Computational Measurement of Social Gaze During Naturalistic Conversations in Autism

Lisa D. Yankowitz*[1], Mohan Kashyap Pargi[1], Ellis DeJardin[1], Casey J. Zampella[1], Whitney Guthrie[1,2], Juhi Pandey[1,2], G. Keith Bartley[1], Darren Chen[1], Denisa Q. McDonald[1], Aashvi Manakiwala[1], Maya Khanna[1], Kelsey Keen[1], Gabriella Buboltz[1], Annie Yang[1], John D. Herrington[1,2], Evangelos Sariyanidi[1], Robert T. Schultz[1,2], Birkan Tunç[1,2]

[1] Center for Autism Research, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

[2] University of Pennsylvania, Philadelphia, PA 19104, USA.

* Correspondence to: Lisa Yankowitz, PhD

Email: yankowitzl@chop.edu

Roberts Center for Pediatric Research

5th Floor, Office 5202

2716 South Street, Philadelphia, PA 19146

ABSTRACT

Standardized, granular measurement of autistic behaviors, such as social gaze during interactions, is needed for a range of clinical applications including diagnosis and detecting clinical change. Computational approaches show promise in automatically measuring social behaviors within natural settings. This study aims to automatically measure social gaze features from videos of dyadic conversations, characterize autism-related differences, and perform individual-level diagnostic classification. 46 autistic Participants and 36 neurotypical Participants, aged 8-29 years, engaged in a brief video-recorded conversation with a research staff member (Partner). A deep learning AI model trained to detect whether each partner was looking at the other achieved 89% cross-validated accuracy. Comparing these automatic gaze measurements, autistic Participants spent less time looking at Partners and engaging in mutual gaze than neurotypical Participants did. They also initiated mutual gaze less frequently and had shorter mutual gaze episodes, but did not differ in mutual gaze counts. An AI-derived social gaze summary score correlated specifically with ADOS-2 Social Affect scores and not Restricted and Repetitive Behavior scores. Cross-validated machine learning using gaze features predicted diagnostic group with 73% accuracy. This study provides a framework for automatically quantifying social gaze behaviors, with potential for enhancing diagnostic precision and tracking therapeutic progress in autism.

Keywords: autism, computer vision, artificial intelligence, social gaze, eye gaze

BACKGROUND

There is a pressing need for scalable biomarkers of autism spectrum disorder (AUT) for diagnosis, monitoring developmental or treatment-related change, parsing heterogeneity, and understanding genetic variants. One promising path forward is to use computational approaches such as artificial intelligence (AI) for automatically and precisely measuring social behaviors within natural, everyday settings where difficulties associated with autism are most salient. One key social behavior that is particularly ripe for AI approaches is eye gaze (Hou et al., 2024; Keehn et al., 2024; Riddiford et al., 2022; Wei et al., 2023). Differences in eye contact are part of core diagnostic criteria (American Psychiatric Association, 2013), and atypical social gaze patterns are among the most widely studied behavioral biomarkers of AUT across the lifespan. Autistic children and adults exhibit atypical social gaze patterns cross-culturally (Ma et al., 2021), including delayed development of joint attention (Koegel et al., 2009), fixation on the nose and mouth rather than the eyes (Klin et al., 2002; Senju & Johnson, 2009; Vacas et al., 2021), reduced fixation on eye regions (Frazier et al., 2017; Papagiannopoulou et al., 2014), and active or unconscious avoidance of eye contact (Kliemann et al., 2012; Madipakkam et al., 2017; Schultz et al., 2000). Differences in social gaze are also one of the earliest signs of autism. In children with elevated familial likelihood for autism, reduced eye contact emerges as early as 6 months (Jones & Klin, 2013; Ozonoff et al., 2014). A fully automated eye-tracking device recently demonstrated sensitivities and specificities of 71-90% in predicting AUT in clinic-based samples of toddlers referred for evaluation (Jones, Klaiman, Richardson, Aoki, et al., 2023; Jones, Klaiman, Richardson, Lambha, et al., 2023), highlighting the potential utility of social gaze as a diagnostic biomarker integrated into clinical practice.

In the lives of autistic people, differences in social gaze are more nuanced and variable than a global reduction in eye contact and unfold within real-world social interactions. Prior studies using wearable eye-tracking devices have demonstrated high accuracy in measuring moment-by-moment eye gaze during semi-naturalistic interactions (Celiktutan et al., 2023; Chong et al., 2020; Li et al., 2018; Shell et al., 2004; Ye et al., 2012, 2015), but such equipment might alter the nature of the interaction while also limiting data collection to a single interaction partner (i.e., the participant). Although bidirectional social gaze can be collected using Zoom-like online conversation paradigms (Ross et al., 2023), it is not yet known whether this would be generalizable to real-life face-to-face interactions, so methods to accurately quantify social gaze in such naturalistic interactions are needed.

Recent developments in AI, specifically in computer vision (Chong et al., 2020), promise the development of noninvasive and scalable approaches for tracking social gaze during face-to-face social interactions that can be used with individuals who cannot tolerate wearables (Alvari et al., 2021; Guo et al., 2021, 2024). Several studies using cameras capturing a side view of interactions have estimated social gaze using computer vision, based primarily on head pose and orientation. These estimates have shown relationships with social ability (Alvari et al., 2021), experimental task(Guo et al., 2024), therapist-coded social gaze behavior(Guo et al., 2021), clinically meaningful subgroups (Alvari et al., 2021), and autism diagnosis (Celiktutan et al., 2023). While head-based estimates have proven useful, incorporating eye movement information is crucial for improving the accuracy and validity of social gaze estimations.

In this study, we used a dyadic data collection paradigm that can capture the social gaze behavior of each interaction partner separately using standard cameras facing each person. By leveraging an advanced AI model to predict social gaze, we demonstrate, for the first time, that it

is possible to automatically and unobtrusively quantify dyadic patterns in social gaze behavior in naturalistic face-to-face interactions. We test how social gaze differed between autistic (AUT) and neurotypical (NT) 8–29-year-olds to characterize the manifestation of social gaze differences associated with AUT at a level of detail and granularity not previously possible. Finally, we use aspects of social gaze alone to classify diagnostic status (AUT versus NT) in a cross-validated framework.

METHODS

Participants

Participants (AUT $n = 46$ [39 male]; NT $n = 36$ [21 male]) were drawn from a larger sample who participated in studies at [details omitted for double-anonymized peer review]. Groups were matched on age, sex ratio, and IQ. Demographic data is presented in Table 1. Autism diagnoses were confirmed through the best clinical judgment of a licensed psychologist using all available information, including administration of the Autism Diagnostic Observation Schedule (ADOS-2), Module 3 or 4 (Lord et al., 2012). Inclusion and exclusion criteria are provided in Supplementary Methods.

[TABLE-1]

Study Procedure

Participants completed the Contextual Assessment of Social Skills(Ratto et al., 2011), a brief 3-4 minute semi-structured "get-to-know-you" conversation with a member of the research staff ("Partner"). Partners were research assistants or students from the lab, assigned based on availability, whom the Participant had not previously met. Participants and Partners were seated across from each other with two video cameras placed in between to record synchronized videos of each person at 30 frames per second (see Figure 1A). Partners were instructed to appear

interested and engaged but not carry the conversation (i.e., speak no more than 50% of the time and wait 5 seconds to re-initiate the conversation after a lapse). After the conversation, the Partner completed a conversation rating scale, which included an item rating the Participant's eye contact on a 1-7 scale ("The other person made appropriate eye contact with me during conversation").

[FIGURE 1]

Annotation of Gaze

Human coders blind to Participant diagnosis annotated social gaze from the videos to create a ground-truth labeled dataset for training the AI algorithm. Each video recording was segmented into 1-second clips, and the last frame of each clip was annotated (see Supplementary Figure 1 for a screenshot of the annotation platform, which raters viewed in a web browser on a monitor). Each clip showed the frontal images of the Participant and Partner side-by-side and was played in a continuous loop with the last frame frozen for one second. Two raters made a yes/no judgment of whether each member of the dyad was looking at the other person, with disagreements resolved by a third rater. We operationalize this as 'social gaze' and not eye contact, as raters were not asked to judge whether the person was looking specifically at the other person's eyes. Reliability between the first two ratings was excellent for both Participants ($\kappa = 0.80$) and Partners ($\kappa = 0.83$).

