

## Understanding the nature of face processing in early autism: A prospective study

Charlotte Tye<sup>1†</sup>, Giorgia Bussu<sup>2†</sup>, Teodora Gliga<sup>3</sup>, Mayada Elsabbagh<sup>4</sup>, Greg Pasco<sup>1</sup>, Kristinn Johnsen<sup>5</sup>, Tony Charman<sup>1</sup>, Emily J.H. Jones<sup>3,#</sup>, Jan Buitelaar<sup>2,6#</sup>, Mark H. Johnson<sup>3,7,#</sup> and the BASIS team

<sup>1</sup>King's College London

<sup>2</sup> Department of Cognitive Neuroscience, Radboud University Medical Center

<sup>3</sup>University of London

<sup>4</sup>McGill University

<sup>5</sup>Mentis Cura

<sup>6</sup>Karakter Child and Adolescent Psychiatry University Centre

<sup>7</sup>University of Cambridge

<sup>†</sup> shared first author

<sup>#</sup> shared last author

### Author Note

Charlotte Tye, PhD, Department of Child & Adolescent Psychiatry and MRC Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London; Giorgia Bussu, Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center; Teodora Gliga, Centre for Brain and Cognitive Development, Birkbeck College, University of London; Mayada Elsabbagh, Department of Psychiatry, McGill University; Greg Pasco, Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London; Kristinn Johnsen, Mentis Cura, Reykjavík, Iceland; Tony Charman, Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London; South London and Maudsley NHS Foundation Trust (SLaM); Emily J. H. Jones, Centre for Brain and Cognitive Development, Birkbeck College, University of London; Jan K. Buitelaar, Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center; Karakter Child and Adolescent Psychiatry University Centre; Mark H. Johnson, Centre for Brain and Cognitive Development, Birkbeck

College, University of London; Department of Psychology, University of Cambridge; the BASIS team is listed in alphabetical order as follows: Anna Blasi, Simon Baron-Cohen, Rachael Bedford, Patrick Bolton, Susie Chandler, Celeste Cheung, Kim Davies, Janice Fernandes, Isobel Gammer, Holly Garwood, Jeanne Giraud, Anna Gui, Kristelle Hudry, Michelle Lieu, Evelyne Mercure, Sarah Lloyd-Fox, Helen Maris, Louise O'Hara, Andrew Pickles, Helena Ribeiro, Erica Salomone, Leslie Tucker, Agnes Volein.

This study complies with APA ethical standards in data treatment. All procedures were in agreement with ethical approval granted by the London Central NREC (approval codes 06/MRE02/73, 08/H0718/76), and one or both parents gave informed consent to participate in the study. The datasets analysed in this study are subject to the BASIS data sharing policy.

Previous studies have reported on partially overlapping datasets (Bedford et al., 2017; Mayada Elsabbagh et al., 2012; Salomone et al., 2018). This study is available as preprint on MedRxiv (<https://medrxiv.org/cgi/content/short/2020.05.06.20092619v1>) and was presented as a poster at the Donders Discussions in Nijmegen, 2017; and as an oral presentation at INSAR in Rotterdam, 2018, at ETADE Brainview conference in London, 2018, at the International Organization of Psychophysiology World Congress in Lucca, 2018, and at vICIS, 2020.

We wish to thank all the individuals and families who participated in this research. This work was supported by MRC Programme Grant no. G0701484 and MR/K021389/1, the BASIS funding consortium led by Autistica ([www.basisnetwork.org](http://www.basisnetwork.org)), EU-AIMS (the Innovative Medicines Initiative joint undertaking grant agreement no. 115300, resources of which are composed of financial contributions from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in-kind contribution), and AIMS-2-TRIALS (the Innovative Medicines Initiative 2 joint undertaking grant agreement no. 777394). Charlotte Tye is supported by the Tuberous Sclerosis Association and the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. Giorgia Bussu has received funding from the Marie Skłodowska Curie Actions of the European Community's Horizon 2020 Program under grant agreement

n°642996 (Brainview). Charlotte Tye and Giorgia Bussu share first authorship of this paper, while Emily Jones, Mark Johnson and Jan Buitelaar share last authorship.

Correspondence concerning this article should be addressed to Dr. Giorgia Bussu, Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Kapittelweg 29, 6525 EN Nijmegen, The Netherlands. Tel: 024-3668236. Fax: 024-3610989. Email: [g.bussu@donders.ru.nl](mailto:g.bussu@donders.ru.nl)

## Abstract

Dimensional approaches to psychopathology interrogate the core neurocognitive domains interacting at the individual level to shape diagnostic symptoms. Embedding this approach in prospective longitudinal studies could transform our understanding of the mechanisms underlying neurodevelopmental disorders. Such designs require us to move beyond traditional group comparisons and determine which domain-specific alterations apply at the level of the individual, and whether they vary across distinct phenotypic subgroups. As a proof of principle, this study examines how the domain of face processing contributes to the emergence of Autism Spectrum Disorder (ASD). We used an event-related potentials (ERPs) task in a cohort of 8-month-old infants with ( $n=148$ ) and without ( $n=68$ ) an older sibling with ASD, and combined traditional case-control comparisons with machine-learning techniques for prediction of social traits and ASD diagnosis at 36 months, and Bayesian hierarchical clustering for stratification into subgroups. A broad profile of alterations in the time-course of neural processing of faces in infancy was predictive of later ASD, with a strong convergence in ERP features predicting social traits and diagnosis. We identified two main subgroups in ASD, defined by distinct patterns of neural responses to faces, which differed on later sensory sensitivity. Taken together, our findings suggest that individual differences between infants contribute to the diffuse pattern of alterations predictive of ASD in the first year of life. Moving from group-level comparisons to pattern recognition and stratification can help to understand and reduce heterogeneity in clinical cohorts, and improve our understanding of the mechanisms that lead to later neurodevelopmental outcomes.

**Keywords:** autism; ERP; face processing; machine learning; prospective longitudinal study.

**General Scientific Summary:** This study suggests that neural processing of faces is diffusely atypical in Autism Spectrum Disorder, and that it represents a strong candidate predictor of later social skills and diagnosis at an individual level in the first year of life. Furthermore, a subgroup of infants developing ASD showed reduced neural responses to faces in infancy but increased sensory sensitivity in toddlerhood.

## Introduction

Autism spectrum disorder (ASD) is defined on the basis of social and communication impairment, restricted patterns of behaviours and interests, and sensory anomalies in early childhood (American Psychiatric Association, 2013). ASD is characterised by high heterogeneity, expressed as considerable variability across individuals in terms of both clinical manifestations and underlying biology (E. J. Jones, Gliga, Bedford, Charman, & Johnson, 2014; Lai, Lombardo, Chakrabarti, & Baron-Cohen, 2013; Vorstman et al., 2017). Parsing this heterogeneity is a main theme of theoretical initiatives in mental health research, such as the Research Domain Criteria (RDoC) Framework (Insel et al., 2010). Theoretical models like RDoC propose a shift from unitary diagnostic labels towards determining how underlying impairments in a set of core domains contribute to diagnostic phenotypes. This may facilitate greater individualization of treatment options by enabling individuals to be characterized by dimensional scores reflecting domain functioning, hypothetically facilitating neurobiologically-informed treatments. Studying RDoC domains in early development, prior to the onset of behavioural symptoms, might be particularly critical for understanding how alterations in these domains contribute to symptom emergence. To do this, it is fundamental to adopt analytic strategies that profile a selected domain at the individual level, and understand its developmental link to ASD.

