Title Page

"Temporal Profiles of Social Attention are Different Across Development in Autistic and Neurotypical People"

Teresa Del Bianco[1*], Luke Mason[1], Tony Charman[2], Julian Tillman[2], Eva Loth[2], Hannah Hayward[2], Frederick Shic[3], Jan Buitelaar[5], Mark H. Johnson[1,4], Emily J.H. Jones[1], & the EU-AIMS LEAP Group

[1] Centre of Brain and Cognitive Development, Birkbeck College, University of London, UK

[2] Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

[3] School of Medicine, University of Washington, Seattle, WA

[4] Department of Psychology, University of Cambridge, Cambridge, UK

[5] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Njmegen, NL

*Corresponding author.

Address: Centre for Brain and Cognitive Development, Birkbeck University of London, Malet Street, London, WC1E 7HX. Phone: +44 (0)20 7079 6322. Email: t.delbianco@bbk.ac.uk

Keywords: autism spectrum disorder, social attention, social motivation, eye-tracking, growth-curve analysis, stratification

Abstract

Background

Socio-communicative difficulties, including atypicalities in eye contact, are core diagnostic features of Autism Spectrum Disorder (ASD). Many studies have used eye-tracking to measure reduced attention to faces in autistic people, however, most of this work has not taken advantage of eye-tracking temporal resolution to examine temporal profiles of attention.

Methods

We used Growth Curve Analysis to model attention to static social scenes as a function of time in a large (N=650) sample of participants diagnosed with ASD, and neurotypical (NT) participants across a wide age range (6-30 years).

Results

The model yielded distinct temporal profiles of attention to faces in the groups. Initially, both groups showed a relatively high probability of attending to faces, followed by decline after several seconds. The NT group, however, were significantly more likely to return their attention to faces in the latter part of each 20s trial, with increasing probability with age. In contrast, the probability of returning to the face in the autistic (AUT) group remained low across development. In the AUT group, more atypical profiles of attention were associated with lower Vineland Communication scores, and a higher curvature in one data-driven cluster correlated with symptom severity.

Conclusions

These findings show that social attention is not only reduced in ASD, but that it differs in its temporal dynamics. The NT group became more sophisticated in how they deployed their social

attention across age, a pattern that was significantly reduced in the AUT group, possibly reflecting delayed acquisition of social expertise.

Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder affecting 13.1–29.3 on 1000 children worldwide (1). Core diagnostic criteria include socio-communicative impairments, such as difficulties in nonverbal communication and reduced social interactions (2). Several theoretical frameworks have emphasized the role that reduced social attention may play in the early development of ASD (3–6) – particularly the possibility of cascading effects across development, whereby less attention to faces leads to fewer rewarding social interactions, and a reduced acquisition of social expertise (7–10). Indeed, many studies have reported a reduction in attention to social cues, such as faces and eyes (11).

Despite a broad consensus in the literature for reduced social attention in ASD, there is nevertheless significant heterogeneity in findings (11,12). This may reflect heterogeneity between autistic people (34), but may also be an artefact of small sample sizes, narrow age ranges, and high variability of stimuli selection and pre-processing choices (12). Furthermore, the majority of studies do not exploit the temporal precision afforded by eye-tracking and simply examine averaged looking time over the duration of a stimulus. In this paper we examine social attention in a large and diverse cohort of autistic (AUT; i.e., with a clinical diagnosis of ASD) and neurotypical (NT; i.e., without a clinical diagnosis of ASD) 6 to 30-year-olds. We preserve the temporal dimension of social attention by performing a Growth Curve Analysis (GCA; 13) on the time course of social attention within each trial. This allows us to detect transient changes in social attention between groups that may be lost in more traditional approaches where variables are collapsed over time (4,14–16)[1].

Endogenously-driven changes in attention may be best evaluated during static scene viewing, where exogenous changes in the stimulus do not confound results (17). Major models of attention distinguish orienting and maintenance stages that occur sequentially (18). On presentation of a static social scene, initial fixations to the face are likely driven relatively more by stimulus-driven, perhaps subcortically-mediated orienting processes (19,20); as time elapses, attentional selection likely becomes increasingly driven by intrinsic interest and motivation (14,15). A difference in the average looking time to faces in ASD may thus reflect a failure of immediate social orienting at the start of a scene (21), or a reduction in sustained focus to social information after a period of time (16). Distinguishing these possibilities will illuminate the mechanisms underlying atypical social attention in ASD. Indeed, an emerging line of evidence from both infants and adults suggests that subcortically-mediated, rapid orienting to faces is intact in ASD (4,22). In contrast, more 'top-down', cortically-driven aspects of social attention may be altered, because of reduced social engagement (23), social motivation (24), social reward (10), or altered learnability of social information (25). Distinguishing these stages by examining the temporal profile of social attention within a trial should thus: 1) increase separation between group distributions and afford greater statistical power; 2) allow us to identify developmental changes in different components of social attention; 3) more precisely delineate the mechanisms underlying atypical social attention through separating their temporal dynamics.

In the present study, we test whether our large cohort of autistic individuals look less at faces; whether this varies over the temporal course of the stimulus presentation; and whether this changes across our wide age range. To strengthen the specification of the fixed effect, we carried out a pre-specified stepwise process of selection (26). Furthermore, we combine a categorical

group comparison approach with correlation analysis of dimensional data and a clustering analysis to test whether atypicalities in social attention (either absolute measures of temporal trends or degree of atypicality relative to the cohort profile) relate concurrently and/or predict clinical characteristics (27) across the AUT group in general, or within data-driven subtypes.

Methods and Materials

Sample

We used the eye-tracking data from the multi-site EU-AIMS Longitudinal European Autism Project, *LEAP* (48). The LEAP study involved 453 autistic participants diagnosed with ASD and 311 neurotypical participants; for the list of the sites, and the inclusion and exclusion criteria, see the Supplementary Material (SM), section 1.1 and 1.2. For the full protocol details, see (28). The main characteristics of the participants included in the analysis are reported in Table 1. Of the total 764 participants, 86.91% (664) had data for the specific task targeted by the current study (see SM 1.3). 5 of these participants were excluded for having > 75% missing gaze samples in any of the presented trials (see Table 2 of the SM for number of excluded trials); 9 others were excluded for not having a record of the Full Scale Intelligence Quotient (FSIQ). The final dataset consisted of 650 participants (age range 6-30, mean FSIQ = 100.46, standard deviation = 19.65). Clinical assessments of autistic traits (*Social Responsiveness Scale, 2nd Edition*, SRS-2) and adaptive communication and socialisation (*Vineland Adaptive Behavior Scales,* VABS) were collected concurrently with the eye tracking data (Time 1), and at a second visit after 12-24 months (Time 2).

[Table 1]

Eye-trackers, Stimuli and Procedure

Stimuli were 6 validated photographs of people in natural settings (28,29; full description SM 2.1; original images available for consultation upon request). Participants viewed the photographs for 20 seconds each in pseudorandom order while gaze was recorded with Tobii eye-trackers (SM 2.4). The stimuli were presented full-screen on the T120 (17", 1280x1024 pixels, aspect ratio 5:4), and with black borders corresponding to a 17", 5:4 display on the TX-300 (23", 1920x1080, aspect ratio 16:9). Participants were approximately 60 cm from the screen; stimuli covered approximately 33*18 visual degrees of angles.

Areas of Interest (AOIs) were manually drawn on each image (see Figure 1). Given our focus on social attention, analyses were focused on the proportion of looking on the "head" - encompassing hair, upper and lower face (53) - relative to the whole scene.

[Figure 1]

The eye tracking session began with online feedback as to the participant's position in front of the eye tracker. The experimenter positioned the participant in the center of the eye tracker headbox (31). A five-point calibration sequence was run (see SM 2.2), followed by the task battery; the progression between stimuli was gaze-contingent (see SM 2.3).

Data Aggregation and Analysis

As a sanity check on the distribution of data quality, we tested for group and age differences in the percentage of missing values with a mixed model including the trial time course with a GCA approach; additionally, we tested differences in accuracy and the number of AOIs sampled by trial with multiple linear regression (see SM 3.3 for details).

Trials and time bins with >75% missing data was excluded (see Table 2 of the SM for sample size before/after exclusion); the Proportional Looking Time (PLT; equals to Samples in AOI /

Total Valid Samples) on the head was calculated in 1 second bins for each stimulus (see SM 3.2 for details).

To present the findings of a traditional analytical approach for comparability to other literature, we first conducted an ANOVA by age and group to examine group differences in head looking (see SM 3.4). Then, we used mixed GCA to investigate temporal profiles of attention (see SM 3.5). Briefly, our base model included linear, quadratic and cubic terms as fixed effects, and varying intercepts and slopes per participants; we then tested the progressive increase in fit of other models (see SM 3.5.1), and selected the model with a significant decrease in AIC but no increase in BIC for the rest of the analyses with the Likelihood Ration test (see SM 3.5.1.3). The final model included group in interaction with continuous age, and sex and FSIQ as covariates. Secondly, we examined dimensional relations to concurrent and future clinical traits measured with the SRS-2 and the VABS (see SM 3.9). Specifically, we extracted measures of both individual temporal profiles (beta coefficients), and their degree of deviance from the overall temporal profile of their group (standard deviations of the random effect) from the base model. We associated those beta coefficients with both concurrent and future symptoms, and standard deviations with concurrent symptoms (given the lack of directionality of the standard deviation). Finally, to examine whether there were discrete subgroups within the AUT group, we conducted a hierarchical cluster analysis (see SM 3.10). We compared phenotypic data across clusters. We ran a parallel analysis in the NT group to see whether the clusters were ASD specific. Within the AUT clusters, we reran concurrent and predictive correlations with temporal profiles and degree of deviance from the overall temporal profile. We assumed a significance threshold of p-values <0.05 (adjusted with the Bonferroni correction where appropriate[2]) throughout.

Results

Quality Check

In the uncleaned dataset, the percentage of missing data diminished more with age in the NT group than the AUT group (X^2 (2) = 45.55, p-value < 0.001; see SM 3.3.1 for the full list of coefficients); both groups developed a curved and descending trend with age (quadratic component), with a higher curvature in the NT group compared to the AUT group (X^2 (2) = 30.34, p-value < 0.001; see SM 3.3.1). The groups did not differ in terms of slope nor cubic component configuring a similar ascending trend over the time of the trial (see Figure 1 of SM 3.3.1).

The groups did not differ nor age influenced Accuracy across the whole session, and the average number of AOIs sampled by trial. The full list of coefficients and p-values is available in SM 3.3.2 (Accuracy) and 3.3.3 (Number of AOIs Sampled).

The process of preparation with the threshold of 75% determined the exclusion of 228 individual trials (5.73% of total completed trials), and 5 autistic participants(see Table 2 of the SM for full report). In order to account for the group and age differences and possible effects introduced by thresholding data of differential missingness, we tested the contribution of the covariate Overall Looking Time to the Screen to the final model – that did not result significant in influencing temporal profiles (beta coef. = 0.03, SE = 0.02, p-value = 0.09; see SM 3.5.1.3). The robustness of this finding echoes the small effect size of the group difference in Overall Looking Time to the Screen (0.002; see Table 1).

Case-control comparison

Raw data per AOI is visualized in Figure 1 of the SM, section 3.3. The ANOVA showed greater looking on the head in the NT vs AUT groups (F = 11.17, p-value <0.001, p-value adjusted =

0.001) and greater looking on the head with older age overall (F = 9.70, p-value = 0.001, p-value adjusted = 0.005). An age per group interaction, indicating a developmental change in the NT group that is diminished in the AUT group just below the significance level (F = 4.15, p-value = 0.04) did not resist correction for multiple comparisons (p-value adjusted = 0.08; Table 2 and Figure 2).

[Table 2]

[Figure 2]

Temporal and Developmental Profiles (GCA)

Model Selection

With the Likelihood Ratio Test, we selected a final model with group modulated by age as fixed effect, and sex and the FSIQ as covariates (AIC = 46331.83, BIC = 46588.43, deviance = 46275.83, $X^2 = 5.91$, p-value < 0.001). The formula of the final model is reported in Equation 1. For a full list of the outputs, see SM 3.5.1.

Equation 1: Equation of the final model, with $y_{i,j,s}$ = proportional looking time to the head of the -ith participant (i) nested in the -jth site (j) for the -sth stimulus (s), β_0 = fixed intercept, β_1 = fixed group slope, x_{ij} = polynomials up to degree 3 at site j for the -ith participant, z_i = age of the the -ith participant, k_1; k_2 = covariates (sex and full scale IQ), β_{0ij} = random intercept for the -ith participant at site j, $\beta_{0i|j}$ = random intercept for the -ith participant within one site j, β_{0j} = random intercept at site j, β_{0s} = random intercept for the -sth stimulus, $\beta_{1i}x_{ij}$ = random slope at site j for the -ith participant, $\beta_{1i}x_{i|j}$ = random slope for the -ith participant within one site j, $\beta_{1j}x_j$ = random slope at site j, ε_{ij} = overall variability, $\varepsilon_{i|j}$ = variability within one site j, ε_j = site variability, ε_s = stimulus variability.

$$y_{i,j,s} = \beta_0 + \beta_1 x_{ij} + \beta_1 x_{ij} z_i + k_1 + k_2 + \beta_{0ij} + \beta_{0i|j} + \beta_{0j} + \beta_{0s} + \beta_{1i} x_{ij} + \beta_{1i} x_{i|j} + \beta_{1j} x_j + \varepsilon_{ij} + \varepsilon_{i|j} + \varepsilon_j + \varepsilon_s$$

Growth Curve Analysis

The overall Proportional Looking Time (PLT) on the head was significantly above zero: ~34% on average (SE = 5.71, p-value < 0.001, CI 95% = 23.48 – 45.86; see SM Table 7 – Deviance

Test – and 8 – beta coefficients and other metrics). The significant, negative Slope indicates that the average PLT decreased across the trial (beta coeff. = -16-16, SE = 3.06, p-value < 0.001, CI 95% = -22.17 – -10.16). The significant and positive Quadratic component indicates a U shape, with attention on the head being high initially (beta coef. = 6.94, SE = 2.79, p-value = 0.02, CI 95% = 1.47 – 12.41).

The average height of PLT (i.e., the intercept) varied by group and age ($X^2 = 20.98$, df = 2, p-value < 0.01); there was a greater increase in the NT intercept (overall looking) with age (beta coef. = 0.33, SE = 0.07, p-value < 0.001, CI 95% = 0.18 – 0.48) than in the AUT intercept (beta coef. =0.19, st. er = 0.07, p-value = 0.01, CI 95% = 0.04 – 0.33). The slope and quadratic components did not significantly vary with group or age (slope: $X^2 = 3.56$, df = 2, p-value = 0.16; quadratic component: $X^2 = 1.11$, df = 2, p-value = 0.57), indicating a similar early orienting response to the face and following rate of decay. The groups did differ by age in the Cubic component ($X^2 = 9.76$, df = 2, p-value = 0.007). The negative beta coefficient for the effect of age in the Cubic component in the NT group (beta coef. = -0.36, SE = 0.12, p-value < 0.001, CI 95% = -0.60 – -0.12) indicates an age-related increase in the late-trial rise in social looking (see Figure 5). The beta coefficient for the effect of age in the AUT group is less negative, and non-significant (beta coef. = -0.23, SE = 0.12, p-value = 0.06, CI 95% = -0.47 – 0.01). Note that the standard error (SE) of the effect of age in the Cubic component in the NT group (0.12) is not close to the beta coefficient (-0.36), suggesting a robust and consistent effect; on the other hand, the SE of the effect of age in the Cubic component in the AUT group (0.12) is half the beta coefficient (-0.23). This means that with increasing age, the Cubic component does not increase as consistently in the AUT vs NT group, resulting in overall diminished social attention at the end of the trial (see Figure 3, panel A).

