Feature-based Modeling of Visual Attention in Autism: A Large-scale Online Eye-tracking Study

Qianying Wu* (qwu@caltech.edu)

Na Yeon Kim* (nayeon@caltech.edu)

Audrey K. Lai (aklai@caltech.edu)

Lynn K. Paul (lkpaul@hss.caltech.edu)

Ralph Adolphs (radolphs@caltech.edu)

Division of the Humanities and Social Sciences, California Institute of Technology
Pasadena, CA, USA

* indicates equal contribution

Abstract

Atypical visual attention is one of the most reliable findings in autism spectrum disorder (ASD), with important implications for clinical screening and diagnosis. However, most findings rely on artificial stimuli and small samples, limiting generalizability. In this pre-registered study, we used webcam-based eye-tracking and feature-based computational modeling to characterize visual attention in a broad sample of 336 ASD participants and 304 neurotypical controls. Participants watched videos of group conversations that incorporated controlled social and nonsocial features. Compared to controls, autistic individuals showed reduced attention to speakers, increased sensitivity to distractors, and more frequent gaze shifts. Our study demonstrates the power of scalable online eye-tracking and modeling approaches for capturing individual differences in visual attention and advancing the understanding of heterogeneity in ASD.

Keywords: autism; attention; eye tracking; saliency modeling; naturalistic stimuli

Introduction

Autism spectrum disorder (ASD) features impaired social functioning and stereotyped behaviors. Elucidating the underlying cognitive processes has been challenging due to considerable heterogeneity, yet deficits in visual attention are almost universally reported. These include reduced attention to socially meaningful cues (e.g., faces; Pelphrey et al., 2002), and increased attention to low-level visual saliency (e.g., geometric patterns; Pierce et al., 2011). Eye-tracking provides a quantitative measure to probe such differences and examine the mechanisms behind social and communicative challenges in autism (Wang et al., 2015; Jones et al., 2024). However, most eye-tracking studies are limited to lab-based settings, restricting scalability and sample diversity, posing a critical obstacle for investigating between-individual variability across the autism spectrum.

In this pre-registered study (https://osf.io/sf7wz), we tackle this limitation by utilizing webcam-based eye tracking (WebGazer; Papoutsaki et al., 2016) over the internet in a large sample with a range of autistic traits. We assessed attention to social and nonsocial features while watching naturalistic group conversations. We estimated attention weights from feature-based computational models and compared them to test three pre-registered hypotheses: Compared to neurotypical controls, autistic individuals will exhibit:

H1. smaller weights on socially relevant features, such as who is speaking at the moment and the gaze direction of the speaker

H2. greater weights on irrelevant distractions

H3. less sustained attention in general

Methods & Results

Participants with a diagnosis of ASD were recruited from two databases: Prolific (Prolific ASD; N = 123, age 31.4 ± 9.4, 60 females) and SPARK autism database (SPARK ASD; N = 213, age 35.5 ± 10.6, 126 females). The Control group consisted of 304 neurotypical individuals (age 36.5 ± 11.8, 136 females). Both ASD groups showed greater autistic traits than the Control group (Prolific ASD: 94.9 ± 34.4, SPARK ASD: 102.0 ± 26.1, Control: 44.7 ± 25.7; scores on SRS2-A-SR, Constantino & Gruber, 2012).

Participants watched on their home computer four videos in which 4 people (trained actors) engaged in a conversation over Zoom (5-6 minutes each), along with three experimentally manipulated factors: turn-taking speech (speaker), irrelevant people or objects in the background (distraction), and actor gaze directions (averted gaze, i.e., look at vs. away from the camera). Gaze data were preprocessed and mapped to screen quadrants with 96.7 ± 4.7 % accuracy (Fig. 1A).

To estimate attention weights, we used L1-regularized multinomial logistic regression models with 10-fold cross-validation (Fig. 1B) to model gaze location (in quadrants) based on 4 key features: speaker, distraction, averted gaze, and 1-st order auto-regressor of the gaze (AR1). The model also included 4 categories of control variables: actor face (face size and facial expressions), low level visual saliency (color, intensity, orientation, and motion), low level auditory saliency (frequency decompositions), and actor body motion (head and hand movement).

Our feature-based attention model successfully characterized gaze patterns of the Control group (validation accuracy = 74.05 ± 7.40 %), Prolific ASD group (validation accuracy = 73.97 ± 7.92 %), and the SPARK ASD group (validation accuracy = 72.34 ± 7.25

%; all chance accuracy = 25%). To test our hypotheses, we compared the attention weights of different features among groups (Fig. 2C).

H1: Both Prolific ASD ($t(1316) = 4.97$, $p < 0.001$) and SPARK ASD ($t(1496) = 6.06$, $p < 0.001$) groups had significantly smaller weights for the speaker compared to the Control group. However, weights for the averted gaze did not differ across the three groups.

H2: Both Prolific ASD ($t(1316) = 3.30$, $p < 0.001$) and SPARK ASD ($t(1496) = 8.65$, $p < 0.001$) groups had significantly higher weights for the distraction compared to the Control group.

