



A Functional Model for Studying Common Trends Across Trial Time in Eye Tracking Experiments

Mingfei Dong¹ · Donatello Telesca¹ · Catherine Sugar^{1,2} · Frederick Shic^{3,4} · Adam Naples⁵ · Scott P. Johnson⁶ · Beibin Li^{3,7} · Adham Atyabi^{3,8} · Minhang Xie³ · Sara J. Webb^{3,9} · Shafali Jeste¹⁰ · Susan Faja¹¹ · April R. Levin¹² · Geraldine Dawson¹³ · James C. McPartland⁵ · Damla Şentürk¹ · the Autism Biomarkers Consortium for Clinical Trials

Received: 4 January 2022 / Revised: 8 May 2022 / Accepted: 25 July 2022 /
Published online: 5 September 2022
© The Author(s) 2022

Abstract

Eye tracking (ET) experiments commonly record the continuous trajectory of a subject's gaze on a two-dimensional screen throughout repeated presentations of stimuli (referred to as trials). Even though the continuous path of gaze is recorded during each trial, commonly derived outcomes for analysis collapse the data into simple summaries, such as looking times in regions of interest, latency to looking at stimuli, number of stimuli viewed, number of fixations, or fixation length. In order to retain information in trial time, we utilize functional data analysis (FDA) for the first time in literature in the analysis of ET data. More specifically, novel functional outcomes for ET data, referred to as viewing profiles, are introduced that capture the common gazing trends across trial time which are lost in traditional data summaries. Mean and variation of the proposed functional outcomes across subjects are then modeled using functional principal component analysis. Applications to data from a visual exploration paradigm conducted by the Autism Biomarkers Consortium for Clinical Trials showcase the novel insights gained from the proposed FDA approach, including significant group differences between children diagnosed with autism and their typically developing peers in their consistency of looking at faces early on in trial time.

Keywords Eye tracking · Autism spectrum disorder · Functional data analysis · Functional principal component analysis · Multilevel functional principal component analysis

✉ Damla Şentürk
dsenturk@ucla.edu

Extended author information available on the last page of the article

1 Introduction

Eye tracking (ET) experiments commonly involve presentation of static or continuous stimuli (possibly coupled with audio), where subjects' continuous trajectory of gaze on a two-dimensional screen is recorded. Experiments typically involve repeated presentation of stimuli, referred to as trials. ET approaches have become a popular tool in the study of cognitive, emotional and attention processes, especially in autism research [1, 2], where passive viewing tasks with eye tracking (ET) methods are frequently used to evaluate the extent to which visual attention is affected in autism spectrum disorder (ASD). Evidence shows that ET experiments may be a promising source of biomarkers that can objectively measure core facets of ASD-related deficits [3].

Our motivating ET study is an implicit visual exploration (VE) paradigm conducted by the Autism Biomarkers Consortium for Clinical Trials (ABC-CT) [4]. The experiment included repeated presentations of an array of social and non-social stimuli to evaluate differences in attention between children with ASD and their typically developing (TD) peers and to reveal ASD-specific signatures of visual array processing. Participants were presented with full-color circular arrays consisting of 5 stimuli: face with direct gaze, outline of face filled with blurred pattern, bird, mobile phone, and motor vehicle (Fig. 1a). Each array consisted of a different set of stimuli, and each trial, consisting of presentation of the circular array accompanied with music, lasted 20 s. The ABC-CT study included three visits (at baseline, 6 weeks, and 6 months), where ET data were collected on two consecutive days at each visit with six trials conducted per day.

Even though the complete path of gaze is recorded during each trial, common analysis of ET data is based on data summaries of looking times in regions of interest (ROI), latency to looking at stimuli, number of stimuli viewed, number of fixations, or fixation length, with responses averaged across trials (Fig. 1a). While there may be benefits to these summary measures in terms of simplicity of analysis, collapsing information across trial time leads to loss of valuable information. We propose a functional model for the analysis of ET data, which includes the proposal of novel functional outcomes, referred to as viewing profiles, which capture common trends in trial time (Fig. 1b). In the VE example, the proposed viewing profiles capture an initial gaze toward the social object (face with direct gaze), earlier on in trial time, which is common across trials and subjects. This common trend across trials is lost in analysis of data summaries that collapse trial time.

Even though motivated by the VE paradigm conducted by ABC-CT, the proposed viewing profiles can also be applied to other ET paradigms, to capture common trends across trial time. Consider a visual world paradigm (VWP) for spoken language processing [5], where participants hear an instruction to click on a beaker, while viewing a display containing a beaker, a beetle, a speaker and a carriage. A common trend across trial time is an eye movement to beetle when only the first two segments have been heard, followed by a gaze at the beaker 100 ms later. Extended to non-static viewing experiments, such as dynamic social

video viewing tasks, where participants are presented with a video of social interactions between individuals, trial time analysis can capture common trends of gaze into social interactive activity ROIs, useful in assessing the extent of interest or disinterest in joint activities.

Following the proposal of viewing profiles, we showcase the use of functional principal component analysis (FPCA) in extracting the dominant modes of variation in the functional outcome and modeling variability across subjects. Multilevel FPCA (MFPCA) is also utilized to decompose the total variation in the data across the three visits into within-subject (across visit) and across-subject components. Lower dimensional representations of viewing profiles, obtained via FPCA and MFPCA, are utilized in drawing efficient inference on diagnostic group (ASD vs. TD) differences via bootstrap.

The literature on functional data analysis (FDA) [6, 7] has grown rapidly over the past two decades, with a considerable fraction of the work involving applications to longitudinal data [8–12]. More recently, there has been interest in analyzing multiple trajectories, with dependencies among the curves created by temporal proximity [13–17]. Different from univariate or multivariate statistics where the unit of analysis may be a scalar or a vector, the unit of analysis in FDA is a function. In addition, the functional outcomes may be collected across multiple visits, referred to as longitudinally collected functional data. FPCA for longitudinal and multivariate functional processes have been a major modeling theme in the FDA literature, providing efficient and interpretable representations of functional variability in highly structured settings. For the longitudinally collected functional data, both semi-parametric [14, 18] and non-parametric [19–21] approaches have introduced a robust family of methods dealing with estimation and inference under a varying array of simplifying assumptions. We showcase the use of both FPCA and MFPCA (proposed by Di et al. [14]) in the analysis of ET data in this work.

To the best of our knowledge, the proposed approach is the first functional model in literature for the analysis of ET data. In addition to the proposal of the novel functional outcomes, the current work also provides tools for capturing variation in the proposed viewing profiles, in addition to efficient inference on diagnostic group differences via bootstrap. Note that there has been limited literature on modeling common trends across trial time in ET experiments using growth curves. Mirman et al. [22] have considered modeling trends across trial time in a VWP for spoken language processing, and Del Bianco et al. [23] modeled trends in visual attention to faces from a static stimulus viewing ET paradigm, both using growth curve modeling. The proposed FDA approach is more flexible than growth curve modeling with its lack of reliance on parametric modeling assumptions. FDA allows modeling of trial time trends in its full complexity with minimal modeling assumptions.