Gaze Detection Algorithm

We used a convolutional neural network (CNN) model originally developed for deciding whether the person in a photo is looking at the camera or not (Zhang et al., 2022). The model automatically detects the eye region within an image and learns patterns of interactions between the eye region and other regions of the image. We used the same architecture to detect social

gaze by training the model with the gold-standard human annotations (see Supplementary Methods for full details). Notably, the algorithm incorporates information about both head orientation and eye gaze direction; see Figure 1B and C for an example of detection. The performance of the algorithm was computed using ten-fold cross-validation where folds were defined with respect to people, not frames; that is, all frames of a person were used only in training or testing, but not both. The trained algorithm was then used to detect "gaze" at every frame of the video recordings, including non-annotated frames, yielding frame-by-frame binary (yes/no) social gaze data for the Participant and Partner, which were used to calculate variables of interest.

Social Gaze Features

Using the AI-generated frame-by-frame binary social gaze data, we defined 18 social gaze features (see Table 2 for definitions, Supplementary Methods for full details). Four were proportions: Participant gaze (proportion of time the Participant is gazing at the Partner), Partner gaze (proportion of time the Partner is gazing at the Participant), mutual gaze (proportion of time both are gazing at each other), and neither gaze (proportion of time neither are gazing at the other). Discrete mutual gaze events were defined, from which we calculated mutual gaze number, proportion of Participant initiations, and the minimum, maximum, mean, and standard deviation of: mutual gaze duration, delay to return gaze, and wait time. Group differences were assessed for each of these features (using only the mean value for the latter three), yielding nine variables. For the machine learning analysis, we used feature selection on all 18 features (plus conversation length) to eliminate redundant features (correlation ≥ 0.7 between features), yielding nine non-redundant variables (see Supplementary Methods and Supplementary Table 1).

A summary social gaze score was calculated from the nine non-redundant variables using the PUNCH algorithm (Tunc et al., 2014) to represent overall social gaze behavior associated with AUT. PUNCH is a decision-level fusion algorithm that assigns interpretable weights to a set of scores (here, the nine variables) to best discriminate between two groups (here, AUT and NT). The feature weights produced by PUNCH are then applied to create a social gaze summary score.

Statistical Analysis and Machine Learning

Conversation length between diagnostic groups and looking time of Partners versus Participants were compared using Wilcoxon rank sum and signed rank tests to account for non-normality. We examined diagnostic differences in social gaze using a linear or generalized linear model for each of the nine variables, using a systematic model-building approach (see Supplementary Methods for details). The final models are presented in Table 2. All models included the main effects of diagnosis (AUT or NT) and age. Although the diagnostic groups were matched on sex ratio and IQ, these were included as nuisance covariates in all models to ensure they did not drive results. All models also controlled for conversation length and Partner gaze (except the model predicting Partner gaze). False Discovery Rate correction used the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) across the nine models.

A Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel (Han et al., 2012) was used for classification. We used nested ten-fold cross-validation to train and test the algorithm, with all model parameters for SVM (C and γ) optimized within the inner cross-validation only using the training data (additional details in Supplementary Methods).

RESULTS

Automated Detection of Social Gaze

Our AI-based automated social gaze detection algorithm achieved high cross-validated performance in detecting gaze annotations with 89% accuracy, 87% specificity, 89% sensitivity, 77% negative predictive value (NPV), and 94% positive predictive value (PPV). Partner ratings of Participant eye contact (AUT $n = 32$, $M = 3.97$, $SD = 2.04$; NT $n = 21$, $M = 5.86$, $SD = 1.11$) provided after the conversation were correlated with the AI-detected proportion of time that the Participant engaged in social gaze ($r = 0.43$, $p < 0.01$), but were not correlated with the time the Partner engaged in social gaze ($r = -0.08$, $p = 0.58$).

[FIGURE 2]

Individual Social Gaze Behavior of Participants and Partners

The conversation length did not differ significantly between autistic and neurotypical Participants (Wilcoxon rank sum $W = 895.5$, $p = 0.53$). Partners, on average, spent a higher proportion of time looking at Participants than Participants looked at Partners (Wilcoxon signed-rank test $V = 3006$, $p < 0.001$), with a mean proportion of 0.86 *vs.* 0.61, likely reflecting the Partners' instruction to appear interested and engaged. Partner social gaze, on average, was lower during conversations with autistic Participants than with neurotypical Participants (Cohen's d (unadjusted for covariates) = -0.37, $q < 0.05$, see Table 2, Figure 2A). Autistic Participants' proportion of time engaged in a social gaze toward Partners was significantly lower compared to NT Participants ($d = -0.66$, $q < 0.01$, Figure 2B).

Mutual Social Gaze Patterns

Consistent with individual gaze behavior, the proportion of time in mutual social gaze was lower in dyads with an autistic Participant ($d = -0.88$, $q < 0.01$, Figure 2C). Conversely, the proportion of time in which neither person was gazing at the other was, on average, higher with autistic Participants ($d = 0.55$, $q < 0.01$, Figure 2D). The number of mutual social gaze episodes

did not significantly differ between autistic and neurotypical Participants (d = -0.11, q = 0.26, Figure 2E), but mutual social gaze episodes were, on average, shorter with autistic Participants (d = -0.79, q < 0.01, Figure 2F). Either the Participant or Partner can initiate a mutual social gaze; autistic Participants initiated a lower proportion of mutual gaze episodes than neurotypical Participants (d = -0.28, q < 0.01, Figure 2G). When the Partner initiated mutual gaze, the average delay before the Participant joined the gaze was longer for autistic Participants (d = -0.28, q < 0.01, Figure 2H). When the Participant initiated mutual gaze, the delay before the Partner joined was similar for autistic and NT Participants (d = 0.25, q = 0.66, Figure 2I).

Machine Learning Classification of Diagnosis

A ten-fold cross-validated SVM classifier trained on social gaze features achieved 73% accuracy in predicting diagnostic labels (AUT versus NT). The model achieved 85% sensitivity, 58% specificity, 72% PPV, and 75% NPV.

Social Gaze Summary Score

Within the AUT group, the social gaze summary score was positively correlated with the ADOS-2 Social Affect Calibrated Severity Scores (CSS) (r = 0.47, p < 0.01, Supplementary Figure 2). In contrast, there was no relationship between the social gaze summary score and ADOS-2 Restricted and Repetitive Behavior CSS (r = 0.12, p = 0.54), indicating the social gaze summary score is specifically related to social skills rather than repetitive behaviors.

[TABLE 2]

Developmental Changes in Social Gaze

Unique age-related differences were observed across different features of social gaze (Figure 3). An age-by-diagnosis interaction was significant for the number of mutual gaze episodes, with NT Participants showing more mutual gaze episodes with age, while AUT

Participants did not show age-related changes. Across the full sample, the number of mutual gaze episodes initiated by the Participant showed a quadratic relationship with age, increasing across the teen years until its peak at 20.8 years and decreasing thereafter. Partner gaze showed a linear decrease with Participant age, with Partners spending less time looking at older Participants. To explore whether the association between Partner gaze and Participant age was driven by the nonsignificant increase in Participant gaze with age, Participant gaze was added as a predictor in the Partner gaze model. Both Participant gaze and Participant age were significant predictors of Partner gaze, suggesting that each has a unique effect in reducing Partner gaze.

[FIGURE 3]

DISCUSSION

This study demonstrates the promise of automated, validated measurement of social gaze patterns during naturalistic conversations and its clinical relevance as a potential biomarker for AUT. Our novel approach derived detailed gaze metrics from both conversation partners, which accurately predicted human coding of gaze and correlated with perceptions of eye contact by the conversation Partner. Our approach detected widely reported reductions in social gaze in autism (Frazier et al., 2017; Papagiannopoulou et al., 2014; Riddiford et al., 2022), showed a reduction in the mutual social gaze shared by both members of the interacting dyad when the Participant was autistic, and provided new insights into the precise nature of social gaze reductions, including that they are driven by shorter durations of mutual gazes rather than fewer instances of mutual gazes.

AI-derived social gaze features predicted individual diagnostic status with 73% accuracy, which is comparable to studies that have established eye gaze-based biomarkers of autism in toddlers using specialized eye-tracking devices (Jones, Klaiman, Richardson, Aoki, et al., 2023;

Jones, Klaiman, Richardson, Lambha, et al., 2023). Notably, the accuracy of this single marker is also in the range of community diagnoses confirmed by gold standard evaluation (47-77%) (Duvall et al., 2024; Hausman-Kedem et al., 2018). Our social gaze features showed higher sensitivity (0.85) than specificity (0.58) for classifying autism, demonstrating that this marker is more useful for identifying AUT when it is present than for correctly ruling it out in neurotypical individuals. This pattern – higher sensitivity and lower specificity – is desirable in a biomarker for screening or triaging in specialty care clinics to identify as many potential cases as possible, in combination with other measures. An important next step will be to determine how well this approach can distinguish autism from other conditions for the clinical task of differential diagnosis, and to extend this approach to infants and toddlers to explore the utility of social gaze as an early biomarker of autism that can be targeted for early autism screening.

