THE DEVELOPMENT OF EYETRACKING AS AN OUTCOME MEASURE FOR SOCIAL INTERVENTIONS IN AUTISM

Rachel K. Greene

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements of the degree of Master of Arts in the Department of Psychology and Neuroscience (Clinical Psychology).

Chapel Hill
2017

Approved by:

Gabriel S. Dichter

David L. Penn

Eric A. Youngstrom

©2017
Rachel K. Greene
ALL RIGHTS RESERVED

ABSTRACT

Rachel K. Greene: The Development of Eyetracking as an Outcome Measure for Social Interventions in Autism
(Under the direction of Gabriel S. Dichter)

This project aimed to evaluate a dynamic eyetracking task as a measure of treatment response for autism spectrum disorder (ASD) social skills interventions. Adolescent and young adult participants with ASD completed the eyetracking task, as well as questionnaires and neurocognitive measures of social functioning, before, immediately after, and two months after completing an empirically-validated ASD social skills treatment (SCIT-A). The study compared SCIT-A participants (n = 17) to participants with ASD who received treatment as usual (TAU; n = 22). Reliability of the eyetracking measure was assessed in typically developing controls (n = 22), and results indicated good test-retest reliability (α = 0.86, ICC=0.801). Correlation analyses found no significant relationships between the eyetracking task and measures of social functioning in all individuals with ASD at baseline. Although SCIT-A participants showed a significant increase in visual social attention from baseline to post-treatment, this trajectory was seen the TAU group as well. Findings indicate that the eyetracking task is measuring a unique construct not measured by the comparative social functioning assessments.

ACKNOWLEDGEMENTS

Assistance for this study was provided by the Research Participant Registry Core of the UNC Carolina Institute for Developmental Disabilities [P30 HD03110]. This research was supported by the North Carolina Translational & Clinical Sciences Institute (UL1TR001111).

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASD	Autism spectrum disorder
ToM	Theory of mind
SCIT-A	Social Cognition and Interaction Training for Adolescents with High-functioning Autism
TAU	Treatment as usual
TDC	Typically developing control
AOI	Area of interest
TFD	Total fixation duration
SRS	Social Responsiveness Scale
ER-40	The Penn Emotion Recognition Task
MEDF-36	The Penn Emotion Differentiation Task
SNI	Social Network Index
LSAS	Leibowitz Social Anxiety Scale

THE DEVELOPMENT OF EYETRACKING AS AN OUTCOME MEASURE FOR SOCIAL INTERVENTIONS IN AUTISM

Social communication impairments are a hallmark feature of autism spectrum disorders (ASD) and have been since the earliest descriptions of the disorder (APA, 2013; Kanner, 1943). Leo Kanner (1943) first noted that “the outstanding, ‘pathognomonic,’ fundamental disorder is the children’s inability to relate themselves in the ordinary way to people and situations from the beginning of life” (p. 242). Still today, ASD is largely characterized by deficits in social communication skills, including marked impairments in social cognition, social perception, and social communication (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012; Howlin, Moss, Savage, & Rutter, 2013). These deficits can be seen in the ability to recognize and identify emotions, discriminate between socially relevant and irrelevant stimuli, and adequately identify another person’s thoughts or mental state, otherwise known as theory of mind (ToM; Baron-Cohen, Leslie, & Frith, 1985; Bauminger, 2002; Buitelaar, Van der Wees, Swaab–Barneveld, & Van der Gaag, 1999).

A number of behavioral and pharmacological interventions have demonstrated success in treating core and associated symptoms of ASD (Cappadocia & Weiss, 2011; Dove et al., 2012); however, unlike other psychiatric disorders, where self-report questionnaires are routinely used to evaluate treatment effectiveness, many individuals with ASD lack insight into their socio-emotional states, rendering self-report less than ideal for this population (Hill, Berthoz, & Frith, 2004). The three most common domains of ASD clinical trial outcome measures include: caregiver-report instruments that have limited sensitivity to change (e.g. Autism Behavior

Checklist (ABC)); self-report inventories that can be especially challenging for younger or lower-functioning individuals (e.g. Pediatric Quality of Life Inventory (PedsQL)); and global provider assessments (e.g., the Clinical Global Impressions Scale), which are only coarse measures of treatment efficacy (Busner, Targum, & Miller, 2009; Payakachat, Tilford, Kovacs, & Kuhlthau, 2012). Currently, the field lacks valid measures of treatment efficacy, which hampers the development of effective treatments for ASD-associated social impairments.

Behavioral and Pharmacological Interventions for ASD

At present, both behavioral and pharmacological interventions are commonly used to treat ASD. It is estimated that approximately 64% of individuals with ASD between the ages of 12 and 17 are taking at least one psychotropic medication (Coury et al., 2012), and 35% of individuals with ASD under the age of 20 are taking medications of two or more drug classes concurrently (Spencer et al., 2013). However, none of the FDA-approved medications currently used to treat ASD were approved to address impairments in social communication or restricted and repetitive behaviors, the core deficits of ASD (Dove et al., 2012). Instead, medications commonly prescribed to individuals with ASD target associated behaviors such as irritability, hyperactivity, and aggression. Additionally, a variety of behavioral treatments (e.g. social skills training) target core social impairments in adolescents and young adults (Kaat & Lecavalier, 2014; Reichow & Volkmar, 2010). Although behavioral treatments show promise, the field still lacks an objective and valid outcome measure to evaluate the efficacy of such treatments. A reliable measure of treatment response is needed to assess and develop effective pharmacological and behavioral interventions for core ASD impairments.

Outcome Measures for ASD Interventions

Although numerous treatment studies have evaluated the efficacy of interventions for ASD, current outcome measures often rely on self- and caregiver-reports, which may be biased or lack sensitivity to change. Many currently used outcome measures were originally designed as diagnostic tools (e.g. ADI-R, ADOS-2, CARS2) and were not meant to serve as measures of subtle behavioral change over time (Kanne et al., 2014; Payakachat et al., 2012). Similarly, continuous measures of ASD symptoms that are commonly used as treatment outcome measures have not been shown to be sensitive to treatment-induced change in symptom severity (e.g. Autism Behavior Checklist; Aman et al., 2004).

Additionally, each informant introduces distinct methodological challenges. Theory of mind deficits and difficulty with insight into socioemotional states make self-report measures problematic for individuals with ASD (Payakachat et al., 2012). Furthermore, younger and lower-functioning individuals with ASD are unable to answer questions about their mental states. Often, caregiver- or parent-report measures (examples) are completed as an alternative to self-report measures. This reporting method, however, lacks sensitivity to change within short periods of time and is limited in its capacity to evaluate symptoms that are not visible to the parent or caregiver, such as subtle changes in social perception or theory of mind (Payakachat et al., 2012). Additionally, clinician reports and observations of treatment response are more likely to indicate stronger placebo effects, and, alternatively, greater response to active treatment conditions, compared to caregiver report (Masi, Lampit, Glozier, Hickie, & Guastella, 2015).

Finally, global provider assessments (e.g. Clinical Global Impressions; Busner et al., 2009) are also routinely used to measure treatment outcomes in ASD, yet they are primarily focused on overall levels of functioning rather than indicating specific impairments. Sensitive and unbiased measures of treatment response would allow for more accurate evaluation of novel ASD behavioral and pharmacological treatments

Eyetracking in ASD

Eyetracking has the potential to function as valid outcome measures for use with individuals with ASD of varying ages and levels of impairment. Eyetracking paradigms have been used for decades with typically developing adults to study gaze behavior (i.e.,where individuals are looking to gain more information about a scene or environment (Boraston & Blakemore, 2007). Eyetracking has also been used with individuals with ASD as a measure of visual social attention and, more specifically, attentional biases to orient towards social stimuli (Dawson, Meltzoff, Osterling, Rinaldi, & Brown, 1998). In 2002, Klin and colleagues showed that individuals with ASD focused their gaze significantly less on the eye regions of a face compared to typically developing controls, and, instead, showed greater visual preference for the mouth region. This seminal paper led to a multitude of eyetracking studies showing abnormalities in social gaze behaviors in individuals with ASD (Chawarska, Macari, & Shic, 2012; Jones, Carr, & Klin, 2008). For example, Chawarska and colleagues (2013) found that 6-month-old infants who were later diagnosed with ASD showed diminished social monitoring toward the individual in the stimulus, with particular visual disinterest in her face, compared to developmentally delayed and typically developing infants who did not develop ASD. Similarly, Shic and colleagues (2011) reported that, while viewing videos of an adult-child play interaction, 20 month-old toddlers with ASD showed reduced attention to the heads and activities of others and focused more on background objects such as toys.