Section 2 outlines examples of traditional analyses of ET data via linear mixed-effects modeling of data summaries. In applications to VE paradigm, the analysis of diagnostic group differences in mean latency to social objects is presented using mixed-effects modeling. Analysis finds that ASD children have a longer mean latency to social objects. Section 3 introduces the novel functional outcome of viewing profiles, followed by Sects. 4 and 5, outlining the use of FPCA and MFPCA in analysis of viewing profiles. Proposals in both Sects. 4 and 5 are followed by ways

of drawing efficient inference, utilizing the resulting lower dimensional representations of ET data. The proposed inference includes bootstrap tests for group mean differences or visit- and group-specific mean differences in the derived functional trajectories. In applications to VE data, the FDA approach shows that compared to their TD peers, ASD children initiate gaze at the social object at similar trial times, but at a lower frequency across trials, leading to a longer mean latency to social objects. Hence the proposed FDA approach uncovers additional information on trends across trial time that leads to novel interpretations of findings. We conclude with the Discussion Sect. 6 on extensions to the proposed methodology.

2 Traditional Analysis of Eye Tracking Data

2.1 Derived Outcomes and Statistical Methods

Traditional analysis of ET data collapses information from gaze path of participants across trial time, and models data summaries such as looking times in ROI, latency to looking at stimuli, number of stimuli viewed, number of fixations, or fixation length (Fig. 1a). Data summaries are typically modeled via analysis of variance (ANOVA) or linear mixed-effects models (LMM) where tests for diagnostic group differences can also be formulated. LMM is a very popular approach for modeling scalar outcomes repeatedly observed within subjects (e.g., at different study visits). Within-subject correlation is often modeled via subject-level random effects (e.g., y-intercept or slope) where goodness of fit criteria such as the Akaike information criteria (AIC) or the Bayesian information criteria (BIC) can be utilized in the choice of the dependency structure.

2.2 Application to Visual Exploration Paradigm

Our motivating ET study, the VE paradigm, was conducted by the ABC-CT on 399 children (280 ASD and 119 TD) between the ages of 6 and 11. Participants were presented with an array of social and nonsocial stimuli for 20 s each, where their complete path of gaze on a two-dimensional screen was recorded within each of the six trials conducted on two consecutive days during the three visits of the study at baseline, 6 weeks, and 6 months (referred to as T1, T2 and T3). Each array consisted of a different set of stimuli from the 5 categories including: face with direct gaze, outline of face filled with a pattern, bird, mobile phone, and motor vehicle, where the location for the stimulus from each category varied between arrays (Fig. 1a). The VE paradigm (adapted from Gliga et al. [24]) was intended to unveil the distinct visual array processing patterns of ASD children and assess the differences in their attention to social and nonsocial objects compared to children in the TD group. Details of the ABC-CT protocol and data availability are given in McParland et al. [4] (preliminary results from a feasibility study on this task involving 46 children can be found in Tsang et al. [25]).

LMMs were used in modeling of multiple outcomes, including looking times at social (face with direct gaze), nonsocial objects (rest of the 4 stimuli categories, including outline of face filled with blurred pattern), and the whole screen (proportional to trial length), number of stimuli viewed during a trial (social, nonsocial), fixation length on ROI (social, nonsocial), and log latency to ROI (social, nonsocial) [26]. LMMs included diagnostic group (ASD and TD) and study visit (T1, T2, T3) as fixed effects (second order interaction between group and visit was not found significant) and subject-level y-intercept as the random effect. TD subjects yielded a higher percentage of valid data (the percentage time of valid data relative to the total trial length, denoted by Valid %) on average compared to the ASD group, hence reported results control for this variable while modeling group differences. The subject-level y-intercept was selected as the optimal random effects structure using BIC among models which included subject- and visit-level random effects, and subject-, visit-, and day-level random effects. While the main focus in this work will be on results from the modeling of latency to social object, as it relates to trial time trends later modeled with FDA, we present summary of results from modeling of the other outcomes as well (Table S1). Briefly, there were significant group differences detected in looking times at nonsocial objects and the whole screen and fixation length on nonsocial objects, where the ASD children displayed shorter mean looking times to nonsocial objects (p value < 0.001) and the whole screen (p value < 0.001), and fixated on nonsocial objects for a shorter duration, compared to TD children (p value < 0.001), adjusted for the percentage of valid data provided. In addition, there was a significant effect of study visit detected, with children from both diagnostic groups displaying longer looking times to social objects (p value < 0.001), and the whole screen (p value < 0.001), lower number of nonsocial stimuli viewed (p value $= 0.001$), and longer fixation duration on both social (p value < 0.001) and nonsocial objects (p value < 0.001) on average in the later visits (T2 and T3) compared to baseline.

Finally, significant diagnostic group differences were detected regarding social objects with the outcome of latency to social subject (Table 1) (diagnostic group was not found significant in modeling latency to nonsocial objects). ASD children had longer average latencies to social object compared to TD children (p value $= 0.001$), controlling for the percentage of valid data provided (children with higher percentages of valid data tend to have shorter latencies to social object). In addition, there was a significant effect of visit on latency to social object as well, similar to results on the other outcomes, with shorter average latencies at the latter visits (T2, T3) (p value $= 0.048$). The interaction between group and visit was also found significant for this outcome, where the diagnosis group difference was slightly smaller in T3 compared to baseline (p value $= 0.008$). Connected to results on longer average latency to social object among ASD children, the FDA modeling below provides additional insights into group differences related to viewing of social objects. It will be shown that while ASD children initiate gaze at the social object, at similar trial times as TD children, they do so at a lower frequency across trials which contribute to longer latencies detected in the ASD group. Note that adjustments to other covariates, such as full scale IQ (FSIQ) or age in the above LMM did not change the conclusions on group differences, where age was found to be significantly

