Journal of Autism and Developmental Disorders (2021) 51:1173–1187
https://doi.org/10.1007/s10803-020-04594-1

ORIGINAL PAPER

Dynamic Eye Tracking as a Predictor and Outcome Measure of Social Skills Intervention in Adolescents and Adults with Autism Spectrum Disorder

Rachel K. Greene[1,9] · Julia Parish-Morris[2] · Miranda Sullivan[3,10] · Jessica L. Kinard[3,4] · Maya G. Mosner[1,11] · Lauren M. Turner-Brown[5] · David L. Penn[1] · Christopher A. Wiesen[6] · Ashley A. Pallathra[7,12] · Edward S. Brodkin[7] · Robert T. Schultz[2] · Gabriel S. Dichter[3,8]

Published online: 12 July 2020

Abstract
To evaluate an eye tracking task as a predictor and outcome measure of treatment response for autism spectrum disorder (ASD) social skills interventions, adolescents and young adults with ASD completed the eye tracking task before, immediately after, and two months after completing Social Cognition and Interaction Training for Autism (SCIT-A). The study compared SCIT-A participants ($n=20$) to participants with ASD who received treatment as usual (TAU; $n=21$). Overall, increased visual attention to faces and background objects and decreased attention to hands playing with toys at baseline were associated with improved social functioning immediately following intervention, suggesting this eye tracking task may reliably predict ASD social intervention outcomes.

Keywords Social skills intervention · Treatment outcome · Autism spectrum disorder · Eye tracking

Autism spectrum disorder (ASD) is characterized by deficits in social skills, including marked impairments in social cognition, social perception, and social communication (Chevallier et al. 2012; Howlin et al. 2013). A number of psychosocial and pharmacological interventions have demonstrated moderate success in ameliorating core and associated symptoms of ASD. However, despite progress in developing novel interventions for ASD, the field still lacks objective and valid measures to accurately evaluate and predict the efficacy of such treatments.

Unlike other psychiatric disorders, where self-report questionnaires are routinely used to evaluate treatment effectiveness, many individuals with ASD have limited insight into their socio-emotional states, rendering self-report

✉ Gabriel S. Dichter
dichter@med.unc.edu

1 Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

2 Center for Autism Research, Children's Hospital of Philadelphia, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

3 Carolina Institute for Developmental Disabilities, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA

4 Division of Speech and Hearing Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

5 TEACCH Autism Program, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA

6 The Odum Institute, The University of North Carolina, Chapel Hill, NC, USA

7 Center for Neurobiology and Behavior, Department of Psychiatry, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

8 Department of Psychiatry, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA

9 Present Address: Institute on Development and Disability, Oregon Health & Science University, Portland, OR, USA

10 Present Address: School of Medicine, Virginia Commonwealth University, Richmond, VA, USA

11 Present Address: Center for Autism Research, Children's Hospital of Philadelphia, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

12 Present Address: Department of Psychology, The Catholic University of America, Washington, DC, USA

treatment outcome measures poorly suited to evaluate treatment outcomes within this population (Hill et al. 2004; Payakachat et al. 2012). Instead, caregiver- or parent-report measures are often used as an alternative to self-report measures. This reporting method, however, typically 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).

Another method that has been used to evaluate ASD treatment effectiveness is clinician reports of changes in symptoms and level of functioning (Arnold et al. 2003; King et al. 2009). However, clinician reports are more susceptible to placebo effects, including greater response to active treatment conditions, compared to caregiver reports (Masi et al. 2015). 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 impairments in specific domains targeted by an intervention. Finally, many diagnostic tools that are commonly used as outcome measures (e.g., Autism Diagnostic Interview-Revised (ADI-R; Lord et al. 1994), Autism Diagnostic Observation Schedule, Second Edition (ADOS-2; Lord et al. 2012), Childhood Autism Rating Scale-Second Edition (CARS-2; Schopler et al. 2010)) were not designed to serve as measures of subtle behavioral change over time (Kanne et al. 2014; Payakachat et al. 2012). This lack of validated measures of treatment response hampers the development and evaluation of effective treatments for ASD-associated social impairments.

Eye tracking has the potential to function as a valid predictor and treatment outcome measure for use with individuals with ASD of varying ages and levels of impairment. Eye tracking has been used extensively as a measure of visual social attention and attentional biases to orient towards or away from social stimuli with individuals with ASD (Dawson et al. 1998), and eye tracking studies have shown abnormalities in social gaze behaviors in individuals with ASD (Chawarska et al. 2012; Jones et al. 2008). For example, Chawarska et al. (2012) found that 6-month-old infants who were later diagnosed with ASD showed diminished social monitoring, with particular visual disinterest in an actor's face, compared to developmentally delayed and typically developing infants who did not develop ASD. Similarly, Shic et al. (2011) reported that, while viewing videos of an adult–child play interactions, 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 that explain the development of social impairments in ASD as a reflection of decreased preference for social stimuli in early life that leads to downstream social communication deficits that form the core of ASD (Chevallier et al. 2012; Dawson et al. 1998; Parish-Morris et al. 2019; Schultz 2005).

Similarly, eye tracking has been used in studies with adolescents and adults with ASD to examine visual attention to social versus non-social objects in the broader environment (for a review see Guillon et al. 2014). Findings from these studies have been mixed, with some showing adolescents and adults with ASD demonstrate decreased or slower gaze to faces compared to typically developing peers (Bird et al. 2011; Riby et al. 2012; Riby and Hancock, 2008, 2009; Sasson et al. 2008). Other studies have suggested findings of reduced visual attention to be conditional, such that individuals with ASD may focus on faces less within social scenes but show no differences in gaze patterns when examining isolated images of faces (Hanley et al. 2013). Individuals with ASD have also demonstrated commensurate visual exploration of both faces and objects, whereas typically developing individuals showed a significant preference for faces (Hanley et al. 2013). Importantly, however, there are also findings that do *not* support this hypothesis of reduced or slower visual attention to social stimuli in ASD (Kuhn et al. 2010; Nakano et al. 2010).

The goal of the present study was to evaluate the validity of an eye tracking task as an outcome measure and as a predictor of treatment outcomes for adolescents and adults with ASD. To accomplish this, test–retest reliability of an established dynamic social eye tracking task was evaluated in typically developing controls (TDCs), with the hypothesis that eye tracking metrics would show at least acceptable test–retest reliability. Second, correspondence between baseline eye tracking metrics and measures of social functioning was assessed in participants with ASD, with the hypothesis that eye tracking metrics would be associated with measures of social functioning, including measures of emotion recognition and theory of mind. Changes in social visual attention were evaluated before, immediately after, and 8-weeks after an 8-week group-based empirically validated psychosocial intervention, Social Cognition and Interaction Training for Autism (SCIT-A). Changes in social attention measured by eye tracking were compared between a group of individuals with ASD who completed SCIT-A and a group of individuals with ASD who solely received treatment as usual (TAU) within the community. Changes in social attention were also compared to changes in neurocognitive and self-report measures of social impairment to evaluate the concurrent validity of the eye tracking measure. It was hypothesized that the magnitude of change in measures of social cognition and social functioning would correspond with the magnitude of change in eye tracking metrics to a greater extent in the treatment group than in the TAU group, given that only

38% of TAU participants were receiving any behavioral intervention at the time of the study. Additionally, baseline social attention was examined as a predictor of treatment outcomes in the SCIT-A group. In exploring this aim, it was hypothesized that increased gaze preference for faces at baseline would predict greater social improvement. This hypothesis supposes that individuals who show heightened visual attention to social stimuli would, therefore, show greater response to the social intervention.

