Title: Detection of idiosyncratic gaze fingerprint signatures in humans

Authors: Sarah K. Crockford[1], Eleonora Satta[1], Ines Severino[1], Donatella Fiacchino[1], Andrea Vitale[1], Natasha Bertelsen[1,2], Elena Maria Busuoli[1,2], Veronica Mandelli[1,2], & Michael V. Lombardo[1]

Affiliations:

1 Laboratory for Autism and Neurodevelopmental Disorders, Center for Neuroscience and Cognitive Systems @UniTn, Istituto Italiano di Tecnologia, Rovereto, Italy

2 Center for Mind/Brain Sciences, University of Trento, Rovereto, Italy

Abstract

Variation in human gaze can be explained by a variety of factors. Within an individual, gaze patterns can be quite reliable and can be explained in part by heritable common genetic mechanisms. If individual gaze patterns are strongly idiosyncratic, could they enable highly accurate detection of individuals via a 'gaze fingerprint'? And are gaze fingerprints linked to variation in phenotypes such as autistic traits, which are underpinned by heritable common genetic mechanisms and cause atypical and idiosyncratic gaze to manifest in early development? To answer these questions we utilized a stimulus-rich design where participants viewed 700 stimuli of complex natural scenes, repeated in two sessions separated by ~1-2 weeks. Across independent discovery (n=105) and replication (n=46) datasets, we find that gaze fingerprinting occurs at high rates (50-65%) relative to chance (1-2%) when gaze similarity is averaged across stimuli. However, such averaging procedures hide a unique non-random individuating code represented as a multivariate 'gaze fingerprint barcode' pattern across all 700 stimuli. Within such barcodes, around 94-95% of all individuals had fingerprintable stimuli counts that were significantly higher than situations where identity is randomly permuted. Data-driven clustering of such barcodes results in each individual as its own unique cluster. A small subtype of individuals can be replicably identified, whose fingerprintable stimuli are highly enriched for human, social, face, and emotion features. Finally, increased 'gaze fingerprintability' is associated with decreased levels of autistic traits. Overall, this work showcases that gaze fingerprints are important precision biomarkers and may have translational relevance to conditions like autism.

Human gaze patterns provide one of the most important signals regarding what an individual finds most salient about their environment. Many factors outside of the individual have been theorized to explain variance in gaze patterns, such as context, task, novelty, and characteristics of the stimuli themselves[1–7]. However, because gaze patterns can also be quite stable and systematic[6–12], this suggests that there must also be a strong within-individual factor that explains variance in gaze patterns. Recent twin studies have revealed heritable common genetic variation to be one such powerful within-individual mechanism constraining gaze patterns[13,14]. These insights suggest the possibility that gaze patterns may be an important individuating signature or 'fingerprint' of high neurobiological and developmental significance. With the advent and uptake of technology that could routinely collect eye gaze data at scale in very large quantities, the ability to identify individuals by what they look at merged with other personal data could have very large impact and raise issues regarding privacy[15–18]. Gaze fingerprint biomarkers may also be of high translational relevance for precision medicine approaches applied to autism. Atypical gaze patterns are evident in autism both in early development[13,19] and at older ages[20]. Such atypical social visual engagement in autism is theorized to be a manifestation of genetically constrained, yet atypical, biological niche construction[21–25]. Gaze fingerprinting biomarkers could provide an individualized metric for assessing change as a function of manipulations to an individual's environmental niche (e.g., early intervention).

In this study we investigate the phenomena of 'gaze fingerprinting' – that is, identification of an individual based on repeatable gaze patterns that are unique to the individual. While prior work has shown that individual gaze patterns can be stable and systematic[6–11], such work does not address whether the 'target' individual can be accurately identified when compared to gaze patterns of many other 'distractor' individuals. As far as we know, only two prior studies have attempted to address this question[12,14]. These studies suggest that some, though not all, individuals may be gaze fingerprintable and that such gaze fingerprints are potentially driven in part by genetic similarity. For example, Kennedy and colleagues[14] examined how well twin pairs could be identified from stimuli-averaged gaze similarity computed from gaze heatmaps of about 80 static pictures of complex scenes that varied by social and non-social content. Around 29% of monozygotic twins could be identified from the gaze of the other twin. In contrast, dizygotic twins can be identified at much lower, though statistically significant, rate of 7%, compared to non-twin pairs and chance accuracy of 1%[14]. Since this study examined identification of twins, it is conceptually related to 'gaze fingerprinting' by demonstrating that degree of genetic similarity between-individuals is related to degree of gaze similarity. However, because it does not directly assess identifiability of the same target individual, it is unclear how well gaze fingerprinting could work in the population.

In subsequent work, Keles and colleagues[12] showed autistic and non-autistic participants two different movie clips and measured summary gaze features such as the duration of looking at specific objects (e.g., faces, body parts, non-social content), as well how deviant their gaze heatmaps were from normative reference gaze heatmaps. Distances were computed between these multivariate gaze vectors for each individual and movie and identity was predicted based on the gaze vector with the minimal pairwise distance to the target individual. Target individuals were correctly identified at higher rates than chance (~2-6%) - ~30% accuracy in non-autistic and ~35% accuracy in autistic individuals. However, further analysis showed that identification accuracy drops when duration of the epoch sampled in the movie is reduced. This suggests that identification accuracy may hinge critically on the amount

of data sampled per individual. In summary, Keles et al., shows that gaze fingerprinting is potentially possible for a small (e.g., ~30-35%) but above chance percentage of individuals[12]. However, some caveats of this work are that the amount of gaze data sampled per individual matters. Second, the fingerprinting paradigm utilized viewing of different video stimuli rather than repeat viewing of the same videos. Thus, while intra-subject and inter-subject gaze similarity can be computed in this type of paradigm, it does not hold constant stimulus type and leaves open the possibility that gaze to the same stimulus may not necessarily be a stable individualized feature. Because the sampling of gaze was limited to two movie clips, such work may be limited in revealing what aspects of a particular stimulus (e.g., low-level visual and/or semantic features) may be most salient and important for gaze fingerprinting different types of individuals. Finally, the relatively similar fingerprint accuracy rates in autistic and non-autistic individuals alongside only slightly above chance accuracy (~61%) in correctly classifying diagnosis based on gaze patterns, may suggest that gaze fingerprint patterns may be less sensitive to the categorical difference signified by diagnosis. Rather, if gaze fingerprinting is linked to heritable common genetic architecture[13,14] and individualized phenotypic variance relevant to autism[13,19,20], it may be better explained by continuous/dimensional factors sensitive to such heritable common genetic architecture such as autistic traits[26].

In this study we fill these important gaps in gaze fingerprinting. Here we measured gaze patterns in a relatively large sample of participants across discovery and replication sets (discovery n=105; replication n = 46) viewing a large richly-annotated stimulus set of 700 complex scenes[5], repeated in two sessions separated by about 1-2 weeks. The 'stimulus-rich' feature of this study allows for dense sampling of gaze patterns across a large range of complex scenes depicting a variety of semantic features that have been heavily studied regarding visual salience[5,6,20]. This dense sampling of gaze across such a feature-rich stimulus set allows for gaze fingerprinting based on stimuli-averaged similarity (as is done in Kennedy et al., [14]), but also at scale per each individual stimulus. Gaze fingerprinting at scale on individual stimuli also allows for 'barcoding' individuals by fingerprintable stimuli sets and semantic features describing those sets. Finally, because gaze similarity is linked to genetically sensitive phenotypes such as autism[11–14,20], we assessed whether gaze fingerprinting metrics are related to phenotypes underpinned by heritable common genetic mechanisms, such as autistic traits.

Materials and Methods

Discovery dataset

The discovery dataset (labeled henceforth as 'IIT') comprised individuals tested in an eye tracking study carried out within the IIT Laboratory of Autism and Neurodevelopmental Disorders (IIT-LAND). This study was approved by the local ethics committee (Comitato Etico per le sperimentazioni cliniche dell'Azienda Provinciale per Servizi Sanitari della Provincia Autonoma di Trento; APSS). All participants gave written informed consent and were reimbursed for their time participating in the study. Participants were primarily recruited through social media and convenience sampling within the University of Trento Center for Mind and Brain Sciences (CIMeC) department. A total of n=120 typically-developing adult participants (age range 18-50 years) with normal or corrected-to-normal vision took part in this study. All participants took part in two sessions, whereby the separation between the two sessions was on-average 7 days (SD = 1.46 days). Due to missing gaze data for some individuals on some stimuli (e.g., participant looked away from screen), n=13 individuals were excluded from analysis if they had more than 15% of the 700 stimuli missing

gaze data for either session 1 or session 2. A further n=2 participants were excluded due to delays between session 1 and session 2 that were over 20 days. The final sample utilized in all further analyses was n=105 (n=52 female, n=53 male). The sample had a mean age of 26.21 years (SD = 6.30) and was within the normal range on full-scale IQ (mean = 108.65, SD = 13.09, range = 75-138). While the main analyses on the IIT dataset use all individuals, in follow-up analyses (see Supplementary Figure 1) we split the IIT discovery dataset by sex in order to evaluate how fingerprint accuracy varies in smaller sample sizes (n=52 female, n=53 female) that are on-par with the size of the Gi replication dataset (e.g., n=46).