[Figure 3]

Association between Inter-Individual Differences in Temporal Profiles and Symptoms

We found no significant correlations in the NT group (see Table 2 and 4). Three significant correlations of moderate size remained in the AUT group after Bonferroni correction for multiple comparisons within each hypothesis (see SM 3.9 for details). Specifically, estimates of conditional standard deviations (degree of atypicality) in the quadratic and cubic components of the dynamic temporal profile of attention correlated negatively with the VABS Communication score (Quadratic: r = -0.17, adjusted p-value = 0.03; Cubic: r = -0.17, adjusted p-value = 0.03; see Table 3; Hypothesis 2 of SM 3.9). For higher scores on the VABS Communication, the degree of atypicality for each component was smaller, i.e. more consistent with the average looking behavior of the overall sample. Conversely, lower-scoring individuals tended to have larger standard deviations, thus showing more dispersed temporal profiles (see Figure 3 - panel B, and Figure 4). However, the lack of a significant relation to the beta coefficients (Hypothesis 1 of SM 3.9) indicates that this correlation is not specific to low or high levels of social attention - rather it is the degree of deviation from the overall sample that is most informative. We found no significant correlations with the change in the VABS Communication and Socialization scores between T2 and T1 and neither the beta coefficient and the conditional standard deviations (Hypothesis 3 of SM 3.9; Table 2, 4 and 5).

[Figure 4]

Cluster Analysis and Partial Correlations within the Clusters

The separation score between clusters (silhouette width) justified 2 clusters with maximum separation across all 4 components (Figure 5, panel A, and SM Table 14). In terms of temporal profiles of social attention, Cluster 1 (92% of the AUT group) showed a slight U-shaped profile

(Figure 5, panel B). In contrast, Cluster 2 (8% of the AUT group) markedly showed a steep decrease in head-looking as a function of time (Figure 5, panel B). The average age, FSIQ, SRS T-score, VABS Scores and proportion of valid trials largely overlapped between clusters (see Table 15 of the SM). Performing the same steps in the NT subset did not identify a cluster that resembled Cluster 2 (see Figure 5, Panel C).

We also performed within-cluster age-corrected partial correlations in the AUT group between the beta coefficients of each component and concurrent and prospective SRS-2 and VABS scores (see Figure 5, panel D for SRS-2). We found no significant correlation in Cluster 1, for either concurrent or prospective clinical scores (see Table 16 of the SM). In Cluster 2, the SRS-2 scores at Time 1 were positively correlated with the quadratic component ($r = 0.71$, adjusted p-value = 0.01), indicating a higher curvature in the likelihood of fixating the head as a function of symptom severity.

[Figure 5]

Discussion

In the current study, we investigated how the temporal profile of social attention varies in typical and atypical development across a wide age range in a large group of individuals with and without AUT. The analysis of overall looking time indicated the expected pattern of less looking to social content (faces) in the autistic group as a whole, with a developmental increase in the NT group that was just below the significance level, but did not survive correction for multiple comparisons.

By applying a growth curve analysis (GCA) to spontaneous looking, we were able to tie this change in overall looking to two aspects of the temporal profile that emerged over developmental time in the NT but not the AUT group, enhancing the separation between the groups. The GCA

confirmed greater overall looking that increased with age in the NT relative to the AUT group (effect of intercept, and interaction with age). Further, it indicated that transient orienting to the face (the quadratic component) did not differ between the groups. Finally, older participants with NT were more likely to return to the head after the decay of the initial attention-grabbing effect (represented by the cubic component), possibly reflecting successive components of social attention. This effect was substantially diminished in autistic adults, indicating the involvement of different attentional components separated by distinct temporal dynamics. We exclude that this may be explained by data missingness, given its non-significant contribution as covariate, and its pattern of variation over time, non-overlapping with the main model (age by group interaction with the quadratic component in the former, age by group interaction with the cubic component in the latter). Rather, this pattern is consistent with the view that stimulus-driven social orienting is intact (32), but endogenous deployment of attention to social features is diminished in ASD (33). In a dimensional analysis, the degree to which autistic individuals were more divergent in their profile of social attention correlated to poorer skills in adaptive communication. Further, the cluster analysis identified a smaller subgroup of autistic participants who showed particularly atypical social attention and a strong correlation with symptom severity. Taken together, our results provide mechanistic insights into the nature of social visual attention alterations in ASD, and indicate methodologies that can improve our ability to detect meaningful links with core symptomatology.

Early orienting to faces

The initial high proportion of orienting to the head in both the AUT and NT groups is likely to be related to the transient effect of a stimulus-driven shift of attention. This rapid orienting to social stimuli could be mediated by both domain-specific and domain-general mechanisms.

Specifically, initial fixations to a static scene can be driven by bottom-up regulation guided by salience, i.e., the distribution of local feature differences (34,35). The head is naturally rich in high-contrast features – i.e., zones of differential luminance and color –, and elements of different orientation – e.g., nose and mouth (36). Alternatively, initial face-orienting may be driven by domain-specific innate biases that channel emerging specialization under the evolutionary pressure of orienting towards conspecifics/face-like stimuli (19). In line with previous studies, our finding points in the direction of a preserved rapid orienting to faces in ASD (37).

Later Trial Dynamics

Later in a trial, fixations are more heavily influenced by factors such as motivation, relevance and experience (38). For example, Hedger et al. (16) found that prolonged observation of social information was associated with a higher load of empathic traits in a typical population. Within our study, the developmental pattern observed in the NT group may thus reflect motivation to re-engage with faces. The lag of this developmental progression in the AUT group may reflect altered motivation to engage with social stimuli (3), and may relate to ASD-specific alterations in a wider spectrum of social behaviors, e.g., gaze-cueing and joint attention (39). Indeed, we observed concurrent relations with Communication skills (see below). Alternatively, our results may reflect more general difficulties with the control of oculomotor function (40) or other domain-general features of attention (41). Habituation or boredom is less likely to explain the pattern of results, since there were no group differences in overall looking times to the stimuli. Also, the average PLT at the body, the background people and the non-AOI areas is mostly flat in both groups (see Figure 1 of the SM), probably due to inconsistent exploration that flattens out with averaging. Independent replications will be required to dissect these possibilities.

Individual differences and Stratification

When we examined individual differences within the autistic cohort, progressively greater atypicality of social attention within a trial was associated with poorer VABS Communication scores. In terms of fixation behavior, this means that individuals who tend to be more consistent with the average looking behavior of the overall sample possess more adaptive communication skills. The VABS Communication scale focuses on skills that involve verbal communication (e.g., “Says both the month and day of his birthday”, “Gives complex directions with three or more steps”). Therefore, higher scores on the communication scale imply that the individual is verbal and uses language for everyday communication. Previous research has highlighted that autistic people with stronger language skills have better outcomes including better social adaptive skills (42), providing a possible link to social attention. Significant relations between communication skills and *degree of atypicality* in social attention are difficult to mechanistically interpret because atypicality can reflect both less or more looking to faces, and eye tracking measures did not predict change in Communication scores over time. However, they may hold promise for the use of refined temporally-resolved measures of social attention to help with symptom tracking in observational or intervention studies.

The characterization of the data-driven clusters may help understand whether heterogeneity in the *nature of atypicality* is meaningfully related to symptom variation. This revealed a distinct subgroup of autistic individuals who consistently showed a steeper decline of the fixation probability to the head; within this group, a higher curvature strongly correlated with a dimensional measure of social symptom severity, the SRS-2. The SRS-2 measures a wide range of symptoms (e.g., from reduced eye-contact, to poor coordination and lacking self-confidence) that may significantly impact social functioning. These characteristics may make social

encounters difficult and sporadic, providing less experience with others (and their faces). The attentive pattern towards the face observed in this subgroup could reflect social withdrawal (14) or a more pronounced reduction self-directed experiences; dissecting these possibilities deserves further investigation.

Limitations

The use of static images enables the temporal profile of attention to be driven by intrinsic motivation rather than extrinsic cues, a valuable approach given our findings. However, static images do not necessarily capture highly dynamic, everyday social encounters. Furthermore, it has been reported that dynamic scenes produce larger group differences when studying typical and atypical development (43). Future research could use the GCA to examine the effect of specific events (e.g., verbal and gaze cues) that have been found to influence looking behavior in AUT (44), and define time-windows of influence in terms of latency, duration and synchrony. Another limitation was that our sample size was not large enough to model variable slopes for each stimulus within the random effect in addition to our inclusion of variable intercepts; this choice implies the assumption that the designed AOI, presented in a slightly different context, would elicit similar trends of visual attention over time. This assumption should be tested in future research with larger samples. Another notable limitation is that whilst measures of visual attention were related to concurrent aspects of communication skill and autistic symptoms, we did not identify any predictive relations to change in clinical symptomatology. It may be more likely that atypicalities in social attention track behavioral symptomatology, rather than mechanistically contribute to its emergence or consolidation. Furthermore, a promising future avenue of research resides in determining the combined and/or specific contribution of comorbidities, such as ADHD, to social attention in ASD (45). However, there was no evidence

that general attentiveness or the presence of ADHD group differences confounded the present results. Finally, sex differences will be important to pursue in future work; here sex was proved as a significant covariate, but the interaction with factors such as age and pubertal status should be examined across multiple settings. Addressing this question fully requires longitudinal research earlier in development.

Conclusion

This study tested the application of GCA to understand temporal profiles of social attention in autism over a wide age range, uncovering a pattern of a late occurring return of attention to the face emerging over developmental time in neurotypical adults, but not the AUT group. Variations in motivated social attention were related to adaptive communicative skills in the wider AUT group, and were correlated with symptom severity within a data-derived subgroup. These results resonate with a model of ASD that implicates altered exploration of the social environment, but are less consistent with the proposal that early-emerging differences in social attention contribute to later behavioral trajectories (10,46). Results raise the possibility that temporally-sensitive eye-tracking-based measures have the potential to provide objective measures of symptom profiles in clinical contexts.

Acknowledgements

We thank the participants and the families that took part in the LEAP Study.

We report hereafter the complete list of the authors included in the EU-AIMS Leap Group: Jumana Ahmad, Sara Ambrosino, Tobias Banaschewski, Simon Baron-Cohen, Sarah Baumeister, Christian F. Beckmann, Sven Bölte, Thomas Bourgeron, Carsten Bours, Michael Brammer, Daniel Brandeis, Claudia Brogna, Yvette de Bruijn, Ineke Cornelissen, Daisy Crawley, Flavio Dell'Acqua, Guillaume Dumas, Sarah Durston, Christine Ecker, Jessica Faulkner, Vincent Frouin, Pilar Garcés, David Goyard, Lindsay Ham, Joerg Hipp, Rosemary Holt, Meng-Chuan Lai, Xavier Liogier D'ardhuy, Michael V. Lombardo, David J. Lythgoe, René Mandl, Andre Marquand, Maarten Mennes, Andreas Meyer-Lindenberg, Carolin Moessnang, Nico Mueller, Declan G.M. Murphy, Bethany Oakley, Laurence O'Dwyer, Marianne Oldehinkel, Bob Oranje, Gahan Pandina, Antonio M. Persico, Barbara Ruggeri, Amber Ruigrok, Jessica Sabet, Roberto Sacco, Antonia San José Cáceres, Emily Simonoff, Will Spooren, Roberto Toro, Heike Tost, Jack Waldman, Steve C.R. Williams, Caroline Wooldridge, and Marcel P. Zwiers.

Financial Disclosures

Dr. Luke Mason, Professor Tony Charman, Professor Mark Johnson and Professor Emily Jones reported having received funding from the UK Medical Research Council. Dr. Luke Mason, Professor Tony Charman, Dr. Julian Tillman, Dr. Eva Loth, Dr. Hannah Hayward, Dr. Jan Buitelaar, Professor Mark Johnson, Professor Emily Jones and the members of the EU-AIMS LEAP Group reported having received funding from the Innovative Medicines Initiative Joint Undertaking, under grant agreement No 115300, for the project EU-AIMS. Dr. Teresa Del Bianco, Dr. Luke Mason, Professor Tony Charman, Dr. Julian Tillman, Dr. Eva Loth, Dr. Hannah Hayward, Professor Jan Buitelaar, Professor Mark Johnson and Professor Emily Jones reported having received funding from the Innovative Medicines Initiative 2 Joint Undertaking, under grant agreement No 777394, for the project AIMS-2-TRIALS. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program, EFPIA, AUTISM SPEAKS, Autistica, and SFARI. Prof. Tony Charman has served as a paid consultant to F. Hoffmann-La Roche Ltd. and Servier and received royalties from Sage Publications and Guilford Publications. Dr. Julian Tillmann is a paid consultant to F. Hoffmann-La Roche AG. Jan Buitelaar has been in the past 3 years a consultant to/member of advisory board of/and/or speaker for Janssen Cilag BV, Eli Lilly, Lundbeck, Shire, Roche, Novartis, Medice and Servier. He is not an employee of any of these companies and not a stock shareholder of any of these companies. He has no other financial or material support, including expert testimony, patents, and royalties. Professor Frederick Shic is a paid consultant and scientific advisor to F. Hoffmann-La Roche AG and Janssen Pharmaceutica. Sven Bölte receives royalties for the German and Swedish KOTDAKT manuals and adaptations of the ADI-R, ADOS, and SRS from Hogrefe Publishers. Bölte has in the last 3 years acted as an author, consultant or lecturer for Shire,

Medice, Roche, Eli Lilly, Prima Psychiatry, GLGroup, System Analytic, Kompetento, Expo Medica, Prophase and receives royalties for text books and diagnostic tolls from Huber/Hogrefe, Kohlhammer and UTB.

Lindsay Ham, Xavier Liogier D'Ardhuy, Joerg Hipp, Pilar Garcés and Will Spooren are employees at F. Hoffmann-La Roche LTD. Gahan Pandina is an employee at Janssen. Andreas Meyer-Lindenberg has received consultant fees and travel expenses from Alexza Pharmaceuticals, AstraZeneca, Bristol-Myers Squibb, Defined Health, Decision Resources, Desitin Arzneimittel, Elsevier, F. Hoffmann-La Roche, Gerson Lehrman Group, Grupo Ferrer, Les Laboratoires Servier, Lilly Deutschland, Lundbeck Foundation, Outcome Sciences, Outcome Europe, PriceSpective, and Roche Pharma and has received speaker's fees from Abbott, AstraZeneca, BASF, Bristol-Myers Squibb, GlaxoSmithKline, Janssen-Cilag, Lundbeck, Pfizer Pharma, and Servier Deutschland. Tobias Banaschewski has served in an advisory or consultancy role for Actelion, Hexal Pharma, Lilly, Medice, Novartis, Oxford outcomes, Otsuka, PCM scientific, Shire and Viforpharma. He received conference support or speaker's fee by Medice, Novartis and Shire. He is/has been involved in clinical trials conducted by Shire and Viforpharma. He received royalties from Hogrefe, Kohlhammer, CIP Medien, and Oxford University Press. The present work is unrelated to the above grants and relationships.

The other authors declare that they have no competing interests.

References

1. Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z, *et al.* (2018): Prevalence of autism spectrum disorder among children aged 8 Years - Autism and developmental disabilities monitoring network, 11 Sites, United States, 2014. *MMWR Surveill Summ*. https://doi.org/10.15585/mmwr.ss6706a1

2. American Psychiatric Association (2013): *Diagnostic and Statistical Manual of Mental Disorders (5th Ed.)*. Washington, DC.

3. Dawson G, Webb SJ, McPartland J (2005): Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies. *Dev Neuropsychol* 27: 403–424.