Methods

Participants

The study included three cohorts: a SCIT-A treatment (SCIT-A) group, a treatment as usual (TAU) group, and a typically developing control (TDC) group (Table 1). The sample included 20 SCIT-A participants (age $M = 17.25$, $SD = 3.58$) from six separate therapy cohorts, 21 TAU participants (age $M = 17.94$, $SD = 4.05$), and 22 TDC participants (age $M = 19.95$, $SD = 4.56$). All participants in the SCIT-A and TAU groups 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). The TAU group did not receive SCIT-A, but some continued to receive clinical services outside of their participation in the current study to evaluate whether the eye tracking task was capable of detecting differential change in social skills over time.

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. Fifty-five percent of the SCIT-A participants were receiving some form of ongoing behavioral intervention (not including SCIT-A), and 80% were taking at least one psychiatric medication over the course of the three study visits (see Table 1). Thirty-eight percent of the TAU group were participating in community- or school-based behavioral therapies, and 48% of TAU participants were taking at least one psychiatric medication. Groups significantly differed with respect to psychotropic medication use ($\chi^2 = 4.63$; $p = 0.031$) at baseline, but not behavioral interventions ($\chi^2 = 1.18$; $p = 0.278$). Participants in all groups met the IQ cutoff of Full-Scale IQ > 70. There were significant differences in IQ between the SCIT-A (FSIQ $M = 95.85$, $SD = 16.53$) and both the TAU (FSIQ $M = 114.90$, $SD = 12.83$) and control (FSIQ $M = 108.23$, $SD = 6.55$) groups, but IQ did not correlate with eye tracking metrics within SCITA ($p = 0.7703$) or across both groups ($p = 0.5506$) at baseline. The control group differed from both ASD groups in sex distribution ($\chi^2 = 15.07$; $p = 0.0005$). Statistical analyses examining group differences included IQ, age, sex, therapy status (yes/no), and psychotropic medication use (yes/no) as predictors.

Table 1 Participant characteristics

	SCIT-A ($n = 20$)	TAU ($n = 21$)	Control ($n = 22$)	Significance
Age [mean (SD)]	17.25 (3.58)	17.90 (4.1)	19.95 (4.6)	$F = 2.52$; $p = 0.089$
FSIQ [mean (SD)]	95.85 (16.53)	114.90 (12.8)	108.23 (6.6)	$F = 12.21$; $p < 0.0001$*
SRS-2 baseline total raw score [mean (SD)]	92.90 (23.44)	94.95 (24.6)	52.18 (21.0)	$F = 23.60$ $p < 0.0001$*
Female [N (%)]	5 (25.0%)	2 (9.5%)	14 (63.6%)	$\chi^2 = 15.07$; $p = 0.0005$*
Race [N (%)]#				$\chi^2 = 8.29$; $p = 0.406$
American Indian or Alaskan Native	0 (0.0%)	0 (0.0%)	1 (5.0%)	
Asian	1 (5.0%)	0 (0.0%)	2 (10.0%)	
Black	4 (20.0%)	2 (9.5%)	2 (10.0%)	
White	13 (65.0%)	19 (90.5%)	14 (70.0%)	
Multiracial	2 (10.0%)	0 (0.0%)	1 (5.0%)	
Ethnicity [N (%)]#				$\chi^2 = 0.50$; $p = 0.778$
Hispanic or Latino	2 (10.0%)	1 (4.8%)	2 (10.0%)	
Non-Hispanic or Latino	18 (90.0%)	20 (95.2%)	18 (90.0%)	
Participating in other behavioral interventions [N (%)]	11 (55.0%)	8 (38.1%)	–	$\chi^2 = 1.18$; $p = 0.278$
Taking psychotropic medications [N (%)]	16 (80%)	10 (47.6%)	–	$\chi^2 = 4.63$; $p = 0.031$*

*$p < 0.05$

#Two control participants declined to identify their race and ethnicity. There was a significant difference in IQ between SCIT-A and both control and TAU participants. There was also a significant difference in number of females between Control and both SCIT-A and TAU. As expected, there was also a significant discrepancy between Control baseline SRS-2 totals and those of both SCIT-A and TAU participants

Recruitment

SCIT-A and TAU participants were recruited primarily from the Carolina Institute for Developmental Disabilities (CIDD) Autism Subject Registry, and recruitment was supplemented by community flyers and online postings. 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. Individuals interested in receiving social skills training but who declined to participate in the present study were still able to receive SCIT-A treatment. SCIT-A and TAU group membership was not randomly assigned. All SCIT-A and TAU participants expressed interest in the social skills training; however, participants who were unable to attend the social skills group because of scheduling or other logistical difficulties were invited to participate within the TAU arm of the current study. This non-random assignment was necessary for adequate study enrollment and ethical concerns surrounding denying services to treatment seeking individuals. TDC participants were recruited from the undergraduate psychology research pool at the University of North Carolina at Chapel Hill.

Procedure

This protocol was approved by the Institutional Review Board at the University of North Carolina at Chapel Hill. Informed consent from participants or parents/guardians was obtained.

TDC Procedures

TDC participants attended two study visits that were separated by approximately 24 h. During the first study visit, participants completed informed consent, a test of cognitive ability, the eye tracking task, clinical report measures, and neurocognitive measures of emotion regulation and theory of mind. At the second study visit, a day later, the eye tracking task was re-administered. No additional study measures were completed at the second visit. TDC participants received course credit for participating.

SCIT-A and TAU Procedures

The study was conducted over three separate testing visits. SCIT-A participants completed one visit within approximately two weeks prior to the first SCIT-A session, one visit within two weeks after completing the final SCIT-A session, and a final visit approximately eight weeks following the final SCIT-A session. Participants in the SCIT-A group attended up to eight social skills group sessions between the first and second testing sessions (sessions attended $M = 7.05$, $SD = 1.10$). TAU participants followed a similar timeline without participating in the SCIT-A group (i.e., approximately eight weeks between each visit). The first testing visit for SCIT-A and TAU participants lasted approximately 3 h, and the second and third visits lasted approximately 1.5 h each. During each testing visit, participants completed the eye tracking 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 a licensed clinical psychologist. All participants received monetary compensation (i.e., $30 for the first visit and $20 for each additional visit).