While most genetic studies treat ASD as a unitary clinical category, the majority of genetic liability to ASD is thought to be spread across many genes with individually small and pleiotropic effects (Huguet, Benabou, & Bourgeron, 2016). A recently proposed framework suggests that these genetic factors act through the critical aggregation of earlier-interacting liabilities in contributing to the later clinical expression of ASD (Constantino, 2018). These liabilities are best described as endophenotypes, quantitative heritable neuropsychiatric alterations that can be identified in the general population as continuously distributed traits (Bearden & Freimer, 2006). A leading candidate domain in the mechanisms underlying ASD development is social cognition, and more specifically, face processing (G. Dawson, Bernier, & Ring, 2012; G. Dawson et al., 2005).

From the first year of life, infants with later ASD demonstrate emerging atypicalities in social-communicative behaviour, such as a declining interest in human faces (W. Jones & Klin, 2013; Maestro et al., 2002; Osterling & Dawson, 1994). These behavioural changes appear to be accompanied by atypical neural responses to faces as measured by event-related potentials (ERPs), which provide the resolution required to investigate different temporal stages of information processing and can be obtained at younger ages than behavioural assessments (de Haan, 2007). The early sensory P1, and later cortical N290 and P400 responses, have been shown to be consistently modulated in face processing tasks in the first year of life, and the N290/P400 are argued to be developmental precursors of the well-established adult N170 response, reflecting processing of semantic and structural aspects of faces (de Haan et al. 2003). While some studies have suggested that low-level sensory responses to faces in infancy, indexed by the P1 ERP, is associated with better social development (E. Jones, G. Dawson, & S. Webb, 2018), atypicalities in higher-level cortical processing of faces and gaze have been reported in toddlers and children with ASD (Geraldine Dawson et al., 2002; Grice et al., 2005; Sara Jane Webb et al., 2011). From 6 months of age, altered later cortical responses to faces versus non-faces, and reduced differentiation of faces that shift gaze towards versus away from the viewer, are observed in infants with later ASD, as indexed by the P400 ERP (Mayada Elsabbagh et al., 2012; E. Jones et al., 2016). In 'social first' theories, atypicalities in social engagement and information processing mutually amplify each other over developmental time, reducing opportunities for social learning and contributing to the atypical development of social communication that is characteristic of ASD (M. Elsabbagh, 2020). Thus, neural responses to faces represents one of the early measurable liabilities that could be aggregated with other factors to ultimately lead to ASD. However, little is known about *how and what* alterations in face processing contribute to ASD emergence.

Whilst previous research converges in identifying face processing as a relevant domain in ASD, studies report greater divergence in the *nature* of ASD-associated alterations. Neural responses to faces can be characterized by many different features of an averaged waveform, which are hypothesized to reflect different underlying cognitive processes. Although the general conclusion is one of altered face processing, this diversity could reflect (a) consistent differences in lots of different specific aspects of face processing that are masked in different studies by their theory-driven focus on one or two components; (b)

individually-specific profiles of specific alterations that will or will not be apparent at the group level depending on its composition; and/or (c) distinct profiles of alterations in coherent subgroups of individuals developing ASD, again apparent or not at the group level depending on the sample composition. To understand these patterns, it is important to couple top-down and data-driven analytic strategies with larger samples and individual-level data analysis.

Here, we used three analytic strategies to ask how atypicalities in the domain of face processing contribute to later ASD. First, we used a prospective approach because it enables the investigation of causal mechanisms. Based on a sibling recurrence rate of around 20% (Ozonoff et al., 2011), prospective studies of infants with an older sibling with ASD (“infant siblings”) represent a powerful research design to identify precursors of symptom emergence in ASD (E. J. H. Jones et al., 2019). Second, we combined group-based comparisons with investigating individual effects through data-driven multi-feature machine learning approaches for prediction of traits and diagnosis. Since domains are multifaceted rather than single cognitive processes, this approach enables an examination of the consistency of results across both top-down prediction and bottom-up discovery and construction of a robust multivariate model for data integration and prediction of outcome at an individual level (Arbabshirani, Plis, Sui, & Calhoun, 2017; Rosenberg, Casey, & Holmes, 2018; Yahata, Kasai, & Kawato, 2017). Third, we decomposed heterogeneity among infants at elevated likelihood for ASD with a clustering approach based on the domain under investigation (here social cognition/face processing) to examine whether we could identify subgroups with qualitatively different face processing alterations (Lombardo et al., 2016; Zhao & Castellanos, 2016). Moving from group-level comparisons, to pattern recognition and stratification, this study promotes the use of novel, individual-level approaches for a dimensional understanding of brain development.

## **Methods and Materials**

### **Participants**

This study included 247 infants at elevated likelihood (EL), based on having an older biological sibling with ASD, and at typical likelihood (TL) of developing ASD, recruited from the British Autism Study of

Infant Siblings ([www.basisnetwork.org](http://www.basisnetwork.org)) across two independent cohorts. Specifically, 54 EL (21 male) and 50 TL infants (21 male) participated in cohort 1 (Mayada Elsabbagh et al., 2012), and 116 EL (64 male) and 27 TL (14 male) in cohort 2. TL controls were full-term infants (gestational age 38-42 weeks) recruited from a volunteer database at the Birkbeck Centre for Brain and Cognitive Development. Infants were seen for the face/gaze ERP task when they were approximately 8 months old (*Table S1/S2*). Subsequently, 237 were seen for assessment around their third birthday by an independent team. Four TL and two EL children were absent for the 36-month visit but were included in the analysis based on assessments at the previous visits. Among the remaining 243 infants, 27 were excluded based on quality of EEG data, resulting in a final sample of 216 infants (TL=68, EL-no ASD=115, EL-ASD=33; see *Table S3* for details). All procedures were in agreement with ethical approval granted by the London Central NREC (approval codes 06/MRE02/73, 08/H0718/76), and one or both parents gave informed consent to participate in the study.

### **Clinical assessment**

The Autism Diagnostic Observation Schedule (ADOS)-generic (Lord et al., 2000), a semi-structured observational assessment, and the Social Communication Questionnaire (SCQ, (Rutter, 2003)), a screening tool for ASD, were used to assess current symptoms of ASD at 36 months. The Autism Diagnostic Interview – Revised (ADI-R), a structured parent interview, was completed with parents of EL infants in cohort 1 and all children in cohort 2. These assessments were conducted without blindness to risk-group status by (or under the close supervision of) clinical researchers with demonstrated research-level reliability. The Mullen Scales of Early Learning (MSEL; Mullen, 1995) and the Vineland Adaptive Behavior Scale [VABS] (Sparrow, Balla, & Cicchetti, 1984) were used to measure, respectively, cognitive abilities and adaptive functioning at each visit.

Experienced researchers and the lead clinician (TC) determined the best estimate clinical outcome by reviewing all available information from visits performed. Of the 148 EL participants included in analyses, 33 [22.3%] participants met criteria for ASD (hereafter EL-ASD) and the remaining 115 [77.7%] participants did not meet criteria for ASD (hereafter EL-no ASD), using ICD-10 criteria (cohort 1) or DSM-5 (cohort 2). There was a significant difference in clinical outcome by gender ( $\chi^2(2) = 13.5, p = 0.001$ ), with



more males receiving an ASD diagnosis than females (odds ratio, OR = 4.84; 95% confidence interval [CI; 1.93 to 12.1];  $p < 0.001$ ). No TL children met criteria for ASD at the 36-month assessment and none had a community clinical diagnosis.