Replication dataset

As a replication dataset we re-analyzed anonymized publicly available test-retest eye tracking data reported from de Haas and colleagues[6] (https://osf.io/n5v7t/). Henceforth, this replication dataset is referred with the original acronym given by de Haas and colleagues of 'Gi'. The experimental design and stimuli for both the IIT discovery and Gi replication datasets were identical in that both utilized the OSIE dataset of 700 stimuli and participants were instructed to freely view the stimuli over a 3 second period in both test and retest sessions. The primary difference between the IIT discovery and Gi replication datasets was the duration of time in between test and retest sessions. While the IIT dataset had a delay of 7 days on-average (SD = 1.46), the delay in the Gi replication dataset a little more than double this duration (mean = 16 days, SD = 7 days). The sample size of the Gi dataset was n=48. However, n=2 individuals were excluded from analysis if they had more than 15% of the 700 stimuli missing gaze data for either session 1 or session 2. The sample characteristics of the Gi replication dataset indicate that the sample size was smaller (n=46) than the IIT discovery dataset (n=105). Furthermore, the Gi dataset was primarily female (e.g., n=11 male), whereas the sexes were relatively balanced in the IIT dataset.

Eye Tracking Task and Stimuli

Participants were asked to freely watch 700 stimuli of complex natural scenes taken from the Object and Semantic Images and Eye-tracking (OSIE) stimulus set[5]. Stimuli in the OSIE set are annotated both for their low pixel- and object-level features as well as higher-level semantic features[5]. Semantic features represented categories such as face, emotion, touched, gazed, motion, sound, smell, taste, touch, text, watchability, and operability. Further annotation beyond these twelve categories was done to label stimuli by presence or absence of 'humans' or 'animals' as well as a categorization of whether the stimuli had 'social' (i.e. presence of human faces) or 'non-social' (i.e. absence of humans or animals) content. While all semantic features can be defined and analyzed independently, they co-occur within the OSIE stimuli in a non-independent manner. Therefore, to observe the emergent structure of co-occurrence of multiple semantic features as clusters, we utilized two-way hierarchical clustering with Jaccard distance matrices and ward.D2 as the clustering agglomeration method to reveal data-driven clusters across all 700 stimuli and across all semantic features (Fig 1A). This analysis reveals that there are 2 primary clusters amongst the 700 stimuli that largely define 'social' versus 'nonsocial' stimuli (Fig 1A). Semantic features can be separated into 6 clusters (Fig 1A).

Procedure

Eye tracking data in the IIT discovery set was collected on a Tobii Pro Spectrum machine with a sampling rate of 150 Hz. Participants sat in front of the eye tracking monitor

at a distance of approximately 60 cm with their chin on a chinrest. Each participant completed 2 eye tracking sessions separated by an average of 7 days. Within each session, participants viewed all 700 stimuli, broken into 7 blocks of 100 stimuli each. Each stimulus was shown for a duration of 3 seconds. Participants were given the instructions to watch the stimuli in a natural fashion. Stimulus presentation was randomized such that within each session participants viewed the stimuli in a random order. Between the presentation of each stimulus was a fixation cross in the center of the screen with a jittered duration between 500 to 1000 milliseconds. At the end of each block participants were given the opportunity to rest before starting the next block. The entire eye tracking session with all 7 blocks lasted approximately 60 minutes. Stimulus presentation was implemented within MATLAB 2019b using Psychtoolbox 3.0.16. Calibration was run using 5 calibration points in the center and at each corner and were presented in a fixed order. We accepted an average calibration accuracy across the 5 points and across the two eyes less than 2 degrees of visual angle.

Behavioral Measures

At the end of the first session in the IIT discovery dataset, participants completed the Wechsler Abbreviated Scale of Intelligence (WASI) in order to characterize full-scale IQ. At the end of the second session, participants were asked to complete two self-report questionnaires measuring autistic traits, translated into Italian - the Autism Quotient[27] and the adult self-report form of the Social Responsivity Scale, Second Edition (SRS-2)[28]. A small subset of individuals (n=9) fell above clinically-significant cut-offs of 26 on the AQ or a 2 SD cutoff from typically-developing norms of the Italian translation of the SRS-2 adult self-report (i.e. total raw score > 80.81). These individuals are annotated in figures showing autistic traits in order to visually mark individuals with or without pronounced elevation in autistic traits.

Gaze fingerprinting analysis

Fixations were defined using velocity-based classification algorithm whereby eye movements are identified based on a threshold when the velocity of directional shifts of the eye exceed 30 degrees of visual angle per second. Data below this velocity-based threshold are considered fixations. Fixations with a duration below 100 milliseconds were excluded. The x and y coordinates within a time-epoch defined as a fixation are then averaged to get the fixation coordinates. The duration of each fixation is the time elapsed between onset of the first gaze data sample and offset of the last data sample included in the fixation epoch. Fixation data coordinates were used to compute gaze heatmaps for each stimulus and participant. Gaze heatmaps were constructed by taking fixation points and smoothing them by a fovea-sized (e.g., 24 pixels on a 800 x 600 image) Gaussian kernel and then normalizing the final heatmap by the maximum value[5]. Heatmaps were then utilized to compute intra-subject and inter-subject gaze similarity. Gaze similarity is generally defined as the Pearson correlation between two vectorized gaze heatmaps[14]. Intra-subject gaze similarity corresponds to the similarity value between session 1 and 2's heatmaps for the target individual. Inter-subject gaze similarity for a particular stimulus is the between-subject similarity when the target individual's session 1 heatmap is compared to a distractor individual's session 2 heatmap.

Gaze fingerprinting analysis utilizes both intra-subject and inter-subject gaze similarity values. A symmetric gaze similarity matrix of size [n_subjects, n_subjects] is first computed whereby along the diagonal of this matrix are the intra-subject gaze similarity values. The off-diagonal values of this matrix represent all other pairwise inter-subject similarity values when the target individual's session 1 heatmap is compared to the distractor individual's session 2

heatmap. Once this matrix has been computed, we implement a 'find-the-biggest' (i.e. argmax) algorithm[14,29,30] whereby we loop through the rows of the gaze similarity matrix and identify the subject index corresponding to the maximum gaze similarity value. A 'hit' is defined as a situation when the max gaze similarity value corresponds to the target individual in question. In contrast, a 'miss' occurs when the max gaze similarity value corresponds to some other distractor individual. Fingerprint accuracy can then be computed as the number of hits divided by the total number of participants. Null distributions for fingerprint accuracy are also computed by first permuting subject identifiers and then re-running the fingerprint analysis pipeline 1000 times. From these null distributions, p-values can be computed by counting the number of times fingerprint accuracy was as large or larger than the real fingerprint accuracy, and then dividing that number by 1001.

Gaze fingerprinting analysis, as defined above, was applied at scale to every stimulus. Additionally, we also applied gaze fingerprinting analysis to averaged similarity values across all or subsets of the stimuli set grouped by semantic features. To mimic the gaze fingerprinting analysis procedure of Kennedy et al.,[14], we first computed median gaze similarity by computing the median over all 700 stimuli. The median was chosen to better reflect central tendency of the distribution than the mean, particularly in situations where the distribution is skewed. When distributions are not heavily skewed, the mean and median should be relatively similar. With a 3-dimensional gaze similarity matrix of size [n_subjects, n_subjects, n_stimuli], we compute the median over the 3rd dimension of the 700 stimuli to get a median gaze similarity matrix of size [n_subjects, n_subjects]. The fingerprint analysis algorithm described above is then used to define 'hits' or 'misses' and compute accuracy. This same procedure of computing a median gaze similarity matrix was also redone on subsets of stimuli that correspond to each of the 12 semantic features labelled in the OSIE dataset[5]. To get 95% confidence intervals (CI) around accuracy when averaging all stimuli, we utilized bootstrapping with 10,000 resamples. This allowed us to compare accuracy rates within semantic feature categories to the lower bound bootstrap CI on accuracy computed when averaging across all stimuli.