SCIT-A Treatment

The SCIT-A program was originally adapted from a social skills group intervention for individuals with psychosis (Roberts et al. 2015), due to overlap in social cognitive impairments in schizophrenia and ASD. SCIT-A was specifically targeted to focus on social cognitive processes and strategies within a structured teaching environment. The SCIT-A program is 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; and (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). A 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 a TAU group (Turner-Brown et al. 2008). Additionally, SCIT-A has been incorporated as a component of a novel, multi-part intervention designed to improve social functioning in adults with ASD, with promising results (Pallathra et al. 2018). The treatment was provided over eight consecutive weekly sessions, and each group consisted of six to ten individuals with ASD (IQs > 70). This intervention was used in the current study to evaluate eye tracking as a treatment prediction and outcome measure, rather than to validate the intervention itself. Attendance was recorded and participants were required to attend at least five out of the eight sessions to be included in analyses (SCIT-A attendance $M = 7.05$, $SD = 1.099$); attendance was not associated with change in eye tracking metrics (i.e., visual social prioritization) from baseline to post-treatment.

Materials and Measures

Cognitive and ASD diagnostic assessments were only completed at the baseline visit, whereas all other measures were completed at each testing visit.

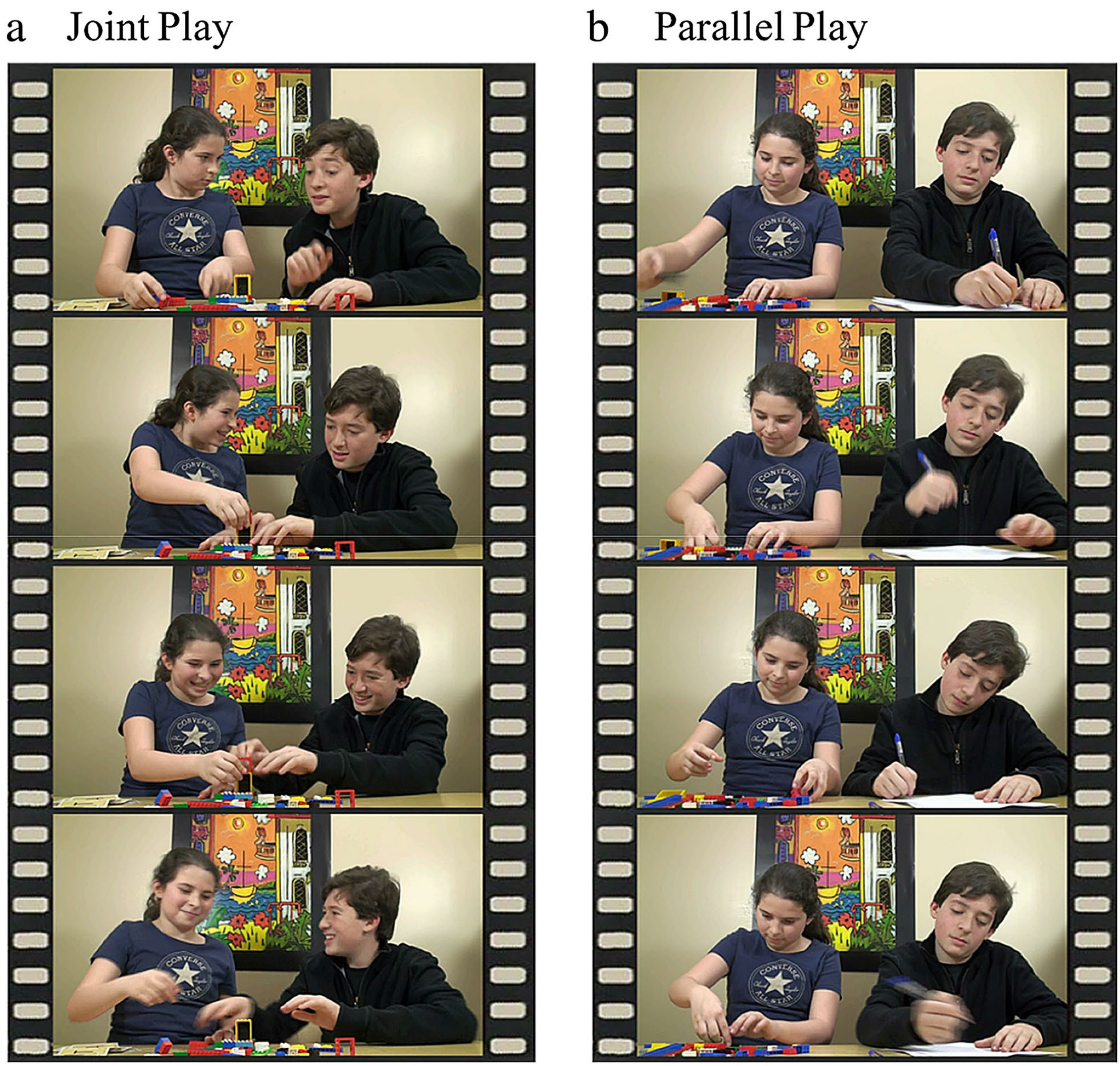


Fig. 1 Stills from the interactive visual exploration eye tracking task. The left panel **a** depicts siblings playing together (joint condition); the right panel **b** depicts siblings playing independently (parallel condition). Figure from Parish-Morris et al. (2019)

Eye Tracking Task

The eye tracking task, the Interactive Visual Exploration (IVE) task, presented 22 silent video clips of 11 sibling pairs each participating in a social (joint condition) and non-social (parallel condition) play activity (see Fig. 1; Chevallier et al. 2015). 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 everyday life outside of the laboratory setting. The actors participated in one of two conditions: social interaction or parallel play. In the condition portraying 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 and 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 min. The IVE task has been found to be more successful in differentiating between ASD and control samples than eye tracking paradigms consisting of static images or videos without ecologically-valid social context, and IVE task outcomes correlate with the magnitude of ASD symptoms (Chevallier et al. 2015).

Each of these video clips contained pre-determined areas of interest (AOIs) that were traced by hand and changed over time with the progression of the dynamic stimulus. These captured faces, background objects, and hands as they moved during the paradigm. The primary metric obtained from this task was a measure of total fixation duration (TFD) for each AOI (e.g., faces, background, hands), representing the total amount of time the viewer directed their eye gaze at that specific AOI. TFD was aggregated for each AOI across all video clips to create a variable summing all fixations made to faces, background objects, and hands.

Eye Tracking System Specifications and Settings

The IVE eye tracking task was administered on a Tobii TX300 eye tracker integrated with a 23″ display monitor located at the CIDD. Before beginning the task, participants' eyes were positioned approximately 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 re-administered until all nine target locations were accounted for accurately. Once all locations were precisely calibrated, as determined by the Tobii system, participants were asked to remain still and silent while they watched the video on the screen.