Taken together, such findings have contributed to the emergence of theoretical frameworks seeking to explain the development of social impairments in ASD. Specifically, it has been suggested that decreased visual preference for social stimuli in early life could lead to the social communication deficits inherent to ASD (Dawson et al., 1998; Schultz, 2005).

The Current Study

Rather than evaluating the efficacy of a treatment itself, the present study evaluated the utility of an eyetracking task as a novel treatment outcome measure for ASD social interventions by using an intervention as a mechanistic probe. Specifically, we examined changes in social visual attention before and after an 8-week group-based empirically validated psychosocial intervention, Social Cognition and Interaction Training for Adolescents with High Functioning Autism (SCIT-A). Changes in social attention due to treatment were compared to changes in a group that receives treatment as usual (TAU). This group did not take part in the social skills group treatment, but some continued to receive clinical services outside of their participation in the current study. Additionally, changes in eyetracking will be compared to changes in neurocognitive and report measures of social impairment in

Previous analysis of the SCIT-A program has shown that individuals with ASD who participated in SCIT-A showed significant improvement in their social cognitive abilities, particularly theory-of-mind skills (Turner-Brown, Perry, Dichter, Bodfish, & Penn, 2008). The SCIT-A program was designed to be presented in two phases: 1) introducing the notions of interest and disinterest in a social partner and how that might affect the trajectory of a social interaction; 2) teaching participants to focus on socially relevant cues within the environment and learning to interpret and plan based on those stimuli (Turner-Brown et al., 2008).

In summary, the current study evaluated the accuracy of a dynamic eyetracking task in detecting change due to the SCIT-A intervention with adolescents and young adults with ASD in a pre-test, post-test, follow-up design, with the third visit examining maintenance of treatment gains. This research design also included an age- and gender-matched TAU comparison group. The SCIT-A and TAU groups were compared to evaluate whether the eyetracking task was capable of detecting change in social skills over time. Additionally, a typically developing control (TDC) control group was included to examine the test-retest reliability of the eyetracking measure. These hypotheses were tested via the following specific aims:

Aim 1: To evaluate test-retest reliability of the eyetracking task in typically developing young adults.

Hypothesis: Eyetracking metrics will show acceptable to good to excellent test-retest reliability between two time points within the typically developing control (TDC) sample.

Aim 2: To evaluate the correspondence between baseline eyetracking and measures of social functioning in participants with ASD.

Hypothesis: Baseline eyetracking metrics of social attention will be correlated with questionnaires and neurocognitive measures of social functioning, including measures of emotion recognition and theory-of-mind.

Aim 3: To examine the use of the eyetracking paradigm as a measure of treatment-related change in social functioning compared to change in a TAU comparison group.

Hypothesis: Changes in measures of social cognition and social functioning will predict the magnitude of change in eyetracking metrics to a greater extent in the treatment group than in the TAU group.

METHODS

The non-biomedical institutional review board at UNC Chapel Hill approved this study.

Participants

Individuals enrolled in the study made up three age-matched cohorts, treatment (SCIT-A), treatment as usual (TAU), and typically developing controls (TDC). We recruited 17 SCIT-A participants (age *M* = 16.23, *SD* = 2.3) and 22 TAU participants (age *M* = 17.77, *SD* = 4.0), all of whom met diagnostic criteria for ASD, confirmed by the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2), a gold-standard Autism diagnostic tool (Lord et al., 2012). Individuals were not excluded based on the presence of other psychiatric or medical comorbidities to increase feasibility and the generalizability of findings. The presence of such diagnoses was, however, examined by collecting a brief medical history and a short questionnaire assessing psychiatric symptomatology. Of the ASD participants, 64% were receiving some form of ongoing therapeutic service (not including those participating in SCIT-A), and 59% were taking at least one psychiatric medication over the course of the three study visits. Additionally, it should be noted that there were significant differences in IQ between the SCIT-A and TAU groups. This difference may be representative of the differences in group recruitment described below. Statistical analyses examine IQ as a covariate to evaluate the impact of this cognitive difference. The TDC cohort consisted of 22 typically developing young adults (age *M* = 19.14, *SD* = 1.1). All participants met the IQ cut off with Full Scale IQs >70. The three groups showed varied differences in IQ, age, and gender (see Table 1). SCIT-A and TAU participants were recruited from the Carolina Institute for Developmental Disabilities

(CIDD) Autism Subject Registry (*n*=5,914 families), and were supplemented by IRB-approved informational flyers and online postings within the community that targeted individuals with ASD interested in participating in a social skills group. The program was offered three times per year through the Carolina Institute for Developmental Disabilities (CIDD), and individuals who expressed interest in this clinical service were also made aware of the accompanying research study. Participants were invited to participate in the social skills group regardless of research participation. To increase feasibility, participants were not randomly assigned to treatment or TAU group membership. Individuals who were willing and able to participate in the 8-week social skills group were invited to do so. This non-random assignment was necessary for adequate study enrollment and ethical concerns of denying treatment. TDC participants were recruited from the undergraduate psychology research pool at UNC Chapel Hill.

Procedure

TDC Procedures. TDC participants attended two study visits, which were separated by approximately 24 hours. During the first study visit, participants completed a test of cognitive ability, the eyetracking task, clinical report measures and neurocognitive measures of emotion regulation and theory of mind. At the second study visit, a day later, the participants were re-administered the eyetracking task. No additional study measures were completed at the second visit. TDC participants received course credit for participating in the research study.

SCIT-A and TAU Procedures. The study was conducted over three separate testing visits. Participants in the treatment group attended eight social skills group sessions between the first and second testing sessions. The first testing visit lasted approximately 3 hours and the second and third visits lasted approximately 1.5 hours each. During each testing visit, participants completed the eyetracking task, clinical report measures, and neurocognitive testing.

In addition to these measures, the first visit included the consent process, diagnostic testing for ASD, and cognitive testing. All testing was administered by a trained, research-reliable graduate research assistant under the supervision of the Principal Investigator (PI), Dr. Gabriel Dichter. All participants received monetary compensation in denominations of $30 for the first visit and $20 for each additional visit.

SCIT-A Group. The SCIT-A program was originally adapted from a social skills group intervention for individuals with psychosis (Roberts, Penn, & Combs, 2015), due to overlap in social cognitive impairments in both schizophrenia and ASD. A small pilot study with adults with ASD indicated that the treatment was feasible, that participants found the intervention helpful, and that they showed improvements in theory of mind skills compared to controls (Turner-Brown et al., 2008). SCIT-A was specifically targeted to focus on social cognitive processes and strategies within a structured teaching environment. The group was conducted over eight consecutive weekly sessions, and typically consisted of six to ten individuals with high-functioning ASD. This intervention was being leveraged for the current study to evaluate eyetracking as a treatment outcome measure, rather than to validate the intervention itself. Research participants were required to attend at least five out of the eight sessions. Weekly attendance and homework completed were recorded for study purposes.

Materials and Measures

Eyetracking Task. The secondary aim of the study involved examining the correspondence between this dynamic eyetracking task and other measures of social cognition and functioning. The eyetracking task, the Interactive Visual Exploration (IVE) task, has previously proved successful in differentiating between ASD and control samples and has been validated in predicting the magnitude of ASD symptoms, compared to eyetracking paradigms consisting of static images or videos without ecologically-valid social context (Chevallier et al., 2015).

This paradigm presented 22 silent video clips of 11 sibling pairs each participating in a social (joint condition) and non-social (parallel condition) play activity. The video clips were filmed in rooms where background objects (e.g., light switches, toys, posters) are clearly visible, thus adding to the ecological validity of the paradigm. This natural setting was intentional, so that the environment within the video seems less artificial and more representative of every-day life outside of the laboratory setting. The actors were seen participating in one of two conditions: social interaction or parallel play. In the condition portraying a social interaction between two children, both actors were seen engaging in a game together in a natural manner. During the parallel play condition, however, the sibling-pairs did not engage one another. For example, the children participating in the social activity may have been playing a card game together, making facial expressions and gestures toward each other, whereas those depicted in the non-social or parallel play condition would have individually participated in their own task (e.g., drawing, "barrel of monkeys" game, etc.), without interacting with their sibling. Actors were school-aged children of both genders, who were instructed not to make direct eye contact with the video camera. All video clips were integrated into one single paradigm lasting just under 7 minutes.

Each of these video clips contained pre-determined areas of interest (AOIs), which were traced by hand and change over time with the progression of the dynamic stimulus. These were mapped out to capture faces, background objects and hands as they moved throughout the paradigm. Data was collected and reported in a measure of total fixation duration (TFD), or the total amount of time the viewer directed their eye gaze at that specific AOI within the video clip.

TFD was aggregated for each AOI across all video clips to create a variable summing all fixations made to faces, background objects, and hands.