The 'hit' or 'miss' nature of the gaze fingerprinting output is a binary output that may hide information regarding the continuous nature of 'fingerprintability' or degree of 'gaze uniqueness'. That is, the fingerprint output may be a 'miss' because the intra-subject similarity is not the maximum gaze similarity among all distractor individuals. However, in the instance where the intra-subject similarity is still ranked as very large (e.g., 2nd highest gaze similarity), this would be a situation whereby the individual's gaze is unique to some high degree, though not maximally unique when compared to all distractor individuals. Thus, to derive a continuous metric of degree of 'gaze uniqueness', we developed a score we call the gaze uniqueness index (GUI) that conveys in a continuous manner the degree of gaze uniqueness that would otherwise be lost in a binary fingerprint output of hits or misses. GUI is computed by dividing intra-subject similarity by average inter-subject similarity. The denominator in this equation of average inter-subject similarity is important for interpretational reasons, as it acts to scale the intra-subject similarity to what can be thought of as a 'norm' or 'baseline' level of similarity for a particular stimulus, represented by the average of all pairwise inter-subject similarities for the target individual. Given that GUI values are ratios of intra-subject/average inter-subject similarity, the null value is 1, indicating equal intra-subject and average inter-subject gaze similarity. Values significantly increased greater than 1 signify a global gaze pattern across all stimuli that is relatively unique to the individual. GUI values are computed at scale per all 700 stimuli and are also summarized as one number per participant by taking the median GUI value across all 700 stimuli. Hypothesis tests on GUI values per each subject are implemented with a one-sample t-test against a null value of 1. From these t-tests, p-values are computed and then

we use false discovery rate (FDR) multiple comparison correction to identify individuals that pass FDR $q<0.05$.

Gaze fingerprint barcoding

Given the stimulus-rich nature of the OSIE dataset (e.g., 700 stimuli), gaze fingerprinting at scale for each subject and stimulus allows for an [nsubs, nstimuli] binary output matrix of hits and misses. The rows of this matrix can be thought of as individualized barcodes that tell us which stimuli are fingerprintable for that individual. Given that there is no requirement for each individual's barcode to be similar or different, it is an open question as to whether such gaze fingerprinting barcodes are unique to individuals or are clustered into subsets of individuals with similar kinds of barcodes. To answer this question, we compute a Jaccard distance matrix to represent the dissimilarity between gaze fingerprinting barcodes and then utilize hierarchical clustering (ward.D2 agglomeration method) on this distance matrix to identify how many clusters are represented in the dataset. This analysis allows us to answer what the optimal number of clusters (k) is using the silhouette coefficient over a range of $k=2$ up to $k = n-1$ as the maximum number of clusters. In a situation where $k = n-1$, such a solution would indicate that each individual's gaze fingerprint barcode is unique and that clusters are individual subjects.

Gaze fingerprint barcodes can also be described by enrichment with semantic feature categories. To achieve these goals, we computed enrichment odds ratios and p-values in a manner identical to an analogous statistical procedure in bioinformatics known as gene set enrichment analysis. In this type of analysis, we examine the extent of overlap between two lists (e.g., fingerprintable stimuli and semantic feature stimuli) whereby effect size is indicated by an odds ratio. This enrichment odds ratio can be interpreted with respect to the null value of 1, whereby values increasingly higher than 1 indicate an over-representation of a specific semantic category within an individual's fingerprintable stimuli set. To compute odds ratios, we first compute a 2x2 contingency table, with cells marked A and B in row 1 and cells C and D in row 2. Counts in cell A reflect the number of fingerprintable stimuli given a semantic feature. Counts in cell B are the number of fingerprintable stimuli minus the number of fingerprintable stimuli given a semantic feature. Counts in cell C are the total number of stimuli within a semantic feature category minus the number of fingerprintable stimuli given a semantic feature. Finally, counts in cell D correspond to the total number of stimuli in the OSIE dataset minus the summed quantity in cells A, B, and C. Such a contingency table can be inserted into the *fisher.test* function in R and get as output an enrichment odds ratio. These enrichment tests were computed per each subject using the subset of fingerprintable for each subject and testing the overlap of that stimulus set with the stimuli within each of the 12 semantic categories. A resulting matrix of [nsubs, nfeatures] can be computed as output of this analysis and then assessed for subtypes via hierarchical clustering analysis, where optimal k is identified by majority vote amongst many clustering metrics computed in the NbClust package in R[31]. This analysis allows for discovery of subtypes characterized by different patterns of semantic feature enrichment. Each semantic feature was also tested with a one-sample t-test to test whether enrichment odds ratios differed from the null value of 1. P-values from these hypothesis tests were used to compute FDR and comparisons that survive FDR $q<0.05$ in both IIT discovery and Gi replication sets are identified.

Association analysis with autistic traits

Robust regression linear modeling was utilized as the modeling strategy to test for association between gaze fingerprinting metrics and autistic traits. Autistic traits were considered the dependent variable in the model. Autistic traits were initially measured via two different questionnaires, the AQ and SRS-2. These measures are highly correlated ($r = 0.77$, $p < 2.2e-16$), indicating that shared variance in both may best account for autistic traits as a latent variable. To isolate the shared component of variation in AQ and SRS-2, we utilized principal components analysis (PCA) and found that the first principal component accounts for 88% of the shared variance. This autistic traits PC1 was used as the dependent variable measuring autistic traits. Two gaze fingerprinting metrics were utilized as independent variables in the model – 1) fingerprintable stimuli count and 2) median gaze uniqueness index computed across all 700 stimuli. Rather than plugging both variables into the model as separate independent variables, we first checked to see if the two fingerprinting metrics were correlated and found that they were ($r = 0.80$, $p < 2.2e-16$). Thus, we instead used PCA to decompose fingerprinting metrics into two orthogonal variables, of which PC1 captured around 90% of the variance and loaded equally onto both metrics, while PC2 accounted for the remaining 10% of the variance. With PC1 and PC2 gaze fingerprinting metrics, these variables were utilized as independent variables in the linear model to test for associations with autistic traits. Finally, the other independent variable used as a covariate of no interest was sex. This was done because prior research has shown robust sex differences in autistic traits on both AQ and SRS-2[26,27,32]. Thus, our modeling strategy uses sex as a covariate before assessing the association of fingerprinting metrics with autistic traits. The linear model was a robust regression linear model and was utilized to guard against the influence of bivariate outliers and was implemented with the *lmrob* function in the *robustbase* library in R. The *lmrob* function was also utilized to draw best fit lines in scatterplots to visualize the results.

We also compared 3 different models predicting autistic traits PC1. All of the models first account for sex as a covariate. Model 1 utilizes gaze fingerprinting PC1. Model 2 utilizes the median intra-subject gaze similarity computed across all 700 stimuli. Finally, model 3 utilizes the median average inter-subject gaze similarity computed across all 700 stimuli. Models 2 and 3 can be thought of as 'null' models with respect to the model of importance for gaze fingerprinting (e.g., model 1). Since gaze fingerprinting metrics (i.e. number of fingerprintable stimuli, median GUI) are computed based on intra-subject and inter-subject similarity values, comparing model 1 to models 2 and 3 allow us to assess the importance of gaze fingerprinting metrics relative to models that predict autistic traits solely on the basis of components used within gaze fingerprinting – that is, intra-subject similarity (model 2) or inter-subject similarity (model 3). Percentage of variance explained (R^2) was computed for each model and compared to each other, with the better model being the model that explains more variance (e.g., higher R^2) in autistic traits PC1.

Data and code availability

All code and data for reproducing analysis can be found within our GitHub repository: https://github.com/IIT-LAND/gaze_fingerprint_osie.

Results

In our first analysis we tested whether gaze fingerprinting is possible, and if so, how good is it at accurately detecting individuals as a function of semantic features within stimuli. As a first approach to answering these questions, we averaged gaze similarity across stimuli.

Here we found fingerprint accuracy of 50.47% in the IIT discovery dataset and 65% in the Gi replication dataset (Fig 1B). Both accuracy rates are well above the theoretical (e.g., ~1-2%) and empirically-derived (via permuting subject identifier) chance accuracy rate ($p < 9.99e-4$). These accuracy rates are also well above the accuracy in identifying twin pairs reported by Kennedy et al.,[14] (i.e. 29%) or in movie stimuli reported by Keles et al.,[12] (i.e. 30%). In sex-stratified analysis on the IIT discovery dataset, sample sizes are comparable to the Gi dataset, and the accuracy rates fall within a similar range of approximately 48-64% (e.g., 48% in females, 64% in males) (see Supplementary Figure 1).

Split into different subsets of stimuli reflecting high-level semantic features, accuracy in all features are well above chance levels in both discovery and replication sets (IIT: 27-52%, all $p < 9.99e-4$; Gi: 34-63%, all $p < 9.99e-4$) (Fig 1B). Notably, accuracy levels for each semantic feature are within the 95% bootstrap confidence intervals computed when averaging across all stimuli in both IIT and Gi datasets. The exceptions to this are smell, sound, emotion, and animal features. These features are the most infrequently occurring stimuli (e.g., comprising <22% of all stimuli) (e.g., Fig 1A). Thus, relatively lower accuracy on these features is likely due to number of stimuli averaged within the semantic feature set. In summary, the relatively high accuracy rates when averaging across all stimuli can also be achieved by averaging across stimuli within semantic features, provided that the semantic feature category contains an adequate number of stimuli (e.g., >154 stimuli).

