

## 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>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

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

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 was presented as a poster at the 10<sup>th</sup> Donders Discussions in Nijmegen, 2017; and as an oral presentation at INSAR in Rotterdam, 2018, and at ETADE Brainview conference in London, 2018.

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 a Tuberous Sclerosis Association fellowship and the NIHR Maudsley Biomedical Research Centre (BRC) at the Institute of Psychiatry, Psychology & Neuroscience and South London and Maudsley NHS Foundation Trust. 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 atypicalities 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 a clinical diagnosis 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 like supervised classification for prediction of clinical outcome at 36 months and Bayesian hierarchical clustering for stratification into subgroups. Our findings converge to indicate that a broad profile of alterations in the time-course of neural processing of faces is an early predictor of later ASD diagnosis. Furthermore, we identified two brain response-defined subgroups in ASD that showed distinct alterations in different aspects of face processing compared to siblings without ASD diagnosis, suggesting that individual differences between infants contribute to the diffuse pattern of alterations predictive of ASD in the first year of life. This study shows that 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 outcome at an individual level in the first year of life.

## 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 outcome.

While most genetic studies treat ASD as a unitary clinical category, the majority of ASD risk 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). Altered face processing has been shown to be a strong candidate precursor of ASD diagnosis and its social features (Mayada Elsabbagh et al., 2012; E. Jones et al., 2016), and 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* atypicalities in face processing contribute to a later ASD outcome.

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). While some studies have suggested that low-level sensory sensitivity to faces in infancy is associated with better social development (E. Jones, Dawson, & 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). Altered responses to faces versus non-faces, and reduced differentiation of faces that shift gaze towards versus away from the viewer from 6 months of age are observed in infants with later ASD, as indexed by higher-level cortical responses (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, examining neural responses to faces in infants with later ASD provides an excellent context in which to interrogate the generalisability or specificity of the mechanisms through which socio-cognitive atypicalities relate to ASD emergence.

Whilst previous research converges in identifying face processing as a relevant domain in ASD, studies have report greater divergence in the *nature* of ASD-associated atypicalities. Neural responses to faces can be characterized by many different features of an averaged waveform, which are hypothesized to reflect different underlying cognitive processes. These ERP components are typically identified by their polarity and timing. 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 atypicality that will or will not be apparent at the group level depending on its composition; and/or c) distinct profiles of atypicality in particular coherent subgroups of individuals with ASD, again apparent or not at the group level depending on the sample composition. To understand these patterns, we need to couple both 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), research on infants with an older sibling with ASD (“infant siblings”) has accumulated over the past decade, with concomitant progress in characterising candidate neural and cognitive precursors of symptom emergence in ASD (E. J. Jones et al., 2014; Szatmari et al., 2016). Prospective studies of infant siblings represent a powerful research design to identify such precursors (E. J. H. Jones et al., 2019), and hold the potential to generate developmental data that could transform our understanding of the nature of the ASD diagnostic category itself. Second, we combine group-based comparisons with investigating individual effects through a data-driven multi-feature machine learning approach. Since domains are multifaceted rather than single cognitive processes, such approach enabled us to examine the consistency of our results across both top-down prediction and bottom-up discovery and build a robust multivariate model for data integration and prediction of later clinical outcome at an individual level (Arbabshirani, Plis, Sui, & Calhoun, 2017; Rosenberg, Casey, & Holmes, 2018; Yahata, Kasai, & Kawato, 2017). Third, we decomposed heterogeneity through stratifying the ASD group with a clustering approach based on the domain under investigation (here social cognition/face processing) to examine whether subgroups showed qualitatively different face processing atypicalities (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, 241 were seen for assessment around their third birthday by an independent team. Two TL children were absent for the 36-month visit but were included in the analysis as they showed typical development at the previous visits. Among the remaining 245 infants, 29 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