The eye tracker 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. Datapoints with less than 50% accuracy were excluded from subsequent analyses. SCIT-A and TAU participants were only included in analyses if they had valid eye tracking data for two out of three timepoints. Single data points from four separate participants were excluded on this basis. All three data points were obtained for 16 SCITA participants and 17 TAU participants. Two valid data points were collected for 4 SCITA and 4 TAU participants. Control participants were only included if they had valid eye tracking data for both timepoints.

Autism Diagnostic Assessment

SCIT-A and TAU participants were administered either module 3 or module 4 of the ADOS-2, as determined by age and verbal ability. The ADOS-2 was administered by a research-reliable graduate level research assistant who was supervised by a licensed clinical psychologist. This portion of testing lasted approximately 45 min, during which participants were asked to complete activities such as telling a story from pictures and giving an account of a routine daily activity. Additionally, participants were asked questions regarding their perceived role in social situations and understanding of personal responsibilities. Standard ADOS-2 algorithm cutoff scores were used to determine whether or not participants met ASD criteria.

Cognitive Assessments

To assess general cognitive functioning and match study groups, participants were administered one of two tests. Because cognitive ability was not centrally relevant to test–retest reliability and to reduce the burden of additional testing, TDC participants completed the National Adult Reading Test-Revised (NART-R; Crawford et al. 1989), a brief assessment 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, which indexed verbal and spatial intelligence respectively. Administration of the two-subtest version took approximately 20 min, and performance on these two subtests were aggregated to provide a Full-Scale IQ (FSIQ).

Symptom Report Measures

Participants and caregivers in both ASD groups completed the Social Responsiveness Scale, Second Edition (Constantino and Gruber 2002). This 65-item measure assesses autism symptom severity. Participants and caregivers answered each question using a 4-point Likert scale. The SRS-2 reliably distinguishes individuals with ASD from individuals with other psychiatric diagnoses (Constantino et al. 2003; Constantino and Todd 2000). In addition to an overall ASD symptom severity score, the SRS-2 also provides subscales indexing social cognition, social motivation, social awareness, social communication, and restricted interests and repetitive behaviors. Caregiver reported SRS-2 scores were used in the final analyses unless a caregiver of an adult participant was not present, in which case self-report was used.

Finally, based on these AOI TFD proportions, a "social prioritization score" was derived 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):

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

Assessments of Social Cognitive Processes

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. Study participants were asked to identify the emotion of each facial expression. A composite score indicating the percentage of correct responses was used in the subsequent analyses.

The Hinting Task (Corcoran et al. 1995) presents ten short vignettes involving social interactions between two characters, one of whom drops a hint for the fictional partner about their desire or intention (e.g., As her birthday approaches, a girl mentions to her father how much she loves dogs, hinting that she wants a pet dog for her birthday). Participants were asked to identify the implicit intention of the hint. The participant was awarded two points for each correct response; however, if they provided an incorrect response initially, they were given additional information about the social interaction and another chance to determine the fictional social partner's meaning for a single point. A total score was then calculated by adding the number of points earned for each of the ten items.

Data Analysis

Data Preparation

Three areas of interest (AOIs) were obtained from the eye tracking data: faces, hands with toys, and background. Total fixation duration (TFD; in milliseconds) on each AOI was calculated using Tobii Pro Studio software. The proportion of total fixation duration (PTFD) 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:

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

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). The following were used as dependent variables in analyses: social prioritization score, PTFD faces, PTFD background, and PTFD hands. All analyses were conducted using SAS software, Version 9.4 (SAS Institute, 1985) or SPSS software, Version 23 (IBM Corp. 2015).

Test–Retest Reliability

Two methods were used to assess the test–retest reliability of four eye tracking dependent measures (i.e. social prioritization total, proportion of total fixation duration (PTFD) to faces, PTFD to background, and PTFD to hands) across the two TDC time points. First, an absolute change in the intraclass correlation coefficient (ICC) was calculated to examine the reliability of the eye tracking metric across timepoints within the TDC sample only. Next, Cronbach's alpha was calculated to further examine consistency across the two eye tracking time points in the TDC sample only. Because there is no definitive consensus regarding acceptable levels of reliability, providing accurate descriptors of reliability is challenging and depends on a number of factors including sample size, setting, and purpose (Charter and Feldt 2001; Skinner et al. 2018). However, it is generally accepted that alpha values between 0.6 and 0.7 are acceptable, and values 0.8 and higher are very good (Ursachi et al. 2015). Regarding intraclass correlation coefficients (ICCs), values less than 0.5 are considered poor, those between 0.5 and 0.75 moderate, those 0.75 to 0.90 good, and those greater than 0.9 excellent (Koo and Li 2016). These are the ranges that will be used in interpreting the current findings.

Correspondence Between Baseline Eye Tracking Metrics and Measures of Social Functioning in the ASD Groups

A Pearson correlation analysis examined the dimensionality of the eye tracking task and its correspondence with measures of social cognition (i.e., the Hinting Task, ER-40) and social functioning (i.e., SRS-2) at baseline in both the SCIT-A and TAU groups. Correlations were used to examine the convergent validity between the eye tracking paradigm and measures of social impairment.

Eye Tracking as a Predictor of Social Improvement Immediately Following SCIT-A

Pearson correlation analyses examined baseline eye tracking metrics as predictors of change in social functioning from baseline to immediately following the SCIT-A intervention (i.e., from baseline to post-treatment). Only SCIT-A participants (i.e., not TDC or TAU) were included within these analyses. Social functioning in these analyses was measured across multiple SRS-2 subscales (i.e., social cognition, social awareness, social communication, social motivation, restricted interests and repetitive behaviors) and the SRS-2 total score. Change scores were then used to represent the level of improvement from baseline to post-treatment.

Eye Tracking as a Treatment Outcome Measure

Multilevel modeling (MLM) was used to assess the utility of the social prioritization eye tracking metric as a measure of change (Bryk and Raudenbush 1987). This metric was chosen given its ability to reliably differentiate ASD from TDC participants (Chevallier et al. 2015). MLM techniques were chosen to account for the dependence of each data point from the same individual (i.e., in contrast to other analytic models, such as repeated measures ANOVA, that assume independence of individual data points).

Because this study included participants who took part in six separate SCIT-A therapy cohorts over the course of several years, a preliminary random effects model was used to calculate the intraclass correlation coefficient (ICC) for eye tracking data across and within therapy cohorts to determine if there was within-group dependence. This was done to confirm that there was not a substantial effect of therapy cohort. This model estimated that a small amount of variance could be attributed to therapy cohort membership. The ICC estimate for therapy cohort was 0.08, indicating that approximately 8% of the variability in individual prioritization of social information versus non-social information could 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. Analyses also compared a linear model structure with a quadratic function and found that the linear model provided the best model fit. A homoscedastic error structure was assumed for the following analyses.