The ability to automatically quantify moment-by-moment social gaze revealed nuanced differences in social patterns associated with AUT, which would have been difficult to observe using more traditional methods. For example, while reduced mutual social gaze was observed in the AUT group, the number of discrete instances of mutual gazes was similar between groups. Clinically, this suggests that a diagnostic approach of counting instances of eye contact would not be fruitful in this context; rather, attending to the duration of mutual gaze episodes and the number of times an individual initiates mutual gaze may be more informative. Future work will harness other important advantages of AI, namely the ability to capture detailed temporal dynamics and multimodal integration to analyze the dynamics of social gaze over the course of the conversation and in relation to other modalities (*e.g.*, speech, facial expressions, gestures).

Social gaze is an inherently dyadic behavior, and these results demonstrate the importance of observing social gaze in the context of a dyadic interaction. Notably, Partners

looked less at autistic Participants despite instructions to appear engaged in all conversations. This finding, as well as the one mentioned previously – that autistic and NT Participants engaged in a similar number of mutual gaze episodes with Partners – could not be observed in a task in which participants do not have a live interactor who responds to their behavior. The ability to extract nuanced features from naturalistic, interactive scenarios using AI-based approaches is an important complement to more tightly controlled paradigms such as eye-tracking performed during computerized tasks (Frazier et al., 2017; Jones, Klaiman, Richardson, Aoki, et al., 2023; Jones & Klin, 2013).

The specific correlation between the "social gaze score," derived from the AI-based measures, and the ADOS-2 Social Affect (but not RRB) calibrated severity score is a clear indicator that our AI approach can yield a simple, interpretable score that aligns with well-established measures, thereby increasing the likelihood of clinical translation. A lack of this type of interpretability and clear relatability to clinical tools is one common criticism of novel AI tools purported to have utility for healthcare. The highly interpretable features produced by our AI approach – both the nuanced social gaze features and the social gaze summary score – hold promise for characterizing heterogeneity within autistic adolescents, which would have applications for precision medicine.

This cross-sectional study revealed nuanced and, in some cases, non-linear associations between social gaze and age, which warrant further exploration. It is notable that the number of mutual gaze episodes – one of the few social gaze features not to show a diagnostic effect – showed a diagnosis-by-age interaction, suggesting that while the number of mutual social gaze episodes may not differ in autism in childhood, differences may emerge by early adulthood due to more mutual gaze episodes among NT adults.

Limitations

In this work, we defined social gaze broadly as looking at the other person, without distinguishing between looking at the other person's mouth versus their eyes, which prior research has shown to differ in AUT (Klin et al., 2002; Senju & Johnson, 2009; Vacas et al., 2021). It is theoretically possible for a computer vision algorithm to detect true eye contact from these videos. However, generating gold-standard labels for training and validation presents a significant challenge without the inclusion of eye-tracking data, which was not collected in our sample. Notably, our approach was able to detect between-group and individual-level differences even without this level of facial region specificity, reflecting the large effect size of social gaze differences in autism and the fact that general social orienting and looking toward the face are clinically meaningful behavioral features. Another limitation is reliance on a specific data collection set-up (two people seated with cameras directly between them). Further study will be necessary to understand how robust this approach is to minor deviations in camera or seat placement and to develop computer vision methods that are more robust to such differences. An important next step will be to validate these results on interactions conducted over video conferencing, which would maximize scalability. Finally, we acknowledge the limitations of our sample, including the small sample size and exclusion of nonverbal Participants and non-English speakers. Future work will investigate how AI-derived gaze features differ in other age ranges (including toddlers) and in comparison with other clinical groups.

Conclusions

An AI approach applied to standard videos can accurately detect social gaze from each of two conversation partners, allowing for automatic and granular quantification of social gaze behavior in naturalistic contexts without specialized equipment. This approach advances

understanding of the nature of social gaze differences in autism. Accurate cross-validated prediction of diagnostic labels using social gaze features demonstrates the potential of this approach as a scalable biobehavioral marker of autism.

REFERENCES

Alvari, G., Coviello, L., & Furlanello, C. (2021). EYE-C: Eye-Contact Robust Detection and Analysis during Unconstrained Child-Therapist Interactions in the Clinical Setting of Autism Spectrum Disorders. *Brain Sciences*, *11*(12), Article 12. https://doi.org/10.3390/brainsci11121555

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. JSTOR.

Celiktutan, O., Wu, W., Vogeley, K., & Georgescu, A. L. (2023). A Computational Approach for Analysing Autistic Behaviour During Dyadic Interactions. In J.-J. Rousseau & B. Kapralos (Eds.), *Pattern Recognition, Computer Vision, and Image Processing. ICPR 2022 International Workshops and Challenges* (pp. 167–177). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-37660-3_12

Chong, E., Clark-Whitney, E., Southerland, A., Stubbs, E., Miller, C., Ajodan, E. L., Silverman, M. R., Lord, C., Rozga, A., Jones, R. M., & Rehg, J. M. (2020). Detection of eye contact with deep neural networks is as accurate as human experts. *Nature Communications*, *11*(1), 6386. https://doi.org/10.1038/s41467-020-19712-x

Duvall, S. W., Greene, R. K., Phelps, R., Rutter, T. M., Markwardt, S., Grieser Painter, J., Cordova, M., Calame, B., Doyle, O., Nigg, J. T., Fombonne, E., & Fair, D. (2024). Factors Associated with Confirmed and Unconfirmed Autism Spectrum Disorder Diagnosis in Children Volunteering for Research. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-024-06329-y

Frazier, T. W., Strauss, M., Klingemier, E. W., Zetzer, E. E., Hardan, A. Y., Eng, C., & Youngstrom, E. A. (2017). A Meta-Analysis of Gaze Differences to Social and Nonsocial Information Between Individuals With and Without Autism. *Journal of the American Academy of Child and Adolescent Psychiatry*, *56*(7), 546–555. https://doi.org/10.1016/j.jaac.2017.05.005

Guo, Z., Chheang, V., Li, J., Barner, K. E., Bhat, A., & Barmaki, R. L. (2024). Social Visual Behavior Analytics for Autism Therapy of Children Based on Automated Mutual Gaze Detection. *Proceedings of the 8th ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies*, 11–21. https://doi.org/10.1145/3580252.3586976

Guo, Z., Kim, K., Bhat, A., & Barmaki, R. (2021). An Automated Mutual Gaze Detection Framework for Social Behavior Assessment in Therapy for Children with Autism. *Proceedings of the 2021 International Conference on Multimodal Interaction*, 444–452. https://doi.org/10.1145/3462244.3479882

Han, S., Qubo, C., & Meng, H. (2012). Parameter selection in SVM with RBF kernel function. *World Automation Congress 2012*, 1–4. https://ieeexplore.ieee.org/abstract/document/6321759

Hausman-Kedem, M., Kosofsky, B. E., Ross, G., Yohay, K., Forrest, E., Dennin, M. H., Patel, R., Bennett, K., Holahan, J. P., & Ward, M. J. (2018). Accuracy of Reported Community Diagnosis of Autism Spectrum Disorder. *Journal of Psychopathology and Behavioral Assessment*, *40*(3), 367–375. https://doi.org/10.1007/s10862-018-9642-1

Hou, W., Jiang, Y., Yang, Y., Zhu, L., & Li, J. (2024). Evaluating the validity of eye-tracking tasks and stimuli in detecting high-risk infants later diagnosed with autism: A meta-analysis.