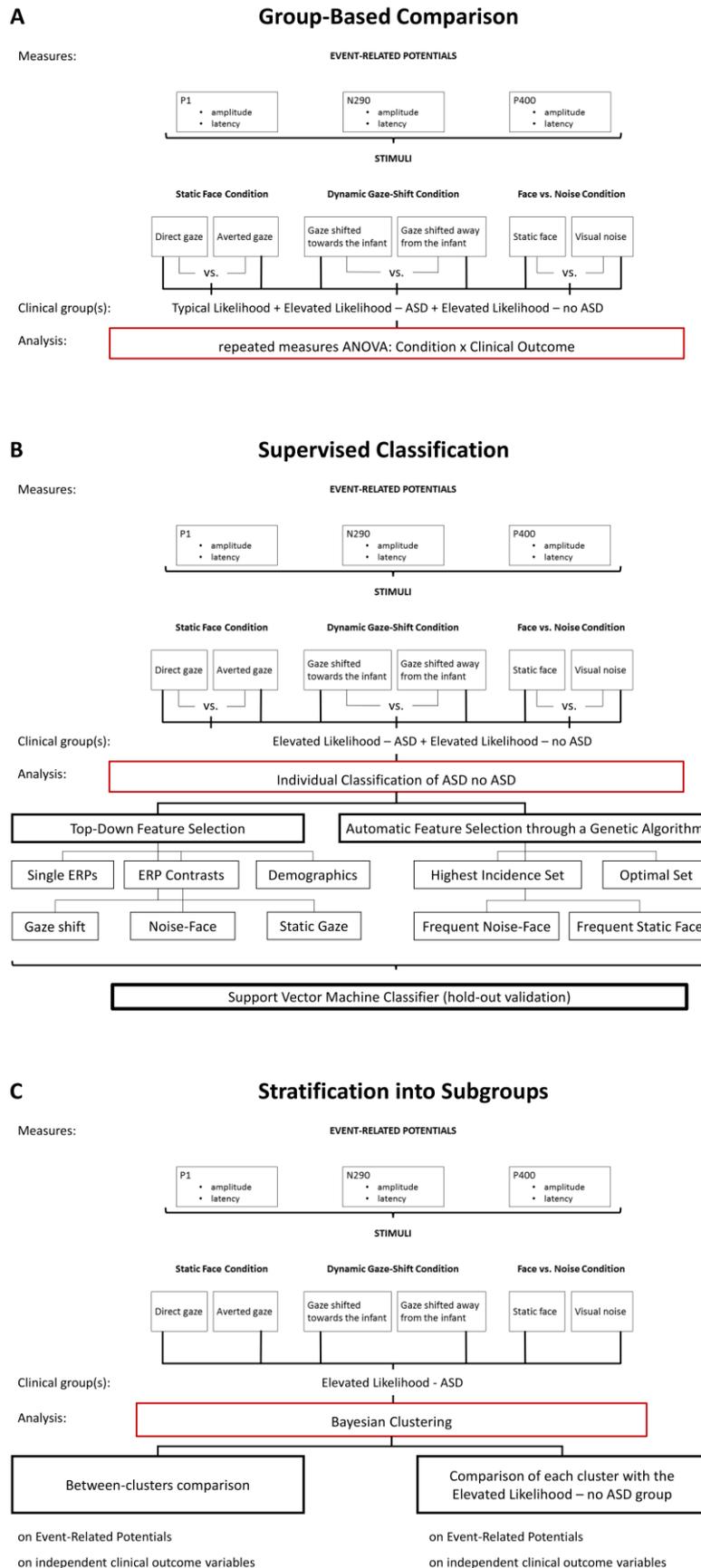
The Autism Diagnostic Observation Schedule (ADOS)-generic(Lord et al., 2000), a semi-structured observational assessment, and the Social Communication Questionnaire (SCQ(Rutter, 2003)), a screening tool for ASD, were used to assess current symptoms of ASD at 36 months. The Autism Diagnostic Interview – 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 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 sex ( $\chi^2(2) = 13.5$ ,  $p = 0.001$ ), with more males receiving an ASD diagnosis than females (odds ratio, OR = 4.84; 95% confidence interval [CI; 1.93 to 12.1];  $p < 0.001$ ).

### **Electrophysiological measures**

The task was the same as in Elsabbagh et al. (2012). It was 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). Components P100, N290, and P400 averaged across occipito-temporal channels were quantified by amplitude and latency in response to the different stimuli and the specified contrasts, and used as input features for subsequent analyses. See *Supplemental Materials* for details.

### **Statistical analysis**



**Figure 1: Flow chart of statistical analysis strategy**

### ***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>).

### ***Supervised Classification.***

Supervised classification was performed on EL infants test classification accuracy of ERP measures to predict clinical outcome at 36 months at an individual level (*Figure 1.B*). A subsample of 144 EL infants was included in this analysis based on having at least 70% of ERP data available. Imputation through expectation maximization was used to handle missing data, which showed a pattern of data Missing At Random. Standardized average and differential ERP responses to the different stimuli conditions (see *Figure 1.B*) were included as input features for the classifier. Gender and age were also included as features to take into account their confounding effect (see *Table S1*).

To investigate whether a subset of ERP measures, based on stimuli conditions or stage of processing, had a higher predictive value compared to the entire set, we performed top-down and bottom-up feature selection and compared classification performance of the entire feature set ('all' set) with that of the selected subsets (see *Figure 1.B*). For top-down feature selection, we manually selected 6 different sets of features: (1) differential ERP responses to all stimuli conditions ('ERP contrasts'); (2) averaged ERP responses to all conditions ('single ERPs'); (3) averaged ERP responses to the Face-vs-Noise condition ('noise-face'); (4) averaged ERP responses to the static face condition ('static face'); (5) averaged ERP responses to the dynamic gaze shift condition ('gaze shift'); (6) age and gender ('demographics'). For bottom-up feature selection, we

used a genetic algorithm ((Johannesson et al., 2002; Snaedal et al., 2012); *Supplementary Material*) based on the optimization of the Area Under the Curve (AUC) as 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).