Given that accuracy tends to drop with number of stimuli present within a semantic feature category, it is likely that identification accuracy applied at scale to each stimulus would result in relatively low accuracy. However, would identification accuracy computed per each stimulus be higher than chance identification rates? Indeed, we find that identification accuracy at the level of individual stimuli is heavily reduced compared to accuracy on averaged gaze similarity (e.g., IIT: ~1.5-12%; Gi: ~3-24%) (Fig 1C). Average accuracy of stimuli grouped by semantic features also shows attenuated accuracy at the individual stimulus level (IIT: 3.9-4.9%; Gi: 5.8-6.5%) (Fig 1D). However, we can evaluate how many of the 700 stimuli show accuracy rates significantly above chance at FDR $q<0.05$. Here we find in the IIT discovery set that around 94.5% (662/700) of stimuli allow for above chance accuracy. This percentage was much lower for the Gi replication set (62.5%) (Fig 1C). These results generally suggest that gaze fingerprinting is possible and that averaging across a range of different stimuli can enhance detection rate. While most individual stimuli show statistically higher than chance levels of accuracy, accuracy for individual stimuli is far lower than averaging gaze similarity across stimuli. This suggests that gaze fingerprinting based on any one stimulus in isolation may not achieve high levels of accuracy across a population.

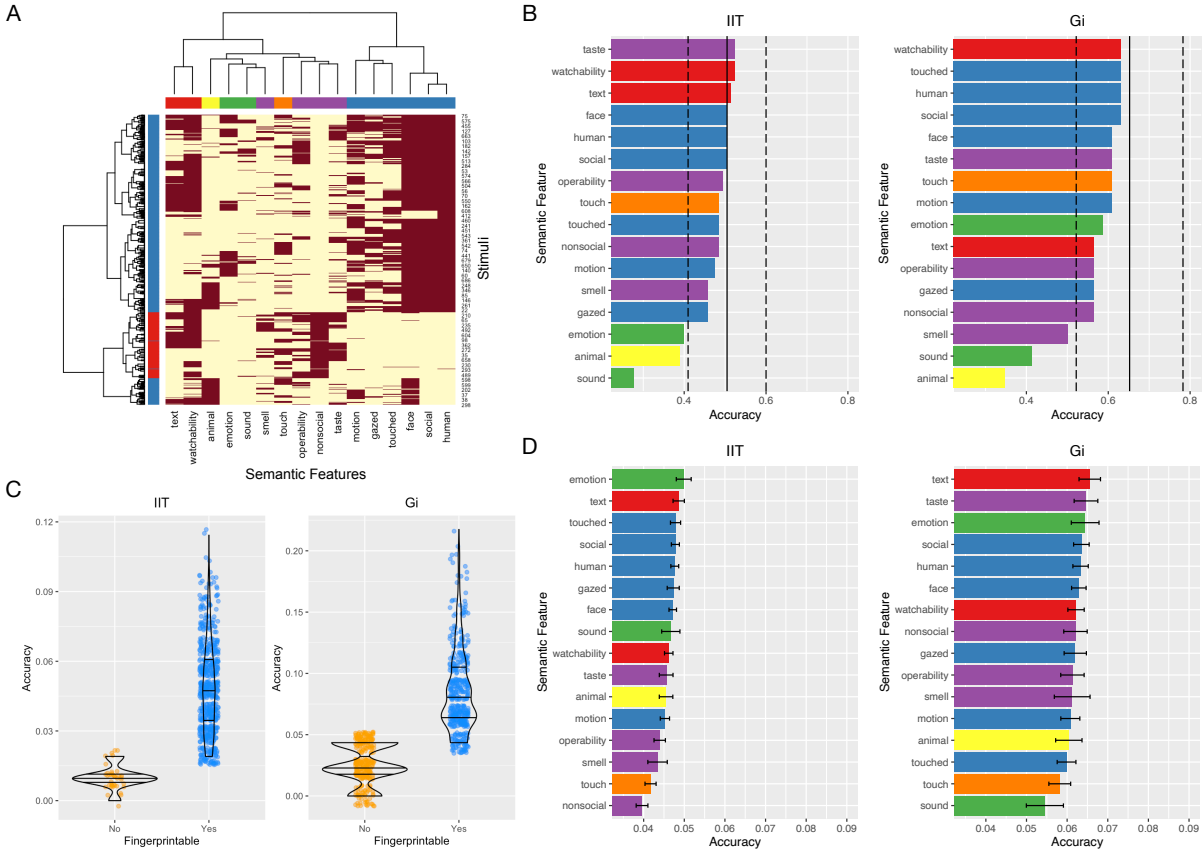


Fig 1: Gaze fingerprint accuracy on averaged gaze similarity across stimuli or for each individual stimulus. *Panel A is a heatmap representation of the co-occurrence of various semantic features (columns) across all 700 stimuli (rows) in the OSIE stimulus set. Dark red indicates presence of a stimulus in a particular semantic feature category, while light yellow indicates absence of a stimulus in a particular semantic feature category. Colors for clusters of semantic features (underneath the column dendrogram) are used in panels B and D. The vertical black line in panel B shows gaze fingerprint accuracy (x-axis) when average gaze similarity is computed across all 700 stimuli (dotted vertical black lines represent bootstrap 95% confidence intervals). Each bar shows gaze fingerprint accuracy when computed as average gaze similarity on subsets of stimuli grouped by semantic features shown on the y-axis. On the left are accuracy values in the IIT discovery set, while on the right are the accuracy values for the Gi replication set. Panel C shows accuracy values (y-axis) for each of the 700 stimuli. The x-axis denotes stimuli that are fingerprintable or not. This distinction is made based on whether the fingerprint accuracy values are higher than chance levels empirically estimated from permutation analysis where subject identifiers are randomized. The plot on the left shows the IIT discovery set, while the plot on the right shows the Gi replication set. Panel D shows accuracy values (IIT discovery, left; Gi replication, right) when fingerprint accuracy is evaluated separately for each individual stimulus. These plots group accuracy values by semantic feature categories. Bars depict the mean accuracy over all stimuli within a semantic feature category, while the error bars depict ±1 standard error of the mean.*

While the accuracy of around 50-65% for averaged gaze similarity across stimuli is already impressive and much higher than previous work on twins or with a small number of movie stimuli (e.g., ~29-30% accuracy[12,14]), it is possible that averaging across stimuli may hide important individuating information within the patterning of fingerprintability across stimuli. Given the unique stimulus-rich characteristic of our dataset, we next utilized a 'gaze fingerprint barcoding' approach to test whether the patterning of fingerprintability across

individual stimuli would allow for more individuation and fingerprint detectability in all participants. The gaze fingerprint barcoding approach uses the binary vector of hits or misses across all 700 stimuli as a potentially unique individuating gaze fingerprint pattern for an individual. As shown in Fig 2C, gaze fingerprint barcodes are relatively sparse due to each individual showing a small, but potentially unique pattern of fingerprintable stimuli across all 700 OSIE stimuli. To begin teasing apart whether this approach could yield further insights beyond those gleaned from fingerprinting based on averaged gaze similarity, we first ask whether the sparse information within barcodes, that is - the number of fingerprintable stimuli detected per individual, is larger than what we would expect at chance when subject identifier is randomly permuted. Here we find that number of fingerprintable stimuli varies widely between individuals, with some individuals showing very minimal levels of fingerprintable stimuli to others who can be identified based on 12-13% of all stimuli (Fig 2A). We replicably find that nearly all individuals (IIT: 99/105 or 94%; Gi: 44/46 or 95.6%) had fingerprintable stimuli counts that were significantly higher than chance when subject identifier is randomly permuted (Fig 2A). Thus, the information present within sparse gaze fingerprint barcodes is non-random and much higher than would be expected at chance when subject identifier is randomly permuted.

As an alternative to a binary vector representing the gaze fingerprint barcode, we introduce the concept of degree of 'gaze uniqueness', via a metric we call the 'gaze uniqueness index' (GUI). GUI can be computed on a per-stimulus basis and is defined as the ratio of intra-subject gaze similarity divided by average inter-subject gaze similarity. Increasing GUI values indicate increasing degree of gaze uniqueness, whereas the null value for GUI of 1 indicates equal intra-subject compared to average inter-subject gaze similarity. Hypothesis tests were done on the set of all 700 stimuli to test whether an individual has a GUI significantly above 1, indicating a relatively unique gaze profile on-average across all stimuli. As shown in Fig 2B, about 80% of IIT (84/105) and 78% (36/46) of Gi individuals show a GUI significantly greater than the null value of 1. Thus, GUI can also capture individuating information for a sizable majority of individuals but may not be as sensitive as the sparser signal present in gaze fingerprint barcodes.