### **Electrophysiological measures**

The task was the same as in Elsabbagh et al. (2012), designed to assess responses to the following contrasts: (1) faces (valid static (irrespective of gaze direction) vs. visual noise stimuli presented at the beginning of each block); (2) static gaze (faces with direct vs. averted gaze); and (3) dynamic gaze shifts (gaze toward vs. away from the infant; see *Supplemental Materials*). EEG was recorded from a 128 channel Hydrocel Sensor Net. Following artifact detection and rejection (see *Supplemental Materials*), stimulus-locked epochs (-200 to 800ms peristimulus window) were averaged for each of the three contrasts. Peak amplitude and latency of the averaged P100, N290, and P400 across occipito-temporal channels for each stimulus/contrast were used as input features for subsequent analyses.

### **Statistical analysis**

<Figure 1: Flow chart of the statistical analysis>

#### **(A) Group-based comparison.**

A repeated measures ANOVA was conducted on each ERP parameter, with contrast as the within-subjects factor and group as the between-subjects factor (*Figure 1.A*). A set of analyses was run with cohort as an additional between-subjects factor and followed up with post-hoc t-tests to compare ERP amplitude and latency of the EL-ASD group against other groups. Sidak correction was used to correct for multiple testing. Covariates (age at time of EEG acquisition, MSEL visual reception and fine motor (non-verbal) t-score at 36 months) were entered into a second round of analyses. Gender was not a significant covariate in any analysis and was not retained. Analyses were performed on SPSS v22 (<http://www.ibm.com/analytics/us/en/technology/spss>).

#### **(B) Supervised Classification**

A subsample of 144 EL infants was selected based on having at least 70% of ERP data available. There were no significant differences between infants with complete and missing data in age ( $t(138)=-1.46$ ,  $p=0.15$ ), gender ( $\chi^2(1)=2.08$ ,  $p=0.15$ ), clinical outcome ( $\chi^2(1)=0.002$ ,  $p=1$ ) or cognitive level at 8 months ( $t(128)=0.07$ ,  $p=0.94$ ), providing reasonable evidence to consider the pattern of missing data as “missing at random”. We used imputation through expectation maximization to handle missing data. The sample was split into a main sample (70% of sample,  $n=101$ ) for model selection and training, and a separate holdout sample (30% of sample,  $n=43$ ) for validation, with stratification for binary outcome (i.e., ASD vs. no-ASD). Standardized average and differential ERP responses to the different stimuli conditions were used as predictive features, together with gender and age to account for their confounding effect (see *Table S1*).

We performed top-down and bottom-up feature selection to extract information about the most relevant features for prediction of ASD (Dash, 1997). For top-down feature selection, we manually selected feature sets based on stimuli and type of measure (*Figure 1.B*). For bottom-up feature selection, we used a genetic algorithm ((Back, 1996; Johannesson et al., 2002; Snaedal et al., 2012); *Supplementary Material*) based on the optimization of the Area Under the Curve (AUC) of a 10-fold cross-validated Support Vector Machine (SVM) classifier with linear kernel. AUC is an effective and combined measure of sensitivity and specificity, which allows to test the inherent ability of the predictor, providing a useful metric to evaluate diffusivity of the predictive features within the examined population (Kumar & Indrayan, 2011). Selection resulted in the feature set providing the highest AUC over repeated ( $n=100$ ) evolution (‘optimal’ set), the set of features with highest incidence ( $>80\%$ , ‘highest incidence’ set) among the best performing feature sets ( $AUC>85\%$ ), and condition-specific subsets of the ‘highest incidence set’ (*Figure 1.B*). We validated the different classifiers on the holdout sample and evaluated performance through AUC, sensitivity, specificity, accuracy, negative predictive power (NPV), and positive predictive power (PPV) from the ROC curve, with bootstrapped ( $n=10000$ ) 95% confidence intervals (CI). We used a shuffle test with  $n=10000$  repetitions (Golland & Fischl, 2003) to test for significant differences of classification performance (AUC) from chance level prediction, and for significant differences in performance between the best performing classifier and the other classifiers. Analyses were completed using the *LIBSVM* toolbox (Chang, 2011) and custom scripts implemented on Matlab R2016b (MATLAB 9.1, The MathWorks Inc., Natick, MA, 2016).

### ***(C) Elastic-net regression***

Regression with elastic-net regularization (Zou & Hastie, 2005) was used to select relevant ERP features for prediction of social skills in toddlerhood (VABS Socialization score at 36 months) on the same subsample of EL infants included in the classification analysis (n=144), using the same ERP variables as predictors (*Figure 1.C*). As pre-processing, skewness of predictors was checked to be lower than 0.7 and none of them needed to be transformed, while all variables were standardized. Leave-one-out cross-validation was used to cross-validate the predictive model, and nested 10-fold cross-validation with 10 repetitions was used for parameter optimization based on minimization of the root mean squared error (RMSE). To evaluate predictive performance, we computed RMSE and the relative error (RMSE/range of outcome scores). Furthermore, we computed Pearson's correlation between predicted and observed variables, and checked residuals for normality and absence of heteroscedasticity. 95% confidence interval (CI) for RMSE was computed using bootstrap with n=1000 repetitions, while the p-value was computed through a shuffle test (Golland & Fischl, 2003).

### ***(D) Stratification into subgroups.***

To test whether neural processing of faces can define meaningful subgroups in infants at elevated likelihood for developing ASD, we used Bayesian hierarchical clustering (Marrelec, Messe, & Bellec, 2015; Savage et al., 2009) on averaged ERP responses (*Figure 1.D*) from infants in the EL group (n=144) as a model-based bottom-up process using marginal likelihoods to decide which clusters to merge. As number of clusters is determined automatically, we tested stability of results through leave-one out cross-validation. To characterize the identified clusters, we investigated differences between clusters in face processing through an analysis of variance (ANOVA). Next, we evaluated clustering performance by examining the association of cluster membership with clinical outcome variables (ADOS, ADI-R, SCQ, MSEL, VABS) at 36 months through ANOVA (see *Table S7*). Holm-Bonferroni correction was used to correct significance of ANOVAs for multiple comparisons *separately* for ERPs and clinical outcome variables. For significant findings, we used post-hoc pairwise Tukey's tests. Analyses were implemented in *R*.

## Results

### (A) Group-level differences

<Figure 2: Grand average ERPs across contrasts and groups, and means and standard errors of amplitude and latency of the three face-sensitive ERP components in each group>

We found a significant condition x outcome interaction on N290 latency in the face-noise contrast. Specifically, the EL-ASD group did not show a stimulus differentiation, while the TL ( $p=.010$ ,  $d=0.61$ ) and EL-no ASD ( $p=.021$ ,  $d=0.52$ ) groups showed longer latency to faces compared to noise, with no difference between TL and EL-no ASD groups ( $p=.551$ ,  $d=0.10$ ; *Figure 2*). This did not vary by cohort ( $F(1,170)=0.76$ ,  $p=.386$ ) and neither of the covariates had a significant interaction ( $ps>.41$ ).