Eyetracking System Specifications and Settings. Participants from all three study groups completed the eyetracking task. The IVE eyetracking task was displayed on a TobiiTX300 eyetracker integrated with a 23” display monitor located at the CIDD. This Tobii system is highly accurate and precise, and has the ability to compensate for relatively large motion of the head, which is extremely valuable in working with an ASD population.

Before beginning the task, participants’ eyes were positioned 60 cm from the monitor, and their eye gaze was calibrated. This calibration procedure showed a red dot moving to nine different locations on a grey screen. Participants were asked to follow this dot with their eyes while remaining as still as possible. Calibration was readministered until all nine target locations were accounted for accurately. Once all locations were precisely calibrated, participants were asked to remain still and silent while they watched the video on the screen.

The eyetracker acquired gaze position at 300 Hz with the following parameters: linear interpolation was enabled with a max gap length of 75 ms; an average of both eyes was taken to determine gaze position; noise reduction was disabled; and the velocity calculator was set to 20 ms. Adjacent fixations were merged when the time between those two fixations was 75 ms or less and when the maximum angle between these fixations did not exceed 0.5°. Finally, fixations under 60 ms were discarded. Participant data with less than 50% accuracy was excluded from the subsequent analyses.

Autism Diagnostic Assessment. SCIT-A and TAU participants were administered either module 3 or module 4 of the ADOS-2 to assess for autism diagnostic criteria. Module administration was determined by age and verbal ability. This measure was administered by a research-reliable graduate level research assistant, who was supervised by Dr. Gabriel Dichter, a licensed psychologist. This portion of testing lasted approximately 45 minutes, throughout which participants were asked to complete activities such as telling a story and giving an account of a routine daily activity. Additionally, they were asked questions regarding their perceived role in social situations and understanding of personal responsibilities. It was determined whether or not participants met ASD criteria based on the ADOS-2 algorithm cut off scores.

Cognitive Assessments. In order to assess general cognitive functioning and match study groups, participants were administered one of two cognitive tests. TDC participants completed the National Adult Reading Test – Revised (NART-R; Crawford, Stewart, Cochrane, Parker, & Besson, 1989), which consists of 61 English words. Participants were asked to read each word and were scored based on correct pronunciation. The NART-R provided a predicted WAIS Full Scale IQ. SCIT-A and TAU participants completed the 2-subtest version of the Weschler Abbreviated Scale of Intelligence (WASI-II; Axelrod, 2002). The two subtests administered were the vocabulary and matrix reasoning sections, representing verbal and spatial intelligence respectively. Administration of the two-subtest version took only about 20 minutes, and performance on these two subtests were aggregated to provide a Full Scale IQ (FSIQ).

Symptom Report Measures.

Social Functioning. Participants and caregivers (when applicable) of all study groups each completed the Social Responsiveness Scale (SRS; Constantino & Gruber, 2002) as an index of change in ASD symptoms before and after treatment. This 65-item measure served to assess the severity of social-communicative autism symptoms as they occur in natural settings. Participants answered each question using a given 4-point Likert scale, which ranged in severity. Questions included content regarding intense interests or preoccupations and perceptions of social ability. The SRS is able to reliably distinguish individuals with ASD from individuals with other psychiatric diagnoses (Constantino et al., 2003; Constantino & Todd, 2014). The SRS also provides subscales, which give detailed insight into an individual's social cognition, social motivation, social awareness, social communication, and autism symptomatology.

Because detecting progress in social functioning was of primary interest in this study, we examined this construct using two distinct measures. The Liebowitz Social Anxiety Scale (LSAS-SR; Heimberg et al., 1999) is a 24-item self-report measure commonly used in with ASD populations to assess social anxiety and takes approximately 5 minutes for participants to complete. This scale allowed for the discrimination between how fearful a participant finds a particular situation versus how often they avoid that same situation.

The Social Network Index (SNI; Cohen, Doyle, Skoner, Rabin, & Gwaltney, 1997) is a 15-item self-report measure used to assess 12 types of social relationships and the frequency of contact within those relationships in the last two weeks. This questionnaire helped to determine the type and size of social network held by each participant, and took approximately 10 minutes to complete. The SNI was completed by the participant at each testing visit.

Neurocognitive Assessments.

Emotion Recognition and Differentiation. The Penn Emotion Recognition Task (ER-40; Gur et al., 2002) is a standardized test of facial emotion recognition ability consisting of 40 color photographs of evoked expressions from adult actors displaying four basic emotions (i.e., happy, sad, angry, fearful) and neutral facial expressions. All study participants were asked to identify the emotion of each facial expression with one of the five choices mentioned previously.

The Measured Emotion Differentiation Task (MEDF-36) assessed the participant's ability to determine the intensity of an emotion through facial expressions. The 36-trial stimulus presents pairs of faces, each expressing the same emotion, one more intense than the other or of equal intensity. Gradations of intensity were obtained by morphing a neutral to an emotionally intense expressions and the difference between pairs of stimuli ranged between 10 – 60% of mixture. Participants were asked to select the face showing the more intense emotion. If both faces appeared to be of equal intensity, the participant was instructed to select a button labeled, "Equal."

Theory of Mind (ToM). Empirical evidence suggests that individuals with ASD struggle with theory of mind (ToM) tasks, which involve the ability to understand others' mental states (Chevallier, Noveck, Happé, & Wilson, 2011). The Hinting Task tests this ability by presenting ten short vignettes involving social interactions between two characters, one of whom drops a hint for the fictional partner about their desire or intention. Participants were asked to identify the implicit intention of the hint. This task was administered at each testing visit.

RESULTS

Data Preparation

Dr. Christopher Wiesen, statistical analyst at the UNC Odum Institute, consulted with respect to data analysis for this project. Three areas of interest (AOIs) were obtained from the eyetracking data: faces, hands, and background. The total fixation duration (TFD; in milliseconds) on each AOI was calculated by the Tobii Pro Studio software. The proportion of TFD was calculated by dividing the fixation time participants devoted to each AOI group by their TFD on the entire screen, thus standardizing the metrics across each individual. For example:

$$Proportion\ of\ Total\ Fixation\ Duration\ Face = \frac{Total\ Fixation\ Duration\ to\ Face}{Total\ Fixation\ Duration\ to\ Entire\ Screen}$$

Finally, based on these AOI TFD proportions, we computed a "social prioritization score" by subtracting the proportion of fixation time devoted to social AOIs (e.g. faces) minus the proportion of fixation time devoted to object AOIs (e.g. background), as seen below:

$$Social\ Prioritization\ Score = (Proportion\ TFD\ Face) - (Proportion\ TFD\ Background)$$

This measure indicates preference for social stimuli across paradigm conditions (e.g. social and non-social conditions). See Chevallier et al. (2015) for similar methods. Only the social prioritization score was used as dependent variables in the following analyses.

Predictor variables included the following: treatment group (SCIT-A or TAU), measures of social cognition (i.e., the Hinting Task, ER-40, MEDF-36, SRS social cognition score), and measures of broader social functioning (i.e., SRS total score, Social Network Index, LSAS). Covariates included age, IQ, current therapy status and current psychotropic medication use. All analyses were conducted using SAS software, Version 9.4 (SAS Institute, 1985) or SPSS software, Version 23 (IBM Corp., 2015).

Test-retest reliability in TDC group

Three methods were used to assess the test-retest reliability of two eyetracking dependent measures (i.e. proportion of total fixation duration to faces and social prioritization total) across the two TDC time points.

Intraclass Correlations. First, an absolute change intraclass correlation coefficient (ICC) was calculated to examine the reliability of the eyetracking metric across timepoints one and two within the TDC sample only. The two-way mixed-effects ICC models were conducted based on a mean-rating (k = 2) and indicated good reliability for the social prioritization score (ICC = 0.801, 95% CI [0.348, 0.927]).

Cronbach's Alpha. Next, Cronbach's alpha was calculated to further examine the consistency across the two eyetracking time points. Based on these analyses, the social prioritization score was found to be highly reliable (α = 0.86).

Box Plots. Test-retest reliability was also examined visually by constructing box plots for eyetracking metrics across each timepoint (see Figure 2). To do this, we calculated mean eyetracking social prioritization scores across all TDC subjects at both time points and subtracted that grand mean from the average eyetracking TDC data at each time point. A score of zero on a given time point indicates that the average result from that time point is equal to the mean score obtained from both time points across individuals, and that there is no change between time points. The grand mean-centered social prioritization scores were visually similar at time point 1 (M = -1.25, SD = 4.45) and time point 2 (M = 1.37, SD = 5.23).

Correlational Analyses

A Pearson correlation analysis was employed to examine the dimensionality of the eyetracking task and its correspondence with measures of social cognition (i.e., the Hinting Task, ER40, MEDF36), and measures of social functioning (i.e., SRS, SRS Social Cognition Subscale, Social Network Index, LSAS) at baseline.