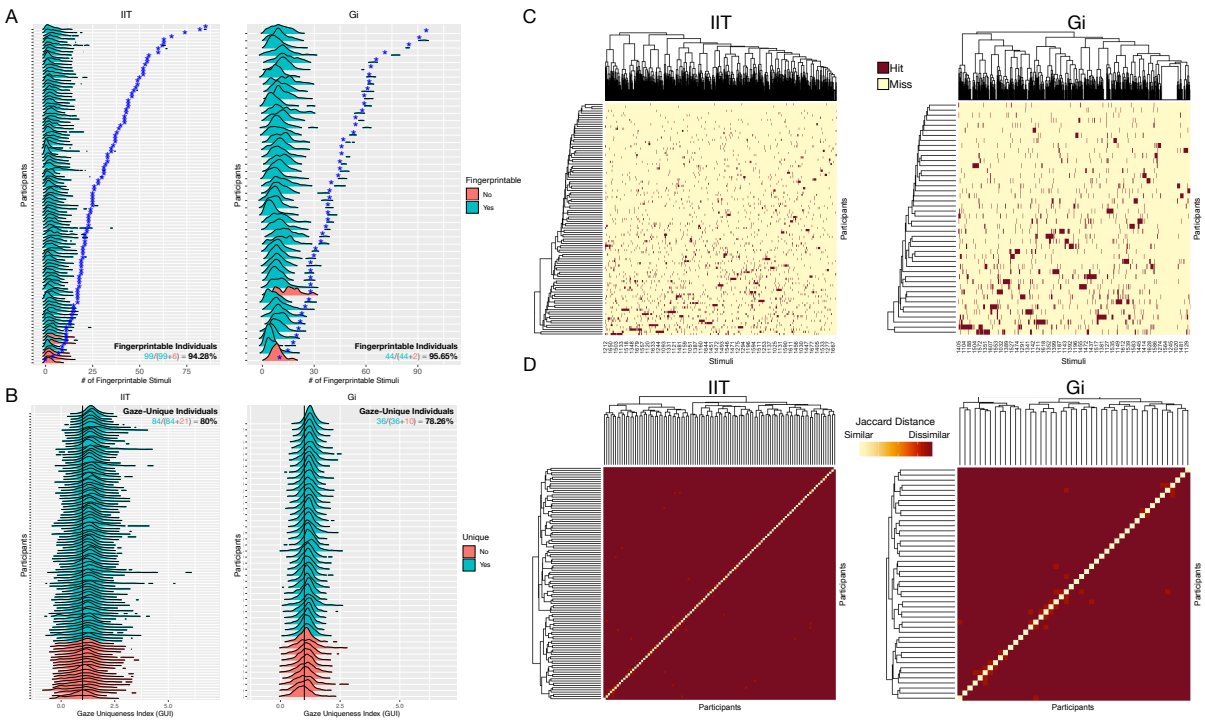


Fig 2: Gaze fingerprint barcoding and degree of gaze uniqueness. *Panel A shows counts of fingerprintable stimuli (x-axis) per each individual (y-axis) (IIT discovery, left; Gi replication, right). The actual count for fingerprintable stimuli per subject is denoted by the blue star. The density plots per each individual represent the null distribution of fingerprintable stimuli count over 1000 permutations where subject identifier is randomly permuted. The colors denote a distinction of whether the individual is fingerprintable (turquoise) or not (pink), with turquoise individuals having an actual fingerprintable stimuli count (blue star) that is statistically significantly elevated relative to their null distribution (density plots) and at a false discovery rate (FDR) of q<0.05. As annotated on the bottom right of the plots, ~94% of IIT discovery and ~95% of Gi replication subjects are fingerprintable in having statistically significant fingerprintable stimuli counts. Panel B depicts a continuous metric indicative of the degree of gaze uniqueness per individual (IIT discovery, left; Gi replication, right). The gaze uniqueness index (GUI) shown on the x-axis is computed as the ratio between intra-subject divided by average inter-subject gaze similarity. Increasing GUI values indicates increasing gaze uniqueness, whereas values near 1 indicate equal levels of intra-subject and average inter-subject gaze similarity. The density plots per each subject shown on the y-axis depict GUI values across all 700 stimuli, and statistical tests were computed per each individual to test whether GUI was significantly elevated relative to the null value of 1 (vertical black line). After FDR correction for multiple comparisons, subjects can be colored according to individuals that have GUI values significantly different from 1 (turquoise) versus those that do not significantly differ from the null value of 1 (pink). Panel C shows a visual representation of 'gaze fingerprint barcode' matrices (IIT discovery, left; Gi replication, right). These matrices have subjects along the rows and stimuli along the columns and depict binary values that indicate 'hit' (dark red) or 'miss' (light yellow) output of the gaze fingerprint classifier. A 'hit' indicates that an individual can be fingerprinted for that particular stimulus. A 'miss' indicates that the individual cannot be fingerprinted for that particular stimulus. Panel D represents between-subject Jaccard dissimilarity matrices computed on binary gaze fingerprint barcode matrices. The diagonal represents an individual's barcode similarity with themselves and is maximially similar (e.g., light yellow). The off-diagonal values depict how similar/dissimilar are gaze fingerprint barcodes between individuals. Rows and columns of this matrix are re-arranged according to hierarchical clustering dendrograms, such that subtypes would appear as blocks along the diagonal. The absence of such subtype blocks along with near maximal dissimilarity for all between-subject comparisons illustrates visually what clustering algorithms identify as the optimal clustering solution of k = n-1. This solution indicates that each individual is its own cluster. This result illustrates the uniqueness represented in the sparse information conveyed by gaze fingerprint barcodes.*

Gaze fingerprint barcodes (Fig 2C) can also be utilized within a clustering framework to identify whether barcodes are unique to an individual or whether they cluster together in subsets of similar types of individuals. Using Jaccard distance to quantify how similar or dissimilar the binary gaze fingerprint barcodes are between-individuals, we can then apply clustering analysis to this between-subject distance matrix and identify how many clusters exist. If barcodes are unique individuating codes, clustering should identify that the maximum number of clusters detectable (e.g., k = n-1) will be the optimal number of clusters. Indeed in both discovery and replication datasets, we find that the optimal number of clusters is k = n-1 (Fig 2D). This result confirms that gaze fingerprint barcodes hold unique individuating information and that such insights can only be gleaned by gaze fingerprint profiling and barcoding of individuals at scale across a large set of stimuli.

Since gaze fingerprint barcodes are non-random individuating sequences of information, we next asked whether there is any systematic pattern of semantic feature enrichment within an individual's fingerprintable stimuli set. To answer this question, we computed semantic feature enrichment odds ratios that confer an effect size relevant to how many fingerprintable stimuli fall within a particular semantic category. These semantic feature enrichment odds ratios can be thought of as 'semantic feature enrichment barcodes' that semantically describe each individual's fingerprintable stimuli set. Thus, we applied a similar clustering approach to assess how unique are these types of barcodes amongst individuals. Here we find replicable evidence of two clusters of participants in both discovery and replication datasets (Fig 3A-B). One of these subtypes can be largely characterized by individuals whose fingerprintable stimuli are largely enriched for primarily social features of stimuli (e.g., social, humans, faces, emotion) and with an under-representation of non-social features (Fig 3C-D). These specific features have semantic feature enrichment odds ratios that significantly deviate from the null value of 1 in both IIT discovery and Gi replication sets. The other subtype did not have a robust and replicable pattern of enrichment with specific semantic features (Fig 3C-D), indicating that information within their gaze fingerprint barcodes may not readily overlap with the semantic features annotated here in this analysis. These results further enhance our understanding of higher-level of meta-information present in gaze fingerprint barcodes, as they represent strata in the population with respect to the types of semantic features that may hold more-or-less individuating information.