Next, trajectories of eye tracking social prioritization scores were examined across study visits while controlling for the following time-invariant covariates: FSIQ, sex, age, treatment group (i.e., TAU or SCIT-A), participation in outside therapies, and use of psychotropic medications. The time variable was defined such that the baseline visit was coded as the intercept (i.e., the reference). Additionally, all visits were represented as individualized variables to account for the unique distribution between time points (e.g., baseline vs. post-treatment, baseline vs. follow-up, post-treatment vs. follow-up). This removed the assumption that there would be commensurate change in scores between each visit, and allowed us to examine distinct differences in the eye tracking metric between varying time points. This model allowed for a random intercept. No other variables were included within the random effects model.

Finally, the correspondence between eye tracking measures and reports of social functioning and neurocognitive measures of social cognition and social functioning was assessed over the three timepoints, again using an MLM framework. Predictors of the eye tracking social prioritization score included: ER-40, SRS-2 total score, and Hinting Task in separate models. Models assumed no within-group dependence within therapy cohorts, and, therefore, represented a 2-level structure, rather than a partially-nested 3-level model. Models also maintained the same coding of the timing variable as described above.

Fig. 2 Prediction of change in social communication impairment (as measured by SRS-2 social communication subscale) using Pearson correlations. The images in the left column of the figure depict the visual areas of interest included in each separate analysis. The participants viewed the videos in color. The current image is presented in black and white, highlighting the areas of interest in yellow, for clarity. Scatterplots in the right-hand column illustrate the associations between baseline visual attention to various stimuli (i.e., background objects, faces, hands and manipulated objects) and changes in social communication impairment from baseline to post-treatment for SCIT-A participants. *PTFD* proportion total fixation duration

Treatment Effects as Measured by SRS-2 in the ASD Groups

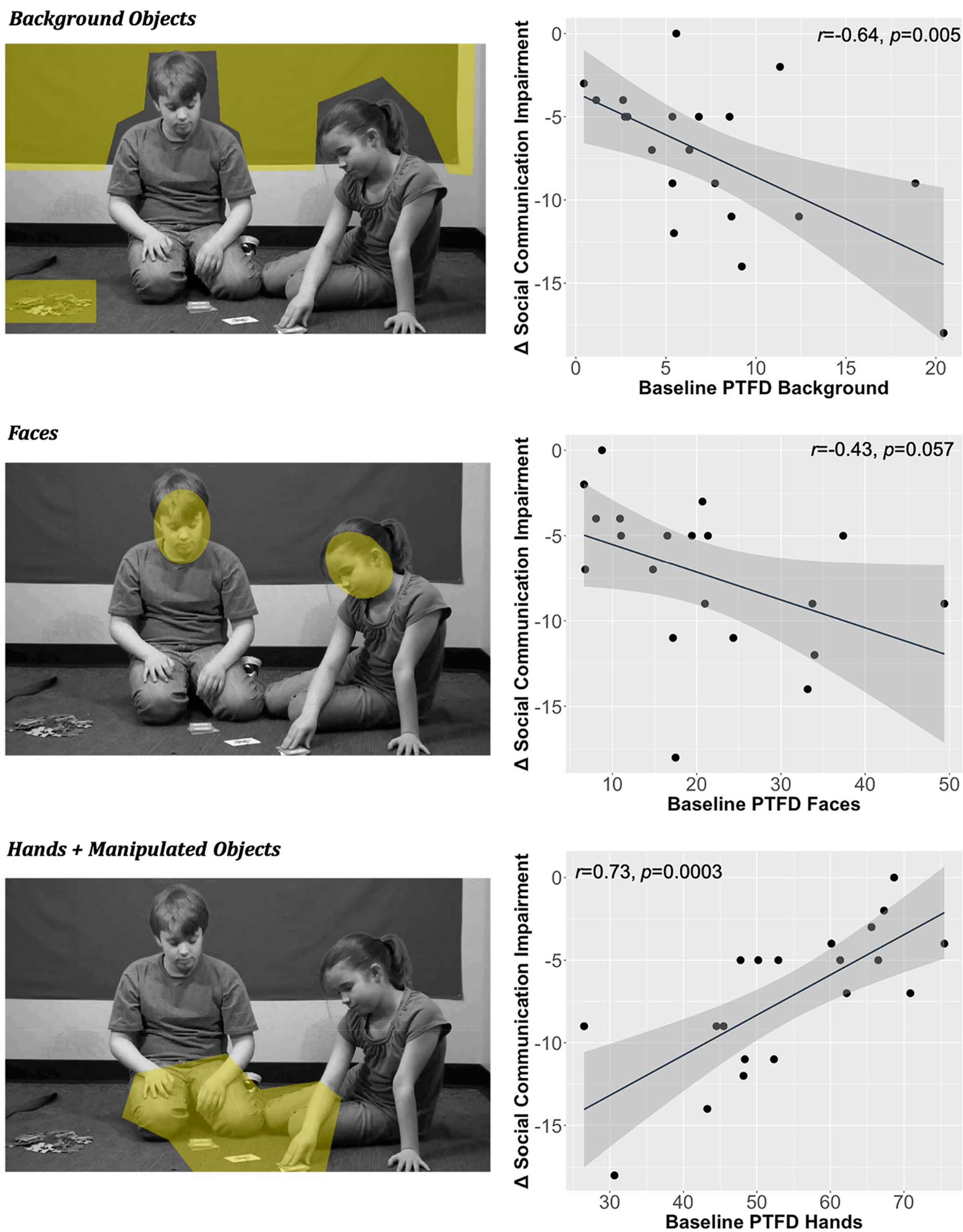
Independent t-test analyses were conducted to examine group differences in SRS-2 scores between pre- and post-treatment.

Results

Test–Retest Reliability

Intraclass Correlations

The two-way mixed-effects ICC models were conducted based on a mean-rating ($k = 2$). Results indicated good reliability for the social prioritization score (ICC = 0.801,


Background Objects
r=-0.64, p=0.005
Δ Social Communication Impairment
Baseline PTFD Background
0
-5
-10
-15
0 5 10 15 20
Faces
r=-0.43, p=0.057
Δ Social Communication Impairment
Baseline PTFD Faces
10 20 30 40 50
Hands + Manipulated Objects
r=0.73, p=0.0003
Δ Social Communication Impairment
Baseline PTFD Hands
30 40 50 60 70

95% CI [0.348, 0.927]), PTFD faces (ICC = 0.859, 95% CI [0.407, 0.952]), and PTFD background (ICC = 0.790, 95% CI [0.489, 0.913]). Moderate reliability was observed for PTFD hands (ICC = 0.671, 95% CI [0.181, 0.866]).

Cronbach's Alpha

Both the social prioritization score (α = 0.859) and PTFD faces measure (α = 0.910) were found to have very good reliability. The PTFD background (α = 0.783) and PTFD hands (α = 0.733) measures both showed acceptable reliability.

Correspondence Between Baseline Eye Tracking Metrics and Measures of Social Functioning in the ASD Groups

Contrary to hypotheses, there were no significant correlations between any of the four eye tracking measures (i.e., social prioritization, PTFD faces, PTFD background, or PTFD hands) and the other measures of social functioning and social cognition at baseline. Additionally, there were no significant group differences (i.e., between SCIT-A and TAU) in gaze metrics at baseline.