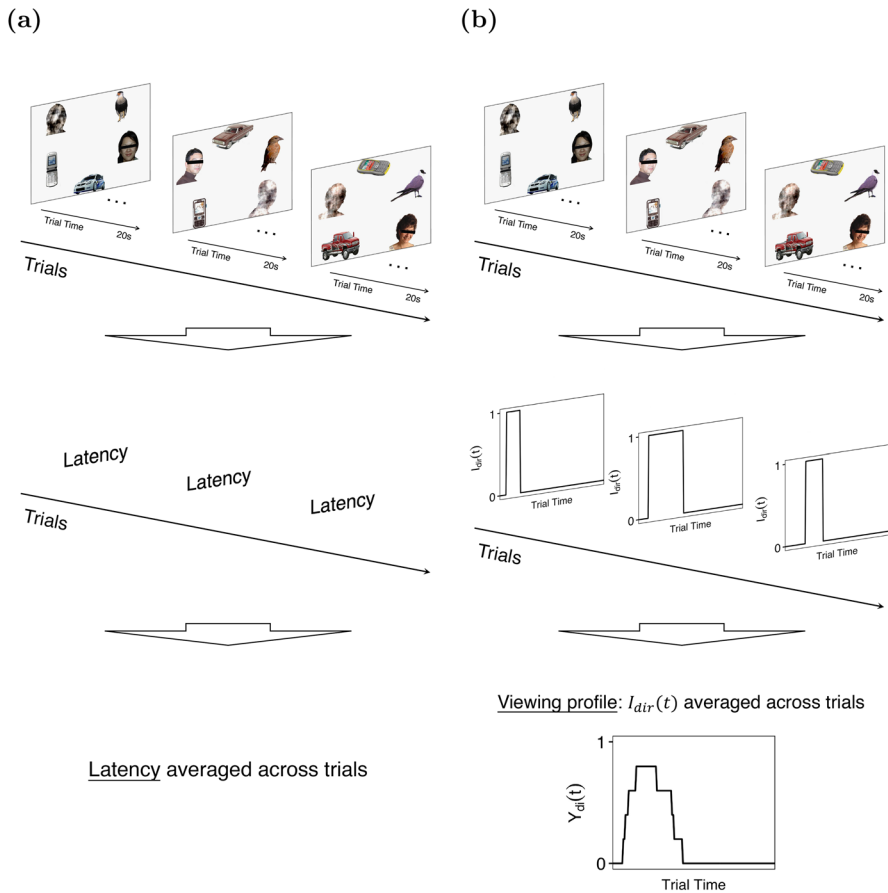


Fig. 1 **a** Derivation of ET outcomes in traditional analyses. Sample arrays from the visual exploration paradigm contained 5 full-color stimuli: face with direct gaze (social object, black bar inserted over eyes to protect identity of actress), outline of face filled with blurred pattern, bird, mobile phone, and motor vehicle. **b** Derivation of functional ET outcomes (viewing profiles) retaining information across trial time

associated with shorter latency to social object, and the effect of FSIQ was not found significant.

3 Proposed Functional Outcomes: Viewing Profiles

A limitation of the traditional analysis outlined in Sect. 2.1 is that the summary measures considered collapse information across trial time. With the goal of uncovering common trends of gaze across trial time, we introduce the ET functional outcome of viewing profiles. Let $\mathbb{I}_{dir}(t)$ denote the indicator of a particular viewing behavior at trial time t , $t \in \mathbb{T}$, for trial r , $r = 1 \dots, R_{di}$, of subject i , $i = 1, \dots, N_d$,

Table 1 Results from linear mixed-effect modeling (LMM) of log latency to social object

Response	Diagnosis (ASD)		Visit(T2)		Visit(T3)		Valid%		ASD × T2		ASD × T3	
	Estimate	p value	Estimate	p value	Estimate	p value	Estimate	p value	Estimate	p value	Estimate	p value
log (latency)	0.159	0.001	− 0.046	0.126	− 0.059	0.048	− 1.357	< 0.001	− 0.050	0.163	− 0.095	0.008

Significant terms are given in bold

in group d , $d = 1, \dots, D$, where D , N_d and R_{di} denote the total number of groups, subjects within groups and trials with valid data within subjects, respectively. In applications to the VE paradigm where the main goal centers around the children's viewing behavior on social objects, we let $\mathbb{I}_{dir}(t)$ denote the indicator for the target behavior of the initial gaze being at the social object. If the social object is the first stimulus a child looks at after the start of the trial, then $\mathbb{I}_{dir}(t)$ will be a step function, equaling one for the duration of the initial gaze at the social object (see Fig. 2a for an example where the initial gaze at social object lasts approximately 400 ms). Figure 2b displays $\mathbb{I}_{dir}(t)$ in another scenario, where a child's initial gaze is on a nonsocial object, where $\mathbb{I}_{dir}(t)$ equals zero for the first 3000 ms of the trial. Note that in this application, we concentrate on the first 3000 ms of the 20 s total trial time, which cover nearly all initial gaze profiles.

The functional outcome, viewing profile for the target behavior of the initial gaze being at the social object for subject i in group d , is then defined as the average of $\mathbb{I}_{dir}(t)$ across all available trials of subject i ,

$$Y_{di}(t) = \sum_{r=1}^{R_{di}} \mathbb{I}_{dir}(t) / R_{di}. \quad (1)$$

Hence, the constructed viewing profile portrays the proportion of times the target behavior is observed across trials, referred to as the consistency of the target behavior. In other words, a higher value on the viewing profile corresponds to trends that

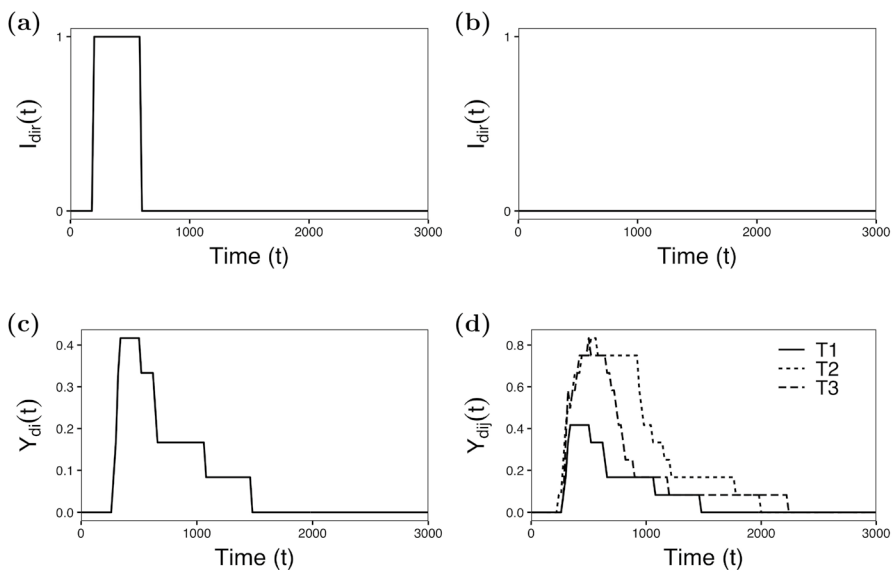


Fig. 2 **a** A trial-level indicator for the initial gaze being at the social object where the initial gaze lasts approximately 400 ms. **b** A trial-level indicator for the initial gaze being at the social object where a child's initial gaze is on a nonsocial object. **c** A sample viewing profile for a subject from baseline in our application to the VE paradigm. **d** The longitudinally observed (T1, T2, T3) viewing profiles of the same sample subject from our application to the VE paradigm

are common across trials (with more of the indicator values across trials equaling one), which is interpreted as the consistency of the target behavior across trials. Note that the proposed viewing profile is a function of trial time and contrary to summary measures of Sect. 2.1, it preserves trends in gaze behavior within each subject that is common across trials. Figure 2c displays a viewing profile for a subject from baseline in our application to the VE paradigm, constructed as the average of the $R_{di} = 12$ trial indicators $\mathbb{I}_{dir}(t)$ available at baseline. For this subject, the consistency of the initial gaze at social object at 340 ms is slightly higher than 40%, which signals that for 40% of the subject's 12 trials available at baseline, the subject started an initial gaze at social object by 340 ms. We also observe that the initial gaze at social object lasted at least for 320 ms in 1/3 of the trials, and at least 760 ms in about 17% of the trials.