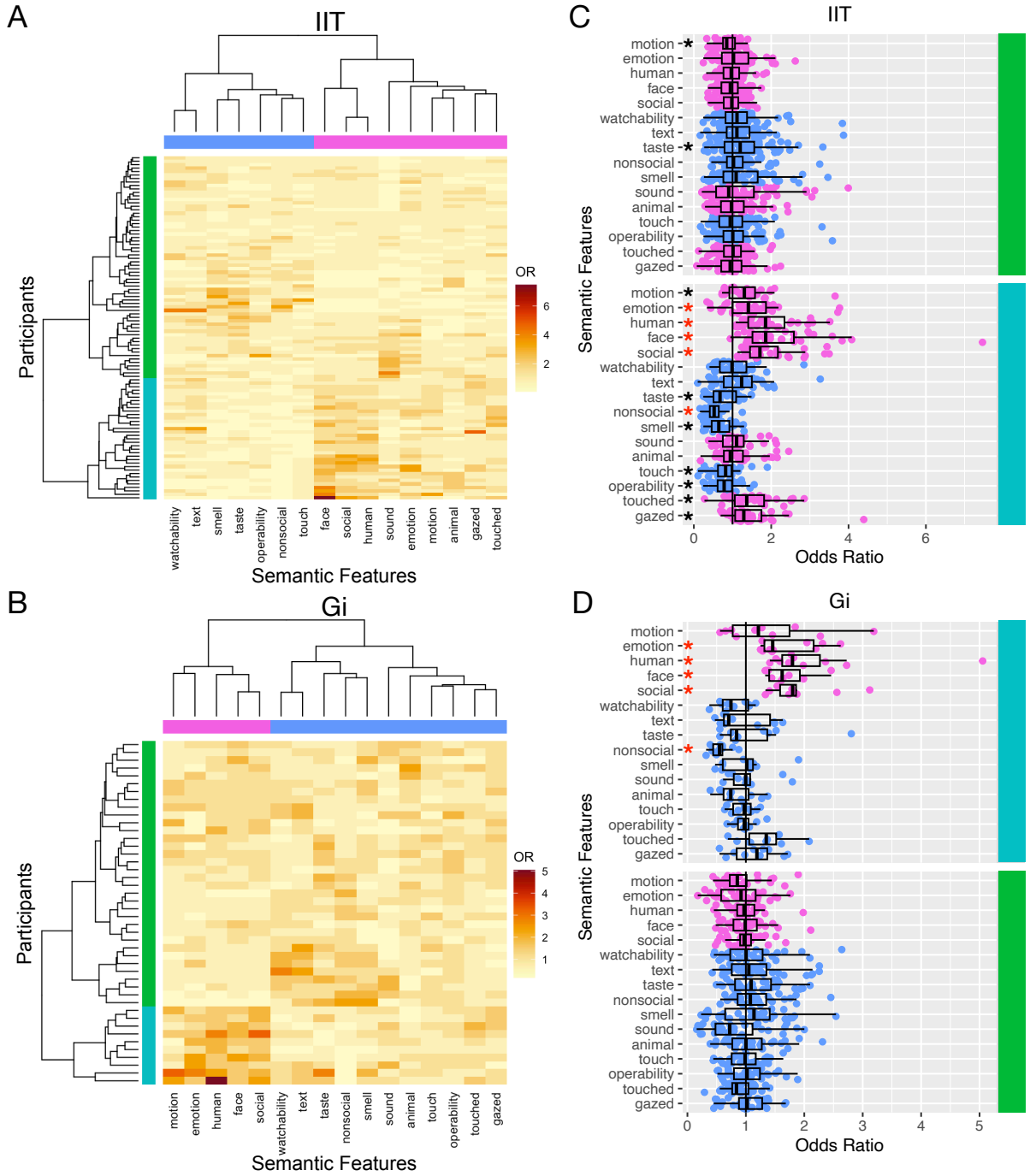


Fig 3: Gaze fingerprint barcodes reveal a subtype enriched for human, social, face, and emotion semantic features. *Panels A (IIT discovery) and B (Gi replication) show heatmaps whereby color represents semantic feature enrichment odds ratios (OR) computed from fingerprintable stimuli sets of each subject. For each subject (rows), we computed enrichment odds ratios, which reflect an effect size relevant to the degree of overlap between a subject's fingerprintable stimuli and stimuli of a specific semantic feature category (columns). Higher enrichment ORs indicate greater overlap between a subject's fingerprintable stimuli and stimuli from a particular semantic feature. Rows of these enrichment OR matrices represent individual subject's 'semantic feature enrichment barcode' and are clustered to identify subtypes with similar patterns of semantic feature enrichment. Columns of this matrix are also clustered to identify collections semantic features with similar enrichment profiles across subjects. Panels A and B show that the turquoise cluster along the subject dimension (rows) can be replicably identified with hotspots that tend to show higher enrichments for motion,*

emotion, human, face, and social features. Panels C (IIT discovery) and D (Gi replication) show odds ratios (x-axis) for each feature (y-axis) and where each individual is a dot. The top and bottom panel of these plots represent the turquoise and green subtypes indicated in panels A and B. The stars next to specific semantic features highlight features that after statistical testing for differences deviating from the null value of 1, are statistically significant at FDR multiple comparison correction q<0.05. The red stars indicate features that are significant in both IIT discovery and Gi replication sets.

In a final set of analyses, we expanded scope on the translational relevance of gaze fingerprinting to autistic traits. Autistic traits represent a continuous and highly heritable phenotype linked back to polygenic or even omnigenic common genetic architecture[26,33]. Degree of gaze similarity is also strongly linked to genetic similarity[13,14], and atypical gaze patterns are key biomarkers linked to autism[13,19,20]. Thus, we reasoned that variation in autistic traits as a genetically sensitive phenotype may be linked to individualized patterns of gaze captured by the gaze fingerprinting approach. Here we used two self-report questionnaires that are widely used to measure autistic traits - the SRS-2 and AQ. These measures are highly correlated ($r = 0.77$, $p < 2.2e-16$) (Fig 4A) and thus, we utilized PCA to capture the first principal component accounting for about 88% of the variance shared between AQ and SRS-2. This autistic traits PC1 was our dependent variable in all further models. Gaze fingerprinting predictors in our model were a combination gaze fingerprinting metrics such as the number of fingerprintable stimuli and median GUI. However, because these variables are highly correlated ($r = 0.80$, $p < 2.2e-16$) (Fig 4B), we orthogonalize them with PCA and use PC1 and PC2 in the model as independent variables. Finally, sex was modeled as a covariate since prior research has shown strong sex differences in autistic traits in the typically-developing population[26,27,32]. Here we find that gaze fingerprinting PC1, reflecting the shared variance between number of fingerprintable stimuli and median gaze uniqueness index, is the primary significant predictor of autistic traits ($\beta = -0.35$, $SE = 0.06$, $t = -5.27$, $p = 7.65e-7$) (see Supplementary Table 1). The relationship between GUI and autistic traits is a negative relationship, whereby increasing autistic traits are observed in individuals with lower gaze fingerprintability (Fig 4C). Importantly, the individuals that heavily deviate from this relationship tend to be individuals with remarkably high autistic traits, beyond cutoffs for typically-developing adult norms on SRS-2 or AQ (turquoise dots in Fig 4C).

Next, we compared models that all incorporate sex as a covariate but differentially utilize either the gaze fingerprinting PC1 as the independent variable or use only median intra-subject or inter-subject gaze similarity. These model comparisons were done to understand whether the variance predicted by gaze fingerprinting PC1 is more than what could be predicted by only using intra- or inter-subject gaze similarity alone. Here we find that the model using gaze fingerprinting PC1 accounts for 21.27% of the variance in autistic traits, whereas median intra-subject gaze similarity accounts 11.67% of the variance and median inter-subject similarity predicts 2.96% of the variance (Fig 4D). Thus, in this direct model comparisons, it is clear that gaze fingerprinting PC1 provides significantly enhanced utility in predicting autistic traits over and above models that only take into account intra- or inter-subject similarity separately.

