mind cognitive model for autism spectrum disorder (ASD). *The Cochrane database of systematic reviews*, 3, CD008785–CD008785.
- Georgiades, S., Szatmari, P., & Boyle, M. (2013). Importance of studying heterogeneity in autism. *Neuropsychiatry*, 3, 123–125.
- Geurts, H.M., Verté, S., Oosterlaan, J., Roeyers, H., & Sergeant, J.A. (2004). How specific are executive functioning deficits in attention deficit hyperactivity disorder and autism? *Journal of Child Psychology and Psychiatry*, 45, 836–854.
- Green, J., Charman, T., McConachie, H., Aldred, C., Slonims, V., Howlin, P., et al. (2010). Parent-mediated communication-focused treatment in children with autism (PACT): A randomised controlled trial. *Lancet*, 375, 2152–2160.
- Gwynette, M.F. (2013). Heterogeneity in autism spectrum disorder(s). *Journal of the American Academy of Child and Adolescent Psychiatry*, 52, 1095–1096.
- 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.
- Hodapp, A.F., & Gerken, K.C. (1999). Correlations between scores for peabody picture vocabulary test-III and the Wechsler intelligence scale for children-III. *Psychological Reports*, 84, 1139–1142.
- Hoddenbach, E., Koot, H.M., Clifford, P., Gevers, C., Clauser, C., Boer, F., et al. (2012). Individual differences in the efficacy of a short theory of mind intervention for children with autism spectrum disorder: A randomized controlled trial. *Trials*, 13, 206.
- Jones, W., & Klin, A. (2009). Heterogeneity and homogeneity across the autism spectrum: The role of development. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48, 471–473.
- Kaale, A., Fagerland, M.W., Martinsen, E.W., & Smith, L. (2014). Preschool-based social communication treatment for children with autism: 12-month follow-up of a randomized trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, 53, 188–198.
- Kaat, A.J., & Lecavalier, L. (2013). Disruptive behavior disorders in children and adolescents with autism spectrum disorders: A review of the prevalence, presentation, and treatment. *Research in Autism Spectrum Disorders*, 7, 1579–1594.
- Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, 2, 217–250.
- Kasari, C., & Patterson, S. (2012). Interventions addressing social impairment in autism. *Current Psychiatry Reports*, 14, 713–725.
- Kohls, G., Chevallier, C., Troiani, V., & Schultz, R.T. (2012). Social ‘wanting’ dysfunction in autism: Neurobiological underpinnings and treatment implications. *Journal of Neurodevelopmental Disorders*, 4, 10.
- Kraemer, H.C., Wilson, G.T., Fairburn, C.G., & Agras, W.S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry*, 59, 877–883.
- Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Leventhal, B.L., DiLavore, P.C., et al. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30, 205–223.
- Lord, C., Rutter, M., & Lecouteur, A. (1994). Autism diagnostic interview-revised—A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24, 659–685.
- Muris, P., Steerneman, P., Meesters, C., Merckelbach, H., Horselenberg, R., van den Hogen, T., et al. (1999). The TOM test: A new instrument for assessing theory of mind in normal children and children with pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 29, 67–80.
- O'Brien, S.K. (1996). The validity and reliability of the Wing subgroups questionnaire. *Journal of Autism and Developmental Disorders*, 26, 321–335.
- Paynter, J., & Peterson, C.C. (2013). Further evidence of benefits of thought-bubble training for theory of mind development in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 7, 344–348.
- Pelham, W.E., Gnagy, E.M., Greenslade, K.E., & Milich, R. (1992). Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 31, 210–218.
- Peterson, C.C., Slaughter, V.P., & Paynter, J. (2007). Social maturity and theory of mind in typically developing children and those on the autism spectrum. *Journal of Child Psychology and Psychiatry*, 48, 1243–1250.
- Scheeren, A.M., de Rosnay, M., Koot, H.M., & Begeer, S. (2013). Rethinking theory of mind in high-functioning autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, 54, 628–635.
- Scheeren, A.M., Koot, H.M., & Begeer, S. (2012). Social interaction style of children and adolescents with high-functioning autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 42, 2046–2055.

- Spence, S.H., Donovan, C., & Brechman-Toussaint, M. (1999). Social skills, social outcomes, and cognitive features of childhood social phobia. *Journal of Abnormal Psychology*, 108, 211–221.
- Sprung, M. (2010). Clinically relevant measures of children's theory of mind and knowledge about thinking: Non-standard and advanced measures. *Child and Adolescent Mental Health*, 15, 204–216.
- Steerneman, P., Jackson, S., Pelzer, H., & Muris, P. (1996). Children with social handicaps: An intervention programme using a theory of mind approach. *Clinical Child Psychology and Psychiatry*, 1, 251–263.
- Storebo, O.J., Gluud, C., Winkel, P., & Simonsen, E. (2012). Social-skills and parental training plus standard treatment versus standard treatment for children with ADHD—The randomised SOSTRA trial. *Plos One*, 7, e37280.
- Wellman, H.M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72, 655–684.
- Wing, L., & Gould, J. (1979). Severe impairments of social-interaction and associated abnormalities in children—Epidemiology and classification. *Journal of Autism and Developmental Disorders*, 9, 11–29.