Correlations were used to examine the convergent validity between the eyetracking paradigm and measures of social impairment. Contrary to hypotheses, there were no significant relationships between the social prioritization eyetracking measure and the other measures of social functioning and social cognition at baseline (see Table 2).

Multilevel Model Analyses

Multilevel modeling (MLM) techniques were employed to assess the eyetracking task as a reliable measure of change (Bryk & Raudenbush, 1987). MLM techniques were chosen over other analyses of repeated measures in order to account for the dependence of each data point collected for the same individual. Other analytic models (e.g., repeated measures ANOVA) assume independence of individual data points from one another, and thus MLM analyses were used because of their ability to account for commonalities between longitudinal data points of the same subject. Additionally, MLM models were preferred because they are able to include participants with missing data points.

First, a preliminary random effects model was used to calculate the intraclass correlation coefficient (ICC) for eyetracking data across and within therapy cohorts (for SCIT-A participants only) to determine if there was within-group dependence. This model estimated that a small amount of variance could be attributed to therapy cohort membership. The ICC estimate for therapy cohort was 0.159, indicating that approximately 16% of the variability in individual prioritization of social information versus non-social information can be attributed to differences between therapy cohorts. Because the ICC indicated that little cohort-level clustering was present, multilevel models were fit as 2-level rather than 3-level partially-nested structures. We also compared a linear model structure with a quadratic function, and found that the linear model provided a better model fit. A homoscedastic error structure was assumed for the following analyses.

Next, we examined the trajectories of eyetracking social prioritization scores across study visits while controlling for the following time-invariant covariates: FSIQ, age, treatment group (i.e. TAU or SCIT-A), participation in outside therapies, and current administration of psychotropic medication. The time variable was defined such that the baseline visit was coded as the intercept (i.e. the reference). Additionally, all visits (i.e. visits 0, 1, and 2) were represented nominally as individualized variables to account for the unique distribution between time points (e.g. visit 0 vs. visit 1, visit 0 vs. visit 2, visit 1 vs. visit 2). This removed the assumption that there would be commensurate change in scores between each visit, and allowed us to examine distinct differences in the eyetracking metric between varying time points.

$$SocialPrioritization_{it} = \gamma_{00} + \gamma_{10}Visit_{it} + \gamma_{20}Group_{it} + \gamma_{30}FSIQ_{it} + \gamma_{40}Age_{it} + \gamma_{50}Therapy_{it} + \gamma_{60}Meds_{it} + r_{it}$$

Results from this model showed that individuals with ASD (regardless of treatment group) taking psychotropic medications attended to social over non-social stimuli significantly more frequently than unmedicated participants ($\beta=2.98$, $t=2.66$, $p<0.01$). Additionally, both nominal variables representing visit 1 (post-treatment visit) and visit 2 (follow up visit) were significant predictors of social prioritization, which indicates that the eyetracking metrics were significantly different at visit 1 ($\beta=2.49$, $t=3.27$, $p<0.001$) and visit 2 ($\beta=2.61$, $t=3.27$, $p<0.001$) compared to the baseline visit. Specifically, there was a 2.49 increase in social prioritization between baseline and visit 1, and a 2.61 increase from baseline to the follow up at visit 2. A contrast between visit 1 and 2, however, showed no significant change (β=-0.12, *t*=-0.15, p<0.88). This clarifies that the change in social prioritization over time was largely driven by the significant increase in visual attention to social stimuli from scores at baseline and visit 1. Additional analyses were completed to examine the apparent group interaction between visit 1 and visit 2 (see Figure 3); however, the findings showed this interaction to be non-significant (β=-1.37, *t*=-0.84, p<0.40). Full scale IQ, age at baseline, treatment group, and current participation in therapy did not significantly affect social prioritization scores across study visits (see Table 3).

Finally, we assessed the correspondence between eyetracking measures and clinical report and neurocognitive measures of social cognition and social functioning over the three timepoints, again using an MLM framework. Predictors of the eyetracking social prioritization score included: ER-40, MEDF-36, SRS Total, SRS Social Cognition Subscale, Hinting Task, LSAS, and the SNI. Similarly, the model presented below assumes no within-group dependence within therapy cohorts, and, therefore, represents a 2-level structure rather than a partially-nested 3-level model. The model also maintains the same coding of the timing variable as described above. Only one social cognition or social functioning predictor was included as a predictor in each model. For example:

$$SocialPrioritization_{it} = \gamma_{00} + \gamma_{10}Visit_{it} + \gamma_{20}Group_{it} + \gamma_{30}\boldsymbol{SRS}_{it} + \gamma_{40}Visit_{it} \text{ x } \boldsymbol{SRS}_{it} + \gamma_{50}Group_{it} \text{x } Visit_{it} + \gamma_{60}Group_{it} \text{ x } \boldsymbol{SRS}_{it} + \gamma_{60}Group_{it} \text{ x } \boldsymbol{SRS}_{it} \text{ x } Visit_{it} + r_{it}$$

These analyses showed significant two-way interactions between visit 2 and ER-40 performance (β=1.22, *t*=2.16, p<0.035) and visit 2 and treatment group (β=54.38, *t*=2.42, p<0.019). These outcomes suggest that emotion recognition ability and treatment group, respectively, serve as moderators of the relationship between visit 2 and social prioritization (see Figure 4).

Additionally, findings showed a significant three-way interaction between visit 2, treatment group, and ER-40 performance (β=-1.71, t=-2.48, $p<0.017$), reflecting the combined contribution of emotion recognition ability and treatment group on social prioritization at visit 2. Models including the remaining measures (i.e., SRS Social Cognition Subscale, ER-40, MEDF-36, Hinting Task, Leibowitz Social Anxiety Scale, and the Social Network Index) did not result in significant findings.

DISCUSSION

The current results explore the potential for the eyetracking task to serve as a measure of treatment response in ASD populations. Analyses revealed that the social prioritization score obtained from the eyetracking task showed good test-retest reliability in a TDC sample. This indicates that fluctuations in these scores over time may be attributed to changes in visual attention to social stimuli rather than to poor reliability of the measure.

Although we expected our eyetracking measures to align with other measures of social cognitive functioning, the social prioritization score used in the analyses by Chevallier et al. (2015) did not significantly relate with other measures of social functioning at baseline in this sample. Although these findings countered the original hypothesis, they do suggest that the constructs measured by the eyetracking metrics are distinct from the clinical report measures and neurocognitive assessments, which evaluate emotion recognition, emotion differentiation, theory of mind, social anxiety, social cognition, and broader social communication. This finding does not necessarily indicate that the eyetracking measure is failing to detect the social abilities and social preferences of an individual, but, rather, that the eyetracking task may be measuring a unique component of an individual's social functioning ability, untapped by other commonly used questionnaires and neurocognitive measures.

Furthermore, it should be noted that few significant correlations were found between all measures at baseline. The few significant findings included the correlations between ER-40 and the Hinting Task, and between the Liebowitz Social Anxiety Scale (LSAS) and the Social Network Index (SNI). This suggests that emotion recognition abilities were closely associated with theory of mind abilities and that social anxiety were inversely related to the number of people in one's social network, respectively. Additionally, the SRS Social Cognition Subscale was, as expected, positively correlated with the SRS total score; however, contrary to what was hypothesized, the SRS Social Cognition Subscale was negatively correlated with the Hinting Task, a measure of theory of mind. Given that theory of mind is a social cognitive skill, this was somewhat surprising. This distinction could highlight differences in task presentation (e.g. questionnaire vs. neurocognitive assessment), or it may be that theory of mind is not well represented in the social cognition subscale of the SRS. All other measure combinations were not significantly related.

While we did not find any significant correlations between social cognition measures and social prioritization, some of the covariates we examined longitudinally were significant. For example, individuals taking psychotropic medication were significantly more likely to prioritize social stimuli in the eyetracking paradigm than those not taking psychotropic medications. This held true for all participants with ASD regardless of treatment group, and it suggests that the eyetracking metric may be picking up on a specific response associated with greater salience to social stimuli produced by the administration of medication. Given that the medication variable was recorded as a binary measure (i.e. psychotropic medication or no psychotropic medication), further research should be done to examine the unique effects of various drug types and dosages. Drug types represented in this population included antipsychotics, stimulants, antidepressants, and antianxiety pharmaceuticals. As previously stated, the only FDA approved medications used to treat ASD address peripheral symptomatology (e.g. irritability, inattention, aggression) and do not address core ASD symptoms (i.e. social communication deficits, restricted and repetitive behaviors; Dove et al., 2012). Because psychotropic medication served as a significant predictor for the prioritization of visual social stimuli in this sample, it would be interesting to learn how specific medications may, in fact, have an impact on social visual preference or salience to social stimuli. Additional studies should be done to examine the utility of this eyetracking task as an outcome measure specific to medication trials in ASD populations, given it's apparent sensitivity to pharmacological treatments.