Eye Tracking as a Predictor of Social Improvement Immediately Following SCIT-A

Heightened visual attention to faces (i.e., PTFD faces) at baseline was associated with improved social awareness ($r(18) = -0.46$, $p = 0.04$) and marginally improved social communication ($r(18) = -0.43$, $p = 0.057$) immediately following SCIT-A intervention (see Fig. 2). Similarly, increased visual attention to background objects (i.e., PTFD background) at baseline was associated with improved social communication scores ($r(18) = -0.64$, $p = 0.005$) and restricted interests and repetitive behaviors ($r(18) = -0.55$, $p = 0.01$), as well as overall social functioning (SRS-2 total score; $r(18- = -0.51$, $p = 0.02$). Alternatively, heightened visual examination of hands and objects manipulated by the hands at baseline was associated with less improvement in social communication ($r(18) = 0.73$, $p = 0.0003$), restricted interests and repetitive behaviors ($r(18) = 0.62$, $p = 0.004$), and overall social functioning (SRS-2 total score; $r(18) = 0.71$, $p = 0.0004$) following SCIT-A intervention. There were no significant associations, however, between the social prioritization eye tracking metric at baseline and changes in SRS-2 pre- and post-treatment.

Eye Tracking as a Treatment Outcome Measure

Concordance Between Eye Tracking and Demographic and Treatment Variables Over Time in the ASD Groups

Results from a multilevel model (MLM) 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 ($\beta = 3.23$, $t = 2.68$, $p = 0.011$). Neither the post-treatment visit nor the follow-up visit were significant predictors of social prioritization, indicating eye tracking metrics for all participants remained stable over time compared to baseline. A contrast between post-treatment and follow-up also showed no significant change ($\beta = -0.85$, $t = -0.15$, $p = 0.75$). Full-scale IQ, age at baseline, sex, treatment group, and current participation in therapy similarly did not significantly impact social prioritization scores across study visits.

Group Differences in Eye Tracking and Social Impairment Over Time in the ASD Groups

Additional MLM analyses revealed a significant two-way interaction between follow-up visit and ER-40 performance on social prioritization scores ($\beta = 1.60$, $t = 2.40$, $p = 0.02$), such that individuals with higher ER-40 scores (i.e., higher emotion recognition abilities) had higher social prioritization at the follow-up visit (Fig. 3). Additionally, a significant three-way interaction on social prioritization scores between follow-up visit, ER-40 scores, and treatment group ($\beta = -1.19$, $t = -2.64$, $p = 0.011$) revealed that this relationship was only true for the SCIT-A group. Models including the remaining measures (i.e., SRS-2 total score and Hinting Task) were not significant, all $ps > 0.05$.

Treatment Effects as Measured by SRS-2 in the ASD Groups

Group differences in SRS-2 total scores were observed from baseline to post-treatment ($\beta = 13.33$, $t = 3.80$, $p = 0.0003$) and baseline to follow up ($\beta = 11.52$, $t = 3.19$, $p = 0.002$), such that individuals who participated in SCIT-A showed significant reductions in overall social impairment immediately following the 8-week social skills group and then again at 2-month follow up visit. Participants in the TAU group did not show these changes.

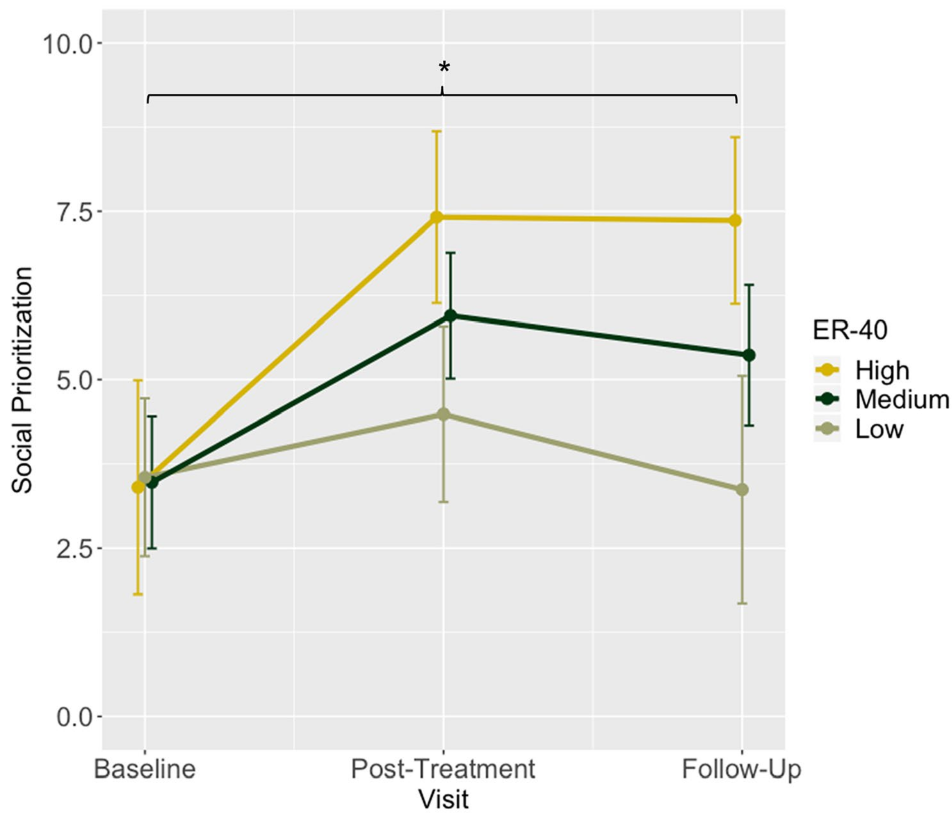


Fig. 3 *p < .05. Error bars represent one standard error (SE) of the estimate. ER-40 = Penn Emotion Recognition Test. ER-40 Low = 1 SE below the mean, ER-40 Medium = mean, ER-40 High = 1 SE above the mean. Emotion recognition serves as a significant moderator of social prioritization of visual stimuli at follow-up. This graph solely visualizes social prioritization estimates within the SCIT-A group.

Discussion

The current study evaluated a dynamic eye tracking task as both a predictor and an outcome measure of treatment response in ASD.

Test–Retest Reliability

The social prioritization score obtained from the eye tracking task showed good to very good test–retest reliability in a TDC sample with two administrations separated by approximately 24 h. This indicates that fluctuations in these scores over time may be primarily attributed to changes in visual attention to social stimuli rather than to poor reliability of the measure. Although the eye tracking measure showed good-test retest reliability in a TDC sample, future studies should examine the psychometric properties of this task in an ASD sample, as well.