4. Jones W, Klin A (2013): Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature* 504: 427–431.

5. Chawarska K, Volkmar F, Klin A (2010): Limited attentional bias for faces in toddlers with autism spectrum disorders. *Arch Gen Psychiatry* 67: 178–85.

6. Pelphrey KA, Sasson NJ, Reznick JS, Paul G, Goldman BD, Piven J (2002): Visual Scanning of Faces in Autism. *J Autism Dev Disord* 32: 249–261.

7. Grossmann T, Johnson MH, Lloyd-Fox S, Blasi A, Deligianni F, Elwell C, Csibra G (2008): Early cortical specialization for face-to-face communication in human infants. *Proc R Soc London B Biol Sci* 275: 2803–2811.

8. Meltzoff AN, Kuhl PK, Movellan J, Sejnowski TJ (2009): Foundations for a New Science of Learning. *Science (80-)* 325: 284–288.

9. Gliga T, Jones EJH, Bedford R, Charman T, Johnson MH (2014): From early markers to neuro-developmental mechanisms of autism. *Dev Rev* 34: 189–207.

10. Chevallier C, Kohls G, Troiani V, Brodkin ES, Schultz RT (2012): The Social Motivation Theory of Autism. *Trends Cogn Sci* 16: 231–239.

11. Chita-Tegmark M (2016): Social attention in ASD : A review and meta-analysis of eye-tracking studies. *Res Dev Disabil* 48: 79–93.

12. Frazier TW, Strauss M, Klingemier EW, Zetzer EE, Hardan AY, Eng C, Youngstrom EA (2017, July 1): 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*, vol. 56. Elsevier, pp 546–555.

13. Mirman D, Dixon JA, Magnuson JS (2008): Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *J Mem Lang* 59: 475–494.

14. Schofield CA, Inhoff AW, Coles ME (2013): Time-course of attention biases in social phobia. *J Anxiety Disord* 27: 661–669.

15. Shimojo S, Simion C, Shimojo E, Scheier C (2003): Gaze bias both reflects and influences preference. *Nat Neurosci*. https://doi.org/10.1038/nn1150

16. Hedger N, Haffey A, McSorley E, Chakrabarti B (2018): Empathy modulates the temporal structure of social attention. *Proc R Soc B Biol Sci* 285: 20181716.

17. Smith TJ, Mital PK (2013): Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *J Vis* 13: 16–16.

18. Posner MI (1980): Orienting of attention. *Q J Exp Psychol* 32: 3–25.

19. Johnson MH, Senju A, Tomalski P (2015): The two-process theory of face processing: Modifications based on two decades of data from infants and adults. *Neuroscience and Biobehavioral Reviews*, vol. 50. pp 169–179.

20. Langton SRH, Law AS, Burton AM, Schweinberger SR (2008): Attention capture by faces. *Cognition* 107: 330–342.

21. Falck-Ytter T, Von Hofsten C, Gillberg C, Fernell E (2013): Visualization and analysis of eye movement data from children with typical and atypical development. *J Autism Dev Disord* 43: 2249–2258.

22. Johnson MH (2014, January 6): Autism: Demise of the innate social orienting hypothesis. *Current Biology*, vol. 24. Cell Press. https://doi.org/10.1016/j.cub.2013.11.021

23. Ami Klin, Warren Jones, Robert T. Schultz, Fred D. Volkmar, F. R. Volkmar (2005): The enactive mind - From actions to cognitions: Lessons from Autism, Hoboken, N.

24. Pelphrey KA, Shultz S, Hudac CM, Vander Wyk BC (2011, June 1): Research review: Constraining heterogeneity: The social brain and its development in autism spectrum disorder. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 52. John Wiley & Sons, Ltd, pp 631–644.

25. Johnson MH (2017, June 1): Autism as an adaptive common variant pathway for human brain development. *Developmental Cognitive Neuroscience*, vol. 25. Elsevier Ltd, pp 5–11.

26. Harrell Jr FE (2015): *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis.* Springer.

27. Pickles A, Angold A (2003): Natural categories or fundamental dimensions: On carving nature at the joints and the rearticulation of psychopathology. *Dev Psychopathol* 15: 529–551.

28. Loth E, Charman T, Mason L, Tillmann J, Jones EJH, Wooldridge C, *et al.* (2017): The EU-AIMS Longitudinal European Autism Project (LEAP): design and methodologies to identify and validate stratification biomarkers for autism spectrum disorders. *Mol Autism* 8: 24.

29. Webb SJ, Shic F, Murias M, Sugar CA, Naples AJ, Barney E, *et al.* (2020): Biomarker Acquisition and Quality Control for Multi-Site Studies: The Autism Biomarkers Consortium for Clinical Trials. *Front Integr Neurosci* 13: 71.

30. Shic F, Naples A, Barney E, Chang SA, Li B, McAllister T, *et al.* (2018): *ABC-CT Data Acquisition and Analytic Core ET Main Study Derived Results Manual*, Version 1. Seattle WA.

31. Niehorster DC, Cornelissen THW, Holmqvist K, Hooge ITC, Hessels RS (2018): What to expect from your remote eye-tracker when participants are unrestrained. *Behav Res Methods* 50: 213–227.

32. Guillon Q, Hadjikhani N, Baduel S, Rogé B (2014, May): Visual social attention in autism spectrum disorder: Insights from eye tracking studies. *Neuroscience and Biobehavioral Reviews*, vol. 42. pp 279–297.

33. Joens EJH, Gliga T, Bedford R, Charman T, Johnson MH (2014): Developmental pathways to autism: A review of prospective studies of infants at risk. *Neurosci Biobehav Rev* 39: 1–33.

34. Amso D, Haas S, Tenenbaum E, Markant J, Sheinkopf SJ (2014): Bottom-up attention orienting in young children with autism. *J Autism Dev Disord* 44: 664–73.

35. Freeth M, Foulsham T, Chapman P (2011): The influence of visual saliency on fixation patterns in individuals with Autism Spectrum Disorders. *Neuropsychologia* 49: 156–160.

36. Simion F, Di Giorgio E (2015): Face perception and processing in early infancy: Inborn predispositions and developmental changes. *Front Psychol*. https://doi.org/10.3389/fpsyg.2015.00969

37. Del Bianco T, Mazzoni N, Bentenuto A, Venuti P (2018): An Investigation of Attention to

Faces and Eyes: Looking Time Is Task-Dependent in Autism Spectrum Disorder. *Front Psychol* 9. https://doi.org/10.3389/fpsyg.2018.02629

38. Orquin JL, Bagger MP, Mueller Loose S (2013): Learning affects top down and bottom up modulation of eye movements in decision making. *Judgm Decis Mak* 8: 700–716.

39. Vivanti G, Fanning PAJ, Hocking DR, Sievers S, Dissanayake C (2017): Social attention, joint attention and sustained attention in autism spectrum disorder and williams syndrome: Convergences and divergences. *J Autism Dev Disord* 47: 1866–1877.

40. Falck-Ytter T, Bölte S, Gredebäck G (2013): Eye tracking in early autism research. *J Neurodev Disord* 5: 28.

41. Elsabbagh M, Volein A, Holmboe K, Tucker L, Csibra G, Baron-Cohen S, *et al.* (2009): Visual orienting in the early broader autism phenotype: disengagement and facilitation. *J Child Psychol Psychiatry* 50: 637–642.

42. Suh J, Eigsti IM, Canfield A, Irvine C, Kelley E, Naigles LR, Fein D (2016): Language representation and language use in children with optimal outcomes from ASD. *Innovative Investigations of Language in Autism Spectrum Disorder*. Washington: American Psychological Association, pp 225–244.

43. Harrop C, Jones D, Zheng S, Nowell S, Schultz R, Parish-Morris J (2019): Visual attention to faces in children with autism spectrum disorder: Are there sex differences? *Mol Autism* 10: 28.

44. Chawarska K, Macari S, Shic F (2012): Context modulates attention to social scenes in toddlers with autism. *J Child Psychol Psychiatry* 53. https://doi.org/10.1111/j.1469-7610.2012.02538.x

45. Johnson MH, Gliga T, Jones E, Charman T (2015): Annual research review: Infant

development, autism, and ADHD - Early pathways to emerging disorders. *J Child Psychol Psychiatry Allied Discip* 56: 228–247.

46. Salley B, Colombo J (2016): Conceptualizing Social Attention in Developmental Research. *Social Development*, vol. 25. NIH Public Access, pp 687–703.

Footnotes

[1] Alternative approaches to examining within-trial temporal dynamics exist, such as modelling categorical time bins in a traditional ANOVA. However, any follow-up tests on individual time bins must be corrected for multiple comparisons, thus imposing a tradeoff between temporal resolution and statistical power. The modelling approach used in GCA allows us to represent attentional dynamics using a mathematical function, and to test differences in the shape of these dynamics in one model.

[2] Adjusted p-value = p-value/number of comparisons (12 = 4 terms * 3 variables for each of the 3 independent hypotheses as specified in SM 3.9)

Figures and Tables

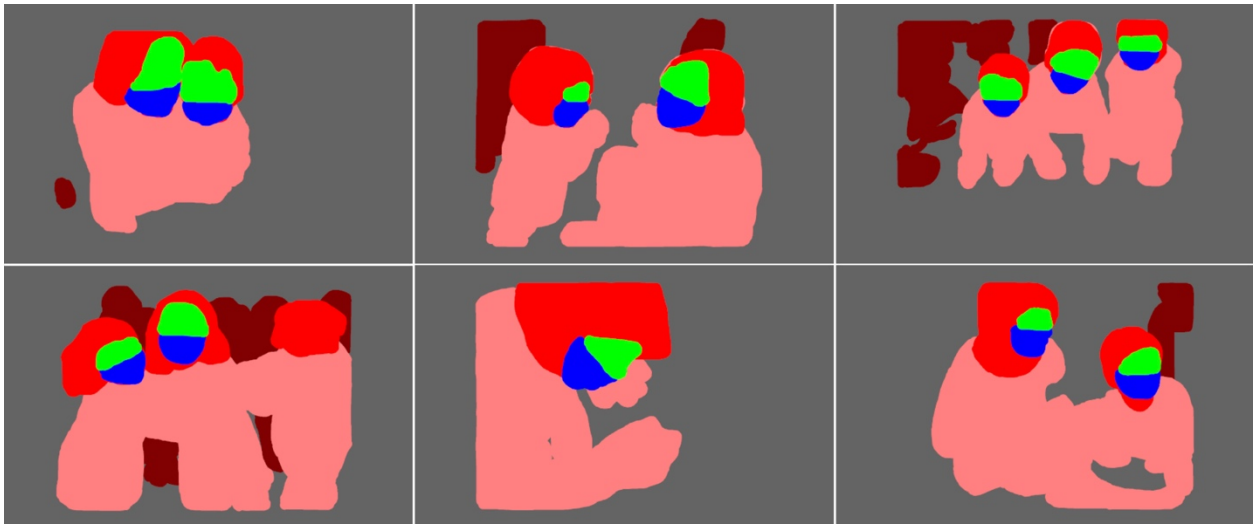

Figure 1: Areas of Interest (AOIs) drawn on the elements of the stimulus. Red = hair, green = upper face, blue = lower face, pink = body, maroon = background people. The red, green and blue areas form the "head".

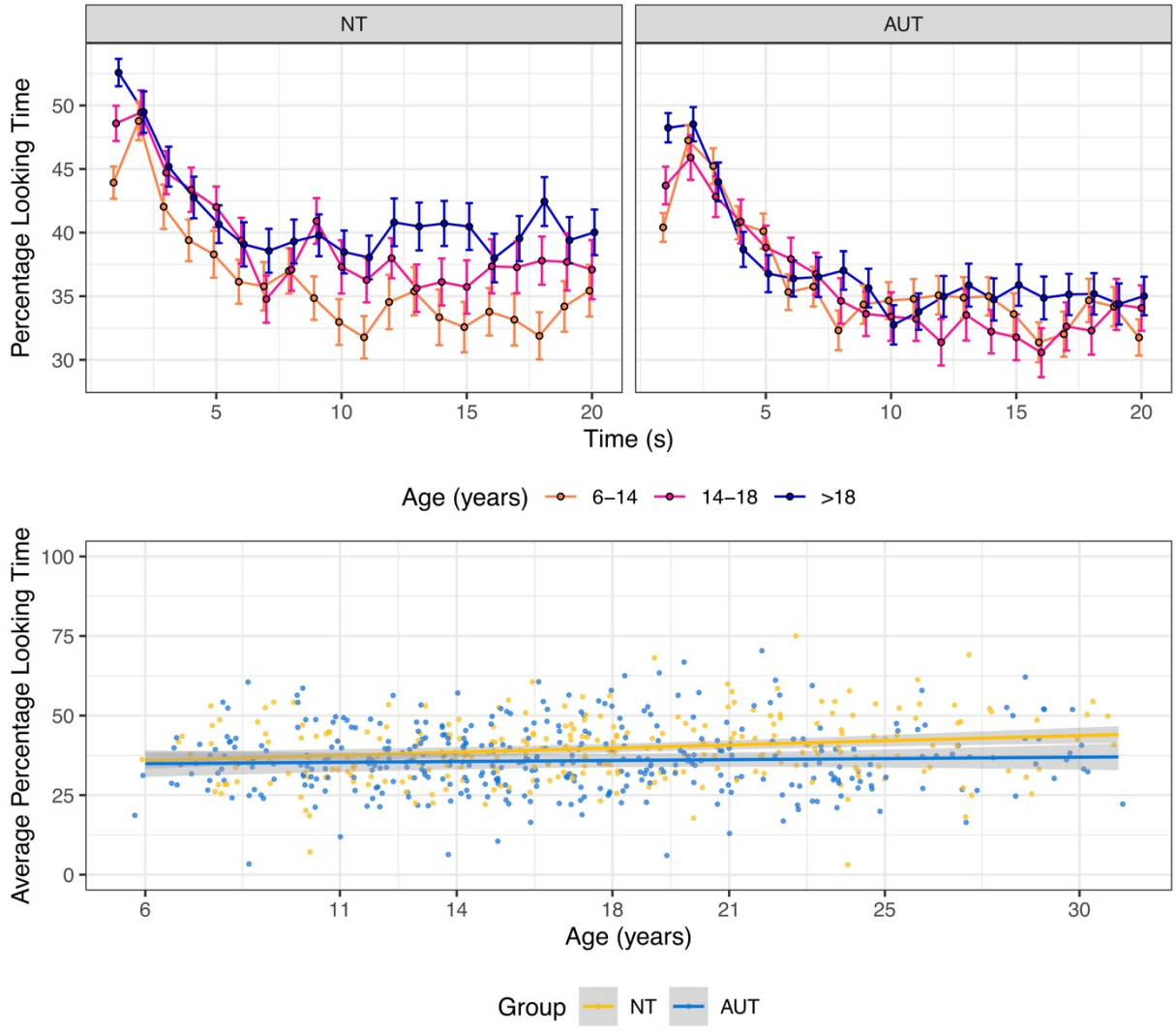

Figure 2: A) group-level averages and 95% confidence intervals of the observed proportional looking time at the head over the course of the trial by group and putative age-group B) change in linear slope by group and age (interaction term) with 95% confidence intervals (line).