Extending the definition of viewing profiles over multiple visits of the study, let

$$Y_{dij}(t) = \sum_{r=1}^{R_{dij}} \mathbb{I}_{dijr}(t) / R_{dij} \quad (2)$$

denote the viewing profile for the j th visit of subject i from group d , equaling the average of visit-specific indicators $\mathbb{I}_{dijr}(t)$ for the viewing behavior at the r th trial, $r = 1, \dots, R_{dij}$ of the i th subject at visit j , $j = 1, \dots, J$. In (2), R_{dij} denotes the total number of valid trials observed for subject i at visit j . Note that the new functional outcome constitutes longitudinally collected functional data, where viewing profiles are collected over J visits for each subject. Figure 2d displays the longitudinally observed viewing profiles of the same sample subject with the baseline viewing profile plotted in Fig. 2c. The longitudinally observed viewing profiles are given from the three visits (T1, T2, T3) of ABC-CT. The consistency of the initial gaze at social object by 300 to 400 ms is much higher (80%) for this subject at the latter visits (T2 and T3), compared to baseline (40%). We also observe that the length of the initial gaze at social object is the longest at T2, followed by T3 and T1 (the initial gaze at social object lasted at least 480 ms in 75% of the trials at T2).

Other viewing profiles can also be defined in applications to the VE paradigm that utilize different target behaviors such as initial social gaze, or more generally, social gaze. The first viewing profile would mark the first gaze at social object regardless of whether the social object was the first stimulus viewed and the latter viewing profile would mark the gaze at social object regardless of order of viewing, providing different insights into the gaze behavior across trial time. In this article, we utilize the initial gaze being at the social object, an indicator for whether or not the initial gaze is at the social object, as the target viewing behavior to showcase the proposed functional approach.

4 Modeling Viewing Profiles

4.1 Functional Principal Component Analysis

One of the most commonly used tools for modeling variation in functional data is the functional principal component analysis (FPCA). Similar to principal component analysis which estimates the dominant modes of variation in multivariate data through eigenvectors, FPCA utilizes eigenfunctions to describe the directions of variation in functional data. Eigenscores capture the subject-specific weights which represent the lower dimensional multivariate representation of functional data along different directions of variation specified by the eigenfunctions. More specifically, consider the following decomposition of the viewing profile $Y_{di}(t)$ in (1), for subject $i, i = 1, \dots, N_d$, from diagnosis group $d, d = 1, \dots, D$, at trial time $t, t \in T$,

$$Y_{di}(t) = \mu(t) + \eta_d(t) + Z_{di}(t) + \epsilon_{di}(t),$$

where $\mu(t)$ and $\eta_d(t)$ denote the overall mean function and group-level mean shifts, $Z_{di}(t)$ denotes the subject-level random deviation from the group-specific mean, and $\epsilon_{di}(t)$ denotes the measurement error with mean zero and variance σ_d^2 . Denote the covariance of $Z_{di}(t)$ by $\Sigma^{(d)}(s, t) = \sum_{k=1}^{\infty} \lambda_k^{(d)} \phi_k^{(d)}(s) \phi_k^{(d)}(t)$, where $\phi_k^{(d)}(t)$ and $\lambda_k^{(d)}$ are the eigenfunctions and eigenvalues, representing the directions of variation and the amount of variation along the identified directions observed across subjects within group d , respectively. Then the Karhunen Loève representation [27–29] of $Y_{di}(t)$ is given as

$$Y_{di}(t) = \mu(t) + \eta_d(t) + \sum_{k=1}^{\infty} \xi_{dik} \phi_k^{(d)}(t) + \epsilon_{di}(t), \quad (3)$$

where $\xi_{dik} = \langle Z_{di}(t), \phi_k^{(d)}(t) \rangle$ are the subject-specific eigenscores, with mean zero and variance $\lambda_k^{(d)}$, representing the projection of Z_{di} onto the set of orthonormal eigenfunctions $\phi_k^{(d)}(t)$. Note that the decomposition in (3) provides a lower dimensional representation of the functional data via projections onto the dominant modes of variation and representation of the sample through the eigenscores. The number of eigencomponents included in each diagnostic group in (3) is truncated in practice at K_d , using fraction of variance explained (FVE). We use scree plots in choosing the number of eigencomponents retained. The number of eigencomponents is selected by locating the elbow of the scree plot (the point after which additional components add relatively reduced gains in terms of the FVE). In applications to the VE data, the leading two eigencomponents ($K_d = 2$) are retained according to the scree plots (Figs. S1a, b), explaining more 80% of the total variation.

FPCA can be fit using the publicly available R package ‘refund’ [30]. Briefly, the overall mean $\mu(t)$, and group-level mean shifts $\eta_d(t)$ in (3) are targeted using cubic B-splines. Due to possible imbalances in sample sizes across diagnostic groups, the smoothed mean functions within each group are averaged to target the overall mean function $\mu(t)$. Next, $\Sigma^{(d)}(s, t)$ is targeted using the empirical covariance estimates of $\text{cov}\{Y_{di}(s), Y_{di}(t)\}$. Two-dimensional thin plate splines are utilized in smoothing the

raw covariance before performing FPCA, where the degree of smoothness is determined using generalized cross validation. Before smoothing, the diagonal elements of the raw covariance, which are prone to measurement error effects, are removed. Measurement error variance σ_d^2 is then targeted as the mean difference between the diagonal elements of the smoothed and unsmoothed covariances. For further details on estimation of the means and covariances in FPCA, see Yao et al. [31] and Goldsmith et al. [32].

4.2 Group-Level inference via Parametric Bootstrap

In most applications, similar to our VE example, it is of interest to draw valid inference on group mean trends. The lower dimensional representation achieved via FPCA, in addition to identifying differences in the leading modes of variation, also allows for formulation of efficient inference on diagnostic group mean differences. More specifically consider the null hypothesis $H_0: \eta_d(t) = 0$ for $d = 1, \dots, D$, corresponding to no diagnosis group mean difference. The proposed parametric bootstrap procedure generates functional outcomes under the null utilizing the estimated FPCA components, $Y_{di}^b(t) = \hat{\mu}(t) + \sum_{k=1}^{K_d} \xi_{dik}^b \hat{\phi}_k^{(d)}(t) + \epsilon_{di}^b(t)$, where ξ_{dik}^b and $\epsilon_{di}^b(t)$, are sampled from $N(0, \hat{\lambda}_k^{(d)})$ and $N(0, \hat{\sigma}_d^2)$, respectively. A total of B bootstrap samples are generated at sample sizes in each group identical to the observed data for testing ($B = 200$ is used in the reporting of the below results). The proposed test statistic, $T_0 = \left[\sum_{d=1}^D \int \{\hat{\eta}_d(t) - 0\}^2 dt \right]^{1/2}$, based on the norm of the sum of the departures of the estimated group-level shifts $\hat{\eta}_d(t)$ from zero, is calculated for the observed data, denoted by T_0 , and each bootstrap sample, T_0^b , $b = 1, \dots, B$. Test statistics computed in the bootstrap samples provide an approximation of the null distribution of the test statistic T_0 , and the p value is estimated by $\sum_{b=1}^B \mathbb{I}(T_0^b > T_0)/B$, where $\mathbb{I}(\cdot)$ denotes the indicator function.