For classification, we used SVM classifiers with linear kernel. The sample was split into a main sample (70% of sample, n=101) for model selection, and a separate holdout sample (30% of sample, n=43) for validation. The sample partition was stratified for binary outcome (i.e., ASD vs. no-ASD). Each classifier was then fully cross-validated via 10-fold cross-validation on the main sample, with sample partitioning into folds stratified for binary outcome. The number of features was selected based on the AUC level reached during the evolutionary process, and stability of the process assessed through visual inspection. Once selected the number of features (n=21), the evolutionary process was repeated n=100 times to investigate the variability in the feature space. Feature sets with highest AUC (>85%) were used as input to a frequency analysis on the selected features, with incidence as an estimate of the relevance of each feature for the classification problem. The feature set providing the highest AUC ('optimal' set), the set of features with highest incidence (>80%, 'highest incidence' set), and condition-specific subsets of the 'highest incidence set' ('frequent noise-face' set, 'frequent static face' set) were selected as input for subsequent classifier analysis.

After feature selection, we performed classification through linear-kernel SVM classifiers built on the entire feature set and subsets of feature selection (see *Figure 1.B*), trained on the main sample, and tested on the holdout validation sample. To evaluate classification performance, we computed AUC, sensitivity, specificity, accuracy, negative predictive power (NPV), and positive predictive power (PPV) from the ROC curve. 95% confidence intervals (CI) for each performance metric were computed using bootstrap with n=10000 repetitions. We tested for significant differences of classification performance, indexed by the AUC, with chance level prediction through a shuffle test with n=10000 repetitions (Golland & Fischl, 2003). The same procedure was used to test 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).

### ***Stratification into subgroups.***

To test whether neural processing of faces can define meaningful subgroups in ASD, we performed a clustering analysis within the EL-ASD group. We used Bayesian hierarchical clustering (BHC, (Savage et al., 2009)) on averaged ERP responses from infants in the EL-ASD group ( $n=32$ ) to each condition (36 features, see Figure 1.C). BHC is a model-based clustering algorithm built on a Dirichlet process mixture to model uncertainty in the data. Compared to other clustering techniques, it overcomes common limitations such as relatively arbitrary selection of number of clusters, or distance metric (Marrelec, Messe, & Bellec, 2015). It uses, in fact, marginal likelihoods to decide which clusters to merge at each step of a bottom-up hierarchical clustering process. The use of Bayesian hypothesis-testing as model-based criterion for merging clusters is also advantageous in terms of quality of the resulting clusters compared to the use of ad-hoc distance metrics. As number of clusters is determined automatically in BHC, we tested stability of results through leave-one out cross-validation. Analysis were implemented using the *bhc* function from the *Bioconductor* package in *R* (Savage R, 2019).

### ***Comparison between clusters***

To characterize the identified clusters, we investigated differences between clusters in face processing through t-tests. Next, we evaluated clustering performance by examining the association of cluster membership with clinical outcome variables at 36 months through t-tests. Specifically, we tested differences between clusters on symptoms, indexed by the ADOS Calibrated Severity Score (CSS) obtained from the raw total scores (CSS-Tot), Social Affect (CSS-SA) and Restricted and Repetitive Behaviors (CSS-RRB) domains; the ADI-R domain scores for the Social (ADI-Soc), Communication (ADI-Comm) and Repetitive Behaviours and Interests domains (ADI-RBI); and the SCQ total score (SCQ-Tot). Furthermore, we tested differences in developmental level, indexed by domain T-scores of MSEL: receptive language (RL), expressive language (EL), fine motor skills (FM) and visual receptive skills (VR). Finally, we tested differences

in adaptive functioning, indexed by domain standard scores of VABS: communication, socialization, motor skills and daily living skills. Comparisons were implemented in *R*. Holm-Bonferroni correction of significance threshold ( $\alpha$ ) was used to correct significance of t-tests for multiple comparisons *separately* for the comparison between clusters on ERPs and on clinical outcome variables.