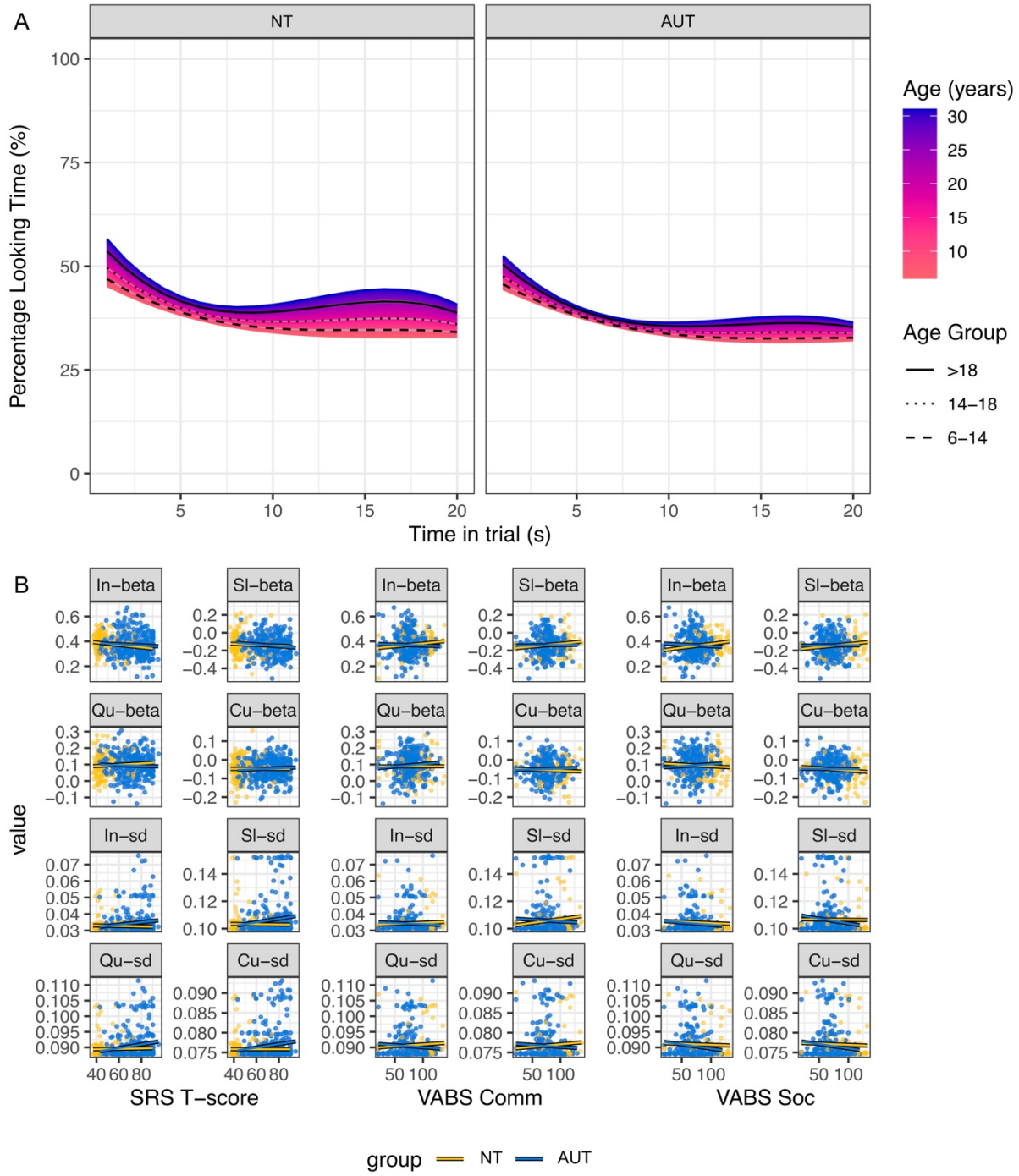

Figure 3: A) fitted curves representing the relationship between PLT, group and age; changes of the curve represent changes of the probability to fixate the head over the course of the trial at each unit of change of age (1 year, 6-30 years). The curves are estimated through a bootstrapped distribution for the random and the fixed effects (1000 simulations). The superimposed lines represent the predicted intervals averaged across three putative age groups with 95% confidence intervals B) scatterplots and concurrent linear relationships between beta coefficients and clinical variables at Time 1 and conditional standard deviations and clinical variables at Time 1.

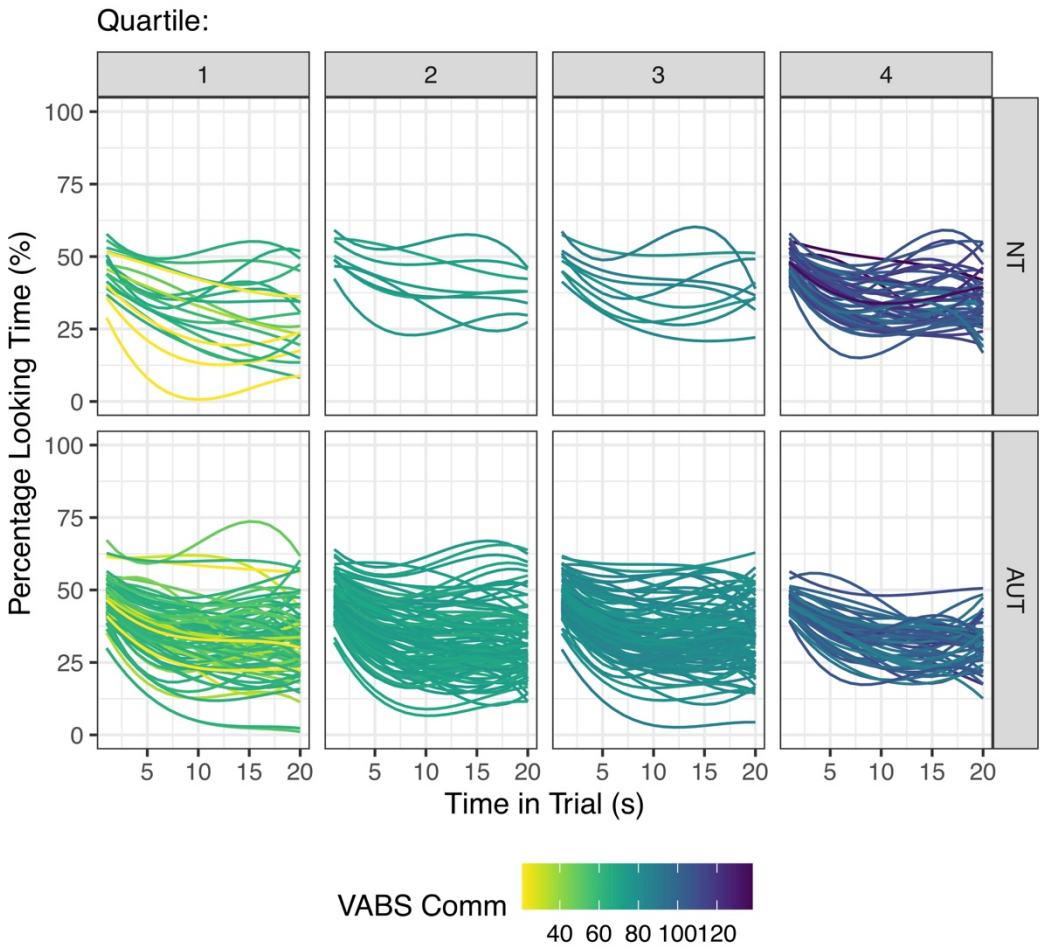

Figure 4: The curves represent the individual predicted values for each autistic participant (bootstrapped distribution for the random and the fixed effects with 1000 simulations), divided and sorted by the quartile proportions of the VABS Communication Scores. The color codes indicate higher/lower VABS Communication Scores.

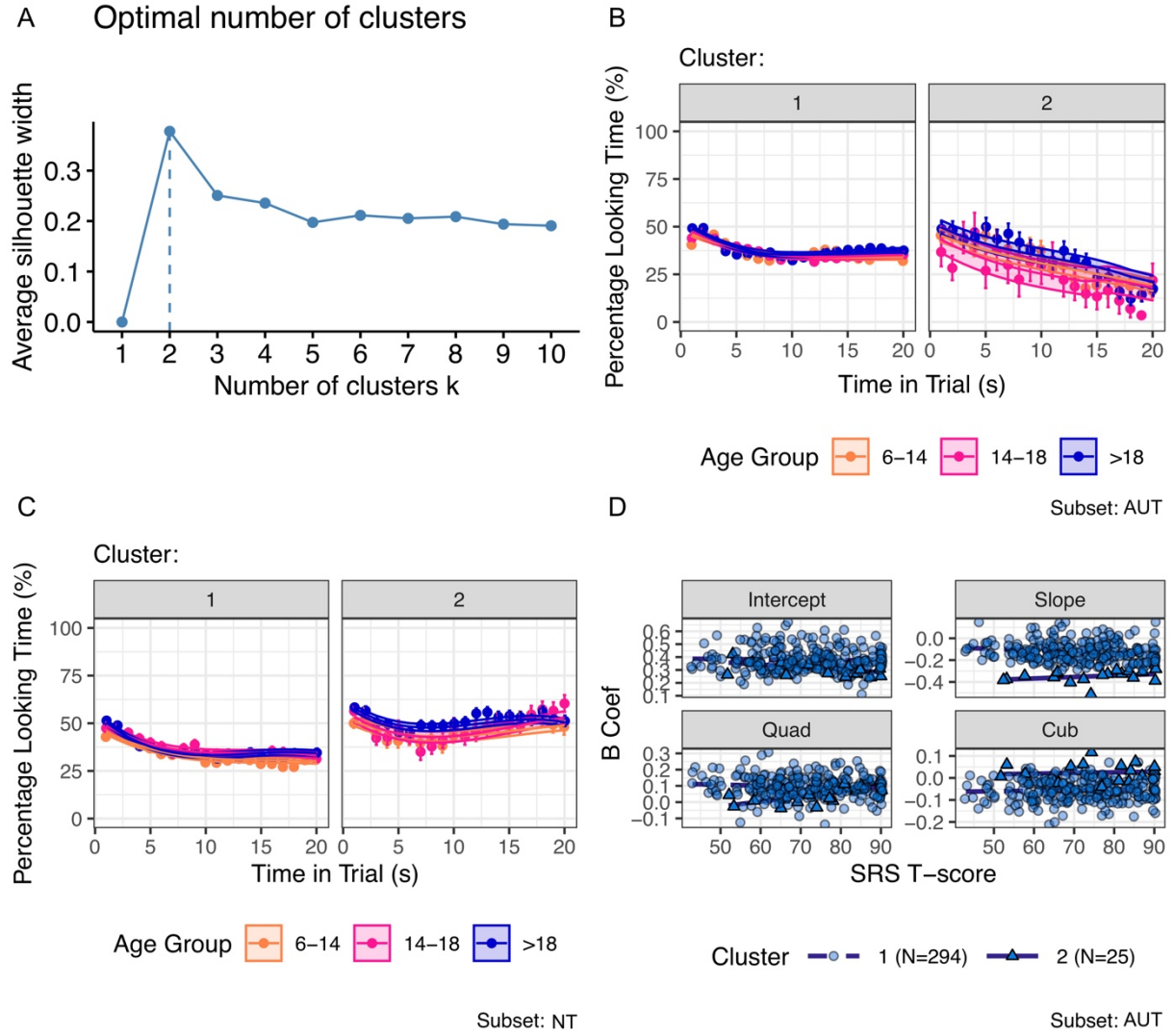

Figure 5: A) silhouette plot B) fitted curves and prediction intervals (bootstrapped distribution for the random and the fixed effects with 1000 simulations) of percentage looking within the two clusters and by age across the duration of the trial – ASD group C) fitted curves and prediction intervals (bootstrapped distribution for the random and the fixed effects with 1000 simulations) of percentage looking within the two clusters and by age across the duration of the trial – NT group D) scatterplots and concurrent linear relationships between the beta coefficients and the SRS-2 within the clusters

Table 1: descriptive statistics of the sample (N = number, SD = standard deviation, FSIQ = full scale intelligence quotient, NT = neurotypical, AUT = autism)

Group			*NT*		*AUT*		*Between Group Chi-Squared Statistics (p-value)*	*Group Difference Cohen's D Effect Size***
		Time						
N total		*1*	*273*		*377*			
		2	*243*		*346*			
N females		*1*	*94*	*M:F*	*105*	*M:F*	*2.92 (0.09)*	
				1.90		*2.5*		
		2	*88*	*1.76*	*99*	*2.49*	*3.46 (0.06)*	
Mean Age (SD)		*1*	*17.35 (5.89)*		*16.75 (5.68)*			*0.03*
		2	*16.78 (5.71)*		*16.26 (5.44)*			*0.03*
Mean FSIQ (SD)*		*1*	*104.28 (18.79)*		*97.64 (19.81)*			*0.28*
Mean Proportional Looking Time on Screen (SD)		*1*	*0.84 (0.16)*		*0.80 (0.18)*			*0.002*
Mean SRS T-Score (SD)		*1*	*48.38 (9.45)*		*72.25 (11.78)*			*1.29*
		2	*46.12 (8.15)*		*73.11 (10.86)*			*1.58*
VABS	*Mean Communication Score (SD)*	*1*	*89.75 (25.79)*		*74.64 (17.37)*			*0.55*
		2	*80.49 (35.58)*		*73.74 (17.22)*			*0.20*
	Mean Socialization Score (SD)	*1*	*95.37 (25.08)*		*69.90 (16.56)*			*0.94*
		2	*90.82 (31.60)*		*74.19 (16.79)*			*0.54*

*Mean Dawba ADHD Score (SD)***	*1*	*0.46 (1.08)*	*2.00 (1.60)*		*0.09*

** The FSIQ was standardized across countries; see the SM for additional information.*

*** The Development and Well-Being Assessment (DAWBA) was used to screen for comorbidities in the current sample, and is reported in terms of effect size given the possible influence of attention (but it is not otherwise mentioned elsewhere in this work).*

**** Formula:* $\frac{m_1 - m_2}{s}$ *with m = group mean, and s = pooled standard deviation =* $\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{(n_1-n_2)-2}}$.
Effect sizes< 0.2, between 0.5 ~ 0.8, >0.8 correspond, respectively, to small, medium and big effects.