H3: The SPARK ASD group had significantly smaller weights for AR1 compared to the Control group ($t(1496) = 2.00$, $p = 0.03$), indicating more frequent gaze shifts, but there was no difference between the Prolific ASD and Control group.

Additionally, we performed variance partitioning to estimate the unique variance explained by the four categories of control variables (Fig. 2D). We found that facial features explained the most variance, followed by the low level visual features, low level auditory features, and body motions. No significant differences were found between the Control group and either ASD group.

Conclusion

Using scalable online eye-tracking and feature-based modeling, we identified robust group differences in attention patterns during naturalistic viewing: reduced attention on the speakers, increased attention on irrelevant distractions, and more frequent gaze shifts in ASD. This framework will allow further investigations into individual variability and potential subgroups in autism.

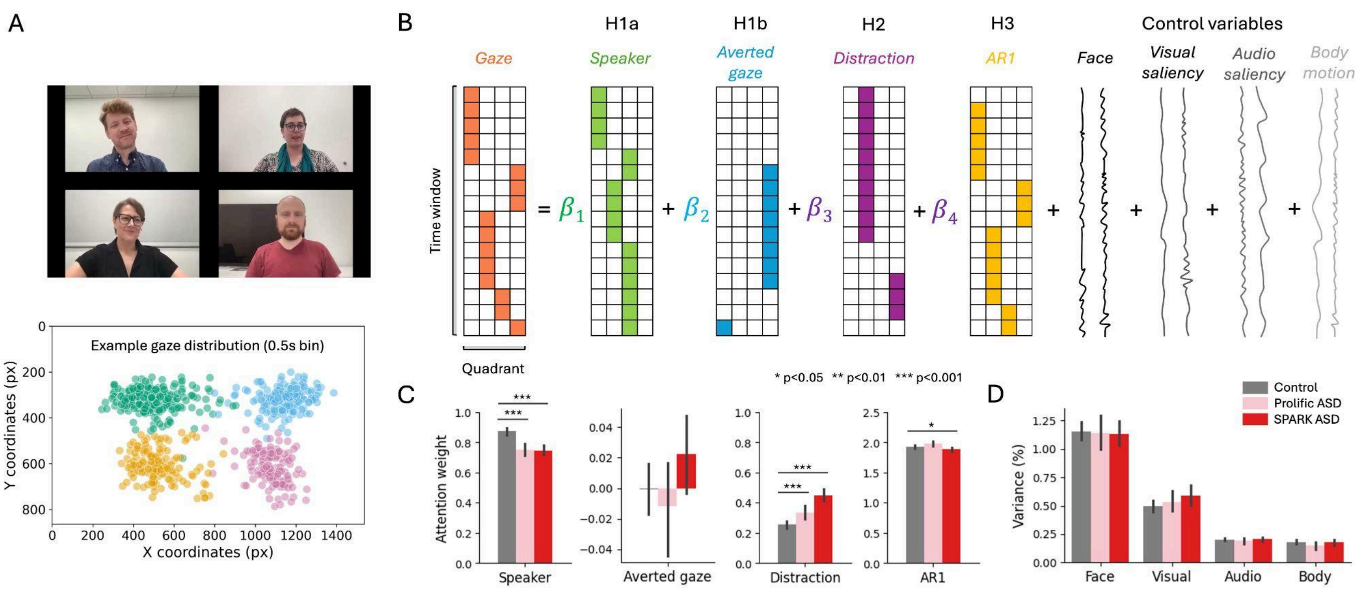

Figure 1: (A) Example stimulus frame and the spatial distribution of one example participant's gaze while watching one video, color coded by quadrant. (B) Model-based estimation of attention weights. (C) Results for the main hypotheses. (D) Exploratory variance partitioning analysis of control variables.

Acknowledgments

This work was supported by the Simons Foundation Autism Research Initiative. We thank Brenna Outten and Ava Barbano for assistance with stimulus generation and all members of the Emotion and Social Cognition Laboratory for helpful comments.

References

Constantino, J.N., & Gruber, C.P. (2012). *Social responsiveness scale*, 2nd ed. Los Angeles, CA: Western Psychological Services.

Jones, W., Klaiman, C., Richardson, S., Aoki, C., Smith, C., Minjarez, M., ... & Klin, A. (2023). Eye-tracking–based measurement of social visual engagement compared with expert clinical diagnosis of autism. *JAMA*, 330(9), 854-865.

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable webcam eye tracking using user interactions. In *Proceedings of the Twenty Fifth International Joint Conference on Artificial Intelligence—IJCAI 2016* (pp. 3839–3845). Arlington, VA: National Science Foundation.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders*, 32, 249-261.

Pierce, K., Conant, D., Hazin, R., Stoner, R., & Desmond, J. (2011). Preference for geometric patterns early in life as a risk factor for autism. *Archives of General Psychiatry*, 68(1), 101-109.

Wang, S., Jiang, M., Duchesne, X. M., Laugeson, E. A., Kennedy, D. P., Adolphs, R., & Zhao, Q. (2015). Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. *Neuron*, 88(3), 604–616.