Clinical Psychology Review, *112*, 102466. https://doi.org/10.1016/j.cpr.2024.102466
Jones, W., Klaiman, C., Richardson, S., Aoki, C., Smith, C., Minjarez, M., Bernier, R., Pedapati, E., Bishop, S., Ence, W., Wainer, A., Moriuchi, J., Tay, S.-W., & Klin, A. (2023). Eye-Tracking-Based Measurement of Social Visual Engagement Compared With Expert Clinical Diagnosis of Autism. *JAMA*, *330*(9), 854–865. https://doi.org/10.1001/jama.2023.13295
Jones, W., Klaiman, C., Richardson, S., Lambha, M., Reid, M., Hamner, T., Beacham, C., Lewis, P., Paredes, J., Edwards, L., Marrus, N., Constantino, J. N., Shultz, S., & Klin, A. (2023). Development and Replication of Objective Measurements of Social Visual Engagement to Aid in Early Diagnosis and Assessment of Autism. *JAMA Network Open*, *6*(9), e2330145. https://doi.org/10.1001/jamanetworkopen.2023.30145
Jones, W., & Klin, A. (2013). Attention to Eyes is Present But in Decline in 2–6 Month-Olds Later Diagnosed with Autism. *Nature*, *504*(7480), 427–431. https://doi.org/10.1038/nature12715
Keehn, B., Monahan, P., Enneking, B., Ryan, T., Swigonski, N., & McNally Keehn, R. (2024). Eye-Tracking Biomarkers and Autism Diagnosis in Primary Care. *JAMA Network Open*, *7*(5), e2411190. https://doi.org/10.1001/jamanetworkopen.2024.11190
Kliemann, D., Dziobek, I., Hatri, A., Baudewig, J., & Heekeren, H. R. (2012). The Role of the Amygdala in Atypical Gaze on Emotional Faces in Autism Spectrum Disorders. *Journal of Neuroscience*, *32*(28), 9469–9476. https://doi.org/10.1523/JNEUROSCI.5294-11.2012
Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual Fixation Patterns During Viewing of Naturalistic Social Situations as Predictors of Social Competence in Individuals With Autism. *Archives of General Psychiatry*, *59*(9), 809–816. https://doi.org/10.1001/archpsyc.59.9.809
Koegel, R. L., Vernon, T. W., & Koegel, L. K. (2009). Improving Social Initiations in Young Children with Autism Using Reinforcers with Embedded Social Interactions. *Journal of Autism and Developmental Disorders*, *39*(9), 1240–1251. https://doi.org/10.1007/s10803-009-0732-5
Li, Y., Liu, M., & Rehg, J. M. (2018). In the Eye of Beholder: Joint Learning of Gaze and Actions in First Person Video. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision – ECCV 2018* (Vol. 11209, pp. 639–655). Springer International Publishing. https://doi.org/10.1007/978-3-030-01228-1_38
Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule, Second Edition: ADOS-2.* Western Psychological Services.
Ma, X., Gu, H., & Zhao, J. (2021). Atypical gaze patterns to facial feature areas in autism spectrum disorders reveal age and culture effects: A meta-analysis of eye-tracking studies. *Autism Research*, *14*(12), 2625–2639. https://doi.org/10.1002/aur.2607
Madipakkam, A. R., Rothkirch, M., Dziobek, I., & Sterzer, P. (2017). Unconscious avoidance of eye contact in autism spectrum disorder. *Scientific Reports*, *7*(1), 13378. https://doi.org/10.1038/s41598-017-13945-5
Ozonoff, S., Young, G. S., Belding, A., Hill, M., Hill, A., Hutman, T., Johnson, S., Miller, M., Rogers, S. J., Schwichtenberg, A. J., Steinfeld, M., & Iosif, A. M. (2014). The broader autism phenotype in infancy: When does it emerge? *Journal of the American Academy of Child and Adolescent Psychiatry*, *53, 53*(4, 4), 398, 398-407.e2. https://doi.org/10.1016/j.jaac.2013.12.020, 10.1016/j.jaac.2013.12.020
Papagiannopoulou, E. A., Chitty, K. M., Hermens, D. F., Hickie, I. B., & Lagopoulos, J. (2014). A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders. *Social Neuroscience*, *9*(6), 610–632. https://doi.org/10.1080/17470919.2014.934966
Ratto, A. B., Turner-Brown, L., Rupp, B. M., Mesibov, G. B., & Penn, D. L. (2011).

Development of the Contextual Assessment of Social Skills (CASS): A Role Play Measure of Social Skill for Individuals with High-Functioning Autism. *Journal of Autism and Developmental Disorders*, *41*(9), 1277–1286. https://doi.org/10.1007/s10803-010-1147-z
Riddiford, J. A., Enticott, P. G., Lavale, A., & Gurvich, C. (2022). Gaze and social functioning associations in autism spectrum disorder: A systematic review and meta-analysis. *Autism Research*, *15*(8), 1380–1446. https://doi.org/10.1002/aur.2729
Ross, A. I., Chan, J., & Ryan, C. (2023). Eye gaze During Semi-naturalistic Face-to-Face Interactions in Autism. *Advances in Neurodevelopmental Disorders*. https://doi.org/10.1007/s41252-023-00378-7
Schultz, R. T., Gauthier, I., Klin, A., Fulbright, R. K., Anderson, A. W., Volkmar, F., Skudlarski, P., Lacadie, C., Cohen, D. J., & Gore, J. C. (2000). Abnormal Ventral Temporal Cortical Activity During Face Discrimination Among Individuals With Autism and Asperger Syndrome. *Archives of General Psychiatry*, *57*(4), 331–340. https://doi.org/10.1001/archpsyc.57.4.331
Senju, A., & Johnson, M. H. (2009). The eye contact effect: Mechanisms and development. *Trends in Cognitive Sciences*, *13*(3), 127–134. https://doi.org/10.1016/j.tics.2008.11.009
Shell, J. S., Vertegaal, R., Cheng, D., Skaburskis, A. W., Sohn, C., Stewart, A. J., Aoudeh, O., & Dickie, C. (2004). ECSGlasses and EyePliances: Using attention to open sociable windows of interaction. *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*, 93–100. https://doi.org/10.1145/968363.968384
Tunc, B., Ghanbari, Y., Smith, A. R., Pandey, J., Browne, A., Schultz, R. T., & Verma, R. (2014). PUNCH: Population Characterization of Heterogeneity. *NeuroImage*, *98*, 50–60. https://doi.org/10.1016/j.neuroimage.2014.04.068
Vacas, J., Antolí, A., Sánchez-Raya, A., Pérez-Dueñas, C., & Cuadrado, F. (2021). Visual preference for social vs. Non-social images in young children with autism spectrum disorders. An eye tracking study. *PLOS ONE*, *16*(6), e0252795. https://doi.org/10.1371/journal.pone.0252795
Wei, Q., Cao, H., Shi, Y., Xu, X., & Li, T. (2023). Machine learning based on eye-tracking data to identify Autism Spectrum Disorder: A systematic review and meta-analysis. *Journal of Biomedical Informatics*, *137*, 104254. https://doi.org/10.1016/j.jbi.2022.104254
Ye, Z., Li, Y., Fathi, A., Han, Y., Rozga, A., Abowd, G. D., & Rehg, J. M. (2012). Detecting eye contact using wearable eye-tracking glasses. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 699–704. https://doi.org/10.1145/2370216.2370368
Ye, Z., Li, Y., Liu, Y., Bridges, C., Rozga, A., & Rehg, J. M. (2015). Detecting bids for eye contact using a wearable camera. *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, *1*, 1–8. https://doi.org/10.1109/FG.2015.7163095
Zhang, D., Wang, B., Wang, G., Zhang, Q., Zhang, J., Han, J., & You, Z. (2021). *Onfocus Detection: Identifying Individual-Camera Eye Contact from Unconstrained Images* (No. arXiv:2103.15307). arXiv. https://doi.org/10.48550/arXiv.2103.15307
Zhang, D., Wang, B., Wang, G., Zhang, Q., Zhang, J., Han, J., & You, Z. (2022). Onfocus detection: Identifying individual-camera eye contact from unconstrained images. *Science China Information Sciences*, *65*(6), 160101. https://doi.org/10.1007/s11432-020-3181-9

Table 1. Sample characteristics. The NT group includes one set of three siblings.

	AUT (N=46)	**NT (N=36)**	**P-value**
Age			
Mean (SD)	14.4 (5.80)	16.0 (6.53)	0.261
Median [Min, Max]	12.2 [8.07, 26.4]	13.9 [8.16, 28.2]	
Sex			
Female	17 (37.0%)	15 (41.7%)	0.837
Male	29 (63.0%)	21 (58.3%)	
IQ			
Mean (SD)	104 (17.4)	107 (12.4)	0.261
Median [Min, Max]	103 [73.0, 137]	108 [86.0, 136]	
ADOS-2 SA CSS			
Mean (SD)	7.30 (2.07)	1.62 (0.805)	<0.001
Median [Min, Max]	8.00 [2.00, 10.0]	1.00 [1.00, 3.00]	
Missing	19 (41.3%)	15 (41.7%)	
ADOS-2 RRB CSS			
Mean (SD)	6.67 (1.88)	1.29 (1.31)	<0.001
Median [Min, Max]	7.00 [1.00, 10.0]	1.00 [1.00, 7.00]	
Missing	19 (41.3%)	15 (41.7%)	
Race			
Asian	0 (0%)	3 (8.3%)	0.107
Biracial	3 (6.5%)	1 (2.8%)	
Black	3 (6.5%)	5 (13.9%)	
White	23 (50.0%)	12 (33.3%)	
Other	1 (2.2%)	0 (0%)	
Missing	16 (34.8%)	15 (41.7%)	

Table 2. Model parameters for the nine main variables examined. FDR correction was applied only to terms of interest, across all models.