4.3 Application to Visual Exploration Paradigm

The estimated FPCA components, in applications to baseline (T1) viewing profiles from the VE paradigm, are given in Figs. 3, 4 and 5. The estimated overall mean function, given in Fig. 3a, signals maximum consistency of the target behavior of the initial gaze being at the social object around 500 ms in 42% of the trials. This shows that the initial gaze at social object is a common behavior across study subjects and trials. The estimated group-level shifts $\eta_d(t)$, $d = 1, 2$, given in Fig 3b, and group-specific mean estimates in Fig. 3c show that even though subjects display an initial gaze at the social object most frequently at 500 ms in both groups, the consistency of the target behavior is lower in the ASD group compared to the TD group. While subjects in the ASD group gaze at the social object by 500 ms in 35% of the trials, TD subjects display the target behavior in 50% of the baseline trials. Hence, while the time to initiate gaze at the social object, is similar across groups, the TD group is more consistent in this behavior. These findings give additional insights on

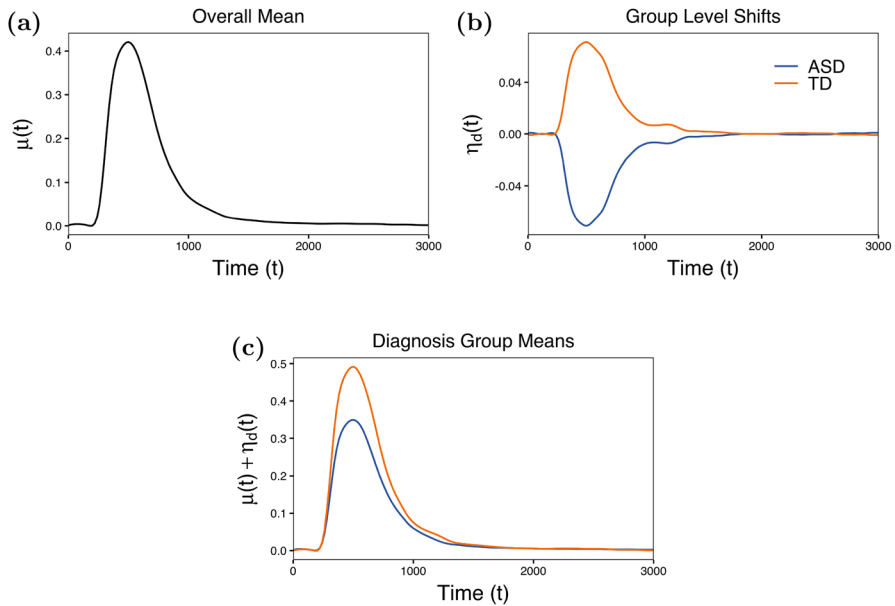


Fig. 3 **a** Estimated overall mean viewing profile, $\mu(t)$. **b** Estimated group-level shifts from the overall mean, $\eta_d(t)$, $d = 1, 2$. **c** Estimated mean viewing profiles, $\mu(t) + \eta_d(t)$, $d = 1, 2$, for the two diagnosis groups (ASD and TD)

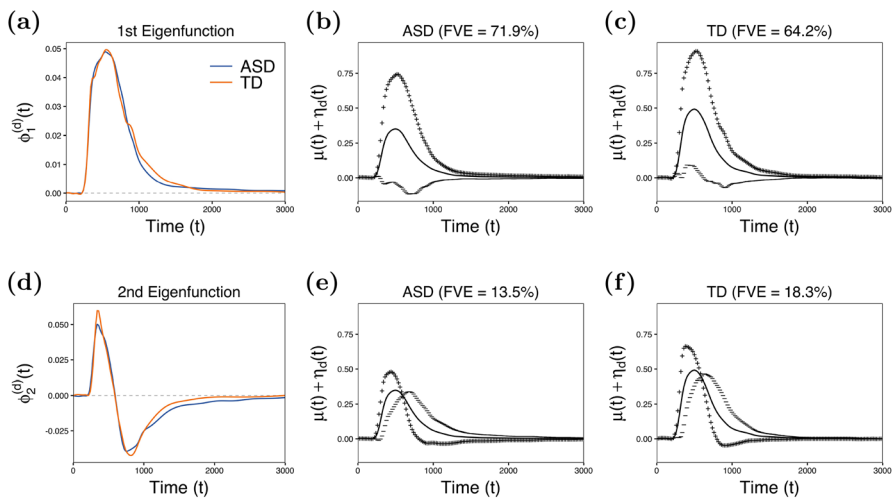


Fig. 4 **a, d** The first two leading eigenfunctions, $\phi_k^{(d)}(t)$, $k = 1, 2$, $d = 1, 2$, estimated from FPCA of the viewing profiles in the VE application. **b, c** Group-specific mean estimates plus and minus two times the amount of variation along the leading eigenfunction, $\mu(t) + \eta_d(t) \pm 2\sqrt{\lambda_1^{(d)}}\phi_1^{(d)}(t)$, $d = 1, 2$. **e, f** Group-specific mean estimates plus and minus two times the amount of variation along the second leading eigenfunction, $\mu(t) + \eta_d(t) \pm 2\sqrt{\lambda_2^{(d)}}\phi_2^{(d)}(t)$, $d = 1, 2$

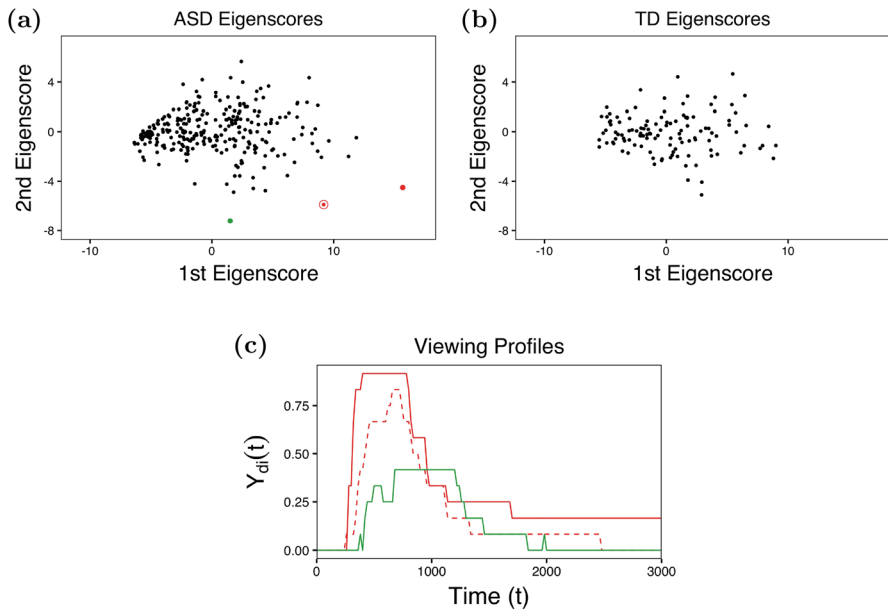


Fig. 5 **a** Estimated first two leading eigenscores for the ASD group. Three ASD subjects with either the first or the second leading eigenscore outside 2 standard deviations from the mean are highlighted in green and red. **b** Estimated first two leading eigenscores for the TD group. **c** Viewing profiles of the three outlying ASD subjects. (Profiles for subjects highlighted by a red dot, circled dot and green dot are given in solid red, dashed red, and solid green.) (Color figure online)

the reasons behind ASD subjects displaying longer latency to the social object, compared to their TD peers.