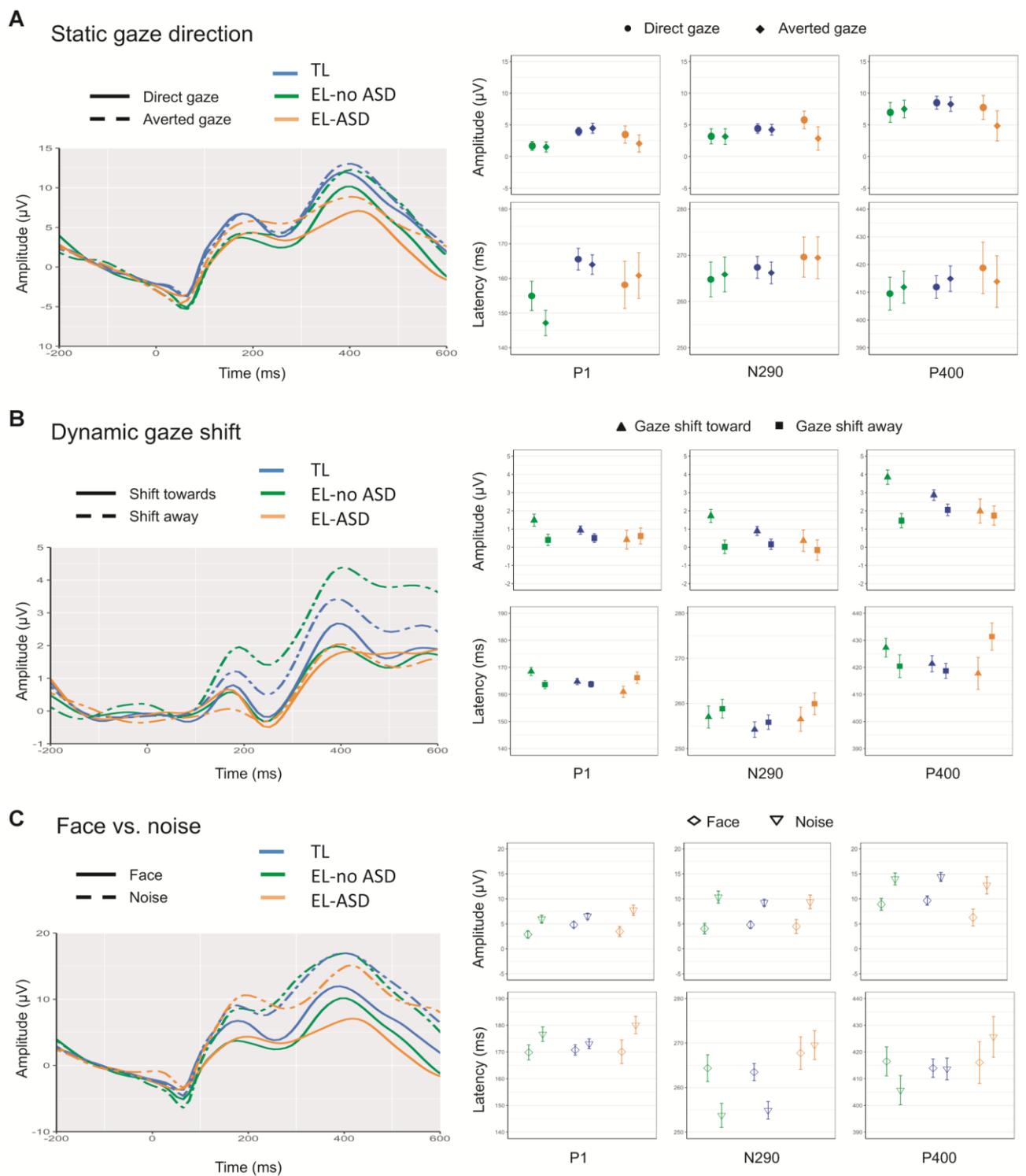
### ***Comparison between clusters and EL-no ASD group***

To understand if the diffuse pattern of face processing alterations identified as predictive of ASD by the supervised classification analysis was determined by subgroups of EL infants developing ASD having different face processing alteration profiles relative to the EL-no ASD group, we compared each cluster to the EL-no ASD group on ERP responses at 8 months. Comparisons were implemented in *R*, with Holm-Bonferroni correction for multiple comparisons.

## **Results**

### **Group-level differences**

We report group-level differences only when statistically significant (*Table S5*).