Outcome	Description	Model Type	Family and Link	Term	Estimate	Std. Error	p-value	q-value
Partner Gaze	Proportion of frames research Partner gazed at Participant	glm	Quasi-binomial logit	Intercept	2.442	1.632	0.139	
				AUT	**-0.291**	**0.124**	**0.022**	**0.045**
				Age	**-0.067**	**0.014**	**<0.001**	**<0.001**
				Age^2	-	-	-	
				AUT*Age	-	-	-	
				Male Sex	0.151	0.178	0.4	
				IQ	0.007	0.005	0.222	
				Length	-0.002	0.008	0.837	

				Partner Gaze	-	-	-	
Participant Gaze	Proportion of frames Participant gazed at research Partner	lm	-	Intercept	1.881	0.469	<0.001	
				AUT	**-0.123**	**0.032**	**<0.001**	**0.001**
				Age	0.009	0.004	0.044	0.074
				Age^2	-	-	-	
				AUT*Age	-	-	-	
				Male Sex	-0.04	0.045	0.383	
				IQ	0	0.001	0.917	
				Length	-0.003	0.002	0.131	
				Partner Gaze	-0.894	0.248	0.001	
Mutual Gaze	Proportion of frames Participant and Partner both gazed at each other	lm	-	Intercept	1.093	0.425	0.012	
				AUT	**-0.109**	**0.029**	**<0.001**	**0.001**
				Age	0.007	0.004	0.088	0.125
				Age^2	-	-	-	
				AUT*Age	-	-	-	
				Male Sex	-0.043	0.041	0.298	
				IQ	0	0.001	0.98	
				Length	-0.003	0.002	0.119	
				Partner Gaze	-0.115	0.225	0.61	
Neither Gaze	Proportion of frames neither Participant nor Partner gazed at the other person	lm	-	Intercept	-1.093	0.425	0.012	
				AUT	**0.109**	**0.029**	**<0.001**	**0.001**
				Age	-0.007	0.004	0.088	0.125
				Age^2	-	-	-	
				AUT*Age	-	-	-	
				Male Sex	0.043	0.041	0.298	
				IQ	0	0.001	0.98	
				Length	0.003	0.002	0.119	
				Partner Gaze	1.115	0.225	<0.001	
Number of Mutual Gazes	Number of discrete episodes of mutual gaze	lm	-	Intercept	47.022	23.707	0.051	
				AUT	6.312	4.112	0.129	0.161
				Age	0.31	0.211	0.146	0.171
				Age^2	-	-	-	
				AUT*Age	**-0.573**	**0.253**	**0.026**	**0.048**
				Male Sex	-0.349	2.263	0.878	
				IQ	-0.155	0.071	0.033	
				Length	0.351	0.103	0.001	
				Partner Gaze	-67.055	12.417	<0.001	
Duration Mutual	Mean duration (seconds) of	glm	Gamma inverse	Intercept	0.305	0.407	0.457	
				AUT	**0.092**	**0.028**	**0.001**	**0.004**

Gazes	episodes of mutual gaze			Age	-0.002	0.003	0.465	0.517
				Age^2	-	-	-	
				AUT*Age	-	-	-	
				Male Sex	0.038	0.038	0.329	
				IQ	-0.002	0.001	0.222	
				Length	0.004	0.002	0.019	
				Partner Gaze	-0.556	0.235	0.021	
Participant Initiation	Proportion of mutual gaze episodes in which the Participant gazed at the Partner before the Partner gazed at the Participant	glm	Quasi-binomial logit	Intercept	4.102	1.734	0.021	
				AUT	**-0.431**	**0.109**	**<0.001**	**0.001**
				Age	**0.244**	**0.088**	**0.007**	**0.018**
				Age^2	**-0.006**	**0.003**	**0.022**	**0.045**
				AUT*Age	-	-	-	
				Male Sex	0.144	0.154	0.352	
				IQ	0.005	0.005	0.316	
				Length	-0.005	0.007	0.436	
				Partner Gaze	-7.693	0.848	<0.001	
Delay to Return Gaze	Mean time (seconds) between Partner gazing at Participant and Participant gazing at Partner, of Partner-initiated episodes	glm	Gamma inverse	Intercept	2.942	0.633	<0.001	
				AUT	**-0.136**	**0.04**	**0.001**	**0.003**
				Age	0.002	0.006	0.709	0.709
				Age^2	-	-	-	
				AUT*Age	-	-	-	
				Male Sex	-0.063	0.051	0.222	
				IQ	-0.001	0.002	0.584	
				Length	-0.003	0.003	0.293	
				Partner Gaze	-1.99	0.38	<0.001	
Wait Time	Mean time (seconds) between Participant gazing at Parner and Partner gazing at Partnicipant, of Participant-initiated episodes	glm	Gamma inverse	Intercept	-0.594	0.86	0.492	
				AUT	-0.027	0.062	0.66	0.695
				Age	-0.012	0.008	0.114	0.153
				Age^2	-	-	-	
				AUT*Age	-	-	-	
				Male Sex	-0.038	0.087	0.665	
				IQ	0.001	0.003	0.588	
				Length	0.003	0.004	0.418	
				Partner Gaze	1.428	0.415	0.001	