The leading two eigencomponents are retained in the decompositions, based on the scree plots given in Figs. S1a and b. Figure 4a and d display the estimated first two leading eigenfunctions which explain 71.9% and 13.5% of the total variation in the ASD and 64.2% and 18.3% of the total variation in the TD subjects, respectively. The estimated eigenfunctions signal similar modes of variation across the two diagnostic groups. The first leading eigenfunction captures the variation in consistency of the initial gaze at the social object across trials, while the second leading eigenfunction signals variation in the duration of the initial gaze. This can also be seen in Fig. 4b, c and e, f, which display group-specific mean estimates plus and minus two times the amount of variation along each eigenfunction. While a positive leading eigenscore suggests higher consistency of an initial gaze being at the social object across trials, a negative second leading eigenscore suggests a longer initial gaze. Figure 5a and b display the two leading eigenscores in both groups, where groups display similar variation in both dimensions. In addition, subjects with negative first leading eigenscores tend to vary less in their second leading eigenscores (trend more visible in ASD, since sample size is larger), suggesting that subjects with low consistency of initial gaze at the social object have less variation in their length of gaze, as expected, since they have a lower percentage of trials displaying the target behavior (viewing profiles of ASD subjects with low and high

consistency of initial gaze at the social object displayed in Supplementary Materials Figs. S2 and S3, respectively). Three ASD subjects with either the first or the second leading eigenscores outside 2 standard deviations from the mean are highlighted in green and red (Fig. 5a). The two ASD subjects highlighted in red both display higher consistency of the initial gaze at the social object and longer gaze durations (Fig. 5c). The subject highlighted in green has the lowest second leading eigenscore in the ASD group, outside of 4 standard deviations from the mean, with the longest gaze duration (Fig. 5c). Consistent with the significant correlation between the second eigenscores and age (with kids displaying shorter gaze durations as they age), reported in the Discussion Section, the ASD subjects highlighted in green and circled red are younger children with ages 6.3 and 6.2 years old. Note that the same three subjects are still identified as having higher consistency of the initial gaze at the social object or longer gaze durations across all three visits in Sect. 5.3 using multilevel FPCA. However, the subject highlighted as circled red is now identified as having only higher consistency of the initial gaze at social object across the three visits. Finally, utilizing the parametric bootstrap test outlined in Sect. 4.2, the diagnosis group mean differences observed in Fig. 3c are found to be significant (p value < 0.001), where TD subjects are significantly more consistent in their initial gaze at the social object across trials. The observed group difference in consistency suggests that the ASD subjects have more intra-individual fluctuations from one trial to another with respect to the initial gaze at social object, which aligns with the previously reported greater intra-individual variability across behavioral, fMRI, and EEG responses in ASD [33].

5 Modeling Longitudinally Collected Viewing Profiles

5.1 Multilevel Functional Principal Component Analysis

For longitudinally collected functional data, such as viewing profiles from the VE application collected at T1, T2, and T3, MFPCA [14] is utilized to model the correlation in the data through two uncorrelated stochastic processes $Z_{di}(t)$ and $W_{dij}(t)$, leading to the decomposition of the total variation into between- and within-subject variation. Let $Y_{dij}(t)$ denote the viewing profile, given in (2), for subject i , $i = 1, \dots, N_d$, from diagnosis group d , $d = 1, \dots, D$, visit j , $j = 1, \dots, J$, at trial time t , $t \in T$. A two-way functional ANOVA model can be given as

$$Y_{dij}(t) = \mu(t) + \eta_{dj}(t) + Z_{di}(t) + W_{dij}(t) + \epsilon_{dij}(t),$$

where $\mu(t)$ and $\eta_{dj}(t)$ denote the overall mean function and group-visit-specific mean shifts, respectively, $Z_{di}(t)$ and $W_{dij}(t)$ denote the subject-specific deviation from the group-visit-specific mean and the residual subject-visit-specific deviation, respectively, and $\epsilon_{dij}(t)$ denotes the measurement error with mean zero and variance σ_d^2 . The ANOVA decomposition, where $Z_{di}(t)$ and $W_{dij}(t)$ are assumed to be uncorrelated, leads to the decomposition of the total variation

($\Sigma_T^{(d)}(s, t) \equiv \text{cov}\{Y_{dij}(s), Y_{dij}(t)\} = \text{cov}\{Z_{di}(s), Z_{di}(t)\} + \text{cov}\{W_{dij}(s), W_{dij}(t)\} + \sigma_d^2 \delta_{\{s=t\}} \)$)
 into
 ($\Sigma_B^{(d)}(s, t) \equiv \text{cov}\{Y_{dij}(s), Y_{dij}(t)\} = \text{cov}\{Z_{di}(s), Z_{di}(t)\} = \sum_{k=1}^{\infty} \lambda_k^{(d)} \phi_k^{(d)}(s) \phi_k^{(d)}(t) \)$)
 and
 within-subject variation
 ($\Sigma_W^{(d)}(s, t) \equiv \Sigma_T^{(d)}(s, t) - \Sigma_B^{(d)}(s, t) - \sigma_d^2 \delta_{\{s=t\}} = \text{cov}\{W_{dij}(s), W_{dij}(t)\} = \sum_{\ell=1}^{\infty} v_{\ell}^{(d)} \psi_{\ell}^{(d)}(s) \psi_{\ell}^{(d)}(t) \)$).
 In the above formulation $\phi_k^{(d)}(t)$ and $\psi_{\ell}^{(d)}(t)$ denote the subject- and visit-level eigenfunctions for group d , respectively, $\lambda_k^{(d)}$ and $v_{\ell}^{(d)}$ denote the corresponding eigenvalues and $\delta_{\{s=t\}}$ equals one for $s = t$ and zero otherwise.

Next the data are projected onto the subject- and visit-level eigenfunctions, leading to the multilevel Karhunen Loève decomposition,

$$Y_{dij}(t) = \mu(t) + \eta_{dj}(t) + \sum_{k=1}^{\infty} \xi_{dik} \phi_k^{(d)}(t) + \sum_{\ell=1}^{\infty} \zeta_{dij\ell} \psi_{\ell}^{(d)}(t) + \epsilon_{dij}(t), \quad (4)$$

where ξ_{dik} and $\zeta_{dij\ell}$ denote the subject- and visit-level principal component scores (eigenscores) with mean zero and variances $\lambda_k^{(d)}$ and $v_{\ell}^{(d)}$, respectively. Similar to the FPCA decomposition given in Sect. 4.1, the number of eigenvectors included at the subject and visit levels in each diagnostic group in (4) is truncated in practice at K_d and L_d , respectively, using the FVE. In applications to the VE data, we truncate decompositions to include $K_d = L_d = 2$ components explaining more than 90% of the subject- and 70% of the visit-level variation in both groups.