Table 2: degrees of freedom (DoF), sum of squares (Sum Sq), mean squares (Mean Sq), F-value, p-value, p-value adjusted (with Bonferroni correction), and effect size (Cohen's F) of the analysis of variance of the proportional looking time by group and age

	DoF	Sum Sq	Mean Sq	F-Value	P-value	P-value Adjusted	Cohen's F
Group	1	0.13	0.11	11.17	<0.001	0.001	0.13
Age	1	0.10	0.10	9.70	0.001	0.005	0.12
Sex	1	0.05	0.05	4.99	0.02	0.05	0.10
FSIQ (Scaled)	1	0.06	0.06	6.18	0.01	-	0.10
Group:Age	1	0.04	0.04	4.15	0.04	0.08	0.08
Residuals	644	6.51	0.01				

Table 3: available data points (N), Spearman correlation coefficients (rho), nominal p-values (p-value) and Bonferroni-adjusted p-values (p-adjusted) between the beta coefficients and the clinical variables at Time 1 (see hypothesis 1 in SM 3.9).

variable	**group**	**component**	**N**	**rho**	**p-value**	**p-adjusted**
SRS-2 T-score (Time 1)	NT	Intercept	146	-0.12	0.14	1.00
		Slope	146	-0.07	0.41	1.00
		Quad	146	0.07	0.44	1.00
		Cub	146	<0.001	0.96	1.00
	AUT	Intercept	313	-0.02	0.70	1.00
		Slope	313	-0.09	0.10	1.00

		Quad	313	-0.02	0.71	1.00
		Cub	313	0.05	0.41	1.00
VABS Comm (Time 1)	NT	Intercept	87	0.11	0.31	1.00
		Slope	87	0.1	0.36	1.00
		Quad	87	-0.07	0.54	1.00
		Cub	87	-0.05	0.67	1.00
	AUT	Intercept	325	-0.01	0.91	1.00
		Slope	325	0.06	0.25	1.00
		Quad	325	0.09	0.13	1.00
		Cub	325	0.01	0.87	1.00
VABS Soc (Time 1)	NT	Intercept	86	0.05	0.64	1.00
		Slope	86	0.05	0.62	1.00
		Quad	86	-0.11	0.33	1.00
		Cub	86	-0.08	0.45	1.00
	AUT	Intercept	321	-0.03	0.62	1.00
		Slope	321	0.03	0.58	1.00
		Quad	321	0.03	0.58	1.00
		Cub	321	-0.02	0.77	1.00

Table 4: available data points (N), Spearman correlations coefficients (rho), nominal p-values (p-value) and Bonferroni-adjusted p-values (p-adjusted) between the conditional standard deviations and the clinical variables at Time 1 (see hypothesis 2 in SM 3.9). Significant correlations after Bonferroni correction are marked with ''*

variable	**group**	**term**	**N**	**rho**	**p-value**	**p-adjusted**

SRS-2 T-score (Time 1)	NT	Intercept	146	0.05	0.54	1.00
		Slope	146	0.1	0.21	1.00
		Quad	146	0.12	0.14	1.00
		Cub	146	0.12	0.14	1.00
	AUT	Intercept	313	0.04	0.51	1.00
		Slope	313	0.05	0.35	1.00
		Quad	313	0.08	0.15	1.00
		Cub	313	0.05	0.34	1.00
VABS Comm (Time 1)	NT	Intercept	87	-0.08	0.44	1.00
		Slope	87	-0.1	0.37	1.00
		Quad	87	-0.1	0.36	1.00
		Cub	87	-0.09	0.42	1.00
	AUT	Intercept	325	-0.12	0.03	0.30
		Slope	325	-0.16	< 0.001	0.05
		Quad	**325**	**-0.17**	**< 0.001**	**0.03***
		Cub	**325**	**-0.17**	**< 0.001**	**0.03***
VABS Soc (Time 1)	NT	Intercept	86	-0.08	0.44	1.00
		Slope	86	-0.11	0.34	1.00
		Quad	86	-0.15	0.17	1.00
		Cub	86	-0.11	0.31	1.00
	AUT	Intercept	321	-0.07	0.22	1.00
		Slope	321	-0.11	0.06	0.68
		Quad	321	-0.13	0.02	0.29

		Cub	321	-0.1	0.07	0.85

Table 5: available data points (N), Spearman correlations coefficients (rho), nominal p-values (p-value) and Bonferroni-adjusted p-values (p-adjusted) between the beta coefficients and the conditional standard deviations and the difference in the clinical variables scores between Time 2 and Time 1 (see hypothesis 3 in SM 3.9).

variable	**group**	**term**	**N**	**rho**	**p-value**	**p-adjusted**
VABS Comm (Time 2 – Time 1)	AUT	Beta Intercept	227	-0.08	0.26	1.00
		Beta Slope	227	0.08	0.24	1.00
		Beta Quad	227	0.15	0.03	0.34
		Beta Cub	227	0.02	0.76	1.00
VABS Soc (Time 2 – Time 1)		Beta Intercept	220	-0.04	0.55	1.00
		Beta Slope	220	-0.04	0.59	1.00
		Beta Quad	220	0.05	0.43	1.00
		Beta Cub	220	0.08	0.22	1.00
VABS Comm (Time 2 – Time 1)	AUT	Intercept SD	227	0.11	0.09	0.27
		Slope SD	227	0.08	0.24	0.72
		Quad SD	227	0.08	0.21	0.63
		Cub SD	227	0.09	0.20	0.60
		Intercept SD	220	-0.04	0.56	1.00

VABS Soc (Time 2 – Time 1)		Slope SD	220	-0.03	0.66	1.00
		Quad SD	220	-0.02	0.79	1.00
		Cub SD	220	-0.03	0.62	1.00

Supplementary Material

1. Sample

1.1. Recruitment Sites

The participants were recruited at Cambridge University, Cambridge, United Kingdom (UCAM), King's College London, London, United Kingdom (KCL), the Central Institute of Mental Health, Mannheim, Germany (CIMH), Radboud University Nijmegen Medical Center, The Netherlands (RUNMC), Universitair Medisch Centrum Utrecht, The Netherlands (UMCU) and Università Campus Bio-Medico, Rome, Italy (UCBM).

1.2. Participants Inclusion and Exclusion Criteria

Autistic participants needed a previous clinical diagnosis of ASD (according to DSM-IV/DSM-IV-TR/DSM-V/ICD-10). All participants were in the age range 6-30, and did not report uncorrected hearing or visual impairments, a history of alcohol and/or substance abuse, presence of MRI contraindications (e.g. metal implants, braces, claustrophobia), or failed to give informed written consent. Participants with typical development (NT) did not report any diagnosis of psychiatric disorders, and did not score ≥ 70 on the self-report (adult) or parent-report form (adolescents and children) of the Social Responsiveness Scale (1,3). Individuals with moderate or severe intellectual disability (FSIQ<50) were excluded from both cohorts.

1.2.1. Intellectual ability

FSIQ was assessed through the Wechsler Abbreviated Scales of Intelligence – 2nd Edition, WASI-II, or the four-subtest short forms of the German, Dutch or Italian WISC-III/IV for children or WAIS-III/IV for adults. The FSIQ was standardized across sites from vocabulary, similarities, matrix reasoning and block design (5). The standard estimates were obtained from national population norms for each site.

1.3. Valid/Invalid Eye-Tracking Data

Of the total 764 participants, 13% (across the different sites) did not obtain usable data and were therefore excluded from the analysis. Table 1 contains the counts and percentages of valid/invalid data categorized by the reason reported on the local electronic Case Report Form.

Table 1: count and percentage of participants that had valid/unvalid eye-tracking data, and were therefore excluded from the analysis. We report the main cause for not collecting/analysing the eye-tracking data.

Condition	Reason	Count	Percent

Valid		664	86.91%
Unvalid	Technical Fault	18	2.35%
	Raw missing data	24	3.14%
	Eye-tracking not recorded	34	4.45%
	Task not reached in battery	5	0.65%
	Task not presented	13	1.70%
	Failed to calibrate	2	0.26%
	Corrupted data	4	0.52%
	Total	764	

2. Eye-tracking Protocol

2.1. Eye-trackers

A Tobii (Tobii AB, Sweden) T120 eye tracker (max sample rate 120Hz) was used at the UCAM, RUNMC and UMCU, and a Tobii TX-300 (max sample rate 300Hz) was used at KCL, CIMH and UCBM.

2.2. Eye-tracking Calibration

The calibration sequence was designed to automatically restart in case the eye-tracker could not gather enough samples of gaze data for each calibration point. After the eye-tracker had computed a calibration with sufficient samples, the experimenter visualized the results on the screen (as five average gaze points and 5 target gaze points at the center of the screen, bottom left and right, top left and right corners) and had to indicate whether the calibration was satisfactory (i.e., all 5 calibration points sampled with right and left eye, close overlap of average gaze point and target gaze point), otherwise it would restart. At both checkpoints, the experimenter could adjust the sitting position of the participant. After 3 calibration attempts, the experimenter could choose to skip the battery, e.g., if the participant was non-compliant/restless. The number of batteries not recorded due to failed calibration is reported in Table 1. The battery

also included 6 automatic post-hoc calibrations at predetermined times during the whole session, to ensure continued high-quality tracking of the participant's gaze.

2.3. Task Battery

The task battery included 9 different tasks, divided into four equal sets, interleaved by a break, for a total duration of approximately 28 minutes. During the break, the head position of the participant was overlaid to the screen background, to aid setting up before the next set. The duration of the break was variable to ensure the comfort of the participant, but controlled by the experiment in order to avoid substantial group differences. The static images were part of the first set, thus presented during the first 5 minutes, interleaved with other static images pertaining to different tasks in random order. A central fixation point (a neutral object such as a schematic ball or the sketch of a house) interleaved the stimuli; the progression was gaze-contingent, so the participant had to fixate the central fixation point before the next stimulus was shown.

2.4. Stimuli Images

The images portrayed medium-close ups of:

1) a woman holding a baby in front of a birthday cake, with room furniture and one person in the background

2) one woman and one child facing each other and blowing bubbles, with a lawn and one person in the background

3) 3 children taking chicks from a container, with a pavement and 3 persons in the background

4) 3 children holding hands, with a lawn and 3 persons in the background

5) one girl drawing on a piece of paper, with a table in the background

6) one woman and a child playing with a drum, with a cabinet and one person in the background. The faces of the persons in the background were not visible (turned/out of focus/cut).

2.5. Software

Stimuli were presented on Apple (Apple Inc., USA) Macbook Pro laptops, using our custom-written stimulus presentation framework Task Engine (sites.google.com/site/taskenginedoc/), running in Matlab using Psychtoolbox 3 (7) and the GStreamer (gstreamer.freedesktop.org). The recording and processing of raw gaze data was performed with Tobii Gaze Analytics SDK 3.0.

3. Statistical Analysis

3.1. Complete List of R packages

For all subsequent analysis, we used R version 3.5.1 (2018-07-02), and the following packages:

- For data manipulation: dplyr version 0.8.0.1 (9)
- For mixed modelling: lmerTest version 3.1.0 (11)
- For the Analysis of Deviance: car version 3.0.2 (13)
- For prediction intervals: merTools version 0.5.0 (15)
- For the silhouette analysis: factoextra version 1.0.5 (17)
- For plotting: ggplot2 version 3.0.0 (19)
- For partial correlations: ppcor version 1.1 (20)

For hierarchical clustering, we used the function hclust (part of the base stats of R programming environment).

3.2. Data Processing

Each dataset was up-sampled to 600Hz (to allow integer scaling of samples recorded at either 120Hz or 300Hz). Any gaze samples to locations beyond the extent of the screen were marked as missing. Valid samples were those where the eye tracker detected at least one eye, and gaze was on the screen. Data were processed separately for each participant, for each stimulus image, and for each AOI. Each sample of gaze data was scored 0 or 1, according to whether it was inside or outside of the AOI. Data and sample-level AOI scores were then segmented into time-binned segments with a duration of one second. We calculated Proportional Looking Time for each time bin by calculating Samples in AOI / Total Valid Samples, thus normalizing looking time to the AOI by the proportion of valid, on-screen looking time. We calculated the percentage of missing data by calculating the Samples in AOI with no data / Total Samples. Since the task was part of a 28-minutes battery including other tasks, we also calculated the Accuracy of the calibration across the whole session, as a quality check of the data over time. Accuracy was calculated as the Euclidean distance from the centroid of all gaze samples to the predefined gaze-contingent stimuli that were presented singularly at the center of the screen (i.e., post-hoc calibration, central fixation and gap/overlap stimuli), corresponding to the drift between the eye-tracker gaze location and the true gaze location. We also computed the average number of samples within each AOI and non-AOI area by summing up the number of valid samples with “1” within one of the AOIs (head, body, background people) and non-AOI area, i.e., visible area of the picture on which no AOI was drawn. Any trials or time bins with >75% missing samples (eyes not detected)

were excluded. Table 2 of the main text reports proportional of valid trials before and after the exclusion process.

3.3. Quality Check

The below Table 2 report in full the excluded trials and participants through 75% thresholding of missing data.

Table 2: mean numbers and proportions of completed and valid trials at the various stages of data preparation. Valid trials refer to the trials where both eyes were detected for at least 25% duration of the trial. Each trial corresponds to 1 stimulus displayed for 20 seconds. (N = number, SD = standard deviation, NT = neurotypical, AUT = autism)

	Total	AUT	NT
N	664	388	276
Mean N of completed trials (max = 6), SD (%)	5.99, 0.2 (99.83%)	5.98, 0.26 (99.66%)	6.00, 0.00 (100%)
Mean proportion of Valid Trials in original sample, SD	0.80, 0.25	0.76, 0.26	0.82, 0.22
N of excluded participants with 100% trials with missing data > 75% on completed trials (%)	5 (0.6%)	5 (1.03%)	0 (0%)
Total N of excluded trials with % missing data > 75% on total N of completed trials (%)	228 (5.73%)	165 (7.10%)	63 (3.80%)
Mean N of excluded trials with % missing data > 75% per participant (max 6), SD (%)	0.03, 0.17 (0.5%)	0.43, 1.07 (7.16%)	0.23, 0.72 (3.83%)
N of participants after exclusion of invalid trials, (%)	659 (99.24%)	383 (98.71%)	276 (100%)

Mean proportion of Valid Trials in final sample, SD	0.83, 0.18	0.81, 0.18	0.85, 0.17

3.3.1. Percentage of Missing Data

To investigate the distribution of missing data, we used a mixed model with the percentage of missing data as dependent variable, polynomials up to degree 3 in interaction with group modulated by age, with varying intercept and slope by participant, and varying intercept by stimulus. As reported in the main text, this age-related decrease differed by group, with less missing data for older NT participant; both groups developed a curved and descending trend with age (quadratic component), with a higher curvature in the NT group; see Table 3a for a full report of the coefficients, standard errors, p-values, and Table 3b for X^2 of type III Anova.

Table 3a: beta coefficients, standard errors (SE), t-values and p-values of the multiple regression with the percentage of missing data

Term	Coefficient	t. Er.	df	T -value	P -value
Intercept	35.44	.45	340.08	14.50	< 0.001
Slope	48.26	.20	630.61	11.48	< 0.001
Quadratic	9.20	.18	628.34	4.22	< 0.001
Cubic	21.20	.80	630.58	7.58	< 0.001
Group AUT : age	-0.59	.14	628.84	-4.29	< 0.001
Group NT : age	-0.92	.14	628.89	-6.57	< 0.001
Slope : group AUT : age	-0.44	.25	630.29	-1.75	0.08
Slope : group NT : age	-0.33	.25	630.43	-1.31	0.19
Quadratic : group AUT : age	0.57	.13	628.02	4.44	< 0.001

Quadratic : group NT : age	0.72	.13	628.09	5.51	< 0.001
Cubic : group AUT : age	0.08	.17	630.26	0.50	0.62
Cubic : group NT : age	0.06	.17	630.37	0.38	0.70

Table 3b: X^2 statistic, degrees of freedom and p-values of type III Anova run over the mixed model of the percentage of missing data.