Our longitudinal results did not indicate any significant differences in gaze preference for social stimuli between the two groups (i.e. SCIT-A and TAU). It is possible that the TAU group was too conservative a comparison to detect between-group effects with this measure due to the availability of specialized ASD treatments in this community. This may be reflected in the trajectory of significant improvement in visual social attention from baseline to post-treatment seen in both groups. In the Research Triangle area of North Carolina, where the study took place and recruited participants, there is a high standard of care for individuals with ASD. While the CDC reports 1 out of every 68 children in the U.S. are diagnosed with ASD, the North Carolina state prevalence is even higher at 1 in 58 (Baio, 2012; Kalkbrenner et al., 2011). This increased prevalence may be attributed to families who move to the area for the high quality of care and ASD resources. The state itself has seven regional TEACCH centers, including the flagship TEACCH center in Chapel Hill, which provides individual and group intervention to children, adolescents, and adults with ASD. The Research Triangle is also home to the UNC CIDD in Carrboro, the Duke Center for Autism and Brain Development in Durham, and numerous private practices serving individuals with ASD. Within this sample alone, 64 percent of participants were receiving some form of ongoing therapeutic service (not including those participating in SCIT-A), and 59 percent were taking at least one psychiatric medication over the course of the three study visits.

Beyond the effect of high-quality ASD interventions in the area, the lack of treatment response above and beyond that seen in the TAU group, as measured by the social prioritization eyetracking measure, may also be attributed to the logistics of the SCIT-A group itself, which provides one 60-minute social skills session a week for eight weeks. Other more intensive ASD social skills groups such as PEERS consist of 14 weekly sessions that each last 90-minutes (Laugeson, Frankel, Gantman, Dillon, & Mogil, 2012). Additionally, these groups require that parents participate in a concurrent parent training so that they may better serve as a social skills "coach" for their child. Because SCIT-A was facilitated at a community mental health clinic, a parent training session was not feasible. Lastly, PEERS necessitates that all participants express a personal motivation and a genuine interest in bettering their social skills to enroll and attend the intervention sessions. While SCIT-A encouraged participants to join of their own free accord, there have been participants whose parents insisted they attend even when the child expressed disinterest. Despite these differences, the SCIT-A group consistently receives positive feedback from participants and their families (Turner-Brown et al., 2008), and those who participated in SCIT-A within this study were no exception. On a brief post-treatment survey participants were asked to rate the value of the group on a scale from 1 (*Not Helpful*) to 3 (*Very Helpful*). The questions included: 1) How useful was the group to you?; 2) How much did the group help you in thinking about social situations?; and 3) How much did the group help you in the way you relate to other people? For each of the three questions, average responses indicated that the group was helpful to very helpful.

Not only did the SCIT-A group receive positive feedback from participants, group differences across visits were observed in the SRS total scores (β=16.89, t=4.34, $p<0.0001$) and the SRS social cognition scales (β=2.75, t=2.89, $p<0.005$), such that individuals who participated in SCIT-A showed significant reductions in social communicative and social cognitive impairment just following the 8-week social skills group and then again at 2-month follow up visit (see Figure 5). This may indicate that the SRS has the ability to detect subtler or more immediate changes in social competence than the eyetracking measure. However, the observed SRS scores could also be impacted by reporter bias, since participants and their families were not blind to treatment group membership. The ER-40 also showed significant effects of both group (β=3.54, t=2.53, $p<0.016$) and visit 2 (β=2.85, t=3.23, $p<0.0021$), but did not show a significant interaction between group and visit 2 (β=-1.23, t=-1.06, $p<0.29$), indicating that group did not significantly affect social prioritization scores across time.

Additionally, while these findings do not support our initial hypotheses that this task would differentiate between treatment groups, these outcomes do reflect similar results from studies that found that individuals with ASD spent the same amount of time looking at social versus non-social stimuli as typically developing controls (Fletcher-Watson, Leekam, Benson, Frank, & Findlay, 2009). Instead, more subtle eyetracking measures of social attention (i.e., social or non-social first fixation) differentiated ASD from control participants. This may suggest that proportion of time spent looking at faces may not capture the nuanced social attention preferences that characterize the social difficulties of individuals with ASD. Furthermore, it’s possible that our findings differ from the findings of Chevallier, et al. (2015) given the age difference between the current sample (age M = 18.17, SD = 4.36) and their study sample (age M = 12.2, SD = 3.3). Perhaps these social attention preferences are more subtle for older individuals, and more sensitive or detailed measures should be employed in this age range.

Beyond group differences, it should be noted that Full Scale IQ approached significance in the MLM covariate analyses ($p<0.08$), indicating that individuals with higher cognitive ability may be more inclined to focus on social versus non-social visual stimuli in the eyetracking task. This may have implications for the adaptation of the interactive visual exploration eyetracking task to cognitively impaired populations. Alternatively, age was not a significant predictor of social prioritization in the longitudinal model. Further studies should investigate this eyetracking paradigm in younger and cognitively-impaired individuals with ASD and compare their performance to chronological and developmental age-matched typically developing controls. Additionally, previous studies have shown that individuals with ASD with higher cognitive and verbal abilities show greater gains in response to group social skills interventions (Solomon, Goodlin-Jones, & Anders, 2004); therefore, the association between visual social prioritization and IQ should continue to be examined in future longitudinal studies assessing response to social skills interventions in ASD.

Finally, although the correlational findings showed that measures of social functioning and social cognition did not significantly relate to visual social prioritization at baseline, the MLM analyses found that the ER-40 did have a significant relationship with the eyetracking metric over the course of the three visits. Rather than solely correlating with measures at single visit, the emotion recognition task showed significant interactions with the increase in social prioritization from baseline to visit 2, as well as significant interactions between visit 2 and treatment group. Figure 4 visualizes the moderation effect of emotion recognition abilities on the prioritization of visual social information within the eyetracking task. Additionally, results revealed a significant three-way interaction between visit 2, treatment group, and ER-40 performance. Because emotion recognition is specifically referenced numerous times within the SCIT-A protocol, it is reasonable to attribute these improvements to the social skills treatment. Overall, these findings may suggest that having greater knowledge of the information that can be gained from examining a facial expression may prompt individuals to visually seek out faces to better understand a social situation. This may also confirm the clinical utility of teaching emotion recognition abilities in an attempt to also indirectly increase gaze preference for social stimuli, thus creating a cascading effect on social skills progress.

In conclusion, this eyetracking measure shows good test-retest reliability within a control sample, but does not differentiate between those enrolled in the social skills treatment and those receiving treatment as usual across a longitudinal trajectory. It is important to note that the SCIT-A treatment group was compared to a group receiving a high level of care within the community, and, therefore, may have utility as a treatment outcome measure that is not reflected in these analyses. Additionally, the prioritization of social stimuli, as measured by the eyetracking task, showed no significant correlations with questionnaires and neurocognitive measures of social ability at baseline, indicating that it may be detecting a unique aspect of social attention or ability that is undetected by other measures. Emotion recognition abilities did, however, prove to be a significant moderator of social prioritization over time. This finding reiterates the importance of teaching emotion recognition skills to individuals with ASD. Although the social prioritization score did not show group differences between SCIT-A and TAU, the measure's good test-retest reliability, the significant change in preference for social stimuli across visits, and the association with emotion recognition ability warrant future studies to examine the utility of this eyetracking paradigm as both a diagnostic tool and a measure of treatment response for behavioral and pharmacological interventions for individuals with ASD.

Table 1

Participant Characteristics

	SCIT-A (*n*=17)	TAU (*n*=22)	Control (*n*=22)	
	Mean (SD)	Mean (SD)	Mean (SD)	Significance
Age	16.23 (2.3)	17.77 (4.0)	19.14 (1.1)	$F(59)=5.26$; $p>0.007$
FSIQ	93.76 (15.9)	115.27 (12.6)	107.83 (6.6)	$F(59)=15.96$; $p>0.001$
Number of Females	5	2	14	$F(59)=9.23$; $p>0.0003$

Note. *=p<0.05. There was a significant difference in IQ between SCIT-A and TAU participants, a significant difference in number of females between Control and both SCIT-A and TAU, and a significant difference in age between the Control and SCIT-A groups.