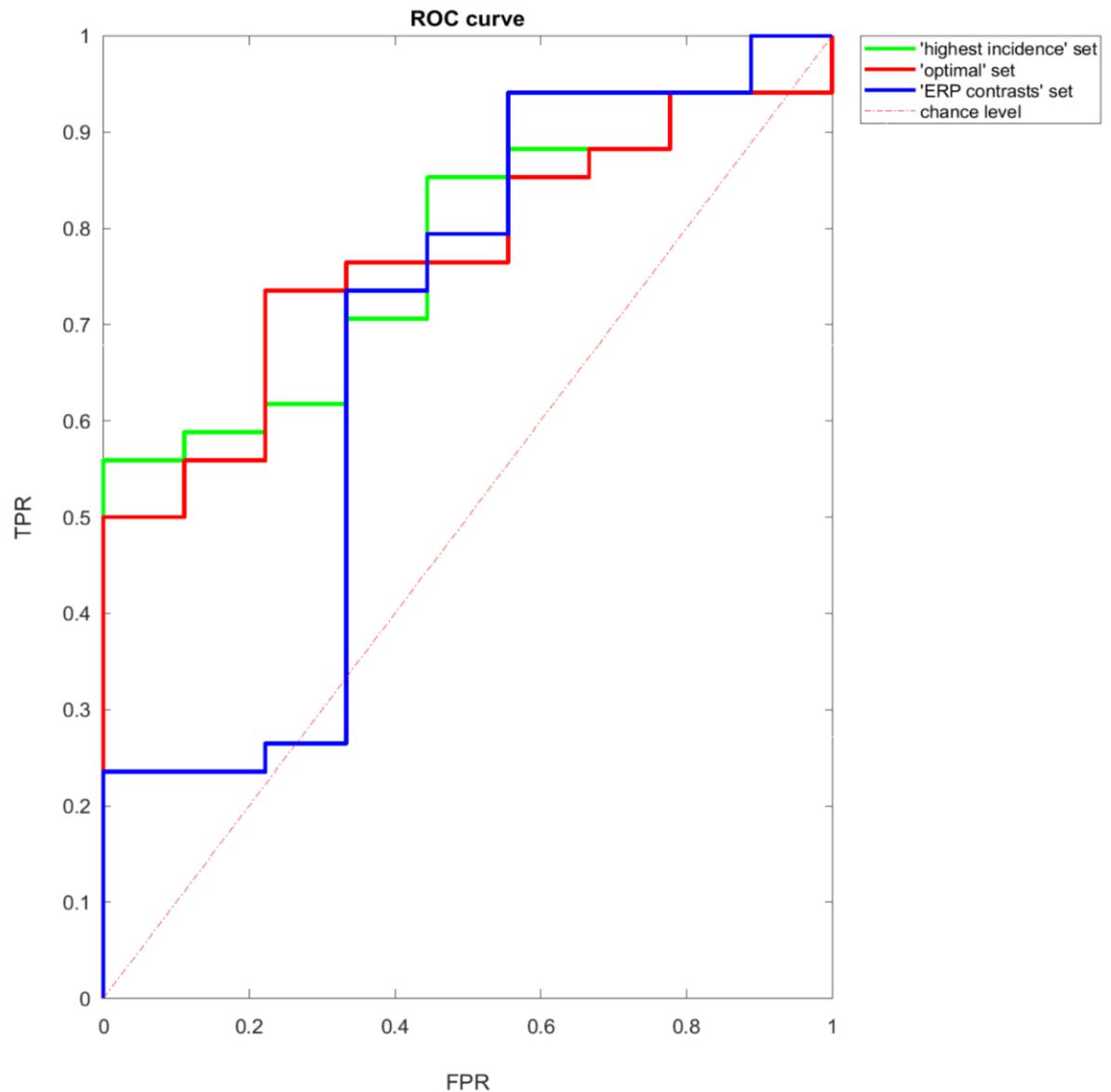


**Figure 2: Grand average event-related potentials.** Grand average ERPs across contrasts and groups over task-sensitive occipito-temporal channels (left) and means and standard errors of amplitude and latency of the three face-sensitive ERP components in each group (right).

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 ( $F(2,212)=1.48$ ,  $p=.226$ ), but 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 ( $ps>.19$ ).

### **Individual-level prediction**

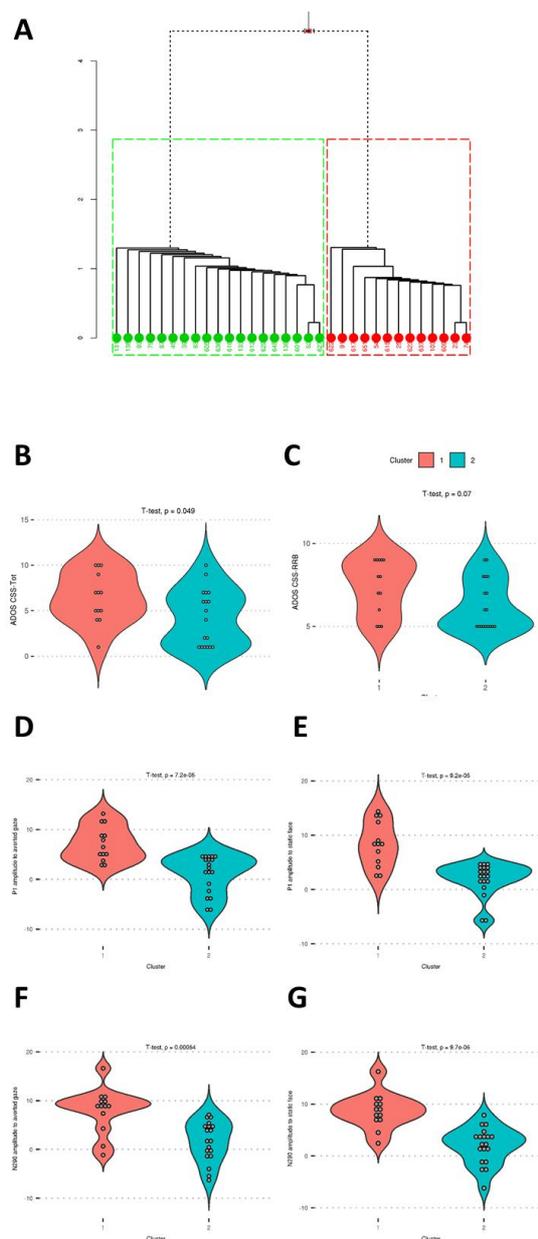


**Figure 3: Classification performance.** Receiver Operating Characteristic (ROC) curve for classifiers using different set of features to classify EL-ASD among EL siblings. Random predictors result in bisecting lines as ROC curves (red dashed line), while deviations in the upper hemifield indicate an 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.