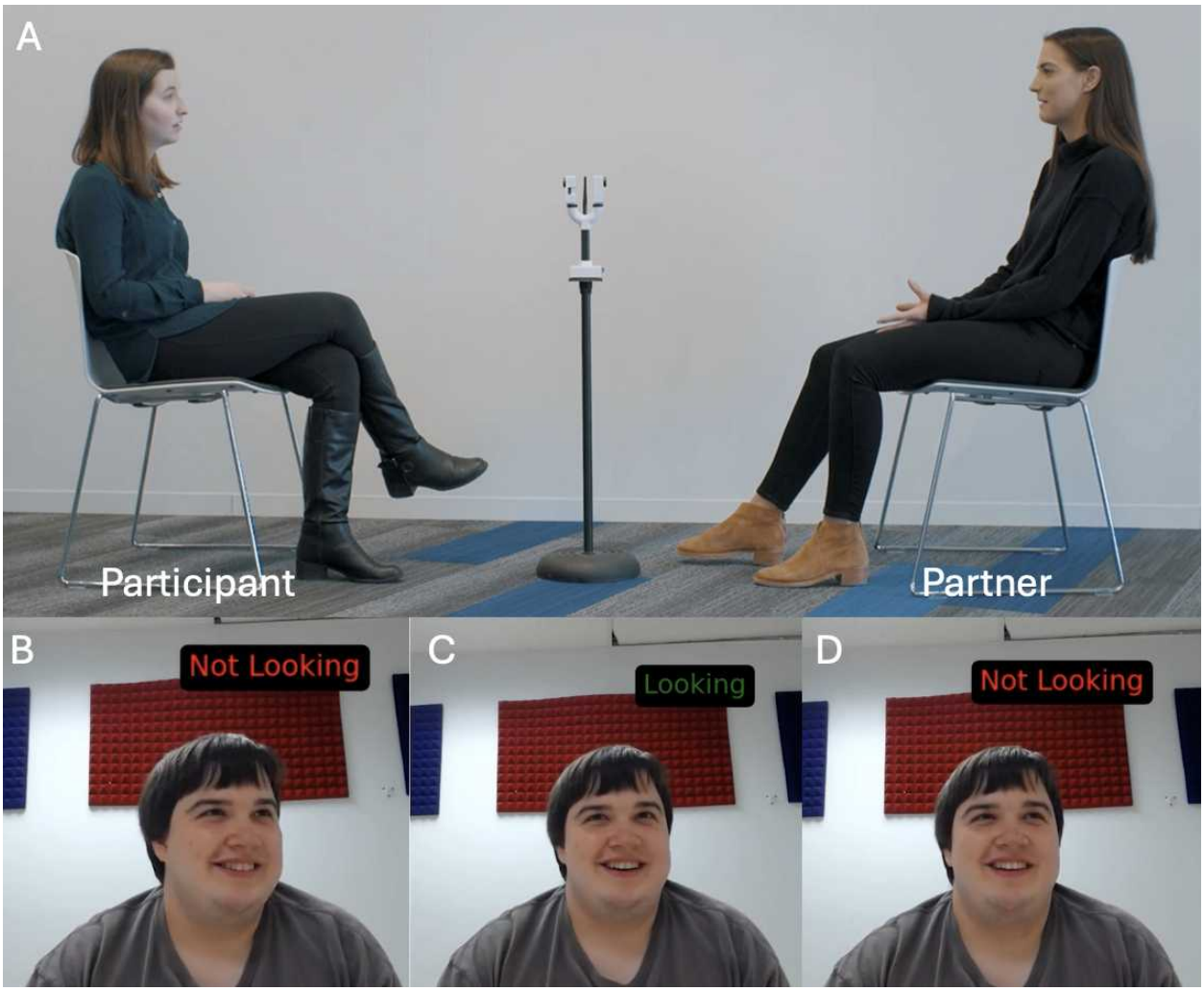


Figure 1. (A) Recording set-up, with Participant and Partner seated facing one another with two synchronized cameras between them. Participants were seated approximately 29 inches from the camera, which was adjusted to approximately chest level, and asked to stay seated in their chair facing forward and to ignore the camera as much as possible. The trained social gaze detection algorithm leverages head, face, and eye information and successfully distinguishes not looking (B, D) from looking (C), even when the head pose is identical and only the eyes have moved (C and D). Images have been cropped vertically for presentation, see Supplementary Figure 1 for full image used for annotation.

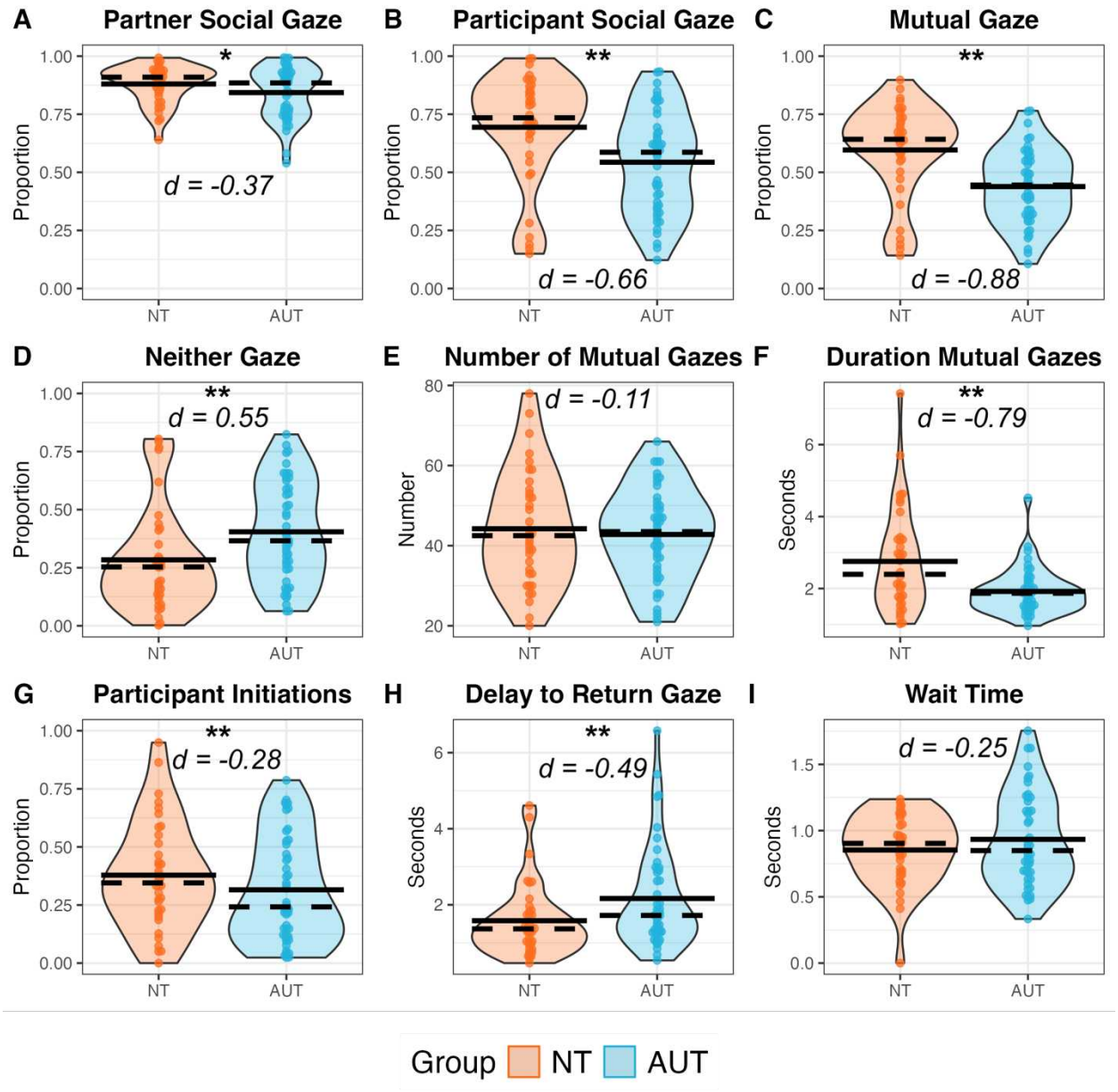


Figure 2. Violin plots of main variables. The solid lines represent the mean and dashed lines represent median values. *FDR-corrected $q < 0.05$, **$q < 0.01$ from models presented in Table 1. Cohen's d values are not adjusted for covariates.