Table 2

Correlations between Eyetracking Social Prioritization and Measures of Social Functioning

Variable	*1*	*2*	3	4	5	6	7	8
1. SRS	---							
2. SRS Social Cognition	0.88*	---						
3. ER-40	-0.05	-0.32	---					
4. MEDF-36	0.06	-0.00	0.17	---				
5. Hinting Task	-0.24	-0.42*	0.47*	0.10	---			
6. LSAS	0.07	-0.05	0.06	0.25	0.03	---		
7. SNI	0.08	0.10	-0.20	0.05	-0.03	-0.41*	---	
8. Social Prioritization Score	-0.09	-0.07	0.01	-0.11	-0.11	-0.08	-0.02	---

Note. * indicates $p \leq .05$. SRS: social communication impairment; ER-40: emotion recognition; MEDF-36: emotion differentiation; Hinting Task: theory of mind; LSAS: social anxiety; Social Network Index: number of people in social network.

Table 3

Results from the multilevel modeling analyses showing the effect of IQ, Age, Group, Therapy, and Medication on Eyetracking Measures across visits

Predictor	Social Prioritization Score			
	Estimate (β)	95% CI	t value	p value
Fixed Effects				
Intercept	1.81	-6.81 – 10.42	0.43	0.67
Visit 1	2.50	0.97 – 4.02	3.27	0.0017*
Visit 2	2.61	1.02 – 4.21	3.27	0.0017*
Group	-1.23	-3.75 – 1.28	-1.00	0.33
FSIQ	0.06	-0.009 – 0.13	1.77	0.08
Age	-0.14	-0.43 – 0.15	-0.95	0.35
Current Therapy	-2.07	-4.81 – 0.68	-1.53	0.14
Current Medication	2.98	0.70 – 5.27	2.66	0.01*
Contrast Estimates				
Visit 1 vs. Visit 2	-0.12	N/A	-0.15	0.88

Note. *indicates $p \leq .05$

Table 4

Results from the multilevel modeling analyses showing the effect of ER-40 Total on Social Prioritization across visits

Predictor	Social Prioritization Score			
	Estimate (β)	95% CI	t value	p value
Fixed Effects				
Intercept	3.50	-9.39 – 16.39	0.55	0.59
Visit 1	-7.11	-22.49 – 8.26	-0.93	0.36
Visit 2	-36.06	-72.48 – 0.36	-1.99	0.052
Group	0.57	-18.48 – 19.62	0.06	0.95
ER-40	0.003	-0.44 – 0.44	0.01	0.99
Visit 1 * ER-40	0.32	-0.20 – 0.84	1.25	0.22
Visit 2 * ER-40	1.22	0.09 – 2.36	2.16	0.035*
Group * ER-40	0.009	-0.61 – 0.62	0.03	0.97
Group * Visit 1	6.56	-18.06 – 31.18	0.53	0.60
Group * Visit 2	54.38	9.37 – 99.39	2.42	0.02*
Group * ER-40 * Visit 1	-0.22	-1.01 – 0.56	-0.57	0.57
Group * ER-40 * Visit 2	-1.711	-3.10 – -0.33	-2.48	0.017*

Note. *indicates $p \leq .05$

Figure 1

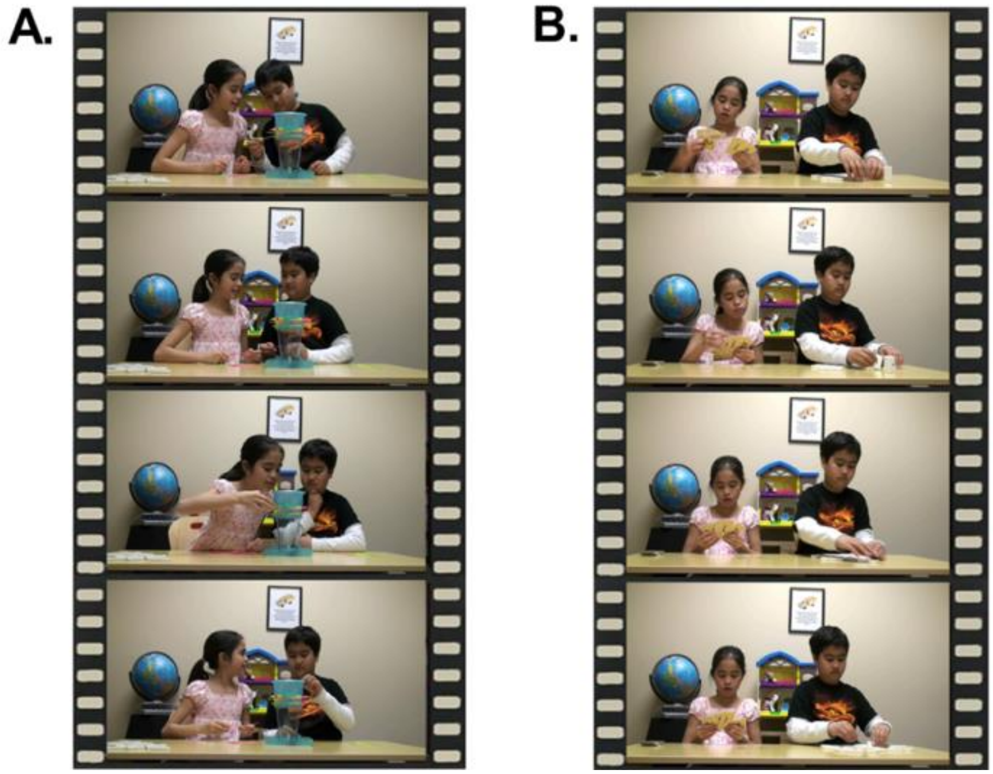

Figure 1. Stills from the dynamic stimuli used during eyetracking. The left panels (A) depict siblings playing together (joint condition); the right panels (B) depict siblings playing independently (parallel condition).

Figure 2

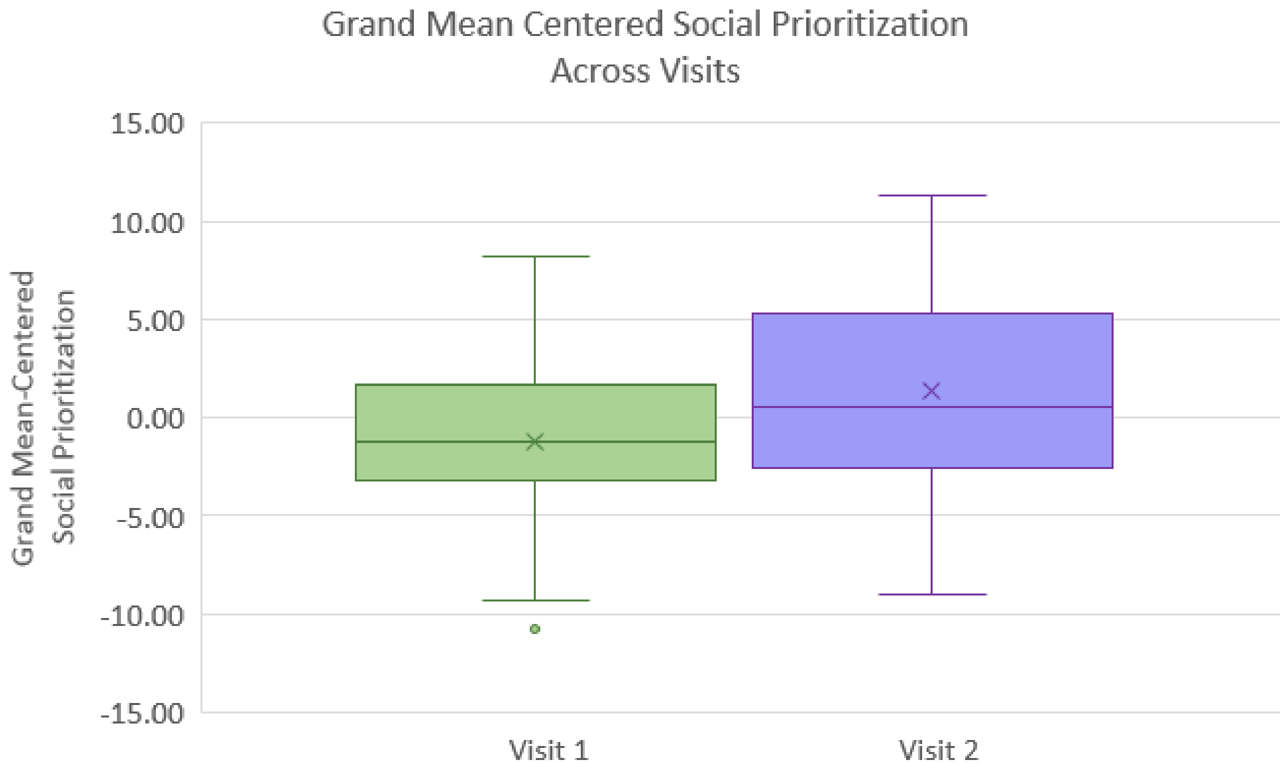

Figure 2. Group mean centered social prioritization scores at visit 1 (M = -1.25, SD = 4.45) and visit 2 (M = 1.37, SD = 5.23).