There was a significant condition x outcome group interaction on P1 latency ( $F(2, 213)=4.95$ ,  $p=.008$ ), P400 amplitude ( $F(2,212)=4.13$ ,  $p=.017$ ) and latency ( $F(2, 208)= 3.51$ ,  $p=.032$ ). Specifically, the EL-ASD group had longer P1 latency to gaze shifting away versus towards, while the opposite effect was observed in the TL ( $p=.002$ ,  $d=0.74$ ) and EL-no ASD groups ( $p=.047$ ,  $d=0.43$ ), with no difference between TL and EL-no ASD ( $p=.122$ ,  $d=0.26$ ). There was no significant interaction with cohort ( $F(2,212)=1.48$ ,  $p=.226$ ). The condition x outcome interaction became a trend when age and non-verbal ability were entered as covariates ( $F(2,201) = 2.45$ ,  $p=.089$ ); lower non-verbal ability was associated with longer P1 latency to gaze shifting towards versus away ( $r=-.18$ ,  $p=.008$ ), with no association with age ( $r=.07$ ,  $p=.32$ ). Next, the EL-ASD group showed longer P400 latency to gaze shifting towards versus away from the viewer, with an opposite effect in TL ( $p=.011$ ,  $d=0.55$ ) and EL-no ASD ( $p=.021$ ,  $d=0.47$ ), and no significant difference between TL and EL-no ASD ( $p=.572$ ,  $d=0.10$ ). This did not vary by cohort ( $F(2,207)=0.91$ ,  $p=.342$ ) and was not influenced by covariates ( $ps>.24$ ). Post-hoc t-tests revealed significant differences between EL-ASD and TL ( $p=.009$ ,  $d=0.60$ ) and EL-no ASD ( $p=.019$ ,  $d=0.47$ ), but not between TL and EL-no ASD ( $p=.488$ ,  $d=0.12$ ). See *Supplementary Material* for an additional analysis on the association between findings for P1 and P400 latency. Finally, there was enhanced P400 amplitude to gaze shifting towards versus away in the EL-ASD group ( $p=.016$ ,  $d=0.46$ ) and EL-no ASD group ( $p=.014$ ,  $d=0.41$ ), with no difference between EL groups

( $p=.482$ ,  $d=0.12$ ) and an opposite effect in the TL group. There was no interaction with cohort ( $F(1,211) = 0.27$ ,  $p=.605$ ) nor with covariates ( $ps>.19$ ).

### **(B) Individual-level prediction of diagnosis**

*<Figure 3: Classification performance>*

Combined sets of brain responses including both early sensory (P1) and later higher-level ERP components (N290 and P400) in response to direct and averted gaze, static faces and visual noise (see *Figure 3*) provided the best predictive accuracy for ASD category at an individual level, while classification performance for condition-specific subsets of features showed largely poor predictive power. Specifically, the ‘Optimal’ set provided the best predictive algorithm for ASD outcome at 36 months (*Table 1*), with an AUC of 77.1% (95% CI: [61.1, 90.5],  $p=0.01$ ), significantly higher than classifiers built on condition-specific subsets selected *top-down*. Compared to the ‘Highest Incidence’ set, which provided a nominally but not significantly higher classification accuracy (AUC=77.5%, 95% CI: [61.8, 90.2],  $p=0.02$ ;  $p=0.15$ ), the ‘Optimal’ set provided a better sensitivity to the ASD cases (73.5% compared to 55.9%). Of note, gender and age were selected less than 6% of the times among best performing classifiers, suggesting that prediction did not depend on these confounding variables. *Table S4* shows details on classification performance.

### **(C) Individual-level prediction of social traits**

*<Figure 4: Prediction of social skills in toddlerhood.>*

Prediction of stronger social skills in toddlerhood using neural responses to faces and visual noise in infancy was promising (see *Figure 4.A*), with an average RMSE=13.17 (95% CI; [11.7; 14.6];  $p=0.002$ ) corresponding to 20.9% error relative to the score range in the sample, and a correlation between predicted and observed values  $\beta=0.24$  ( $t(137)=2.86$ ,  $p=0.005$ ). Of the 27 features relevant to prediction (*Figure 4.B*), N290 latency response to the face-noise contrast had the largest negative correlation to outcome, with faster N290 response to noise compared to faces associated to better social functioning in toddlerhood. There was a large overlap between predictive sets of features at the dimensional and

categorical levels. Of note, gender had the largest correlation to later social skills level, indicating that females tend to have higher social scores at age 3. We also found a negative association between age of ERP assessment and level of social functioning at age 3, which likely indicates a bias in the voluntary recruitment for the study.

#### **(D) Stratification into subgroups of EL infants**

*<Figure 5: Subgroups of EL-infants based on face processing in infancy.>*

We identified five stable clusters among EL infants indicating a diffuse pattern of differences in ERP components between subgroups for all stimuli except for dynamic gaze shifts. Cluster 1 (n=23, 3 EL-ASD) showed slower P400 to faces and reduced N290 amplitude to visual noise; cluster 2 (n=17, 4 EL-ASD) showed no specific differences compared to other clusters; cluster 3 (n=37, 12 EL-ASD) showed reduced P1, N290 and P400 amplitude to faces and visual noise, and longer P400 latency to eye-gaze; cluster 4 (n=25, 2 EL-ASD) showed increased N290 amplitude and shorter P400 latency to visual noise, and shorter N290 latency to faces and visual noise; cluster 5 (n=42, 11 EL-ASD) showed increased P1 and N290 amplitude, and longer P1 latency to faces and visual noise (see *Figure 5, Table S7*).

The identified clusters did not clearly map into clinical categories, and there were no significant differences between clusters on ASD symptoms, cognitive development and adaptive functioning at 36 months (see *Table S7*). However, the majority of infants in the EL-ASD group were in cluster 3 and 5, which respectively suggest hypo- and hypersensitivity to faces and visual noise. To test this hypothesis, we compared clusters on sensory sensitivity total scores from the Infant-Toddler Sensory Profile (Dunn & Daniels, 2002) at 36 months, reflecting low neurological threshold to stimuli and passive self-regulation. We found a significant main effect of clusters through ANOVA ( $F(4,139)=3.99, p=0.004$ ); however, post-hoc Tukey's tests showed higher sensory sensitivity in infants from cluster 3 (*mean [standard deviation] = 38.36 [11.48]*) compared to cluster 4 (*45.67 [8.92], p=0.006*) and cluster 5 (*44.06 [8.75], p=0.039*).

## **Discussion**

We present a set of analytic approaches to investigate *whether* and *how* a particular RDoC-defined domain is linked to later psychopathology in the context of a prospective design. Specifically, we focused on face processing as a putative precursor of the social features emerging later in development that are characteristic of ASD. Findings indicate a diffuse pattern of alterations in both early sensory and later higher-level stages of neural processing of social (face/gaze) and visual noise stimuli in infancy that strongly converged to predict ASD diagnosis and social traits in toddlerhood across categorical and dimensional analyses (*Table 1*). These findings support a theoretical framework in which diffuse and individually heterogeneous anomalies in social and perceptual processing converge to contribute to the later emergence of autism.

*<Table 1: Summary of findings across different analyses>*

### **Neural processing of faces is diffusely atypical in ASD**

Our analyses converged to support the contention that the socially relevant processes of detecting a face and a shift in gaze are altered in emerging ASD across a long time-course of information processing from the shortest latency components, and across multiple stimuli conditions (*Table 1*). Our data-driven analyses identified a specific pattern of interactions between these different ERP components, rather than specific components, to be predictive of individual-level outcome. This supports hypothesis (*a*) on diversity of altered face processing to reflect consistent differences in different specific aspects of face processing that are masked in previous studies by their theory-driven approach on one or two components.