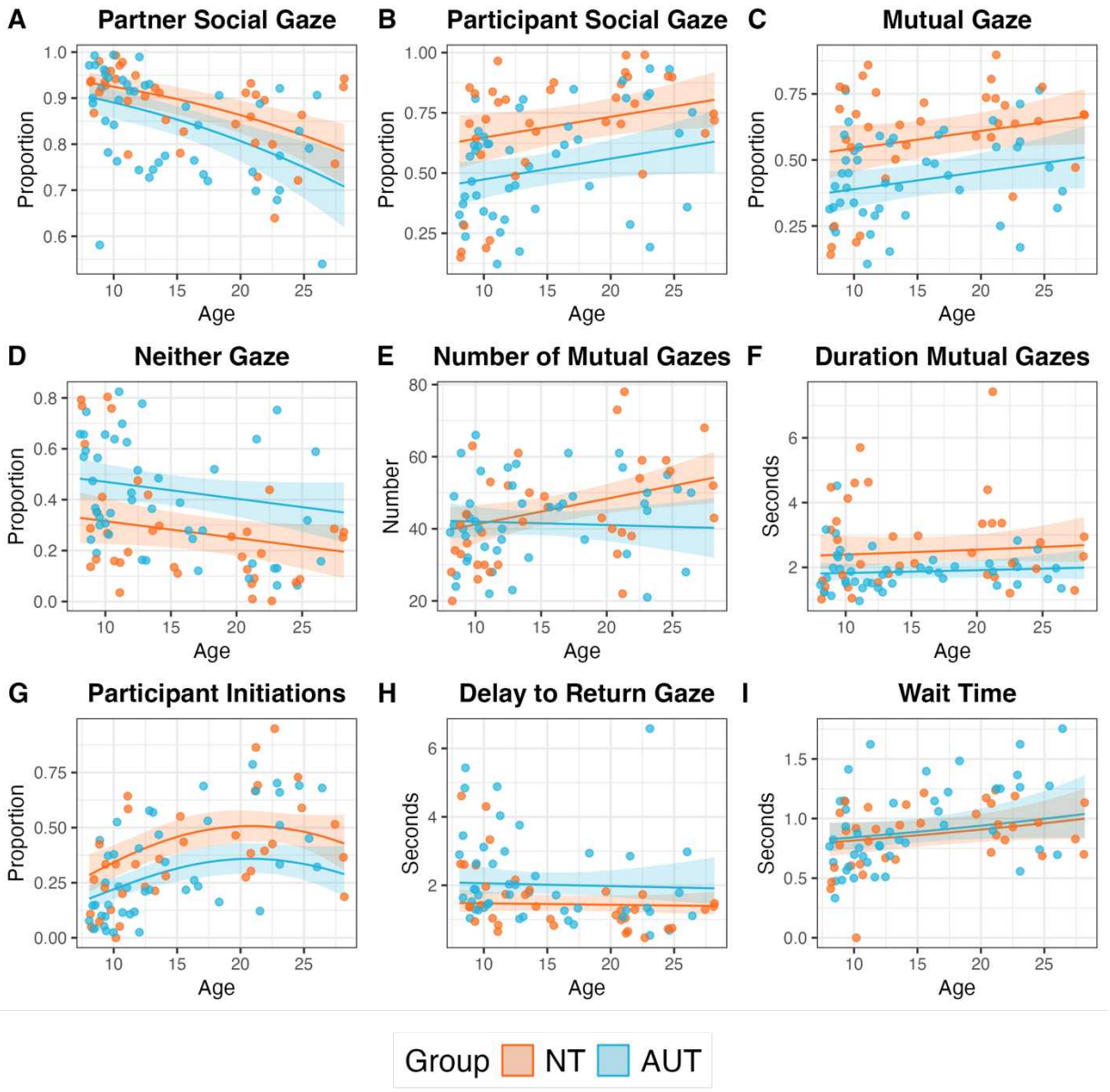


Figure 3. Developmental changes in social gaze patterns. Age-by-diagnosis interaction was observed only for the number of mutual gaze episodes. A quadratic effect was observed for number of mutual gaze episodes initiated by the Participant. Partner looking decreased with Participant age.

Supplementary Materials

Supplementary Methods: Sample, Annotation of Social Gaze, Gaze Detection Algorithm, Social Gaze Features, Statistical Analysis, Machine Learning Classification

Supplementary Figure 1: A screenshot from the video annotation interface.

Supplementary Table 1: List of social gaze variables defined in the study

Supplementary Figure 2: Social gaze summary score and ADOS scores.

Sample

The full possible sample included 114 participants (57 AUT, 57 NT, age = 7-49 years) who participated in studies at the Center for Autism Research at the Children's Hospital of Philadelphia which included the CASS task. We created the subsample used in this analysis by restricting the age to 8-29 years, the most densely and uniformly represented range in our sample (i.e., younger or older age bins were very sparsely represented), to increase the robustness of our findings. The inclusion of younger or older participants negatively affected the fitness of outcome variables to standard distributions due to sparse representations and, possibly, significant age effects. The final sample was created by matching AUT and NT groups on age, sex ratio, and IQ. This sample includes one set of three siblings (all NT). Autism diagnoses were confirmed through the best clinical judgment of a licensed psychologist using all available information, including administration of the Autism Diagnostic Observation Schedule (ADOS-2), Module 3 or 4, parent- and self-report questionnaires, and brief unstructured interviews.

Inclusion criteria for the larger study included English as a primary language. Exclusion criteria included a known disorder, injury, or medication use which affects motor functioning. For the present study, additional selection criteria included the availability of videos of acceptable quality (i.e., working video/audio for both participant and partner) and compatibility with the automated method (i.e., both individuals positioned appropriately, directly facing the camera, and seated at an appropriate distance).

Annotation of Social Gaze

We segmented each video recording into 1-second clips. Human coders blind to participant diagnosis annotated social gaze for the last frame of each clip. Thus, for a video recording with a

length of three minutes (180 seconds, 5400 frames), 180 frames were annotated, with the remaining 5220 not annotated.

Raters used a web app that we built in-house, specialized for behavioral annotations of video recordings. When raters log in to the system, a random clip not annotated by them is shown for annotation. Each clip showed the frontal images of the participant and partner side by side and was played in a continuous loop with the last frame frozen for one second. eFigure1 shows a screenshot from the web app. Two raters made a yes/no judgment of whether each member of the dyad was looking at the other person, with disagreements resolved by a third rater. Reliability between the first two ratings was excellent for both Participants ($\kappa = 0.80$) and Partners ($\kappa = 0.83$).

Training for the annotation task was minimal, and raters were simply instructed to indicate whether the individual was "looking at the other person." The task was intentionally kept broad to maximize reliability between raters, as it was deemed unlikely that raters would be able to determine where, specifically, the individual was looking (e.g., to distinguish between looking at the other person's eyes versus mouth). This choice, however, prevented us from knowing what part of the face the person was looking at. For that reason, we refer to these annotations as "social gaze" and not "eye contact". For both the Participant and Partner, the final annotation labels were determined by the majority vote of the two or three ratings and used for training and testing our AI algorithm.

Gaze Detection Algorithm

We used a deep learning architecture that has been developed for detecting whether an individual in a photo is looking at the camera or not(Zhang et al., 2021). The method was developed to work effectively with unconstrained "in the wild" images (*i.e.*, no restrictions on where, when, and how the photos were taken). It uses a sophisticated convolutional neural network

(CNN) architecture that automatically detects the eye region in the photo and then processes both the eye region and the remaining parts ("context") through two separate streams and, finally, merges the learned representations of the two to model the interactions between them. This architecture allowed us to model both the eye behavior and the context (in our case, face and head behavior) together to make the final decision on whether the person was looking at the other person or not. This is a clear novelty compared to earlier works that used solely head and body posture to make a similar decision.

To adapt the model to our needs, we re-trained the system using data from our video recordings and human annotations. That is, the system learned to detect social gaze, not looking at the camera. Video frames to include only the head region in training and testing, following the specifications of the original model.

The performance of the algorithm was computed using ten-fold cross-validation where folds were defined with respect to individuals, not frames; that is, all frames of an individual were used only in training or testing, but not both. For each fold, the algorithm was trained with a balanced subsample including the same amount of looking and not looking instances, selected randomly. Therefore, reported sensitivity, specificity, PPV, and NPV stats reflect the cross-validated performance of a perfectly balanced training sample. After the performance of the model was computed using cross-validation, we used all annotated data to train the algorithm for a final time and used it to detect social gaze at every single frame of the video recordings, including both frames that were annotated and not annotated, yielding frame-by-frame binary (yes/no) social gaze data for the whole video recordings of the Participant and Partner.

Social Gaze Events

To define social gaze events, we applied noise removal over the binary social gaze data (*i.e.*, time series of looking/not looking) to eliminate possible false positives and negatives. Specifically, we split the binary time series into "looking" and "not looking" blocks by simply splitting at transition points. For example, if the original time series was 000000011111000001100111, then the blocks were "0000000", "11111", "00000", "11", "00", "111". We then assumed that a person would not change their looking state faster than one-third of a second (*i.e.*, ten frames), and we merged any blocks of "looking" or "not looking" with its surrounding blocks to get a smoothed time series. We used simple heuristics to make the merger decisions instead of using, for instance, a Gaussian smoothing kernel or a median-based smoothing since such methods cannot guarantee the minimum block size of the output and may result in overly smoothed outputs. After noise removal, the binary social gaze data for the Participant and Partner were merged. "Mutual gaze" events were defined where both individuals had 1 values, and "neither gaze" events were defined where both individuals had 0 values. It is also possible to define events where only the Participant or only the Partner was looking, but these events were not directly analyzed in this study.

Statistical Analysis

We used a linear or generalized linear model for each of the nine primary variables of interest (see eTable 1), using a systematic model building approach. First, a linear model was constructed. All models included the main effects of diagnosis (AUT or NT) and age. Although the diagnostic groups were matched on sex ratio and IQ (as well as age), these two variables were included as nuisance covariates in all models to ensure they did not drive results. All models were also controlled for conversation length and Partner looking (except the model predicting Partner looking). A quadratic term for age and an age-by-diagnosis interaction term were then added to

the models, and subsequently retained in the final model if they significantly improved model fit as assessed by a likelihood-ratio test. A Shapiro-Wilks test was then applied to the residuals; if they were not normally distributed, a generalized linear model was constructed instead using the same model building approach described above, with either a quasinomial family and logit link function (proportion variables) or Gamma family and inverse link function. False Discovery Rate correction using the Benjamini-Hochberg procedure was applied to the terms of interest (i.e., not nuisance covariates) study-wide across the nine models. Final models are presented in Table 2 in the main text.

Machine Learning Classification

We trained a machine learning model to classify diagnostic labels (AUT versus NT) using all 18 variables plus the conversation length as potential features. Some of these features are highly correlated by design, and thus would be highly redundant if included together in a machine learning model. Therefore, we first applied a simple, unsupervised (*i.e.*, without using any diagnostic labels) feature selection step. Correlations between all pairs of features were computed, and only one feature (randomly selected) was retained if the correlation was higher than 0.7. This procedure yielded a set of nine non-redundant variables, of which five were variables used in the main analysis (Partner Gaze, Participant Gaze, Neither Gaze, Number of Mutual Gazes, Wait Time), three additional gaze variables (Minimum Mutual Gaze Duration, Minimum Delay to Return Gaze, Minimum Wait Time), and conversation length. It is acknowledged that this simple feature selection approach does not account for potential multilinear relationships among features. Future research will explore more advanced feature selection methods to address this limitation.