Figure 3

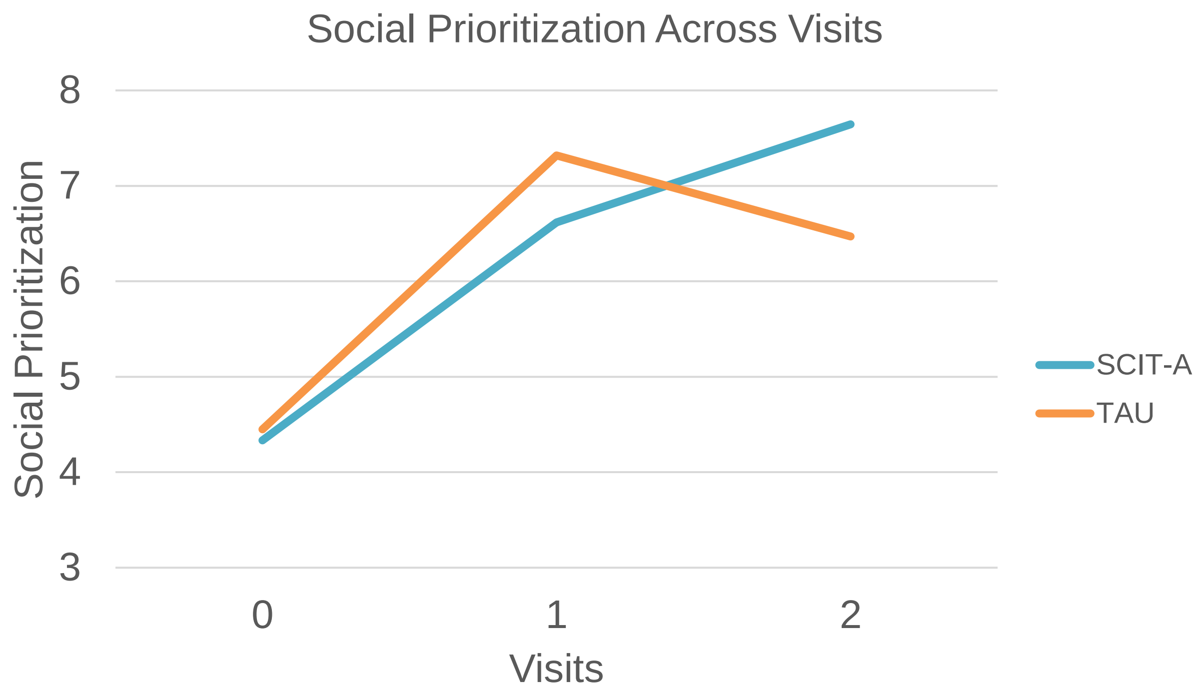

Figure 3. Social prioritization, as measured by the eyetracking task, across all visits and treatment groups.

Figure 4.

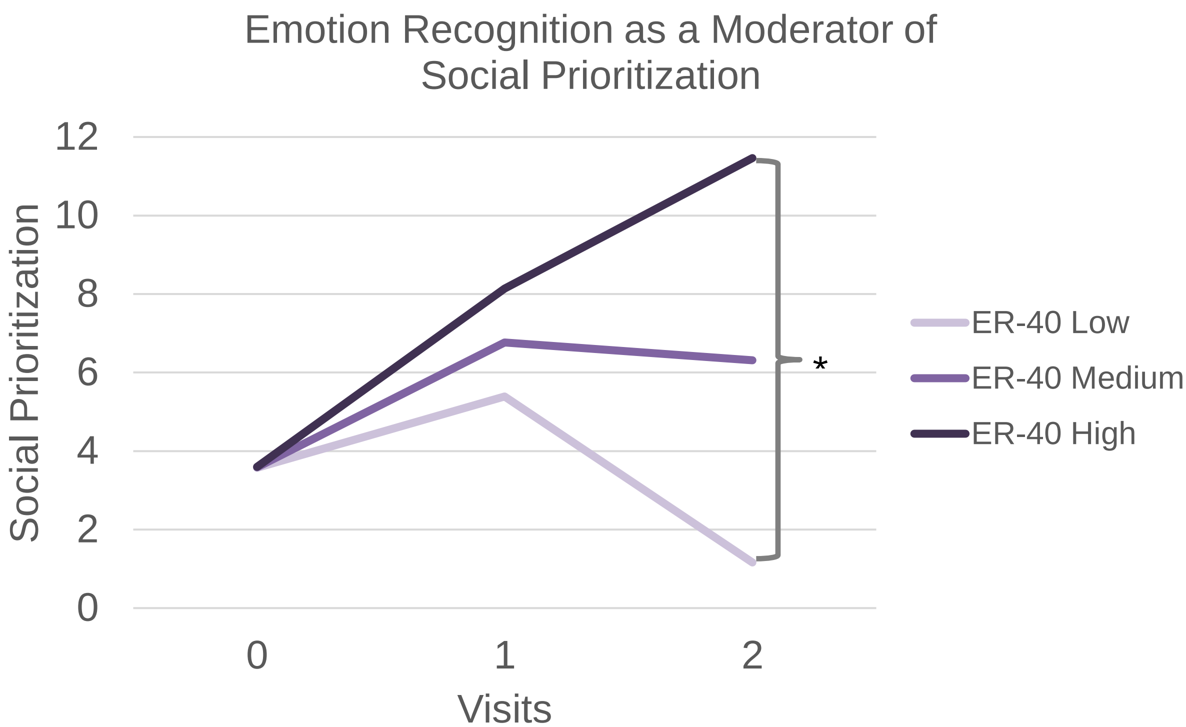

Figure 4. *=p<.05. Emotion recognition serves as a significant moderator of social prioritization of visual stimuli at visit 2.

Figure 5

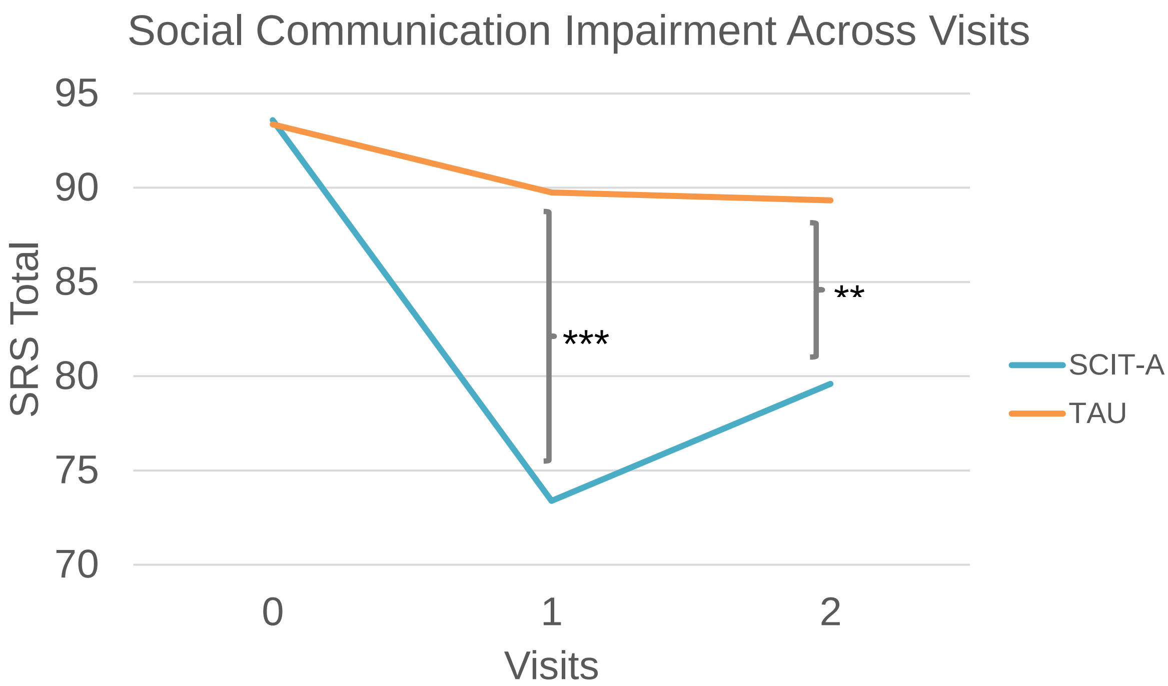

Figure 5. *** = p<.001; **=p<.01. These symbols indicate the significance of the interaction between group and visit. Social communication impairment (as measured by the SRS) examined across all visits and treatment groups.

REFERENCES

Aman, M. G., Novotny, S., Samango-Sprouse, C., Lecavalier, L., Leonard, E., Gadow, K. D., . . . Chez, M. (2004). Outcome measures for clinical drug trials in autism. *CNS spectrums, 9*(01), 36-47.