Combined sets of brain responses to different conditions of face stimuli provided the best predictive accuracy for ASD at an individual level. These combined sets provided, in fact, a predictive accuracy significantly different from chance level ('ERP contrasts', 'Optimal' and 'Highest Incidence' classifiers; see *Table S6/Figure 3*), while classification performance for condition-specific subsets of features showed largely poor predictive power (see 'Noise-Face', 'Static Face', 'Gaze Shift', 'Frequent Noise-Face' and 'Frequent Static Face' classifiers in *Table S6*). Specifically, the 'Optimal' set provided the best predictive algorithm for ASD outcome at 36 months, 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 (see *Table S6*). This suggests that automatic feature selection through the genetic algorithm outperforms top-down selection to indicate that combined sets of responses to different face stimuli improved predictive accuracy compared to condition-specific responses. Furthermore, 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 (see *Table S6*). This suggests that, while the 'Highest Incidence' set indicates the most relevant features based on frequency of selection, there is additional useful information in the 'Optimal' set for prediction of ASD at an individual level (see *Table 1* for details on the features selected). Of note, sex was never selected by the genetic algorithm among sets providing the highest AUC, and age was selected less than 6% of the times, suggesting that the predictive model did not depend on the confounding variables. Details on classification performance from the different classifiers can be found in *Table S5*.

### **Stratification into subgroups of ASD**



**Figure 4. Stratification into subgroups of ASD based on face processing in infancy.** Two clusters were identified among infants who meet diagnostic criteria for ASD at 36 months (A). Differences between clusters are shown in ASD symptoms severity (B), restricted repetitive behaviours (C), and ERP responses to face stimuli (D-to-G). Significance of the t-test for between-group differences is shown on each plot, not adjusted for multiple comparisons.

*Abbreviations:* ERP = event-related potentials; ASD = autism spectrum disorder; ADOS = autism diagnostic observation schedule; CSS-Tot = total composite score from the ADOS; CSS-RRB = composite score for the restricted repetitive behavior subscale of the ADOS.

The cluster analysis on average ERP responses to the different stimuli conditions (see *Figure 1.C*) identified two stable clusters among infants with later ASD diagnosis (*Figure 4*): cluster one consisted in 13 EL infants developing ASD (40.6% of the EL-ASD sample); cluster two consisted in 19 EL infants developing ASD (59.4%). Descriptive statistics for the different clusters are shown in *Table S7*.

### ***Comparison between clusters***

Compared to cluster two, cluster one was characterized by higher P1 amplitude ( $t(21)=4.83$ ,  $p<0.001$ ) and N290 amplitude to static face, regardless of gaze direction ( $t(27)=5.45$ ,  $p<0.001$ ), and to faces with averted gaze (respectively  $t(27)=4.67$  and  $t(24)=3.99$ ,  $ps<0.001$ ). Furthermore, infants from cluster one had significantly higher levels of ASD symptoms (ADOS CSS-Tot:  $t(27)=2.06$ ,  $p=0.049$ ), and marginally higher levels of restricted repetitive behaviours (ADOS CSS-RRB:  $p=0.07$ ). However, differences were not significant after correcting for multiple comparisons (*Figure 4*).

### ***Comparison between clusters and EL-no ASD group***

Differences on face processing between ASD subgroups and EL infants who did not develop ASD were not significant after Holm-Bonferroni correction for multiple comparisons. Nevertheless, we found trend-level differences between groups indicating different alterations in neural processing of faces in infants from different ASD subgroups compared to siblings who did not develop ASD. In particular, compared to the EL-no ASD group, cluster one had reduced P400 amplitude in response to faces with direct gaze (unadjusted  $p=0.036$ ), faces with averted gaze (unadjusted  $p=0.038$ ), and irrespective of gaze direction (unadjusted  $p=0.049$ ). Cluster two had, instead, reduced P1 amplitude to faces, irrespective of gaze direction (unadjusted  $p=0.032$ ), reduced P400 amplitude to faces with direct gaze (unadjusted  $p=0.036$ ), and longer N290 latency to faces with direct gaze (unadjusted  $p=0.026$ ) and to visual noise (unadjusted  $p=0.006$ ).

## **Discussion**

We present a set of analytic approaches to investigate *whether* and *how* a particular RDoC-defined domain is linked to a later diagnostic outcome (here ASD) 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 alterations in both early sensory and later higher-level stages of neural processing of social (face/gaze) and scrambled face stimuli that differentiate infants with later ASD from those without later ASD across both group-based comparison, individual-level prediction and stratification (*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 emergence of later categorical diagnosis of ASD.