An advantage of MFCA is the ability to compare the variability explained at each level of the data (subject vs. visit) via the functional intra-class correlation coefficient (ICC) defined as the proportion of the total variation in the multilevel signal that is explained at the subject level,

$$\rho_d = \frac{\sum_{k=1}^{K_d} \lambda_k^{(d)}}{\sum_{k=1}^{K_d} \lambda_k^{(d)} + \sum_{\ell=1}^{L_d} v_{\ell}^{(d)}}.$$

A higher functional ICC signals a better agreement among visits within a subject, where more of the total variation is explained by the heterogeneity at the subject level.

MFPCA can be fit using the publicly available R package ‘refund’ [30]. Similar to estimation of FPCA, outlined in Sect. 4.2, the overall mean $\mu(t)$, and group-visit-specific mean shifts $\eta_{dj}(t)$ in (4) are targeted using a spline smoother. Next, the raw total, $\Sigma_T^{(d)}(s, t)$, and between-subject covariances, $\Sigma_B^{(d)}(s, t)$, are targeted using the empirical covariance estimates within the same subject and repetition, $\text{cov}\{Y_{dij}(s), Y_{dij}(t)\}$, and within the same subject across different repetitions, $\text{cov}\{Y_{dij}(s), Y_{dij'}(t)\}$, respectively. The raw within-subject covariance $\Sigma_W^{(d)}(s, t)$ is estimated as the difference between the total and between-subject covariances. The raw between- and within-covariances are smoothed via bivariate penalized spline smoothers, where the degree of smoothness is determined by generalized cross validation, before performing FPCA. Before smoothing the within-subject covariance, the diagonal elements of the raw covariance, which are prone to measurement error effects (similar to the diagonal

of the raw total covariance), are removed. Measurement error variance σ_d^2 is then targeted as the mean difference between the diagonal elements of the smoothed and unsmoothed within-subject covariances. For further details on estimation of the means and covariances in MFPCA, see Di et al. [14].

5.2 Group-Level Inference via Parametric Bootstrap

For longitudinally collected functional data, as in our VE example, multiple hypothesis tests of interest may be formulated about mean trends. The first one considered tests whether the groups have the same average mean trajectory over the longitudinal visits, $H_0: \eta_1(t) = \dots = \eta_D(t) = 0$, where $\eta_d(t) = \sum_{j=1}^J \eta_{dj}(t)/J$ denotes the group average for group d over the visits. Similarly, it might be of interest to test whether all visits have equal means across groups, $H_0: \eta_1(t) = \dots = \eta_J(t) = 0$, where $\eta_j(t) = \sum_{d=1}^D \eta_{dj}(t)/D$ denotes the point-wise visit average across diagnostic groups. Finally, if a group mean difference is detected via the first test, it might be of interest to test whether the detected mean difference between two groups d_1 and d_2 stays the same across visits. Next, bootstrap procedures are outlined for the three null hypotheses considered which are all of interest in our VE example. Analogous to the proposals in Sect. 4.2, the three proposed bootstrap tests all utilize estimated MFPCA components from Sect. 5.1, and target the null distributions of the proposed test statistics using the distribution of the test statistics in the bootstrap samples.

5.2.1 Overall Group Difference

To test the null hypothesis that all diagnosis groups have equal means over the visits, i.e., $H_0: \eta_d(t) = 0$ for $d = 1, \dots, D$, where $\eta_d(t) = \sum_{j=1}^J \eta_{dj}(t)/J$ denotes the point-wise group average over the visits, the proposed bootstrap procedure generates B bootstrap samples under the null utilizing the estimated MFPCA components,

$$Y_{dij}^b(t) = \hat{\mu}(t) + [\{\hat{\eta}_{dj}(t) - \hat{\eta}_d(t)\}] + \sum_{k=1}^{K_d} \xi_{dik}^b \hat{\phi}_k^{(d)}(t) + \sum_{\ell=1}^{L_d} \zeta_{dij\ell}^b \hat{\psi}_\ell^{(d)}(t) + \epsilon_{dij}^b(t),$$

where $\hat{\eta}_d(t) = \sum_{j=1}^J \hat{\eta}_{dj}(t)/J$, $d = 1, \dots, D$, $i = 1, \dots, N_D$, $b = 1, \dots, B$. The subject- and visit-level eigenscores as well as measurement errors are sampled from $\xi_{dik}^b \sim N(0, \hat{\lambda}_k^{(d)})$, $\zeta_{dij\ell}^b \sim N(0, \hat{v}_\ell^{(d)})$ and $\epsilon_{dij}^b(t) \sim N(0, \hat{\sigma}_d^2)$, respectively. The proposed

test statistic, $T_0 = \left[\sum_{d=1}^D \int \{\hat{\eta}_d(t) - 0\}^2 dt \right]^{1/2}$, measuring deviations of the group averages over the visits from zero, is computed for the observed data, denoted by T_0 , and each bootstrap sample, T_0^b , $b = 1, \dots, B$. The p value is estimated by $\sum_{b=1}^B \mathbb{I}(T_0^b > T_0)/B$, where $\mathbb{I}(\cdot)$ denotes the indicator function.

5.2.2 Overall Visit Difference

To test the null hypothesis that all visits have equal means across groups, i.e., $H_0: \eta_j(t) = 0$ for $j = 1, \dots, J$, where $\eta_j(t) = \sum_{d=1}^D \eta_{dj}(t)/D$ represents the point-wise visit

average across diagnostic groups, the proposed bootstrap procedure generates B bootstrap samples under the null,

$$Y_{dij}^b(t) = \hat{\mu}(t) + [\{\hat{\eta}_{dij}(t) - \hat{\eta}_j(t)\}] + \sum_{k=1}^{K_d} \xi_{dik}^b \hat{\phi}_k^{(d)}(t) + \sum_{\ell=1}^{L_d} \zeta_{dij\ell}^b \hat{\psi}_\ell^{(d)}(t) + \epsilon_{dij}^b(t),$$

where $\hat{\eta}_j(t) = \sum_{d=1}^D \hat{\eta}_{dij}(t)/D$, $d = 1, \dots, D$, $i = 1, \dots, N_D$, $b = 1, \dots, B$. The subject- and visit-level eigenscores and measurement error are sampled from $\xi_{dik}^b \sim N(0, \hat{\lambda}_k^{(d)})$, $\zeta_{dij\ell}^b \sim N(0, \hat{v}_\ell^{(d)})$ and $\epsilon_{dij}^b(t) \sim N(0, \hat{\sigma}_d^2)$, respectively. The proposed test statistic,

$T_0 = \left[\sum_{j=1}^J \int \{\hat{\eta}_j(t) - 0\}^2 dt \right]^{1/2}$, based on the norm of the visit average across diagnostic groups from zero, is computed for the observed data (denoted by T_0), and each bootstrap sample, T_0^b , $b = 1, \dots, B$. The p value is estimated by $\sum_{b=1}^B \mathbb{I}(T_0^b > T_0)/B$, where $\mathbb{I}(\cdot)$ denotes the indicator function.