Term	X^2	df	P-value
(Intercept)	210.13	1	<0.001
Slope	131.82	1	<0.001
Quadratic	17.80	1	<0.001
Cubic	57.50	1	<0.001
Group : age	45.55	2	<0.001
Slope : group : age	3.09	2	0.21
Quadratic : group : age	30.34	2	<0.001
Cubic : Group : age	0.25	2	0.88

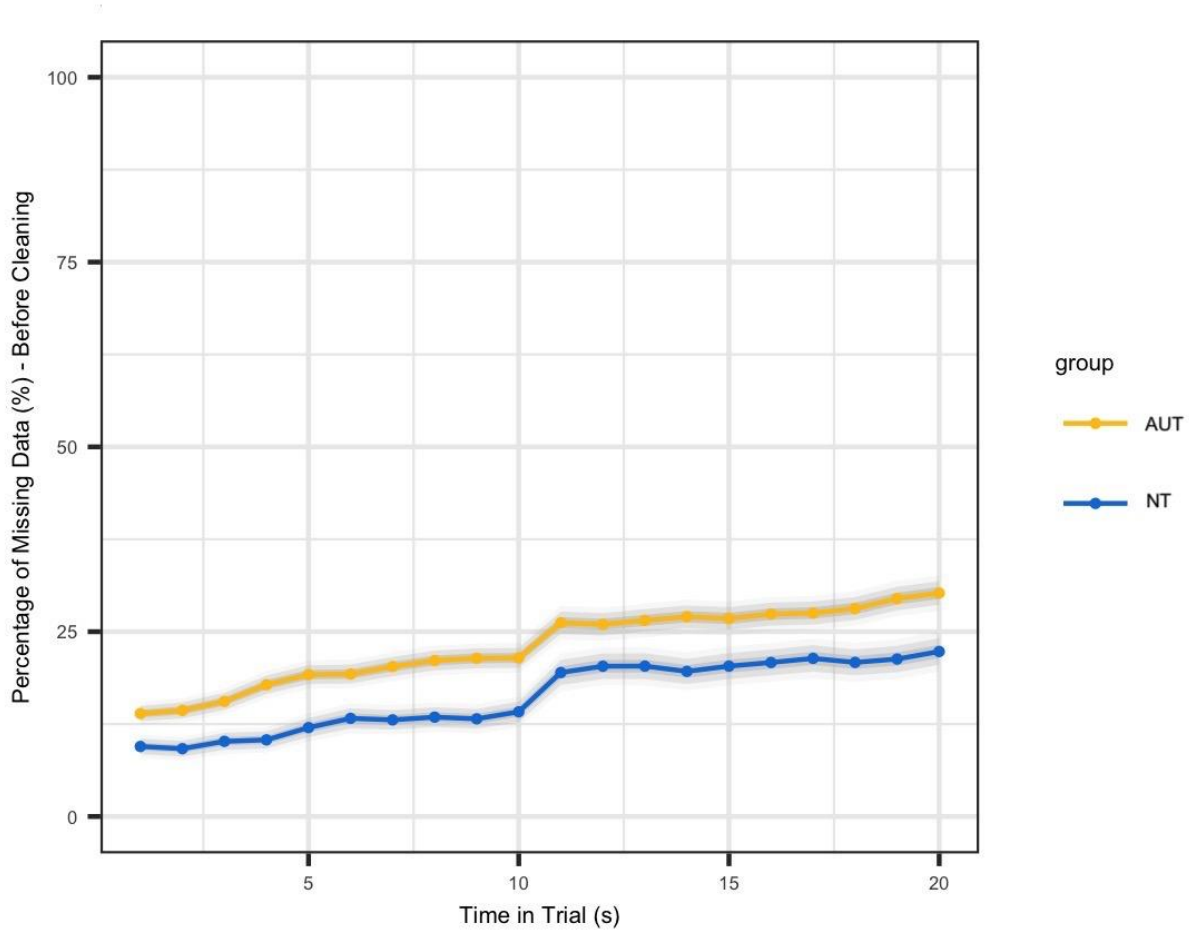

Figure 1: percentage of missing data, averaged over 1 second time bins, of the two groups. The shaded areas mark 2 standard deviations from the mean.

3.3.2. Accuracy

To investigate differences in the accuracy of the eye-tracking across the whole session, we used a multiple linear regression with Accuracy as dependent variable, and group in interaction with age as predictors. We found no differences by group/age, see Table 4 for a full report of the coefficients, standard errors and p-values.

Table 4: beta coefficients, standard errors (SE), t-values and p-values of the multiple regression with accuracy across the session

Term	Co efficient	t. Er.	S tatistic	P -value
Intercept	0.0 486	.0020	2 4.12	< 0.001

Group: AUT	0.0019	(.0025	0.75	0.44
Age Group: 14-18	0.0032	(.0030	1.06	0.28
Age Group: >18	0.0035	(.0027	1.29	0.19
Group: AUT * Age Group: 14-18	-0.0012	(.0039	-0.31	0.75
Group: AUT * Age Group: >18	-0.0013	(.0035	-0.36	0.71

3.3.3. Number of AOIs sampled

To investigate differences in the the number of AOIs sampled during the trials, we used a multiple linear regression with the average N of sample in AOI as dependent variable, and group in interaction with age and AOI (head, body, background people and non-AOI area) as predictors. The background people had significantly less samples, while non-AOI areas significantly more (not surprisingly, given that the AOIs are smaller). Adolescents tended to sample more AOIs across time, with no differences by group. See Table 5 for a full report of the coefficients, standard errors and p-values.

Table 5: beta coefficients, standard errors (SE), t-values and p-values of the multiple regression with the average number of AOI sampled

Term	Coefficient	t. Er.	Statistic	P-value
Intercept	38.90	.97	40.30	< 0.001
Group: AUT	-0.33	.23	-0.27	0.79
Age Group 14-18	3.06	.44	2.12	0.03*
Age Group >18	2.37	.29	1.83	0.07
AOI: background people	-27.50	.37	-20.14	< 0.001
AOI: head	-0.94	.37	-0.69	0.49

AOI: non-AOI	4.23	.37	3.10	< 0.001
Group: AUT * Age Group 14-18	-1.17	.88	-0.62	0.53
Group: AUT * Age Group >18	-0.15	.69	-0.09	0.93
Group: AUT * AOI: b. people	-0.30	.74	-0.17	0.86
Group: AUT * AOI: head	-1.10	.74	-0.63	0.53
Group: AUT * AOI: non-AOI	-0.97	.74	-0.56	0.58
Age Group 14-18: * AOI: b. people	-1.49	.04	-0.73	0.47
Age Group >18 * AOI: b. people	-0.23	.83	-0.13	0.90
Age Group 14-18 * AOI: head	0.23	.04	0.11	0.91
Age Group >18 * AOI: head	0.95	.83	0.52	0.60
Age Group 14-18 * AOI: non-AOI	0.67	.04	0.33	0.74
Age Group >18 * AOI: non-AOI	1.75	.83	0.96	0.34
Group: AUT * Age Group 14-18 * AOI: b. people	1.80	.66	0.68	0.50
Group: AUT * Age Group >18 * AOI: b. people	0.57	.39	0.24	0.81
Group: AUT * Age Group 14-18 * AOI: head	-0.12	.66	-0.05	0.96
Group: AUT * Age Group >18 * AOI: head	-0.51	.39	-0.21	0.83
Group: AUT * Age Group 14-18 * AOI: non-AOI	0.80	.66	0.30	0.76
Group: AUT * Age Group >18 * AOI: non-AOI	0.38	.39	0.16	0.87

3.4. Case-control comparison

We performed a 3-way analysis of variance (ANOVA) of the Proportional Looking Time (PLT) over the head averaged over the entire experiment including group (AUT, NT) and age (as a continuous variable) and their reciprocal interaction as between-subject factors, and the FSIQ

normalized on the mean of 3 age groups (school aged children: 6-13, adolescents: 14-18, young adults: >18) (56,57), and sex as covariates. We included FSIQ as covariates because of 1) the moderate group-difference in terms of FSIQ (see Table 1 of the main text, Group Difference Cohen's D Effect Size) 2) the known influence of sex on social attention (2). We calculated the Cohen's F effect size (4) and report nominal and Bonferroni adjusted p-values (6). Figure 1 illustrates the temporal trends for all AOIs plus looks outside of the AOIs in the raw data.

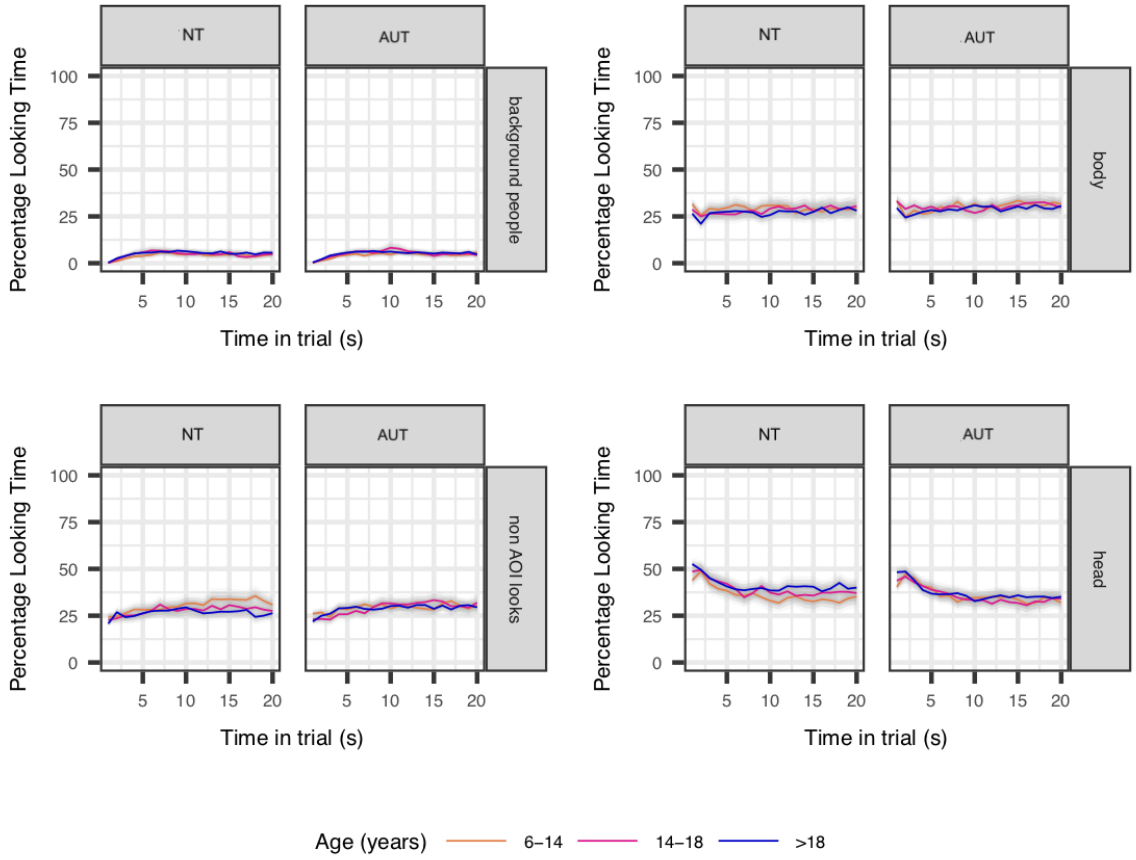

Figure 2: raw data, averaged over 1 second time bins, on the AOIs background people, body, non-AOI area of the photograph and head. The shaded areas mark 2 standard deviations from the mean.

3.5. Growth Curve Analysis (GCA)

We applied GCA (8) to investigate the change in PLT to the head across time through the fixed effects, with a mixed regression model approach, thus accounting for the inter-individual, inter-site and inter-stimuli variability through the random effect (10).

First, we entered the PLT (i.e., the proportional looking time at the head averaged over 1 second time bins) as the dependent variable. Secondly, we entered time (as a sequence of numbers from 1 to 20, each corresponding to 1 second of time passed) and its orthogonal polynomial derivatives up to degree 3 (Linear, Quadratic and Cubic components, see section 3.5.1.1) as fixed effects/independent variables, to describe non-linear trends in the data. This is the key feature of GCA: the beta coefficients of time and its derivatives, as found in the output of the regression, describe the non-linear probability distribution of PLT over time (see Figure 2). Third, we allowed correlated slopes and intercepts per participant nested in the study site, and varying intercepts per stimulus in the random effect, to account for the inter-individual, inter-site and inter-stimuli variability - the key feature of a mixed model (12). We termed this model 'the base model'.

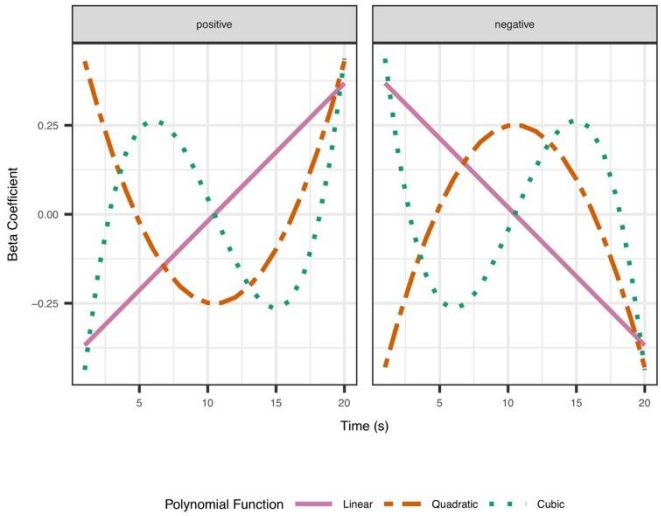

Figure 3: visual representation of temporal trends as indexed by the orthogonal polynomial of time: linear slope (pink line), quadratic component (orange dashed), and cubic component (green dotted). A significant beta coefficient (positive or negative) for any of these 3 polynomials is representative of a predominant trend as represented in the figure.

3.5.1. Model Selection

3.5.1.1. Time Components

The data showed an asymmetrical rise and fall profile. Therefore, we selected a 3rd order polynomial to describe the trend. In order to rule out overfitting, we compared this model with a model including only linear and quadratic polynomials as fixed and random effects with the likelihood ratio test (LRT). The model with 3rd order polynomials significantly differed from the base model and scored lower on the AIC and Deviance metrics thus showing a good fit (see Table 6).

Table 6: LRT between the models including 2nd and 3rd order polynomials and the base model (df = Degrees of Freedom, AIC = Akaike Information Criteria, BIC = Bayesian Information Criteria, LogLik = Logarithmic Likelihood)

Nth Order	f	AIC	BIC	logLik	deviance	χ^2	χ^2 Df	P-value
1		56259.82	56342.40	-28120.91	56241.82	NA	NA	NA
2	6	55918.80	56065.61	-27943.40	55886.80	355.02		< 0.001
3	5	55837.83	56067.22	-27893.91	55787.83	98.97		< 0.001

3.5.1.2. Random Effect

Because of the differences in terms of subjects between the stimuli, we tested the 3rd polynomial model against a model including a random intercept per stimulus as well and opted for this extended random structure (see Table 7).

Table 7: LRT between the model including 3rd order polynomials in the random effect and random intercept for each stimulus and the model including only the 3rd order polynomials

(df = Degrees of Freedom, AIC = Akaike Information Criteria, BIC = Bayesian Information Criteria, LogLik = Logarithmic Likelihood)

RE	f	AIC	BIC	logLik	deviance	X^2	^{2} Df	-value
Full Time Components	5	5837.83	6067.22	-27893.91	5787.83	NA	A	A
Full Time Components + Stimulus	6	6818.81	7057.38	-23383.40	6766.81	9021.02		0.001

3.5.1.3. Fixed Effects

We tested the base model (i.e., including the fixed and random effects of the 3rd order polynomials, as described in the previous paragraph) against a series of successive models with the Likelihood Ratio Test. All models included correlated slopes and intercepts per participant nested in the study site, and varying intercepts per stimulus. The models were set up to investigate group differences in dynamical temporal profiles, the effect of age and relevant covariates, such as sex (14) and overall attention to the screen (16). Therefore, the successive models included the fixed effects: 1) group (AUT/NT), 2) group, and continuous age as covariate, 3) group in interaction with continuous age, 4) group, continuous age as fixed effects, and their interaction 4) covariates: sex, overall looking time to the screen (OLT) and scaled FSIQ.

The Likelihood Ratio Test showed that the deviance and the Akaike information criteria (AIC) significantly dropped when group modulated by age was added as a fixed effect (AIC = 46857.88, BIC = 47096.45, deviance = 46805.88, $X^2 = 16.17$, p-value < 0.001), compared to the base model. Adding sex (AIC = 6852.75, BIC = 47100.50, deviance = 46798.75, $X^2 = 7.12$, p-value = 0.008) and the FSIQ as covariates further improved the fit of the model (AIC =

46331.83, BIC = 46588.43, deviance = 46275.83, X^2 = 5.91, p-value < 0.001), while covarying with the Overall Looking Time to the screen did not significantly differ from the final model (AIC = 46331.08, BIC = 46596.85, deviance = 46273.08, X^2 = 16.17, p-value = 0.09). For a full list of the outputs of the Model Selection process, see Tables 8-10.