Correspondence Between Baseline Eye Tracking Metrics and Measures of Social Functioning in the ASD Groups

Although it was hypothesized that eye tracking measures would correspond with other measures of social cognitive functioning, the social prioritization score used by Chevallier et al. (2015) did not significantly correlate with other measures of social functioning at baseline in the current sample. These findings were inconsistent with original hypotheses, and suggest that the constructs measured by the eye tracking metrics are distinct from the symptom measures and neurocognitive assessments used here, which evaluated emotion recognition, theory of mind, and broader social functioning. This finding does not necessarily indicate that the eye tracking measure does not capture the social abilities and social preferences of an individual, but, rather, that the eye tracking task may measure a unique component of social functioning that is not measured by these other commonly used questionnaires and neurocognitive measures. Additionally, the current sample included adolescents and young adults, whereas the (Chevallier et al. 2015) study examined eye gaze in children. Therefore, the differing results between these studies may reflect developmental effects.

Eye Tracking as a Predictor of Social Improvement Immediately Following SCIT-A

Increased gaze to faces and background objects within the stimulus at baseline was associated with improved social communication and interaction in individuals with ASD who participated in the social skills intervention. In contrast, increased percent time spent looking at the hands or objects manipulated by the hands was associated with decreased therapeutic benefit. These findings align with original hypotheses that individuals who demonstrate greater visual preference for faces and decreased visual preference for hands would show the most improvement following intervention, and are consistent with recent findings that the IVE task differentiated individuals with ASD from their TDC peers based on the amount of time spent looking at hands and toys during social, rather than non-social or parallel, interactions (Parish-Morris et al. 2019). However, it was not hypothesized that increased gaze time to background objects would also be associated with social communication improvements. The SCIT-A curriculum teaches individuals not only to examine facial expressions and body language to understand social intent but also to scan the surrounding environment for socially-relevant clues. For example, one session teaches participants to consider whether they should approach a group of strangers seated in the center of a public space versus a group convening privately behind a closed door. Therefore, it follows that a broader visual scanning of the environment may be a necessary component of social understanding and improved social outcomes for individuals with ASD. Overall, these results suggest that metrics from this dynamic eye tracking stimulus are sensitive predictors of social skills treatment response.

Eye Tracking as a Treatment Outcome Measure

Concordance Between Eye Tracking and Participant Characteristics and Treatment Variables Over Time in the ASD Groups

We examined relations between eye gaze preferences and demographic (i.e., age, IQ) and treatment (e.g., psychiatric medication and behavioral intervention status) variables across all three timepoints. Individuals taking psychotropic medication were more likely to prioritize social stimuli in the eye tracking paradigm than those not taking psychotropic medications. This held true for all participants with ASD regardless of treatment group. Future research should examine this association more closely to see if it is replicated in another sample and to determine whether specific medications could increase attention to social stimuli. Given that the medication variable was recorded as a binary measure (i.e., present or absent), future research should examine the unique effects of various drug classes as well as the utility of this eye tracking task as an outcome measure in ASD medication trials.

Full-Scale IQ approached significance in these multilevel model analyses ($p < 0.08$), indicating that individuals with higher cognitive ability may focus more on social versus non-social visual stimuli in the eye tracking task. This may have implications for the potential adaptation of the interactive visual exploration (IVE) eye tracking task to cognitively impaired populations. Additionally, previous studies have shown that individuals with ASD with higher cognitive and verbal abilities show greater gains during group-based social skills interventions (Solomon et al. 2004). In contrast, age was not a significant predictor of social prioritization in the longitudinal model, suggesting that age does not influence this measure of visual social attention. However, future studies should investigate the use of this eye tracking paradigm in younger individuals and individuals with significant developmental and cognitive delays.

Group Differences in Eye Tracking and Social Impairment Over Time in the ASD Groups

The longitudinal results did not indicate any significant group differences in visual social attention from baseline to post-treatment across both ASD groups, nor did they reveal group differences in gaze preference for social stimuli over time. This could mean that the IVE task is particularly stable and, thus, insensitive to changes over short periods of time. Although group differences in SRS-2 parent-report ratings were observed across visits, suggesting social improvement following SCIT-A intervention, this same result was not found when using the eye tracking metric as the outcome variable. This may indicate that the SRS-2 detected subtler or more immediate changes in social competence than the eye tracking measure. However, the observed SRS-2 scores could also be impacted by reporter bias, as participants and their families were not blind to treatment group membership.

Finally, although measures of social functioning and social cognition did not significantly relate to visual social prioritization at baseline, MLM analyses found that baseline emotion recognition abilities did have a significant relationship with this eye tracking metric at the follow-up visit: the emotion recognition task (ER-40) showed significant interactions with the increase in social prioritization from baseline to follow-up, as well as significant interactions between follow-up visit and treatment group. Figure 3 visualizes the moderation effect of emotion recognition abilities on the prioritization of visual social information within the eye tracking task for the SCIT-A group. Additionally, there was a significant three-way interaction between follow-up visit, 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 suggest that having greater knowledge of the information that can be gained from examining facial expressions may prompt individuals to visually seek out faces to better understand a social situation. Similar associations between emotion recognition ability and preferential gaze to social stimuli in individuals with ASD have been previously reported (Wieckowski and White 2020). Together, this suggests the clinical utility of teaching emotion recognition abilities to increase gaze preference for social stimuli, potentially creating a cascading effect on social skills progress.

Future Directions

Based on the current findings, future studies may investigate eye tracking as a baseline measure of social attention used to individualize social skills treatment protocols based on initial social gaze preference and correspondence with unique behavioral outcomes. Additionally, a future line of inquiry should integrate eye tracking with other behavioral social skills assessments pre- and post-intervention, as well as mid-treatment, to track response to treatment in individuals with ASD. More broadly, the utility of this eye tracking measure as a treatment outcome measure should be further evaluated in larger samples receiving a variety of treatments and evaluate: (1) test–retest reliability, (2) sensitivity to treatment effects, (3) relationship to global functioning, (4) practicality, (5) tolerability, (6) sensitivity to group differences, and (6) internal consistency.

Limitations

The current study was limited by the non-randomization of the study groups. Additionally, the groups differed on a number of demographic variables and this was accounted for within the statistical analyses. Finally, future studies with larger sample sizes should be conducted to replicate these findings.

Conclusions

In conclusion, this dynamic eye tracking measure shows good to very good test–retest reliability within a control sample. Additionally, increased gaze to faces and background objects, as well as decreased visual preference for hands and objects manipulated by hands, was associated with greater social improvement following a social skills intervention. However, changes in eye tracking metrics did not differentiate between those enrolled in the SCIT-A social skills treatment and those receiving treatment as usual within a community with significant ASD treatment resources. Additionally, the prioritization of social stimuli, as measured by the eye tracking task, showed no significant correlations with questionnaires and neurocognitive measures of social ability at baseline, suggesting that eye tracking may index unique aspects of social attention or ability relative to other measures. The eye tracking measure did, however, prove useful for predicting improvement following social skills intervention. Predicting responses to ASD interventions is an urgent public health need, and these results suggest that eye tracking may hold promise as a predictor of ASD treatment outcomes. Emotion recognition abilities also proved to be a significant moderator of visual social prioritization over time. This finding reiterates the importance of teaching emotion recognition skills to individuals with ASD. The measure's good test–retest reliability, predictive ability, and association with emotion recognition ability warrant future studies to examine the utility of eye tracking as a predictor and an outcome measure of treatment response to interventions that target social skills for individuals with ASD.