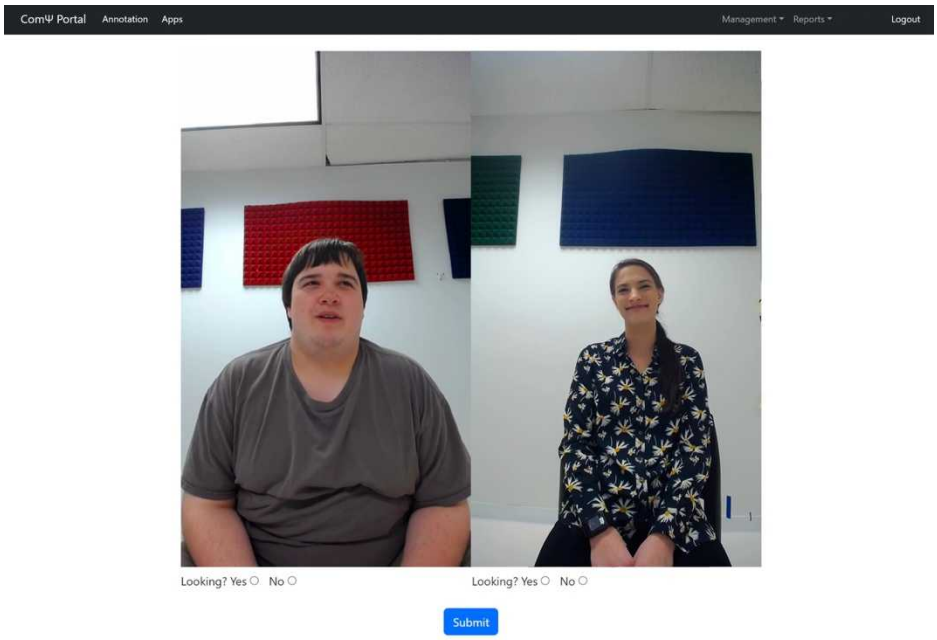
We used a Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel for machine learning classification. We used nested ten-fold cross-validation to train and test the

algorithm, with all model parameters for SVM (C and γ) optimized within the inner cross-validation only using the training data. We used grid search for optimizing C and γ parameters using following ranges

$$C = 2^k \text{ for } k = -5, -4, \ldots, 4, 5$$

$$\gamma = 2^k \text{ for } k = -5, -4, \ldots, 4, 5$$

To account for the slight class imbalance in our sample, we used sample weights proportional to sample sizes (using *class_weight='balanced'* option in Python, Scikit-Learn library).



Supplementary Figure 1. A screenshot from the video annotation interface. Raters were instructed to indicate whether the person was looking at the other person or not, both for the Participant (always on the left of the screen) and the Partner (always on the right of the screen). After the rater

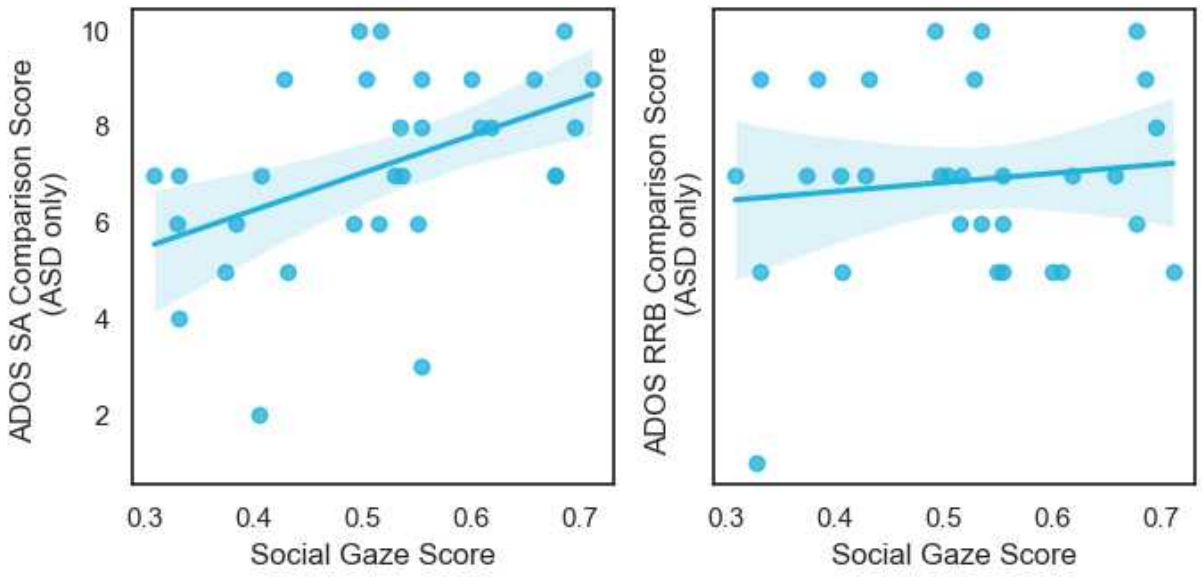
logged into the platform, a random clip that the rater had not annotated before was shown, the rater provided both ratings and submitted, and a new random clip (from any participant, in any order) was shown.

Supplementary Table 1. List of social gaze variables defined in the study. The second column indicates whether the variables were used in the main analysis or not. All variables were used in the machine learning model.

ID	Main Analysis	Selected for Machine Learning/ PUNCH	Variable Name	Description
1	Yes	Yes	Partner Gaze	Proportion of frames research partner gazed at participant
2	Yes	Yes	Participant Gaze	Proportion of frames participant gazed at research partner
3	Yes	No	Mutual Gaze	Proportion of frames participant and partner both gazed at each other
4	Yes	Yes	Neither Gaze	Proportion of frames neither participant nor partner gazed at the other person
5	Yes	Yes	Number of Mutual Gazes	Number of discrete episodes of mutual gaze
6	Yes	No	Participant Initiation	Proportion of mutual gaze episodes in which the participant gazed at the partner before the partner gazed at the participant
7	Yes	No	Mutual Gaze	Mean duration (seconds) of episodes of mutual gaze

			Duration	
8	No	Yes	Mutual Gaze Duration (Minimum)	Minimum duration (seconds) of episodes of mutual gaze
9	No	No	Mutual Gaze Duration (Maximum)	Maximum duration (seconds) of episodes of mutual gaze
10	No	No	Mutual Gaze Duration (SD)	Standard deviation of duration (seconds) of episodes of mutual gaze
11	Yes	No	Delay to Return Gaze	Mean time (seconds) between partner gazing at participant and participant gazing at partner, of partner-initiated episodes
12	No	Yes	Delay to Return Gaze (Minimum)	Minimum time (seconds) between partner gazing at participant and participant gazing at partner, of partner-initiated episodes
13	No	No	Delay to Return Gaze (Maximum)	Maximum time (seconds) between partner gazing at participant and participant gazing at partner, of partner-initiated episodes
14	No	No	Delay to Return Gaze (SD)	Standard deviation of time (seconds) between partner gazing at participant and participant gazing at partner, of partner-initiated episodes
15	Yes	Yes	Wait Time	Mean time (seconds) between participant gazing at partner and partner gazing at participant, of participant-initiated episodes
16	No	Yes	Wait Time (Minimum)	Minimum time (seconds) between participant gazing at partner and partner gazing at participant, of participant-

				initiated episodes
17	No	No	Wait Time (Maximum)	Maximum time (seconds) between participant gazing at partner and partner gazing at participant, of participant-initiated episodes
18	No	No	Wait Time (SD)	Standard deviation of time (seconds) between participant gazing at partner and partner gazing at participant, of participant-initiated episodes



Supplementary Figure 2. Social gaze summary score and ADOS scores. Within the AUT group, the overall "social gaze summary score" derived from nine looking features was significantly associated with the ADOS-2 Social Affect Calibrated Severity Score, but not the Restricted and Repetitive Behavior Score.

Supplementary References

Zhang, D., Wang, B., Wang, G., Zhang, Q., Zhang, J., Han, J., & You, Z. (2021). *Onfocus Detection: Identifying Individual-Camera Eye Contact from Unconstrained Images* (No. arXiv:2103.15307). arXiv. https://doi.org/10.48550/arXiv.2103.15307