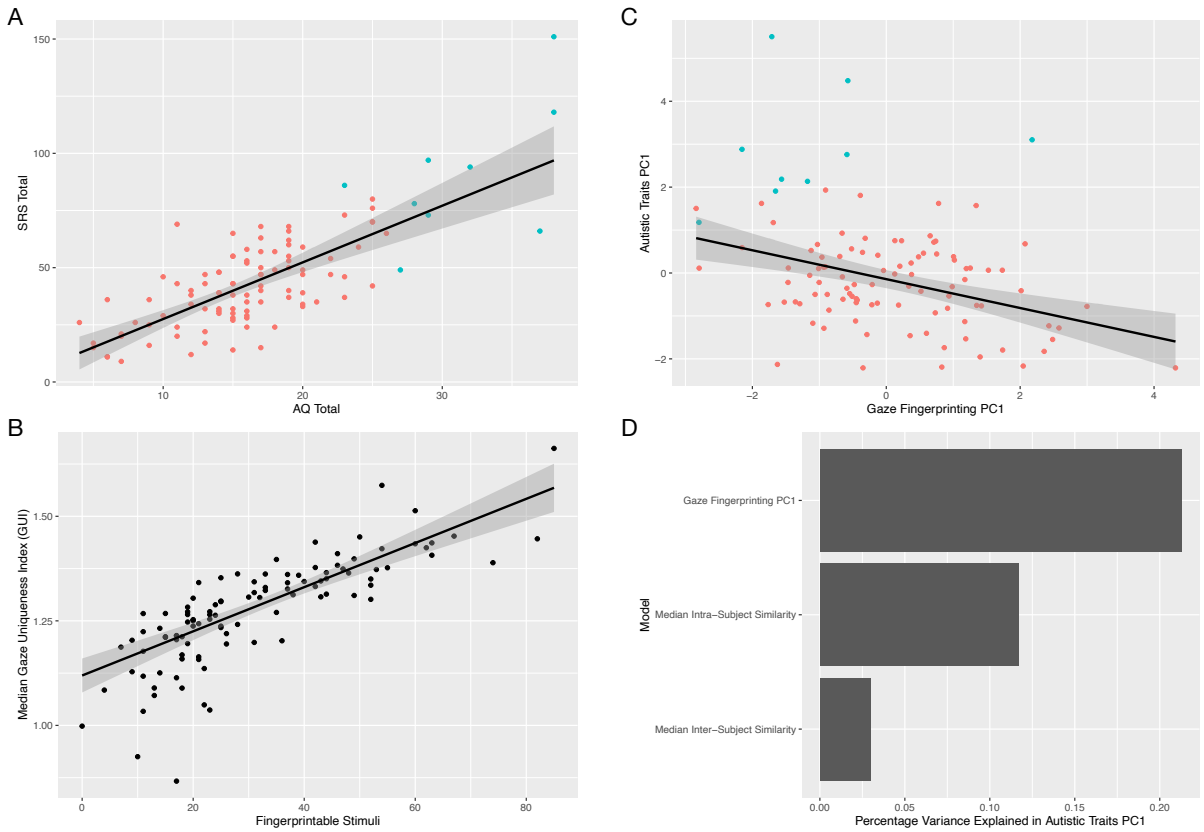


Fig 4: Gaze fingerprinting association with autistic traits. *This figure showcases how gaze fingerprinting metrics are associated with variation in autistic traits. Panel A shows the two autistic trait measures utilized in this study – the AQ (x-axis) and SRS-2 (y-axis). SRS-2 and AQ are highly correlated and thus PCA is used to isolate the first PC representing 88% of the variance shared between AQ and SRS-2. This measure, called 'Autistic Traits PC1', is used as the dependent variable in our model and is plotted on the y-axis in panel C. The coloring of the dots in panel A represents two classes of individuals – individuals within the normative range expected for AQ and SRS-2 in typically-developing adults (n=96, pink) and individuals that are highly deviant from typically-developing adults (n=9, turquoise) (e.g., >26 on the AQ or >2SD (80.81) from typically developing norms on the SRS-2). Panel B shows the 2 gaze fingerprinting measures utilized in the model – that is, the number of fingerprintable stimuli per subject (x-axis) and the median gaze uniqueness index (GUI) computed across all 700 stimuli (y-axis). The scatterplot in panel B shows that these two measures are highly correlated. Thus, PCA was used to orthogonalize the data as two principal components to use as independent variables in a linear model predicting autistic traits. Gaze fingerprinting PC1 accounts for 90% of the variance in the gaze fingerprinting metrics. Panel C shows a scatterplot depicting the association between autistic traits PC1 and gaze fingerprinting PC1. Individuals that are highly deviant from typically-developing norms are plotted in turquoise. The best fit line and confidence bands are computed from a linear robust regression model on all individuals in pink and turquoise. Panel D plots percentage of variance explained in autistic traits PC1 (x-axis) for 3 different models labeled on the y-axis. All 3 models utilize sex as a covariate and primarily differ in terms of the primary independent variable of interest, which is labeled on the y-axis. The model explaining more variance is interpreted as the better model. These model comparisons directly show that gaze fingerprinting PC1 is better at explaining variance in autistic traits compared to utilizing only information about intra-subject or inter-subject gaze similarity alone.*

Discussion

In this work we discovered that identification of individuals is possible simply by using gaze patterns sampled from 3-second viewing epochs on a large and diverse range of static visual stimuli depicting complex natural scenes. By sampling gaze patterns across 700 stimuli, we replicably show that we can gaze fingerprint individuals at levels much higher than chance (50-65% vs 1-2% chance levels), and at rates much higher than prior research identifying twin pairs (29%)[14] or identifying target individuals with a small number of movie clips (~30%)[12]. These higher accuracy rates are achieved by averaging gaze similarity values across large sets of stimuli and we find that for some semantic features with relatively less numbers of stimuli to average over, that accuracy drops. This effect is a methodological point of interest for future work as identifiability likely requires sampling gaze information from an individual across a variety of different stimuli and/or contexts. Indeed, to underscore this point more dramatically, when gaze fingerprinting is done at scale to each and every stimulus, accuracy rates are considerably attenuated compared to analytic strategies such as averaging gaze similarity over large sets of stimuli. Thus, the more information sampled about how an individual looks at the world, the better chance of being able to identify that individual based on gaze patterns.

Despite the added utility of averaging gaze similarity over stimuli, there is the risk with any averaging procedure that the extraction of central tendency estimates may hide unique and important information for the context of gaze fingerprinting. This point has been underscored in the past with analogous literature in functional neuroimaging that contrasts mass univariate analysis techniques to multivariate pattern decoding techniques[30]. Analogous to the stimulus-rich experimental design strategies of representational similarity analysis (RSA)[34], the stimulus-rich aspect of our gaze fingerprinting design allows us to fingerprint individuals on each stimulus. While accuracy rates for any one stimulus are generally much lower than averaging across stimuli, the gaze fingerprinting output per subject and stimuli represent potentially unique 'gaze fingerprint barcodes' that when viewed as a pattern, rather than in isolation, may have important utility in the context of identifying individuals. We find that these 'gaze fingerprint barcodes' are sparse, but are far from random. We replicably find that around 94-95% of all individuals have a fingerprintable stimuli count that is significantly higher than the null distribution for fingerprintable stimuli count when subject identifier is randomly permuted. Clustering these gaze fingerprint barcode patterns shows that each individual is its own unique cluster, and thus confirms that these non-random sparse codes are indeed an identifiable signature per individual when taken into account as a pattern across a stimulus-rich design. Partitioning the fingerprintable signal in these barcodes by semantic features present within visual stimuli showed that we can also replicably identify a subtype of individuals whose fingerprintable stimuli largely represent social, human, face, and emotion stimuli. The other subtype we identify is not systematically enriched for any one particular semantic feature analyzed in this work. Further work examining other attributes of stimuli may identify systematic elements that allow for gaze fingerprinting is these other types of individuals.

Finally, we have shown that gaze fingerprint metrics are not only relevant for identifying individuals, but can also be used to predict variance in other important phenotypic traits of relevance to conditions like autism. Atypical gaze patterns have long been studied in autism[13,19,20] and may represent a key example of how genetic starting points linked to autism affect brain development in ways to create atypical niche construction via how individuals sample the world around them from very early points in development[21–25]. Both individual-

sensitive gaze patterns and autistic traits are likely phenotypes sensitive to heritable polygenic or omnigenic common genetic architecture[13,14,26,33] and support the idea that gaze fingerprinting metrics may be a genetically sensitive signal that can predict variation in autistic traits. Here we find evidence supporting the utility of gaze fingerprinting metrics as predictive of variation in autistic traits such that higher gaze fingerprintability is linked to lower levels of autistic traits. This finding matches up to some evidence supporting that gaze patterns are more idiosyncratic in autism and linked to phenotypic variability in symptom severity[11]. Importantly, we also show that gaze fingerprinting metrics are better at predicting variation in autistic traits than simply using intra-subject or inter-subject gaze similarity. These findings suggestion potential for future work utilizing stimulus-rich designs along with gaze fingerprinting approaches as a potentially more sensitive eye tracking biomarker linked to phenotypic variation indexed by autistic traits.

In summary, we show that gaze fingerprinting is possible both by averaging gaze similarity across stimuli, but also by preserving the gaze fingerprint information held within multivariate gaze fingerprint barcodes. The insights of this study were enabled by the use of a stimulus-rich design and are supported by replicable evidence across discovery and replication sets. Such work is important for future work decomposing individually-sensitive information within gaze patterns and may aid translational work related to understanding atypical gaze patterns in conditions like autism.

Acknowledgments

This project was supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 755816 (AUTISMS) (ERC Starting Grant to MVL).

Author Contributions

Conceptualization: MVL. Methodology: MVL, SKC, ES. Formal analysis: MVL, SKC, ES. Investigation: MVL, SKC, ES, IS, DF, AV, NB, EMB, VM. Writing - original draft preparation: MVL, SKC. Writing - review and editing: MVL, SKC, ES, IS, DF, AV, NB, EMB, VM. Visualization: MVL, SKC. Supervision: MVL. Project administration: MVL. Funding acquisition: MVL.