Table 8: Comparison of increasingly complex models versus the base model with the LRT. '' designs reciprocal interactions; ':' designs modulation or unidirectional interaction (df = Degrees of Freedom, AIC = Akaike Information Criteria, BIC = Bayesian Information Criteria, LogLik = Logarithmic Likelihood)*

Main Effects	f	AIC	BIC	logLik	deviance	X^2	^{2} Df	p-value
Base	8	46888.87	47054.03	-23426.43	46852.87	NA	NA	NA
* Group	2	46880.04	47081.90	-23418.02	46836.04	16.83		< 0.001
* Group + Age	3	46868.05	47079.09	-23411.02	46822.05	13.98		< 0.001
* Group : Age	6	46857.88	47096.45	-23402.94	46805.88	16.17		< 0.001
* Group * Age	0	46862.22	47137.49	-23401.11	46802.22	3.65		0.26

Table 9: Comparison of the models including sex as a covariate versus the model selected in the previous step, with the LRT. '' designs reciprocal interactions; ':' designes modulation or unidirectional interaction (df = Degrees of Freedom, AIC = Akaike Information Criteria, BIC = Bayesian Information Criteria, LogLik = Logarithmic Likelihood)*

Main Effects	f	AIC	BIC	logLik	deviance	^{2}	^{2} Df	p-value
* Group : Age	6	46857.88	47096.45	-23402.94	46805.88	*NA*	*NA*	*NA*
* Group : Age	7	46852.75	47100.50	-23399.38	46798.75	.12		0.008

+ Sex								

Table 10: Comparison of the models including the covariates FSIQ and OLT versus the model selected in the previous step, with the LRT. '' designates reciprocal interactions; ':' designates modulation or unidirectional interaction (df = Degrees of Freedom, AIC = Akaike Information Criteria, BIC = Bayesian Information Criteria, LogLik = Logarithmic Likelihood)*

Main Effects	f	AIC	BIC	logLik	deviance	χ^2	χ^2 Df	P-value
* Group : Age + Sex	7	46340.88	46588.32	-23143.44	46286.88	*NA*	*NA*	*NA*
* Group : Age + Sex + OLT	8	46337.75	46594.34	-23140.87	46281.75	5.138		0.02
* Group : Age + Sex + FSIQ	8	46331.83	46588.43	-23137.92	46275.83	5.914		< 0.001
* Group : Age + Sex + OLT + FSIQ	9	46331.08	46596.85	-23136.54	46273.08	2.750		0.09

3.6. Deviance Test

We evaluated the explanatory power of the fixed effects/independent variables by computing the Analysis of Deviance (Type III Wald Chi-square test) of the model with the best

fit (see Table 11). The X^2 indicates the test significance, associated with the corresponding p-value with an alpha level of 0.05.

Table 11: results of the Analysis of Deviance. ':' designates modulation or unidirectional interaction (X^2 = Chi-Squared Statistics, df = Degrees of Freedom)

Parameter	^{2}	f	P-Value
Intercept	6.89		< 0.001
Slope	7.81		< 0.001
Quadratic Component	.17		0 .01
Cubic Component	.01		0 .89
Sex	.01		0 .01
FSIQ (scaled)	1.31		< 0.001
Group : Age	3.34		< 0.001
Group : Age : Slope	.56		0 .16
Group : Age : Quadratic Component	.11		0 .57
Group : Age : Cubic Component	.76		0 .007

3.7. Effect Sizes, standard errors, confidence intervals

To better understand the magnitude and the direction of the fixed effects, we report the beta coefficients, the standard error and the 95% confidence intervals (Table 12). The beta coefficient may be regarded as an unstandardized effect size of the effect. The standard error and the confidence interval are inversely correlated with the precision of the beta coefficient.

Table 12: decomposed contribution of each main effect. Coefficients, standard error and confidence intervals have been scaled to percentages. '' designates reciprocal interactions; ':' designs modulation or unidirectional interaction (Fix. Ef. = fixed effect, St. Err = standard error, df = Degrees of Freedom, CI = Confidence Interval)*

Parameter		Interaction	Coefficient	t. Er.	df	- value	- value	I 2.5%	I 97.5%
Intercept			34.67	.71	5.85	.07	.00	3.48	5.86
Slope			-16.16	.06	1 47.52	5.27	.00	22.17	-10.16
Quadratic Component			6.94	.79	4 1.56	.49	.02	.47	2.41
Cubic Component			0.28	.08	6 9326.70	.14	.89	3.79	.36
Sex			-1.94	.79	6 42.20	2.45	.01	3.48	-0.39
FSIQ (scaled)			1.27	.38	6 21.57	.36	.00	.53	.02
ge :	roup								
	T	Intercept	0.33	.07	5 44.42	.43	.00	.18	.48
		Slope	0.26	.18	5 30.86	.45	.15	0.09	.60
		Quadratic Component	0.14	.15	4 48.22	.93	.35	0.16	.44
		Cubic Component	-0.36	.12	6 9345.70	2.99	.00	0.60	-0.12
	UT	Intercept	0.19	.07	5 47.95	.49	.01	.04	.33
		Slope	0.08	.18	5 48.29	.47	.64	0.26	.43
		Quadratic Component	0.16	.15	4 76.79	.05	.30	0.14	.46
		Cubic Component	-0.23	.12	6 9254.57	1.90	.06	0.47	.01

3.8. Random effects and Individual Deviances

In the random effect, we accounted for systematic variation between and within participants nested in the 6 collection sites, and between stimuli. This method allows to estimate the deviation of each participant and site from the average intercept and polynomial components, and of the individual stimulus from the average intercept.

In order to evaluate the variance accounted by measures repeated within participant (nested in different sites) and stimuli, we extracted the conditional standard deviation specified in the random effect of participant, site and stimulus (18). We derived the average percentage of deviance from the fixed effect (Intercept, Slope, and Quadratic component). Note that the average standard deviation has been calculated from the final model, that did not include a random effect for the Cubic Component to prevent overfitting (Table 13).

The full set of standard deviations is reported in Table 13. The average standard deviation of the Intercept (i.e., the average height of the probability curve) for each stimulus is approximately 13%, indicating that the attention to the head may vary depending on the picture. A significant degree of variability is also accounted for by the idiosyncratic inter-individual variation, especially in terms of net decrease over time (i.e., the linear degree of change; 15.56%). In other words, there is an overall tendency in the sample to decrease attention to the face across the trial (i.e., -16% on average, see Table 13); however, we may expect participants to fall within a 16% range of variation around that value. Therefore, at the extremes of the distribution, participants may show a net decrease of 0% or -32%. Such individual differences are often defined “idiosyncratic”, i.e., due to unmeasured variance, nonetheless, the paragraph “Association between inter-individual differences in temporal profiles and AUT symptoms” of

the main text aims to investigate the clinical variables as possible sources of this individual variation.

Table 13: standard deviation of the random effects specified in the final model.

Parameter	Level	Intercept	Slope	Quadratic
St. dev.	Participant / Site	9.44	17.64	11.86
	Site	1.65	1.14	2.11
	Stimulus	13.445	NA	NA
	Residual	33.08	NA	NA

3.9. Association between inter-individual differences in temporal profiles and symptoms

We planned a correlation analysis to test 3 hypotheses: 1) the association between individual temporal profiles of attention (linear, quadratic and cubic beta weights) and concurrent behavioral phenotypes; 2) the association between the degree of deviation of each individual from the overall temporal profile of attention (linear, quadratic and cubic standard deviations) and concurrent behavioral phenotypes; 3) the predictive potential for later behavioral phenotypes of individual temporal profiles of attention (linear, quadratic and cubic beta weights) and the degree of deviation of each individual from the overall temporal profile of attention (linear, quadratic and cubic standard deviations). Concurrent behavioral phenotypes were measures of autistic traits (SRS-2) and of adaptive Socialization and Communication skills (VABS); longitudinal behavioral phenotypes were taken from the VABS but not the SRS-2 as it is validated as a stable measure.

First, we extracted the beta coefficient for each individual (fixed effect plus the conditional mode – the difference between population-level fixed effect and the individual-level effect size) (63,64), whose relation with the clinical variables may indicate that symptom severity/communication impairment correlates with low/ high levels of social attention. Second,

we extracted the conditional standard deviations for each individual from the random effects. These standard deviations represent the degree of dispersion around the mean of the beta coefficients for each individual (i.e. the degree of relative atypicality), whose relation with the clinical variables may indicate that symptom severity/communication impairment correlates with the degree of atypicality from the overall sample (with either low or high levels of social attention). Third, we tested the correlation between beta coefficients and standard deviations and the difference between the VABS scores at time point 1 and 2, whose relation may indicate a predictive potential of the gaze profiles.

We used Spearman partial correlations, controlling for age, and Bonferroni-corrected for 12 comparisons, 4 for each of the 3 independent hypotheses as highlighted above (21).

3.10. Cluster Analysis and Partial Correlations within the Clusters

In order to stratify the AUT group by their temporal profile of attention, we performed a cluster analysis on the beta coefficients for Intercept, Slope, Quadratic and Cubic components extracted from the base model (10). We selected beta coefficients to reflect interpretable attention profiles, and we describe the two clusters in terms of their differing temporal profiles of social attention. We estimated the optimal number of clusters with the silhouette method (the absolute distance between a set number of clusters) (22) and applied the agglomerative hierarchical clusterization with complete calculation of Euclidean distance (23). In order to understand whether the clustering identified any groups with strong relationships between visual attention profiles and symptomatology, we examined the partial correlations (controlling for age) within each cluster between the individual beta coefficients and the SRS T-score and the VABS at Time 1, and change between Time 1 and Time 2. We applied a Bonferroni correction to all tests; in order to

determine the size of the family of comparisons, we used the same approach described in the previous paragraph (21).

In order to establish whether the clusters identified in the AUT subsample would be generalizable to the typical population, we ran the same algorithm on the NT subsample.

Table 14: average coefficients and 95% standard errors based on the individual coefficients fed to the clustering algorithm.

cluster	Intercept Mean (SE)	Slope Mean (SE)	Quad Mean (SE)	Cub Mean (SE)
1	0.38 (0.01)	-0.12 (0.01)	0.1 (0.01)	-0.05 (0.01)
2	0.31 (0.02)	-0.35 (0.02)	0.05 (0.02)	0.02 (0.02)

Table 15: sample size, percentage on the total, mean and standard deviations (SD) of age, full scale intelligence quotient (FSIQ), SRS T-score, VABS socialization (Soc) score, VABS communication (Comm) score of the two clusters.

a. AUT subset

Cluster	(n, %)	M:F	Age	FSIQ	SRS T-score	VABS Comm	VABS Soc	Proportion of Valid Trials
			Mean (SD)					
1	(294, 92%)	2.87	16.43 (5.84)	103.33 (15.92)	71.01 (11.95)	77.61 (15.69)	72.08 (14.48)	0.82 (0.16)
2	(25, 8%)	2.00	16.03 (6.43)	104.14 (10.51)	74.83 (12.08)	77.60 (23.11)	70.80 (20.83)	0.86 (0.12)

b. NT subset

Cluster	(n, %)	M:F	Age	FSIQ	SRS T-score	VABS Comm	VABS Soc	Proportion of Valid Trials
			Mean (SD)					

1	171 (70%)	.29	6.24 (5.70)	108.32 (12.78)	6.04 (6.54)	103.75 (14.49)	108.65 (10.99)	0.84 (0.16)
2	72 (30%)	.48	9.06 (6.50)	110.84 (13.77)	3.96 (4.25)	104.73 (16.53)	110.64 (12.53)	0.87 (0.16)

Table 16: available data points (N), Spearman correlations coefficients (rho), nominal p-values and Bonferroni-adjusted p-values between the beta coefficients and the clinical variables at Time 1, and the difference of the VABS Communication (Comm) and Socialization (Soc) scores between Time 2 and Time 1 within cluster 1 and 2. Significant correlations after Bonferroni correction are marked with ''*

variable	cluster	term		rho	p-value	p-adjusted
SRS-2 T-score (Time 1)	1	Intercept	42	< 0.001	0.99	1.00
		Slope	42	-0.11	0.08	0.92
		Quad	42	-0.07	0.27	1.00
		Cub	42	0.01	0.85	1.00
	2	Intercept	9	-0.45	0.06	0.75
		Slope	9	0.33	0.19	1.00
		Quad	9	0.71	< 0.001	0.01*
		Cub	9	< 0.001	0.99	1.00
VABS Comm (Time 1)	1	Intercept	50	-0.01	0.86	1.00
		Slope	50	0.09	0.18	1.00
		Quad	50	0.12	0.07	0.79
		Cub	50	0.03	0.63	1.00
	2	Intercept	0	0.28	0.25	1.00
		Slope	0	0.11	0.64	1.00
		Quad	0	-0.17	0.48	1.00
		Cub	0	0.01	0.98	1.00

VABS Soc (Time 1)	1	Intercept	46	-0.07	0.28	1.00
		Slope	46	0.05	0.48	1.00
		Quad	46	0.10	0.12	1.00
		Cub	46	-0.02	0.79	1.00
	2	Intercept	0	0.28	0.25	1.00
		Slope	0	-0.38	0.11	1.00
		Quad	0	-0.36	0.13	1.00
		Cub	0	<0.001	0.99	1.00
VABS Comm (Time 2 – Time 1)	1	Intercept	86	-0.05	.5	1.00
		Slope	86	0.08	0.29	1.00
		Quad	86	0.16	0.03	0.38
		Cub	86	0.06	0.41	1.00
	2	Intercept	1	-0.16	0.65	1.00
		Slope	1	-0.31	0.39	1.00
		Quad	1	-0.09	.8	1.00
		Cub	1	-0.14	0.71	1.00
VABS Soc (Time 2 – Time 1)	1	Intercept	79	<0.001	0.95	1.00
		Slope	79	-0.07	0.32	1.00
		Quad	79	0.04	0.63	1.00
		Cub	79	0.15	0.05	0.54
	2	Intercept	1	-0.44	0.21	1.00
		Slope	1	-0.21	0.57	1.00
		Quad	1	-0.04	0.91	1.00

		Cub	1	-0.31	0.39	1.00

4. Appendix

4.1. Validation

Our dependent variable is a proportion that is naturally distributed in the interval between 0 and 1. This fact may create instability in the model fit by particularly failing the assumption of the normal distribution of the residuals. Therefore, we applied two alternative methods to validate our conclusions.

4.1.1. Logistic Regression

As our dependent variable is a proportion, the most fit distribution of the residual is the binomial. In order to validate our findings in terms of fixed effects, we run a logistic regression assuming a binomial distribution of the proportion data. The results of the Deviance Test and the coefficients of the fixed effects are shown in Table 10-11. Even though this method does not account for the repeated measures and the nesting, the results fairly overlap (i.e., average decrease across the trial and significant Cubic Component in the NT group, modulated by age, but not in the AUT group). The predicted values in terms of probability (extracted from the model coefficients and the standard errors) also show largely the same pattern of results as the main analysis (see Figure 3).