Prediction of dimensional (social functioning) and categorical outcome (ASD diagnosis) strongly converged in terms of patterns of neural responses to face/noise stimuli that were predictive of later ASD (*Table 1*). This contributes to validate the hypothesis of altered face processing as a candidate neurocognitive process underlying the emergence of ASD over development. Although categorical variables are less compelling to validate underlying mechanisms for psychopathology from an RDoC perspective, a diagnostic outcome retains a high clinical interest for decision-making in a medical system (Casey et al., 2013), and thus provided an opportunity to explore predictive power of altered face processing for ASD at different, complementary levels of analysis. In particular, both data-driven approaches identified the exact

same responses to the face-noise contrast to be most predictive of later ASD, indicating an effect of social content of the stimulus on early alterations of face processing in ASD and a robust association to autism diagnosis and social traits in toddlerhood. Furthermore, neural responses to visual noise were selected among the most predictive features for socialization skills and ASD diagnosis in toddlerhood, suggesting that whilst alterations across different stages of face processing are the most salient features in emerging ASD, alterations in neural responses to non-social stimuli may also play a role. This is consistent with previous evidence of altered response both to social and non-social stimuli in emerging ASD (E. J. Jones et al., 2014), and extends recent work using nonlinear features of EEG signals to support widespread dysfunction not specific to social processing (Bosl, Tager-Flusberg, & Nelson, 2018).

Some components appeared to be more relevant to understanding the mechanisms underlying emergence of ASD social symptoms based on pervasiveness of effects across analyses. Specifically, N290 latency to visual noise versus faces emerged as a relevant feature across all levels of analysis (see *Table 1*), with infants in the EL-ASD group showing a diminished effects of longer N290 latency to faces compared to visual noise compared to TL and EL-no-ASD groups. N290 latency response to the face-noise contrast was selected as part of the “optimal” and “highest incidence” feature sets by the genetic algorithm, providing the best classification accuracies for ASD vs. no-ASD infants (approximately 77% accuracy); and it had the largest negative correlation to dimensional outcome, with faster N290 response to noise compared to faces associated to better social functioning in toddlerhood. The reduced N290 latency difference between the social versus non-social stimuli in EL-ASD is in line with previous studies of EL siblings (E. Jones et al., 2016), and in the comparison of familiar and unfamiliar faces (Geraldine Dawson et al., 2002; Key & Stone, 2012). The N290/P400 complex is thought to be a developmental precursor to the N170 (de Haan, Johnson, & Halit, 2003), an established marker for social functioning (Neuhaus, Kresse, Faja, Bernier, & Webb, 2016) with a long history of research for alterations in ASD (Kang et al., 2018; McPartland, Dawson, Webb, Panagiotides, & Carver, 2004). Early-stage differences in neural processing, as we found here, may subsequently trigger a cascade of events that result in symptoms characteristic of ASD (Mark H Johnson, Gliga, Jones, & Charman, 2015). Reduced depth of processing for social stimuli may result, in fact, in failure to develop expertise in processing faces along a cumulative risk pathway.



Longer P1 latency to gaze shifting towards versus away from the infant could predict later ASD diagnosis at a group level, and could predict reduced social functioning in toddlerhood at an individual level. Similarly, longer P400 latency to gaze shifting towards versus away from the infant could predict later ASD diagnosis at a group level, and was included in the optimal feature set predicting individual ASD diagnosis with approximately 77% accuracy among siblings at elevated likelihood for ASD. Previous work on a subset of this cohort showed that reduced P100 latency differentiation of dynamic gaze predicts ASD outcome in toddlerhood (Mayada Elsabbagh et al., 2012), while reduced P400 differentiation of dynamic gaze predicts ASD outcome in middle childhood (Bedford et al., 2017). Here, neural responses to dynamic gaze shifts alone performed as best candidate precursors of ASD at a group level but did not provide sufficient predictive value at an individual level due to a significant overlap in individual variation. Nevertheless, neural responses to dynamic gaze were selected among relevant features for prediction of later ASD diagnosis and social traits, suggesting that alterations in neural processing of dynamic gaze are not sufficient but necessary conditions for emerging ASD.

### **Stratification into subgroups of face processing**

We identified five different subgroups of EL-infants differing in intensity and latency of response in early sensory (P1) and later higher-order stages (N290 and P400) of processing faces with direct and averted gaze and visual noise.

Infants with later ASD outcome were mainly partitioned into two subgroups characterized by lower (cluster 3) or higher (cluster 5) amplitude of neural responses to faces and visual noise, which was associated to respectively increased and reduced sensory sensitivity at 36 months. A previous study has found sensory sensitivity at age 2 years to predict greater attention capture by faces and more optimal social behaviour at 4 years (E. J. H. Jones, G. Dawson, & S. J. Webb, 2018). Here, we looked at a younger age and found that infants who are less responsive to face-noise stimuli are more passive but also more sensitive responders to environmental stimuli in toddlerhood. In line with the idea of ASD as the a result of early life adaptation to altered neural processing (M. H. Johnson, Jones, & Gliga, 2015), increased sensory sensitivity may result from a system that was initially under-reactive but then amplifies sensory input over

development. This could also explain why relations between neural responses to faces and sensory sensitivity seem to change over development, and highlights the need to investigate mechanisms across RDoC domains. Specifically, future studies should investigate the mechanisms underlying the association between neural sensitivity to social stimuli and sensory alterations observed in ASD. Nevertheless, sensory sensitivity may index a child's likelihood to benefit from supportive environments during development. Thus, our findings indicate subgroups among infants developing ASD that could benefit most from different tailored intervention programs based on their sensory response to the environment.

Of note, differences between clusters in clinical outcome variables were not significant. While statistical power might improve with increased sample size, this can be interpreted in light of a conceptualization of ASD as an epiphenomenon of earlier-interacting susceptibilities (Constantino, 2018). Although altered face processing represents one of these early measurable liabilities to ASD, it might not fully capture alone the heterogeneity in clinical expression of the disorder. Future work should integrate longitudinal measures to further refine our understanding of subgroups of face processing in emerging ASD. Furthermore, future work might employ a different data-driven approach to parse individual heterogeneity and identify more homogeneous and replicable subgroups, which would potentially improve predictive accuracy and allow the identification of more specific physiological mechanisms (Bussu et al., 2019; Loth et al., 2017).

### **Limitations and future directions**

Early intervention for infants at elevated likelihood for ASD, prior to the emergence of core ASD features, is supported and conducted at the group level (Green et al., 2015; Green et al., 2017; S. J. Webb, Jones, Kelly, & Dawson, 2014). Yet, individual prediction of ASD in the first year of life might be crucial to enable targeted intervention within a critical developmental window. ERPs represent a cost-effective, mobile, and infant-friendly neuroimaging technology, providing potential utility for inclusion as proxy outcome markers for intervention trials. Future work should determine whether these parameters are sensitive to the effects of early intervention (E. J. H. Jones, Dawson, Kelly, Estes, & Jane Webb, 2017). It is important to consider methodological limitations in machine learning related to relatively small sample size

and model reliability (Yahata et al., 2017). Currently, there is no good theoretical justification for features selected, and results may vary accordingly. Although we validated our results on a separate, holdout sample, generalizability of the identified model must be tested through replication on an independent sample. Furthermore, machine-learning algorithms were significantly but not strongly predictive of dimensional outcome, thus future work should explore incorporation of other measures (e.g. genetic factors, brain imaging, parent-child interaction and behavioural measures) to capture multidimensional profiles.