Acknowledgments The authors would like to thank the families who participated in this research. Support for this project was provided by the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, through Grant Award Number 2KR631405. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. GSD was supported by U54 HD079124 and MH110933. JLK was supported by T32-HD40127. Data analytic support was provided by the University of North Carolina at Chapel Hill Odum Institute.

References

Arnold, L. E., Vitiello, B., McDougle, C., Scahill, L., Shah, B., Gonzalez, N. M., et al. (2003). Parent-defined target symptoms respond to risperidone in RUPP autism study: Customer approach to clinical trials. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*(12), 1443–1450.…

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.

Bird, G., Press, C., & Richardson, D. C. (2011). The role of alexithymia in reduced eye-fixation in autism spectrum conditions. *Journal of Autism and Developmental Disorders, 41*(11), 1556–1564. https://doi.org/10.1007/s10803-011-1183-3.

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

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.

Charter, R. A., & Feldt, L. S. (2001). meaning of reliability in terms of correct and incorrect clinical decisions: The art of decision making is still alive. *Journal of Clinical and Experimental Neuropsychology, 23*(4), 530–537.

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., Parish-Morris, J., McVey, A., Rump, K., Sasson, N. J., Herrington, J., et al. (2015). Measuring social attention and motivation in Autism Spectrum Disorder using eye-tracking: Stimulus type matters. *Autism Research, 8*(5), 620–628. https://doi.org/10.1002/aur.1479.

Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., et al. (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. (2000). Genetic structure of reciprocal social behavior. *American Journal of Psychiatry, 157*(12), 2043–2045.

Corcoran, R., Mercer, G., & Frith, C. D. (1995). Schizophrenia, symptomatology and social inference: Investigating "theory of mind" in people with schizophrenia. *Schizophrenia research, 17*(1), 5–13.

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

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.

Guillon, Q., Hadjikhani, N., Baduel, S., & Rogé, B. (2014). Visual social attention in autism spectrum disorder: Insights from eye tracking studies. *Neuroscience & Biobehavioral Reviews, 42*, 279–297. https://doi.org/10.1016/j.neubiorev.2014.03.013.

Gur, R. C., Sara, R., Hagendoorn, M., Marom, O., Hughett, P., Macy, L., et al. (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.…

Hanley, M., McPhillips, M., Mulhern, G., & Riby, D. M. (2013). Spontaneous attention to faces in Asperger syndrome using ecologically valid static stimuli. *Autism, 17*(6), 754–761.

Hill, E., Berthoz, S., & Frith, U. (2004). Brief report: Cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives. *Journal of Autism and Developmental Disorders, 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. *Journal of American Academics and Child Adolescent Psychiatry, 52*(6), 572–581. https://doi.org/10.1016/j.jaac.2013.02.017.

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

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.

Kanne, S. M., Mazurek, M. O., Sikora, D., Bellando, J., Branum-Martin, L., Handen, B., et al. (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.…

King, B. H., Hollander, E., Sikich, L., McCracken, J. T., Scahill, L., Bregman, J. D., et al. (2009). Lack of efficacy of citalopram in children with autism spectrum disorders and high levels of repetitive behavior: Citalopram ineffective in children with autism. *Archives of General Psychiatry, 66*(6), 583–590.…

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163.

Kuhn, G., Kourkoulou, A., & Leekam, S. R. (2010). How magic changes our expectations about autism. *Psychological Science, 21*(10), 1487–1493.

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.

Lord, C., Rutter, M., & Le Couteur, 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*(5), 659–685.

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.

Nakano, T., Tanaka, K., Endo, Y., Yamane, Y., Yamamoto, T., Nakano, Y., et al. (2010). Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour. *Proceedings of the Royal Society B: Biological Sciences, 277*(1696), 2935–2943.…

Pallathra, A. A., Day-Watkins, J., Calkins, M. E., Maddox, B., Miller, J., Parish-Morris, J., et al. (2018). *Improvement in social functioning following participation in TUNE In, a novel cognitive-behavioral treatment program-results from a 2nd cohort of adults with ASD*. Rotterdam, Netherlands: Paper presented at the International Society for Autism Research.…

Parish-Morris, J., Pallathra, A. A., Ferguson, E., Maddox, B. B., Pomykacz, A., Perez, L. S., et al. (2019). Adaptation to different communicative contexts: An eye tracking study of autistic adults. *Journal of Neurodevelopmental Disorders, 11*(1), 5. https://doi.org/10.1186/s11689-019-9265-1.…

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. *Expert Review of Pharmacoeconomics & Outcomes Research, 12*(4), 485–503.

Riby, D. M., Brown, P. H., Jones, N., & Hanley, M. (2012). Brief report: Faces cause less distraction in autism. *Journal of Autism and Developmental Disorders, 42*(4), 634–639.

Riby, D. M., & Hancock, P. J. B. (2008). Viewing it differently: Social scene perception in Williams syndrome and autism. *Neuropsychologia, 46*(11), 2855–2860. https://doi.org/10.1016/j.neuropsychologia.2008.05.003.

Riby, D. M., & Hancock, P. J. B. (2009). Do faces capture the attention of individuals with Williams syndrome or autism? Evidence from tracking eye movements. *Journal of Autism and Developmental Disorders, 39*(3), 421–431. https://doi.org/10.1007/s10803-008-0641-z.

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: Oxford University Press.

Sasson, N. J., Turner-Brown, L. M., Holtzclaw, T. N., Lam, K. S., & Bodfish, J. W. (2008). Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays. *Autism Research, 1*(1), 31–42.

Schopler, E., Van Bourgondien, M., Wellman, J., & Love, S. (2010). *Childhood autism rating scale—Second edition (CARS2): Manual*. Los Angeles: Western Psychological Services.

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. https://doi.org/10.1016/j.brainres.2010.11.074.

Skinner, I. W., Hübscher, M., Moseley, G. L., Lee, H., Wand, B. M., Traeger, A. C., et al. (2018). The reliability of eyetracking to assess attentional bias to threatening words in healthy individuals. *Behavior Research Methods, 50*(5), 1778–1792. https://doi.org/10.3758/s13428-017-0946-y.…

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.

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. *Journal of Autism and Developmental Disorders, 38*(9), 1777–1784. https://doi.org/10.1007/s10803-008-0545-y.

Ursachi, G., Horodnic, I. A., & Zait, A. (2015). How reliable are measurement scales? External factors with indirect influence on reliability estimators. *Procedia Economics and Finance, 20*(15), 679–686.

Wieckowski, A. T., & White, S. W. (2020). Attention modification to attenuate facial emotion recognition deficits in children with autism: A pilot study. *Journal of Autism and Developmental Disorders, 50*(1), 30–41.