Table 17: decomposed contribution of each fixed effect in logit odds ratio. ':' designs modulation or unidirectional interaction (Fix. Ef. = fixed effect, St. Err = standard error, df = Degrees of Freedom, CI = Confidence Interval)

Parameter			Coefficient	St. Error	Z-Value	P-Value	CI 2.5%	CI 97.5%
Intercept			-0.62	0.027	-22.68	< 0.0001*	-0.68	-0.57

Slope			-0.68	0.12	-5.42	< 0.0001*	-0.92	-0.43
Quadratic Comp			0.30	0.13	2.32	0.02*	0.05	0.56
Cubic Comp			0.01	0.13	0.12	0.90	-0.24	0.27
Sex			-0.07	0.02	-4.23	< 0.0001*	-0.10	-0.04
FSIQ (scaled)			0.04	0.01	5.17	< 0.0001*	0.026	0.06
Group: Age	T		0.01	0.001	8.87	< 0.0001*	0.010	0.01
	UT		0.006	0.001	4.41	< 0.0001*	0.003	0.009
	lope	T	0.01	0.01	1.59	0.11	-0.003	0.02
		UT	0.0036	0.0072	0.50	0.62	-0.01	0.02
	uad Comp	T	0.0050	0.0076	0.66	0.51	-0.01	0.02
		UT	0.0061	0.0076	0.81	0.42	-0.01	0.02
	ub Comp	T	-0.0149	0.0075	-1.98	0.05*	-0.03	-0.0002
		UT	-0.0099	0.0075	-1.32	0.19	-0.02	0.004

Table 18: results of the Type III Analysis of Deviance. ':' designs modulation or unidirectional interaction (X2 = Chi-Squared Statistics, df = Degrees of Freedom)

Parameter	X^2	f	P-Value
Intercept	9942.58077		<0.00001*
Slope	20.00126		0.00001*
Quadratic Component	1.47172		0.22
Cubic Component	0.35670		0.55
Sex	5.16688		0.02*
FSIQ (scaled)	14.41084		0.0001*

Group:Age	23.13860		0.00001*
Group:Age:Slope	5.73673		0.05*
Group:Age:Quadratic Component	1.73786		0.41
Group:Age: Cubic Component	5.71363		0.05*

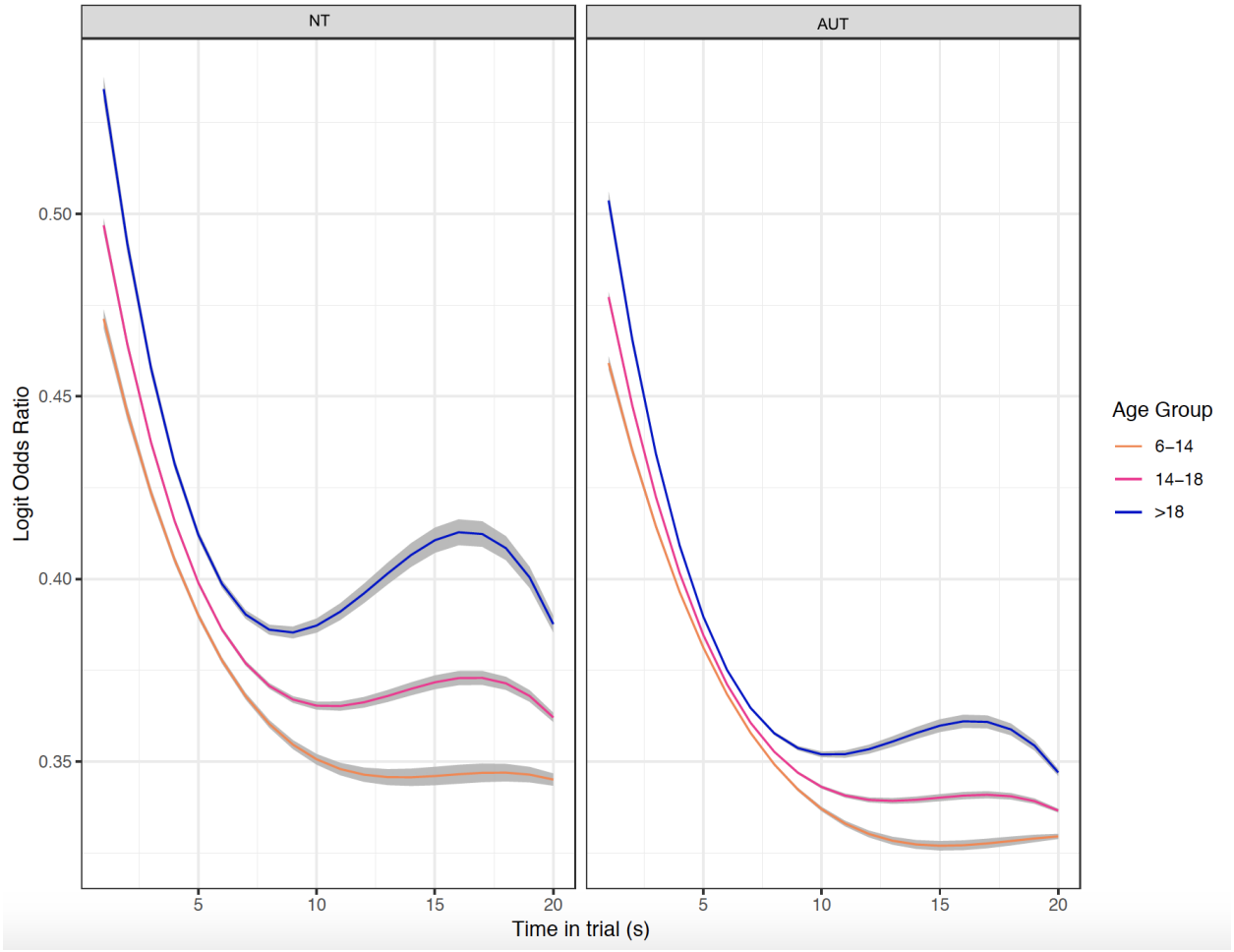

Figure 4: fitted curve of the logit odds ratio of PLT with standard error (shaded area).

4.1.2. GCA with the Empirical Logit Transformation

We run the exact same mixed model as reported in the main section but applied the empirical logit transformation[1] to the dependent variable; the logit transformation maps the proportion data to the line of the real numbers and potentially mitigates patterned residuals (24,25). As shown in Table 5 and 6, we obtained largely the same results. The predicted values also show largely the same pattern of results as the main analysis (see Table 12, and Figure 5); for the fitted curve, we used the same method (calculation of the variation embedded in the fixed effects and the residuals, as well as the conditional modes of the random effects). However, this approach is less interpretable as the coefficients and the confidence intervals are not on the same scale as the dependent variable. Even if a positive/negative Elog means increased/decreased probability, the results are overall less interpretable. Therefore, we decided to maintain untransformed proportional looking time as dependent variable in the main analysis.

Table 19: decomposed contribution of each fixed effect in empirical logits. ':' designs modulation or unidirectional interaction (Fix. Ef. = fixed effect, St. Err = standard error, df = Degrees of Freedom, CI = Confidence Interval)

Parameter			Coefficient	St. Error	DoF	t-Value	p-Value	CI 2.5%	CI 97.5%
Intercept			-650.89	.06	5.80	99.71	0.0001	-6.63	-6.38
Slope			-15.81	.03	100.69	4.47	0.0001	-0.22	-0.08
Quadratic Comp			3.95	.03	34.02	.21	.23	-0.02	0.10

[1] Formula: elog = log((y+0.5)/(N-y+0.5)), with y being the dependent variable, and N being the length of the trial (max 600 samples, in this case).

Cubic Comp			1.38	.02	68912.68	.59	.55	-0.03	0.05
Sex			-1.89	.008	601.91	-2.27	.02	-0.03	-0.002
FSIQ (scaled)			1.52	.004	593.60	.79	.0002	0.007	0.02
Group:Age	T		0.36	.0008	540.57	.33	0.0001	0.002	0.005
	UT		0.19	.0008	546.87	.34	.0195	0.0003	0.003
	lope	T	0.39	.002	457.99	.94	.05	<0.00001	0.007
		UT	0.15	.002	492.45	.77	.43	-0.002	0.005
	uad Comp	T	0.17	.001	395.42	.006	.31	-0.001	0.005
		UT	0.22	.001	445.17	.31	.19	-0.001	0.005
	ub Comp	T	-0.30	.001	68868.91	-2.26	.0236	-0.005	-0.0004
		UT	-0.18	.001	68883.94	-1.39	.16	-0.004	0.0008

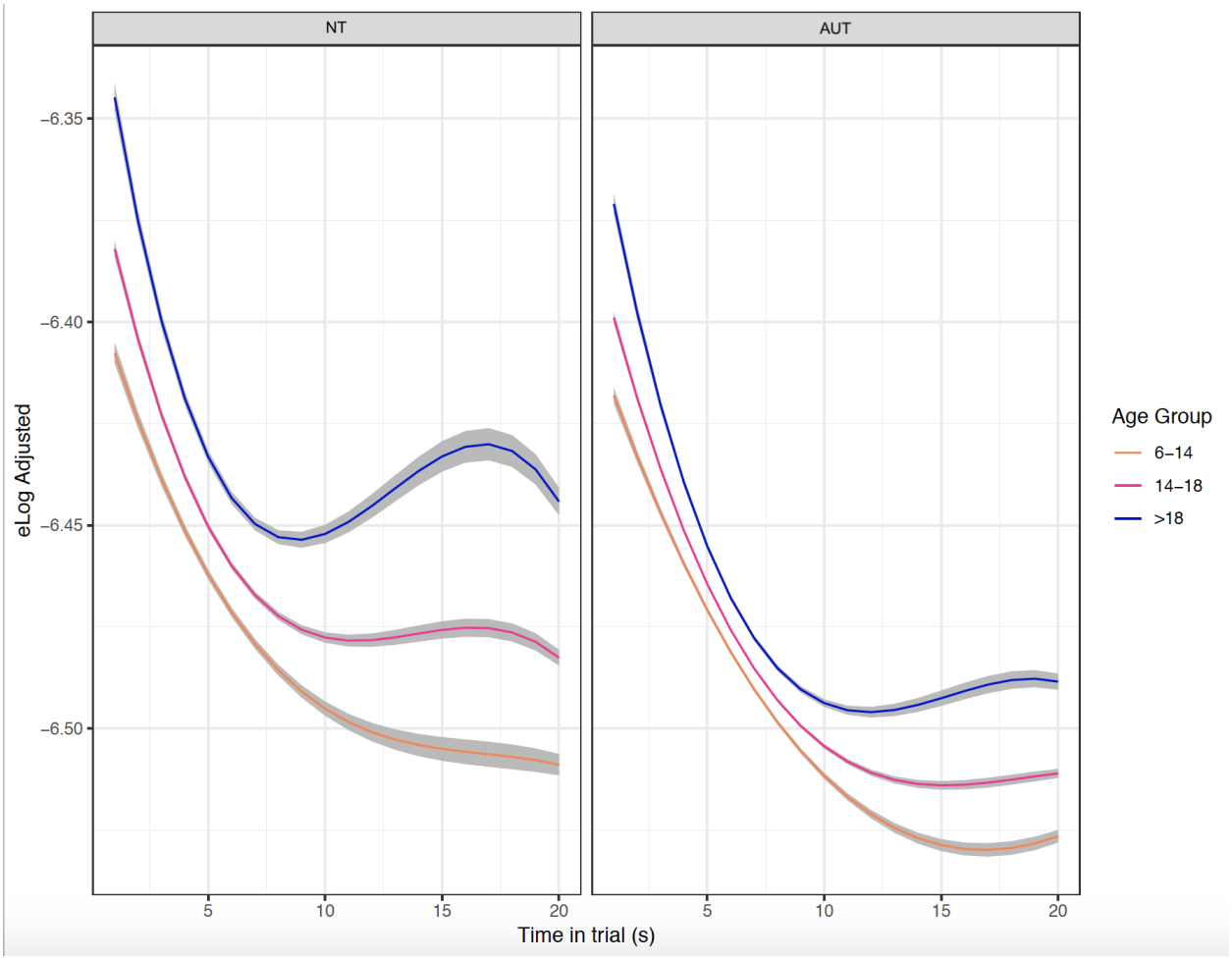

Figure 6: predicted conditional values (scaled by the average intercept) with 95% confidence interval based on 1000 simulations.

5. References

1. Wigham S, McConachie H, Tandos J, Le Couteur AS (2012): The reliability and validity of the Social Responsiveness Scale in a UK general child population. *Res Dev Disabil*. https://doi.org/10.1016/j.ridd.2011.12.017

2. Harrop C, Jones D, Zheng S, Nowell SW, Boyd BA, Sasson N (2018): Sex differences in social attention in autism spectrum disorder. *Autism Res* 11: 1264–1275.

3. Constantino JN, Gruber CP (2005): The Social Responsiveness Scale (SRS). *Los Angeles: Western Psychological Services*.

4. Salkind N (2012): Cohen's f Statistic. *Encyclopedia of Research Design*. SAGE Publications, Inc. https://doi.org/10.4135/9781412961288.n59

5. Sattler JM (1992): *Assessment of Children: WISC—III and WPPSI—R Supplement.*

6. Cramer AOJ, van Ravenzwaaij D, Matzke D, Steingroever H, Wetzels R, Grasman RPPP, *et al.* (2016): Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychon Bull Rev* 23: 640–647.

7. Brainard DH (1997): The Psychophysics Toolbox. *Spat Vis* 433–436.

8. Mirman D, Dixon JA, Magnuson JS (2008): Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *J Mem Lang* 59: 475–494.

9. Wickham H, Francois R, Henry L, Müller K (2019): Package 'dplyr'. A Grammar of Data Manipulation. Retrieved from https://cran.r-project.org/package=dplyr

10. Bates D, Mächler M, Bolker BM, Walker SC (2015): Fitting linear mixed-effects models using lme4. *J Stat Softw*. https://doi.org/10.18637/jss.v067.i01

11. Kuznetsova A, Brockhoff PB, Christensen RHB (2017): lmerTest Package: Tests in Linear

Mixed Effects Models . *J Stat Softw*. https://doi.org/10.18637/jss.v082.i13

12. Bolker BM (2015): Linear and generalized linear mixed models. In: Fox GA, Negrete-Yankelevich S, Sosa VJ, editors. *Ecological Statistics, Contemporary Theory and Application*. Oxford University Press, pp 309–333.
13. Fox J, Sanford W (2011): *An {R} Companion to Applied Regression*, Second. Thousand Oaks, CA: Sage. Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion
14. Harrop C, Jones D, Zheng S, Nowell S, Schultz R, Parish-Morris J (2019): Visual attention to faces in children with autism spectrum disorder: Are there sex differences? *Mol Autism* 10: 28.
15. Knowles JE, Frederick C (2019): merTools: Tools for Analyzing Mixed Effect Regression Models. Retrieved from https://cran.r-project.org/package=merTools
16. Chawarska K, Macari S, Shic F (2012): Context modulates attention to social scenes in toddlers with autism. *J Child Psychol Psychiatry* 53. https://doi.org/10.1111/j.1469-7610.2012.02538.x
17. Kassambara A, Mundt F (2017): factoextra: Extract and Visualize the Results of Multivariate Data Analysis. Retrieved from https://cran.r-project.org/package=factoextra
18. Pelphrey KA, Sasson NJ, Reznick JS, Paul G, Goldman BD, Piven J (2002): Visual Scanning of Faces in Autism. *J Autism Dev Disord* 32: 249–261.
19. Wickham H (2016): ggplot2 - Elegant Graphics for Data Analysis. New York, NY, USA: Springer-Verlag. Retrieved from http://ggplot2.org
20. Seongho K (2015): ppcor: Partial and Semi-Partial (Part) Correlation. R package version 1.1. Retrieved from https://cran.r-project.org/package=ppcor
21. Lakens D (2014): Performing high-powered studies efficiently with sequential analyses. *Eur*

J Soc Psychol 44: 701–710.

22. Rousseeuw PJ (1987): Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20: 53–65.
23. Hastie T, Tibshirani J (2009): Hierarchical clustering. *The Elements of Statistical Learning*, 2nd ed. New York, pp 520–528.
24. Barr DJ (2008): Analyzing "visual world" eyetracking data using multilevel logistic regression. *J Mem Lang* 59: 457–474.
25. Barr D (2012): Walkthrough of an "empirical logit" analysis in R. Retrieved January 21, 2020, from http://talklab.psy.gla.ac.uk/tvw/elogit-wt.html