## **Conclusion**

Our study represents the first attempt to investigate robustness and generalizability of findings on a specific RDoC domain (social cognition/face processing) in infancy across different levels of analysis, from group-based comparisons to individual-level prediction of outcome and stratification of ASD diagnosis and traits in toddlerhood, addressing the emerging need in developmental neuroscience to incorporate RDoC constructs. Prediction of categorical and dimensional outcomes strongly converge on a diffuse pattern of alterations across the time-course of neural processing of face/noise stimuli. Furthermore, we identified subgroups of infants based on neural sensitivity to faces in infancy, that also show different sensory sensitivity in toddlerhood, which might guide tailoring interventions early in development to improve later outcome. This adds to the literature illustrating early structural alterations (Hazlett et al., 2017) by showing early diffuse functional alterations, in line with the idea of ASD as a syndrome emerging from diffuse and interacting liabilities (Constantino, 2018).

## **References**

- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Washington, DC.
- Arbabshirani, M. R., Plis, S., Sui, J., & Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage*, *145*(Pt B), 137-165. doi:10.1016/j.neuroimage.2016.02.079
- Back, T. (1996). Evolution strategies: An alternative evolutionary algorithm. *Artificial Evolution*, *1063*, 3-20.
- Bearden, C. E., & Freimer, N. B. (2006). Endophenotypes for psychiatric disorders: ready for primetime? *Trends in Genetics*, *22*(6), 306-313. doi:10.1016/j.tig.2006.04.004
- Bedford, R., Gliga, T., Shephard, E., Elsabbagh, M., Pickles, A., Charman, T., & Johnson, M. H. (2017). Neurocognitive and observational markers: prediction of autism spectrum disorder from infancy to mid-childhood. *Mol Autism*, *8*, 49. doi:10.1186/s13229-017-0167-3
- Bosl, W. J., Tager-Flusberg, H., & Nelson, C. A. (2018). EEG analytics for early detection of autism spectrum disorder: a data-driven approach. *Scientific reports*, *8*(1), 6828.
- Bussu, G., Jones, E. J. H., Charman, T., Johnson, M. H., Buitelaar, J. K., & Team, B. (2019). Latent trajectories of adaptive behaviour in infants at high and low familial risk for autism spectrum disorder. *Mol Autism*, *10*, 13. doi:10.1186/s13229-019-0264-6
- Casey, B. J., Craddock, N., Cuthbert, B. N., Hyman, S. E., Lee, F. S., & Ressler, K. J. (2013). DSM-5 and RDoC: progress in psychiatry research? *Nat Rev Neurosci*, *14*(11), 810-814. doi:10.1038/nrn3621
- Chang, C.-C. a. L., Chih-Jen. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 27:21--27: 27.
- Constantino, J. N. (2018). Deconstructing autism: from unitary syndrome to contributory developmental endophenotypes. *Int Rev Psychiatry*, *30*(1), 18-24. doi:10.1080/09540261.2018.1433133
- Dash, M. a. L., H. (1997). Feature selection for classification. *Intelligent Data Analysis*, *1*, 131-156.
- Dawson, G., Bernier, R., & Ring, R. H. (2012). Social attention: a possible early indicator of efficacy in autism clinical trials. *J Neurodev Disord*, *4*(1), 11. doi:10.1186/1866-1955-4-11
- Dawson, G., Carver, L., Meltzoff, A. N., Panagiotides, H., McPartland, J., & Webb, S. J. (2002). Neural correlates of face and object recognition in young children with autism spectrum disorder, developmental delay, and typical development. *Child development*, *73*(3), 700-717.
- Dawson, G., Webb, S. J., Wijsman, E., Schellenberg, G., Estes, A., Munson, J., & Faja, S. (2005). Neurocognitive and electrophysiological evidence of altered face processing in parents of children with autism: implications for a model of abnormal development of social brain circuitry in autism. *Dev Psychopathol*, *17*(3), 679-697. doi:10.1017/S0954579405050327
- de Haan, M. (2007). *Infant EEG and event-related potentials*. Hove, England: Psychology Press.
- de Haan, M., Johnson, M. H., & Halit, H. (2003). Development of face-sensitive event-related potentials during infancy: a review. *Int J Psychophysiol*, *51*(1), 45-58.
- Dunn, W., & Daniels, D. B. (2002). Initial development of the infant/toddler sensory profile. *Journal of Early Intervention*, *25*(1), 27-41. doi:Doi 10.1177/105381510202500104
- Elsabbagh, M. (2020). Linking risk factors and outcomes in autism spectrum disorder: is there evidence for resilience? *BMJ*, *368*, l6880. doi:10.1136/bmj.l6880
- Elsabbagh, M., Mercure, E., Hudry, K., Chandler, S., Pasco, G., Charman, T., . . . , B. A. S. I. S. T. (2012). Infant neural sensitivity to dynamic eye gaze is associated with later emerging autism. *Curr Biol*, *22*(4), 338-342.