*<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 face processing is altered in early ASD. Specifically, the socially relevant processes of detecting a face and a shift in gaze are altered across a long time-course of information processing from the shortest latency components, and across multiple stimuli conditions (*Table 1*). N290 latency to visual noise versus faces emerged as a relevant feature across all three levels of analysis (*Table 1*), while P400 amplitude and latency to dynamic gaze shifts appeared to be informative features for prediction of ASD outcome at the level of groups and of individuals. The group of infants with later ASD (EL-ASD) showed a longer P400 latency and smaller P400 amplitude to gaze shifting towards versus away from the viewer and diminished effects of longer N290 latency to faces compared to visual noise compared to TL and EL-no-ASD groups. The same measures were part of a broader pattern predicting individual ASD diagnosis among siblings at elevated likelihood with approximately 77% accuracy. A broad pattern of alterations across the time-course of neural processing contributed to prediction at the individual level, including both shorter and longer latency components in response to different face conditions. Neural responses to dynamic gaze shifts alone, which appeared to be the best candidate precursors of ASD at a group level, did not provide sufficient predictive value on their own at an individual level (*Table S6*). Despite significant differences between groups, there is a significant overlap in individual variation, suggesting that alterations in neural processing of dynamic gaze are neither necessary nor sufficient conditions for ASD development. Taken together, these findings support hypothesis (a) on

diversity of altered face processing, which then likely reflects consistent differences in different specific aspects of face processing that are masked in different studies by their theory-driven focus on one or two components.

Our findings converge with a range of previous work in indicating the importance of early social processing and attention in the emergence of ASD. For example, a previous study of a subset of this cohort has showed that reduced P400 differentiation of dynamic gaze predicts ASD outcome in middle childhood, both in terms of stable diagnosis from 3 to 7 years and a ‘late’ diagnosis made after 3 years of age (Bedford et al., 2017). Furthermore, 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 (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.

### **Stratification within ASD**

The supervised classification analysis identified a diffuse pattern of neural responses to different face/noise stimuli at 8 months as predictive of individual ASD outcome in toddlerhood. To determine whether this pattern was determined by each infant with later ASD diagnosis having a different alteration in face processing at 8 months or by each infant having the same diffuse pattern of alterations, we performed a clustering analysis on ERP responses to the different conditions at 8 months among EL siblings developing ASD in toddlerhood (*Figure 1.C*). We identified two different subgroups in ASD differing in intensity of response in early sensory (P1) and later high-order stages (N290) of processing faces with averted gaze, and static faces in general. Of note, differences between clusters in clinical outcome variables

were not retained when correcting for multiple comparisons. While on the one hand it can be explained by a lack of statistical power, which 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. It is also possible that the different clusters were defined by different patterns of interactions between ERP features, which was uncovered by the data-driven clustering approach, but could not be picked up by traditional group comparisons.

When compared to the broader group of infants who did not develop ASD, Cluster 1 showed reduced engagement to static faces (lower P400 amplitude), irrespective to gaze direction. This was associated with higher levels of ASD symptom severity, of restricted and repetitive behaviours, and of social-communication symptoms than their non-ASD peers, and also reduced motor, communication and daily living functioning skills (*Analysis S3*). Cluster 2, instead, showed reduced perceptual capture to faces (smaller P1 amplitude), in particular to faces with averted gaze; reduced engagement to faces with direct gaze (lower P400 amplitude); and slower processing of faces with direct gaze and visual noise (longer N290 latency) compared to their non-ASD peers. This was associated with higher levels of ASD symptom severity, of restricted and repetitive behaviours, and of social-communication symptoms (*Analysis S3*). Comparisons between ASD clusters and the EL-no ASD group on face processing suggest, therefore, that there may be some divergence within the ASD group. Thus, the diffuse pattern of features identified by the genetic algorithm as predictive of individual ASD outcome likely emerges from different subgroups in ASD being predicted by different features. Specifically, Cluster 1 appeared to be predicted by more specialized alterations in later stages of neural processing of static faces, while the other siblings developing ASD were predicted by a pattern of alterations in face processing that is more diffuse in terms of temporal stages, processing features (speed and/or intensity), and stimuli. Future work should integrate longitudinal measures to further refine our understanding of subgroups of face processing within ASD over development. Furthermore, future work might employ a different data-driven approach to parse individual heterogeneity of ASD 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).

### **Individual-level prediction: towards clinical utility?**

The combination of ERP measures that best predicted ASD outcome at the individual level included responses to face/gaze and visual noise (*Table 1*), extending recent work using nonlinear features of EEG signals to support widespread dysfunction not specific to social processing (Bosl, Tager-Flusberg, & Nelson, 2018). 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. We significantly reduced the age of detection compared to previous classification studies on behavioural measures (Bussu et al., 2018; Chawarska et al., 2014) and ERP parameters (Eldridge, Lane, Belkin, & Dennis, 2014). Individual-level prediction of ASD in the first year of life has shown more than 94% accuracy using functional and structural magnetic resonance imaging (Emerson et al., 2017; Hazlett et al., 2017). However, our results extend these findings by using ERPs to predict a more stable ASD diagnosis at 36 months in a larger EL sample. Furthermore, ERPs represent a more cost-effective, mobile, and infant-friendly neuroimaging technology, providing potential utility for inclusion as proxy outcome markers for intervention trials. To test this, 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), together with tests of integration with other risk markers (e.g. genetic factors, brain imaging, parent-child interaction and behavioural measures) to improve individual prediction of ASD outcome. 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.