Association, A. P. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Washington, DC: American Psychiatric Association.

Axelrod, B. N. (2002). Validity of the Wechsler abbreviated scale of intelligence and other very short forms of estimating intellectual functioning. *Assessment, 9*(1), 17-23.

Baio, J. (2012). Prevalence of Autism Spectrum Disorders: Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States, 2008. Morbidity and Mortality Weekly Report. Surveillance Summaries. Volume 61, Number 3. *Centers for Disease Control and Prevention*.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*(1), 37-46.

Bauminger, N. (2002). The facilitation of social-emotional understanding and social interaction in high-functioning children with autism: Intervention outcomes. *Journal of autism and developmental disorders, 32*(4), 283-298.

Boraston, Z., & Blakemore, S. J. (2007). The application of eye-tracking technology in the study of autism. *The Journal of physiology, 581*(3), 893-898.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*(1), 147.

Buitelaar, J. K., Van der Wees, M., Swaab–Barneveld, H., & Van der Gaag, R. J. (1999). Theory of mind and emotion-recognition functioning in autistic spectrum disorders and in psychiatric control and normal children. *Development and psychopathology, 11*(01), 39-58.

Busner, J., Targum, S. D., & Miller, D. S. (2009). The Clinical Global Impressions scale: errors in understanding and use. *Comprehensive psychiatry, 50*(3), 257-262.

Cappadocia, M. C., & Weiss, J. A. (2011). Review of social skills training groups for youth with Asperger syndrome and high functioning autism. *Research in Autism Spectrum Disorders, 5*, 70-78.

Chawarska, K., Macari, S., & Shic, F. (2012). Context modulates attention to social scenes in toddlers with autism. *Journal of Child Psychology and Psychiatry, 53*(8), 903-913.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in cognitive sciences, 16*(4), 231-239.

Chevallier, C., Noveck, I., Happé, F., & Wilson, D. (2011). What's in a voice? Prosody as a test case for the Theory of Mind account of autism. *Neuropsychologia, 49*(3), 507-517.

Chevallier, C., Parish-Morris, J., McVey, A., Rump, K. M., Sasson, N. J., Herrington, J. D., & Schultz, R. T. (2015). Measuring social attention and motivation in autism spectrum disorder using eye-tracking: Stimulus type matters. *Autism Research*.

Cohen, S., Doyle, W. J., Skoner, D. P., Rabin, B. S., & Gwaltney, J. M. (1997). Social ties and susceptibility to the common cold. *Jama, 277*(24), 1940-1944.

Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., . . . Reich, W. (2003). Validation of a brief quantitative measure of autistic traits: comparison of the social responsiveness scale with the autism diagnostic interview-revised. *Journal of autism and developmental disorders, 33*(4), 427-433.

Constantino, J. N., & Gruber, C. P. (2002). The social responsiveness scale. *Los Angeles: Western Psychological Services*.

Constantino, J. N., & Todd, R. D. (2014). Genetic structure of reciprocal social behavior. *American Journal of Psychiatry*.

Corp., I. (2015). *IBM SPSS Statistics for Windows, Version 23.0*. Armonk, NY: IBM Corp.

Coury, D. L., Anagnostou, E., Manning-Courtney, P., Reynolds, A., Cole, L., McCoy, R., . . . Perrin, J. M. (2012). Use of psychotropic medication in children and adolescents with autism spectrum disorders. *Pediatrics, 130*(Supplement 2), S69-S76.

Crawford, J., Stewart, L., Cochrane, R., Parker, D., & Besson, J. (1989). Construct validity of the National Adult Reading Test: A factor analytic study. *Personality and Individual Differences, 10*(5), 585-587.

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *Journal of autism and developmental disorders, 28*(6), 479-485.

Dove, D., Warren, Z., McPheeters, M. L., Taylor, J. L., Sathe, N. A., & Veenstra-VanderWeele, J. (2012). Medications for adolescents and young adults with autism spectrum disorders: a systematic review. *Pediatrics, 130*(4), 717-726.

Fletcher-Watson, S., Leekam, S. R., Benson, V., Frank, M. C., & Findlay, J. M. (2009). Eye-movements reveal attention to social information in autism spectrum disorder. *Neuropsychologia, 47*(1), 248-257. doi: 10.1016/j.neuropsychologia.2008.07.016

Gur, R. C., Sara, R., Hagendoorn, M., Marom, O., Hughett, P., Macy, L., . . . Gur, R. E. (2002). A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. *Journal of neuroscience methods, 115*(2), 137-143.

Heimberg, R. G., Horner, K., Juster, H., Safren, S., Brown, E., Schneier, F., & Liebowitz, M. (1999). Psychometric properties of the Liebowitz social anxiety scale. *Psychological medicine, 29*(01), 199-212.

Hill, E., Berthoz, S., & Frith, U. (2004). Brief report: cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives. *J Autism Dev Disord, 34*(2), 229-235.

Howlin, P., Moss, P., Savage, S., & Rutter, M. (2013). Social outcomes in mid- to later adulthood among individuals diagnosed with autism and average nonverbal IQ as children. *J Am Acad Child Adolesc Psychiatry, 52*(6), 572-581 e571. doi: 10.1016/j.jaac.2013.02.017

Institute, S. (1985). *SAS user's guide: statistics* (Vol. 2): Sas Inst.

Jones, W., Carr, K., & Klin, A. (2008). Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Archives of general psychiatry, 65*(8), 946-954.

Kaat, A. J., & Lecavalier, L. (2014). Group-based social skills treatment: A methodological review. *Research in Autism Spectrum Disorders, 8*(1), 15-24.

Kalkbrenner, A. E., Daniels, J. L., Emch, M., Morrissey, J., Poole, C., & Chen, J.-C. (2011). Geographic access to health services and diagnosis with an autism spectrum disorder. *Annals of epidemiology, 21*(4), 304-310.

Kanne, S. M., Mazurek, M. O., Sikora, D., Bellando, J., Branum-Martin, L., Handen, B., . . . Warren, Z. (2014). The Autism Impact Measure (AIM): initial development of a new tool for treatment outcome measurement. *Journal of autism and developmental disorders, 44*(1), 168-179.

Kanner, L. (1943). *Autistic disturbances of affective contact*: publisher not identified.

Laugeson, E. A., Frankel, F., Gantman, A., Dillon, A. R., & Mogil, C. (2012). Evidence-based social skills training for adolescents with autism spectrum disorders: The UCLA PEERS program. *Journal of autism and developmental disorders, 42*(6), 1025-1036.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule, second edition (ADOS-2) manual (Part I): Modules 1-4*. Torrance, CA: Western Psychological Services.

Masi, A., Lampit, A., Glozier, N., Hickie, I., & Guastella, A. (2015). Predictors of placebo response in pharmacological and dietary supplement treatment trials in pediatric autism spectrum disorder: a meta-analysis. *Translational psychiatry, 5*(9), e640.

Payakachat, N., Tilford, J. M., Kovacs, E., & Kuhlthau, K. (2012). Autism spectrum disorders: a review of measures for clinical, health services and cost-effectiveness applications.

Reichow, B., & Volkmar, F. R. (2010). Social skills interventions for individuals with autism: Evaluation for evidence-based practices within a best evidence synthesis framework. *Journal of autism and developmental disorders, 40*(2), 149-166.

Roberts, D. L., Penn, D. L., & Combs, D. R. (2015). *Social cognition and interaction training (SCIT): Group psychotherapy for schizophrenia and other psychotic disorders, clinician guide*: Oxford University Press.

Schultz, R. T. (2005). Developmental deficits in social perception in autism: the role of the amygdala and fusiform face area. *International Journal of Developmental Neuroscience, 23*(2), 125-141.

Shic, F., Bradshaw, J., Klin, A., Scassellati, B., & Chawarska, K. (2011). Limited activity monitoring in toddlers with autism spectrum disorder. *Brain research, 1380*, 246-254.

Solomon, M., Goodlin-Jones, B. L., & Anders, T. F. (2004). A social adjustment enhancement intervention for high functioning autism, Asperger's syndrome, and pervasive developmental disorder NOS. *Journal of autism and developmental disorders, 34*(6), 649-668.

Spencer, D., Marshall, J., Post, B., Kulakodlu, M., Newschaffer, C., Dennen, T., . . . Jain, A. (2013). Psychotropic medication use and polypharmacy in children with autism spectrum disorders. *Pediatrics, 132*(5), 833-840.

Turner-Brown, L. M., Perry, T. D., Dichter, G. S., Bodfish, J. W., & Penn, D. L. (2008). Brief report: feasibility of social cognition and interaction training for adults with high functioning autism. *J Autism Dev Disord, 38*(9), 1777-1784. doi: 10.1007/s10803-008-0545-y