- Golland, P., & Fischl, B. (2003). Permutation tests for classification: towards statistical significance in image-based studies. *Inf Process Med Imaging*, 18, 330-341.
- Green, J., Charman, T., Pickles, A., Wan, M. W., Elsabbagh, M., Slonims, V., . . . Gliga, T. (2015). Parent-mediated intervention versus no intervention for infants at high risk of autism: a parallel, single-blind, randomised trial. *The Lancet Psychiatry*, 2(2), 133-140.
- Green, J., Pickles, A., Pasco, G., Bedford, R., Wan, M. W., Elsabbagh, M., . . . Cheung, C. (2017). Randomised trial of a parent - mediated intervention for infants at high risk for autism: longitudinal outcomes to age 3 years. *Journal of Child Psychology and Psychiatry*, 58(12), 1330-1340.
- Grice, S. J., Halit, H., Farroni, T., Baron-Cohen, S., Bolton, P., & Johnson, M. H. (2005). Neural correlates of eye-gaze detection in young children with autism. *Cortex*, 41(3), 342-353.
- Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., . . . Statistical, A. (2017). Early brain development in infants at high risk for autism spectrum disorder. *Nature*, 542(7641), 348-351. doi:10.1038/nature21369
- Huguet, G., Benabou, M., & Bourgeron, T. (2016). The Genetics of Autism Spectrum Disorders. In P. Sassone-Corsi & Y. Christen (Eds.), *A Time for Metabolism and Hormones* (pp. 101-129). Cham (CH).
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry*, 167(7), 748-751. doi:10.1176/appi.ajp.2010.09091379
- Johannesson, G. H., Bligaard, T., Ruban, A. V., Skriver, H. L., Jacobsen, K. W., & Norskov, J. K. (2002). Combined electronic structure and evolutionary search approach to materials design. *Phys Rev Lett*, 88(25 Pt 1), 255506. doi:10.1103/PhysRevLett.88.255506
- Johnson, M. H., Gliga, T., Jones, E., & Charman, T. (2015). Annual Research Review: Infant development, autism, and ADHD—early pathways to emerging disorders. *Journal of Child Psychology and Psychiatry*, 56(3), 228-247.
- Johnson, M. H., Jones, E. J., & Gliga, T. (2015). Brain adaptation and alternative developmental trajectories. *Dev Psychopathol*, 27(2), 425-442. doi:10.1017/S0954579415000073
- Jones, E., Dawson, G., & Webb, S. (2018). Sensory hypersensitivity predicts enhanced attention capture by faces in the early development of ASD. *Developmental Cognitive Neuroscience*, 29, 11-20.
- Jones, E., Venema, K., Earl, R., Lowy, R., Barnes, K., Estes, A., . . . Webb, S. (2016). Reduced engagement with social stimuli in 6-month-old infants with later autism spectrum disorder: a longitudinal prospective study of infants at high familial risk. *Journal of neurodevelopmental disorders*, 8(1), 7.
- Jones, E. J., Gliga, T., Bedford, R., Charman, T., & Johnson, M. H. (2014). Developmental pathways to autism: a review of prospective studies of infants at risk. *Neurosci Biobehav Rev*, 39, 1-33. doi:10.1016/j.neubiorev.2013.12.001
- Jones, E. J. H., Dawson, G., Kelly, J., Estes, A., & Jane Webb, S. (2017). Parent-delivered early intervention in infants at risk for ASD: Effects on electrophysiological and habituation measures of social attention. *Autism Res*, 10(5), 961-972. doi:10.1002/aur.1754
- Jones, E. J. H., Dawson, G., & Webb, S. J. (2018). Sensory hypersensitivity predicts enhanced attention capture by faces in the early development of ASD. *Dev Cogn Neurosci*, 29, 11-20. doi:10.1016/j.dcn.2017.04.001
- Jones, E. J. H., Mason, L., Begum Ali, J., van den Boomen, C., Braukmann, R., Cauvet, E., . . . Eurosibs, T. (2019). Eurosibs: Towards robust measurement of infant neurocognitive predictors of autism across Europe. *Infant Behav Dev*, 57, 101316. doi:10.1016/j.infbeh.2019.03.007
- 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(7480), 427.

- Kang, E., Keifer, C. M., Levy, E. J., Foss-Feig, J. H., McPartland, J. C., & Lerner, M. D. (2018). Atypicality of the N170 Event-Related Potential in Autism Spectrum Disorder: A Meta-analysis. *Biol Psychiatry Cogn Neurosci Neuroimaging*, 3(8), 657-666. doi:10.1016/j.bpsc.2017.11.003
- Key, A. P., & Stone, W. L. (2012). Processing of novel and familiar faces in infants at average and high risk for autism. *Developmental cognitive neuroscience*, 2(2), 244-255.
- Kumar, R., & Indrayan, A. (2011). Receiver operating characteristic (ROC) curve for medical researchers. *Indian pediatrics*, 48(4), 277-287.
- Lai, M. C., Lombardo, M. V., Chakrabarti, B., & Baron-Cohen, S. (2013). Subgrouping the autism "spectrum": reflections on DSM-5. *PLoS Biol*, 11(4), e1001544. doi:10.1371/journal.pbio.1001544
- Lombardo, M. V., Lai, M. C., Auyeung, B., Holt, R. J., Allison, C., Smith, P., . . . Baron-Cohen, S. (2016). Unsupervised data-driven stratification of mentalizing heterogeneity in autism. *Sci Rep*, 6, 35333. doi:10.1038/srep35333
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., . . . Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3), 205-223.
- Loth, E., Charman, T., Mason, L., Tillmann, J., Jones, E. J. H., Wooldridge, C., . . . Buitelaar, J. K. (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. doi:10.1186/s13229-017-0146-8
- Maestro, S., Muratori, F., Cavallaro, M. C., Pei, F., Stern, D., Golse, B., & Palacio-Espasa, F. (2002). Attentional Skills During the First 6 Months of Age in Autism Spectrum Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 41(10), 1239-1245.
- Marrelec, G., Messe, A., & Bellec, P. (2015). A Bayesian Alternative to Mutual Information for the Hierarchical Clustering of Dependent Random Variables. *PLoS One*, 10(9), e0137278. doi:10.1371/journal.pone.0137278
- McPartland, J., Dawson, G., Webb, S. J., Panagiotides, H., & Carver, L. J. (2004). Event-related brain potentials reveal anomalies in temporal processing of faces in autism spectrum disorder. *J Child Psychol Psychiatry*, 45(7), 1235-1245. doi:10.1111/j.1469-7610.2004.00318.x
- Mullen, E. (1995). *Mullen Scales of Early Learning*. Circle Pines MN: American Guidance Service.
- Neuhaus, E., Kresse, A., Faja, S., Bernier, R. A., & Webb, S. J. (2016). Face processing among twins with and without autism: social correlates and twin concordance. *Soc Cogn Affect Neurosci*, 11(1), 44-54. doi:10.1093/scan/nsv085
- Osterling, J., & Dawson, G. (1994). Early recognition of children with autism: A study of first birthday home videotapes. *Journal of Autism and Developmental Disorders*, 24(3), 247-257. doi:10.1007/bf02172225
- Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., . . . Dobkins, K. (2011). Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics*, 128(3), e488-e495.
- Rosenberg, M. D., Casey, B. J., & Holmes, A. J. (2018). Prediction complements explanation in understanding the developing brain. *Nat Commun*, 9(1), 589. doi:10.1038/s41467-018-02887-9
- Rutter, M., Bailey, A., Lord, C. (2003). *The social communication questionnaire: manual.*: Western Psychological Services.
- Salomone, E., Shephard, E., Milosavljevic, B., Johnson, M. H., Charman, T., & Team, B. (2018). Adaptive Behaviour and Cognitive Skills: Stability and Change from 7 Months to 7

- Years in Siblings at High Familial Risk of Autism Spectrum Disorder. *J Autism Dev Disord*. doi:10.1007/s10803-018-3554-5
- Savage, R. S., Heller, K., Xu, Y., Ghahramani, Z., Truman, W. M., Grant, M., . . . Wild, D. L. (2009). R/BHC: fast Bayesian hierarchical clustering for microarray data. *BMC Bioinformatics*, *10*, 242. doi:10.1186/1471-2105-10-242
- Snaedal, J., Johannesson, G. H., Gudmundsson, T. E., Blin, N. P., Emilsdottir, A. L., Einarsson, B., & Johnsen, K. (2012). Diagnostic accuracy of statistical pattern recognition of electroencephalogram registration in evaluation of cognitive impairment and dementia. *Dementia and geriatric cognitive disorders*, *34*(1), 51-60.
- Sparrow, S., Balla, D., & Cicchetti, D. (1984). *The Vineland Adaptive Behavior Scales: Interview edition, Survey form*: Circle Pines, MN: American Guidance Service.
- Vorstman, J. A. S., Parr, J. R., Moreno-De-Luca, D., Anney, R. J. L., Nurnberger, J. I., Jr., & Hallmayer, J. F. (2017). Autism genetics: opportunities and challenges for clinical translation. *Nat Rev Genet*, *18*(6), 362-376. doi:10.1038/nrg.2017.4
- Webb, S. J., Jones, E. J., Kelly, J., & Dawson, G. (2014). The motivation for very early intervention for infants at high risk for autism spectrum disorders. *Int J Speech Lang Pathol*, *16*(1), 36-42. doi:10.3109/17549507.2013.861018
- Webb, S. J., Jones, E. J., Merkle, K., Venema, K., Greenson, J., Murias, M., & Dawson, G. (2011). Developmental change in the ERP responses to familiar faces in toddlers with autism spectrum disorders versus typical development. *Child development*, *82*(6), 1868-1886.
- Yahata, N., Kasai, K., & Kawato, M. (2017). Computational neuroscience approach to biomarkers and treatments for mental disorders. *Psychiatry Clin Neurosci*, *71*(4), 215-237. doi:10.1111/pcn.12502
- Zhao, Y. H., & Castellanos, F. X. (2016). Annual Research Review: Discovery science strategies in studies of the pathophysiology of child and adolescent psychiatric disorders - promises and limitations. *Journal of Child Psychology and Psychiatry*, *57*(3), 421-439. doi:10.1111/jcpp.12503
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net (vol B 67, pg 301, 2005). *Journal of the Royal Statistical Society Series B-Statistical Methodology*, *67*, 768-768. doi:DOI 10.1111/j.1467-9868.2005.00527.x