### **Conclusion**

Our three-pronged approach represents the first attempt to investigate robustness and generalizability of findings on a specific domain (social cognition/face processing) across different levels of analysis, from group-based comparisons to individual-level prediction of outcome and stratification. We focused on a domain in detail to show the analytical approaches necessary to investigate *whether* and *how* a specific domain contributes to the clinical category of ASD over development. Our findings show a diffuse pattern of atypicalities across the timecourse of neural processing in response to face and noise stimuli, suggesting that reported alterations likely reflect a more diffuse dysfunction that broadly affects neural processing of faces. These pattern of diffuse alterations was shared in good proportion across infants developing ASD; however, our clustering approach provided emerging evidence for subgroups of infants developing ASD that share more consistency in face processing. This adds to the literature illustrating early structural atypicalities (Hazlett et al., 2017) by showing early diffuse functional atypicalities, in line with the idea of ASD as a syndrome emerging from diffuse and interacting liabilities (Constantino, 2018). Using traditional and novel analytic approaches, we examined in detail the social cognition domain in infancy and investigated how this relates to the clinical ASD category in toddlerhood. Doing this, we addressed the emerging need in developmental neuroscience to incorporate RDoC constructs to improve our understanding of the mechanisms underlying ASD development.

## 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
- 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. (2018). Prediction of Autism at 3 Years from Behavioural and Developmental Measures in High-Risk Infants: A Longitudinal Cross-Domain Classifier Analysis. *J Autism Dev Disord*. doi:10.1007/s10803-018-3509-x
- 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
- 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.
- Chawarska, K., Shic, F., Macari, S., Campbell, D. J., Brian, J., Landa, R., . . . Bryson, S. (2014). 18-month predictors of later outcomes in younger siblings of children with autism spectrum disorder: a baby siblings research consortium study. *J Am Acad Child Adolesc Psychiatry*, 53(12), 1317-1327.e1311.
- 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
- 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.
- Eldridge, J., Lane, A. E., Belkin, M., & Dennis, S. (2014). Robust features for the automatic identification of autism spectrum disorder in children. *J Neurodev Disord*, 6(1), 12. doi:10.1186/1866-1955-6-12
- 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.
- Emerson, R. W., Adams, C., Nishino, T., Hazlett, H. C., Wolff, J. J., Zwaigenbaum, L., . . . Piven, J. (2017). Functional neuroimaging of high-risk 6-month-old infants predicts a diagnosis of autism at 24 months of age. *Sci Transl Med*, 9(393). doi:10.1126/scitranslmed.aag2882
- 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.
- 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., 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, C. E., Darkins R, Xu Y. (2019). BHC: Bayesian Hierarchical Clustering (Version R package version 1.38.0.).

- 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.
- Szatmari, P., Chawarska, K., Dawson, G., Georgiades, S., Landa, R., Lord, C., . . . Halladay, A. (2016). Prospective longitudinal studies of infant siblings of children with autism: lessons learned and future directions. *Journal of the American Academy of Child & Adolescent Psychiatry*, *55*(3), 179-187.
- 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

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

<u>Static Gaze Condition</u>			<u>Dynamic Gaze Shift Condition</u>			<u>Face vs. Noise Condition</u>			<u>Analysis</u>
Direct Gaze	Averted Gaze	Direct vs. Averted	Shifts Towards	Shift Away	Shift Towards vs. Shift Away	Static Face	Visual Noise	Face vs. Noise	
					P1 latency P400 amplitude P400 latency			<b>N290 latency</b>	Group-Based Comparison
P1 amplitude#	P1 amplitude	P1	N290 amplitude¥	<b>P400</b>	P1 amplitude#	P1 amplitude#	P1 amplitude#	P1 amplitude	Supervised Classification
N290 amplitude#	N290 amplitude	amplitude#	<b>P400 amplitude#</b>	<b>amplitude#</b>	<b>P400 latency#</b>	N290 amplitude	P400 latency#	P1 latency#	
P400 amplitude	P400 amplitude¥	P400				P400 amplitude#		<b>N290 latency</b> P400 amplitude#	
P400 amplitude†‡	P1 amplitude*					P1 amplitude*‡	<b>N290 latency‡</b>		Stratification into Subgroups
N290 latency‡	N290 amplitude*					N290 amplitude*			
	P400 amplitude†					P400 amplitude†			

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 individual prediction of EL-ASD vs. EL-no ASD and/or significant differences between subgroups of the EL-ASD group or between subgroups and the EL-no ASD group. Shaded

---

grey indicates no significant results for the specific stimuli/conditions in a specific analysis. The measures appearing as significant at different levels of analysis are highlighted in bold (2 levels) and bold red (3 levels).

# Features not in the 'highest incidence' set

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

\* Significant difference between clusters among infants in the EL-ASD group

† Significant difference between cluster one and EL-no ASD group

‡ Significant difference between cluster two and EL-no ASD group