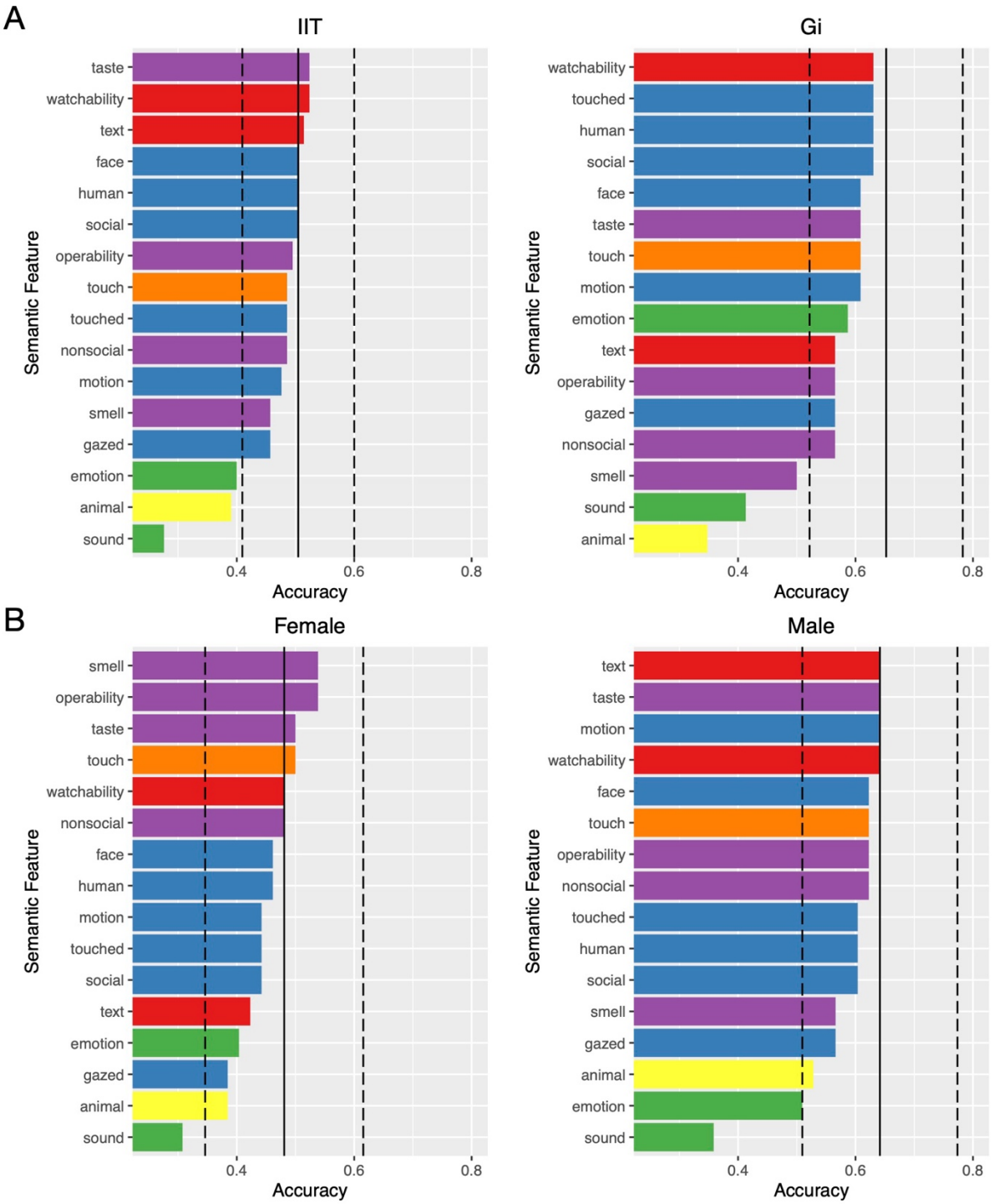
Competing Interests

The authors declare no competing interests.

References

1. Itti, L., Koch, C. & Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 1254–1259 (1998).
2. Einhauser, W., Spain, M. & Perona, P. Objects predict fixations better than early saliency. *J. Vis.* **8**, 18–18 (2008).
3. Borji, A., Sihite, D. N. & Itti, L. What stands out in a scene? A study of human explicit saliency judgment. *Vision Res.* **91**, 62–77 (2013).
4. Henderson, J. M. & Hayes, T. R. Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nat. Hum. Behav.* **1**, 743–747 (2017).
5. Xu, J., Jiang, M., Wang, S., Kankanhalli, M. S. & Zhao, Q. Predicting human gaze beyond pixels. *J. Vis.* **14**, 28 (2014).
6. de Haas, B., Iakovidis, A. L., Schwarzkopf, D. S. & Gegenfurtner, K. R. Individual differences in visual salience vary along semantic dimensions. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11687–11692 (2019).
7. Guy, N. *et al.* A novel perceptual trait: gaze predilection for faces during visual exploration. *Sci. Rep.* **9**, 10714 (2019).
8. Broda, M. D. & de Haas, B. Individual differences in looking at persons in scenes. *J. Vis.* **22**, 9 (2022).
9. Linka, M. & de Haas, B. OSIEshort: A small stimulus set can reliably estimate individual differences in semantic salience. *J. Vis.* **20**, 13 (2020).
10. Linka, M., Broda, M. D., Alsheimer, T., de Haas, B. & Ramon, M. Characteristic fixation biases in Super-Recognizers. *J. Vis.* **22**, 17 (2022).
11. Avni, I. *et al.* Children with autism observe social interactions in an idiosyncratic manner. *Autism Res. Off. J. Int. Soc. Autism Res.* **13**, 935–946 (2020).
12. Keles, U. *et al.* Atypical gaze patterns in autistic adults are heterogeneous across but reliable within individuals. *Mol. Autism* **13**, 39 (2022).
13. Constantino, J. N. *et al.* Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature* **547**, 340–344 (2017).
14. Kennedy, D. P. *et al.* Genetic Influence on Eye Movements to Complex Scenes at Short Timescales. *Curr. Biol.* **27**, 3554-3560.e3 (2017).
15. Kröger, J. L., Lutz, O. H.-M. & Müller, F. What Does Your Gaze Reveal About You? On the Privacy Implications of Eye Tracking. in *Privacy and Identity Management. Data for Better Living: AI and Privacy* (eds. Friedewald, M., Önen, M., Lievens, E., Krenn, S. & Fricker, S.) vol. 576 226–241 (Springer International Publishing, 2020).
16. Liebling, D. J. & Preibusch, S. Privacy considerations for a pervasive eye tracking world. in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* 1169–1177 (ACM, 2014). doi:10.1145/2638728.2641688.
17. Cantoni, V., Musci, M., Nugrahaningsih, N. & Porta, M. Gaze-based biometrics: An introduction to forensic applications. *Pattern Recognit. Lett.* **113**, 54–57 (2018).
18. Rigas, I., Komogortsev, O. & Shadmehr, R. Biometric Recognition via Eye Movements: Saccadic Vigor and Acceleration Cues. *ACM Trans. Appl. Percept.* **13**, 1–21 (2016).
19. Jones, W. & Klin, A. Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature* **504**, 427–431 (2013).
20. Wang, S. *et al.* Atypical Visual Saliency in Autism Spectrum Disorder Quantified through Model-Based Eye Tracking. *Neuron* **88**, 604–616 (2015).

21. Klin, A. *et al.* Affording autism an early brain development re-definition. *Dev. Psychopathol.* **32**, 1175–1189 (2020).
22. Shultz, S., Klin, A. & Jones, W. Neonatal Transitions in Social Behavior and Their Implications for Autism. *Trends Cogn. Sci.* **22**, 452–469 (2018).
23. Johnson, M. H., Jones, E. J. H. & Gliga, T. Brain adaptation and alternative developmental trajectories. *Dev. Psychopathol.* **27**, 425–442 (2015).
24. Johnson, M. H. Autism as an adaptive common variant pathway for human brain development. *Dev. Cogn. Neurosci.* **25**, 5–11 (2017).
25. Lombardo, M. V. *et al.* Default mode-visual network hypoconnectivity in an autism subtype with pronounced social visual engagement difficulties. *eLife* **8**, e47427 (2019).
26. Constantino, J. N. & Todd, R. D. Autistic traits in the general population: a twin study. *Arch. Gen. Psychiatry* **60**, 524–530 (2003).
27. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. & Clubley, E. The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* **31**, 5–17 (2001).
28. Constantino, J. N. & Gruber, C. P. *Social Responsiveness Scale*. (Western Psychological Services, 2012).
29. Finn, E. S. *et al.* Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* **18**, 1664–1671 (2015).
30. Haxby, J. V. *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001).
31. Charrad, M., Ghazzali, N., Boiteau, V. & Niknafs, A. **NbClust** : An *R* Package for Determining the Relevant Number of Clusters in a Data Set. *J. Stat. Softw.* **61**, (2014).
32. Frazier, T. W. *et al.* Confirmatory factor analytic structure and measurement invariance of quantitative autistic traits measured by the social responsiveness scale-2. *Autism Int. J. Res. Pract.* **18**, 31–44 (2014).
33. Robinson, E. B. *et al.* Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat. Genet.* **48**, 552–555 (2016).
34. Kriegeskorte, N. *et al.* Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* **60**, 1126–1141 (2008).



Supplementary Figure 1: *Analysis of IIT and Gi samples and then on IIT sex-stratified samples. Panel A shows results of gaze fingerprint accuracy (x-axis) on IIT (left) and Gi (right) samples. Panel B shows sex-stratified analyses of the IIT sample (females, left; males, right). The vertical black lines in each panel depict gaze fingerprint accuracy when average gaze similarity is computed across all 700 stimuli (dotted vertical black lines represent bootstrap 95% confidence intervals). Each bar shows gaze fingerprint accuracy when computed as average gaze similarity on subsets of stimuli grouped by semantic features shown on the y-axis.*

	Estimate	Std. Error	t-stat	p-value
(Intercept)	-0.3817752	0.16382788	-2.330343	2.18E-02
Sex	0.4328317	0.2007473	2.156102	3.34E-02
Gaze Fingerprinting PC1	-0.3578814	0.06787054	-5.273	7.66E-07
Gaze Fingerprinting PC2	-0.3224181	0.20831048	-1.547777	1.25E-01

Supplementary Table 1: *Results of regression model predicting autistic traits by sex, and fingerprinting PC1 and PC2.*