**Table 1: Summary of findings across different analyses**

Analysis	Static Gaze Condition			Dynamic Gaze Shift Condition			Face vs. Noise Condition		
	Direct Gaze	Averted Gaze	Direct vs. Averted	Shifts Towards	Shift Away	Towards vs. Away	Static Face	Visual Noise	Face vs. Noise
Group-based comparison						P1 latency* P400 amplitude** P400 latency***			N290 latency****
Supervised classification†	P1 amplitude# N290 amplitude# P400 amplitude	<b>P1 amplitude</b> N290 amplitude <b>P400 amplitude</b> ¥	P1 amplitude# P400 latency#	N290 amplitude¥ <b>P400 amplitude</b> #	P400 amplitude#	P1 amplitude# <b>P400 latency</b> #	P1 amplitude# <b>N290 amplitude</b> <b>P400 amplitude</b> #	P1 amplitude# P400 latency#	<b>P1 amplitude</b> <b>P1 latency</b> # <b>N290 latency</b> <b>P400 amplitude</b> #
Prediction of dimensional outcome	P1 latency <b>N290 amplitude</b> N290 latency <b>P400 amplitude</b>	<b>P1 amplitude</b> P1 latency N290 latency <b>P400 amplitude</b>	N290 amplitude P400 amplitude	P1 amplitude <b>P400 amplitude</b> P400 latency	P1 amplitude P1 latency	<b>P1 latency</b> N290 amplitude	N290 latency <b>P400 amplitude</b> P400 latency	P1 latency	<b>P1 amplitude</b> <b>P1 latency</b> <b>N290 latency</b> <b>P400 amplitude</b>
Stratification into subgroups	P1 amplitude P1 latency N290 amplitude N290 latency P400 amplitude P400 latency	P1 amplitude P1 latency N290 amplitude N290 latency P400 amplitude P400 latency					P1 amplitude P1 latency N290 amplitude <b>N290 latency</b> P400 amplitude P400 latency	P1 amplitude P1 latency N290 amplitude <b>N290 latency</b> P400 amplitude P400 latency	

This table shows the ERP measures that emerged as significant across the different analyses, in terms of significant differences between groups and/or being selected by the genetic algorithm for classification of EL-ASD vs. EL-no ASD and/or being selected by the elastic-net regression for prediction of later social skills and/or showing significant differences between subgroups of the EL group. Shaded grey indicates no significant results for the specific stimuli/conditions in a specific analysis. **The measures appearing as significant at group and individual levels of analysis are highlighted in bold; significant at all levels of analysis in bold red; and bold underlined when significant for prediction of both diagnosis and traits.**

\*  $F(2,213)=4.95, p=0.008$ ; shifts towards > shifts away in TL ( $d=0.74; p=0.002$ ) and EL-no-ASD ( $d=0.43; p=0.047$ ).

\*\*  $F(2,212)=4.13, p=0.017$ ; shifts towards > shifts away in EL-ASD ( $d=0.46; p=0.02$ ) and EL-no-ASD ( $d=0.41; p=0.01$ ).

\*\*\*  $F(2,208)=3.51, p=0.032$ ; shifts towards > shifts away in TL ( $d=0.55; p=0.01$ ) and EL-no-ASD ( $d=0.47; p=0.02$ ); with  $p=0.009$  for EL-ASD vs. TL and  $p=0.019$  for EL-ASD vs. EL-no-ASD.



\*\*\*\*  $F(2,171)=3.61$ ,  $p=0.029$ ; *face > visual noise* in TL ( $d=0.61$ ;  $p=0.01$ ) and EL-no ASD ( $d=0.52$ ;  $p=0.02$ ).

† The 'optimal' feature set was the best performing classifier, with AUC = 77.1% [61.1, 90.5],  $p=0.01$ ; accuracy = 75.7% [69.1, 90.0]; sensitivity = 73.5% [41.2, 91.2]; specificity = 77.8% [66.7, 100]; PPV = 76.8% [71.2, 100]; NPV = 74.6% [62.7, 89.5].

# Features in the 'optimal' but not in the 'highest incidence' feature set.

¥ Features selected as part of the 'highest incidence' but not the 'optimal' feature set.

## FIGURES

**Figure 1: Flow chart of statistical analysis strategy.** This figure illustrates the different analyses performed in this study. Specifically, group-level comparison (A); supervised classification for prediction of categorical outcome at 36 months of age (B); elastic-net regression for prediction of dimensional outcome at 36 months of age (B); and stratification into subgroups of infants at elevated likelihood for autism (D).

**Figure 2: Grand average event-related potentials.** Grand average ERPs across contrasts and groups over task-sensitive occipito-temporal channels (left) and means and standard errors of amplitude and latency of the three face-sensitive ERP components in each group (right).

**Figure 3: Classification performance.** Receiver Operating Characteristic (ROC) curve for classifiers using different set of features to classify EL-ASD among EL siblings. Random predictors result in bisecting lines as ROC curves (dashed line), while deviations in the upper hemifield indicate an increase in predictive accuracy. Only classifiers with a classification performance significantly different from chance level (assessed through a shuffle test) are included in this figure.

*Abbreviations:* TPR = true positive rate or sensitivity; FPR = false positive rate, or 1-specificity.

**Figure 4: Prediction of social skills in toddlerhood.** This figure shows results from the elastic-net regression analysis. Specifically, panel A shows predicted against observed values of VABS socialization scores at 36 months. The experimental regression line is depicted with a shaded area for the 95% confidence interval, compared to the chance level regression line shown separately. Panel B shows average and standard deviation of coefficients always selected over cross-validation folds (n=144), indicating the relevant variables for individual prediction of social skills. Coefficients are in standard units.

*Abbreviations:* SD = gaze shifts directed towards the infant; SA = gaze shifts directed away from the infant; FD = direct gaze; FA = averted gaze; N = visual noise; F = static face irrespective of gaze direction.

**Figure 5. Subgroups of EL-infants based on face processing in infancy.** Five clusters were identified among infants at elevated likelihood for ASD. This figure illustrates clusters comparisons on ERPs. Specifically, it shows only differences that were statistically significant after using Holm-Bonferroni correction for multiple

comparisons: P1, N290 and P400 measured in amplitude and latency in response to direct gaze, averted gaze, static faces and visual noise.