5.2.3 Group Difference Across Visits

If a group mean difference is detected via the first test, a follow-up question may be whether the detected mean difference between two groups d_1 and d_2 stays the same across visits. The null hypothesis takes the form $H_0: \Delta_j^{(d_1, d_2)}(t) = \Delta^{(d_1, d_2)}(t)$ for $j = 1, \dots, J$, where $\Delta_j^{(d_1, d_2)}(t) = \eta_{d_1j}(t) - \eta_{d_2j}(t)$ and $\Delta^{(d_1, d_2)}(t) = \sum_{j=1}^J \Delta_j^{(d_1, d_2)}(t)$ represent the group difference at visit j and the average group difference across visits, respectively. The proposed bootstrap procedure generates B bootstrap samples under the null as,

$$Y_{d_1ij}^b(t) = \hat{\mu}(t) + [\{\hat{\eta}_{d_1ij}(t) + \hat{\eta}_{d_2j}(t)\}/2 + \hat{\Delta}^{(d_1, d_2)}(t)/2] + \sum_{k=1}^{K_{d_1}} \xi_{d_1ik}^b \hat{\phi}_k^{(d_1)}(t) + \sum_{\ell=1}^{L_{d_1}} \zeta_{d_1ij\ell}^b \hat{\psi}_\ell^{(d_1)}(t) + \epsilon_{d_1ij}^b(t),$$

$$Y_{d_2ij}^b(t) = \hat{\mu}(t) + [\{\hat{\eta}_{d_1j}(t) + \hat{\eta}_{d_2j}(t)\}/2 - \hat{\Delta}^{(d_1, d_2)}(t)/2] + \sum_{k=1}^{K_{d_2}} \xi_{d_2ik}^b \hat{\phi}_k^{(d_2)}(t) + \sum_{\ell=1}^{L_{d_2}} \zeta_{d_2ij\ell}^b \hat{\psi}_\ell^{(d_2)}(t) + \epsilon_{d_2ij}^b(t),$$

$$Y_{dij}^b(t) = \hat{\mu}(t) + \hat{\eta}_{dij}(t) + \sum_{k=1}^{K_d} \xi_{dik}^b \hat{\phi}_k^{(d)}(t) + \sum_{\ell=1}^{L_d} \zeta_{dij\ell}^b \hat{\psi}_\ell^{(d)}(t) + \epsilon_{dij}^b(t), \text{ for } d \neq d_1, d_2,$$

where $\hat{\Delta}_j^{(d_1, d_2)}(t) = \hat{\eta}_{d_1j}(t) - \hat{\eta}_{d_2j}(t)$ and $\hat{\Delta}^{(d_1, d_2)}(t) = \sum_{j=1}^J \hat{\Delta}_j^{(d_1, d_2)}(t)/J$, $d = 1, \dots, D$, $i = 1, \dots, N_D$, $j = 1, \dots, J$, $b = 1, \dots, B$. The subject- and visit-level eigenscores and measurement error are sampled from $\xi_{dik}^b \sim N(0, \hat{\lambda}_k^{(d)})$, $\zeta_{dij\ell}^b \sim N(0, \hat{v}_\ell^{(d)})$ and $\epsilon_{dij}^b(t) \sim N(0, \hat{\sigma}_d^2)$, respectively. The proposed test statistic

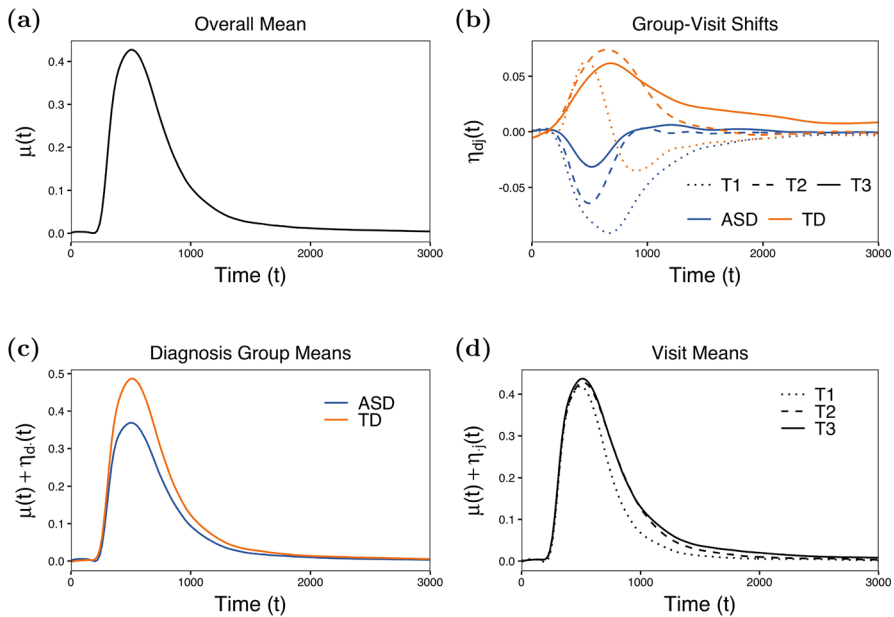


Fig. 6 **a** Estimated overall mean viewing profile, $\mu(t)$. **b** Estimated group-visit shifts from the overall mean, $\eta_{dj}(t)$, $d = 1, 2$, $j = 1, 2, 3$. **c** Estimated mean viewing profiles, $\mu(t) + \eta_d(t)$, $d = 1, 2$, for the two diagnosis groups (ASD and TD). **d** Estimated mean viewing profiles, $\mu(t) + \eta_j(t)$, $j = 1, 2, 3$, at the three visits (T1, T2 and T3) in the VE application

$$T_0 = \left[\sum_{d=d_1, d_2} \sum_{j=1}^J \int \{ \hat{\Delta}_j^{(d_1, d_2)}(t) - \hat{\Delta}^{(d_1, d_2)}(t) \}^2 dt \right]^{1/2}$$

assessing deviations of the visit-specific group differences between groups d_1 and d_2 from the average group difference across visits is computed for the observed data (denoted by T_0), and each bootstrap sample, T_0^b , $b = 1, \dots, B$. The p value is estimated by $\sum_{b=1}^B \mathbb{I}(T_0^b > T_0)/B$, where $\mathbb{I}(\cdot)$ denotes the indicator function.

5.3 Application to Visual Exploration Paradigm

MFPCA is used to decompose the longitudinally observed viewing profiles from the VE paradigm. The estimated MFPCA components are given in Figs. 6, 7 and 8. Similar to the results of FPCA analysis presented in Sect. 4.3, the estimated overall mean function, given in Fig. 6a, also identifies the initial gaze at the social object around 500 ms as a common behavior across subjects, visits and trials, with 43% consistency. In addition, the estimated group-visit mean shifts $\eta_{dj}(t)$, $d = 1, 2$, $j = 1, 2, 3$, given in Fig. 6b, as well as group- and visit-specific mean estimates given in Fig. 6c and d show that even though both groups display the highest consistency of the target behavior at 500 ms, the TD group has higher consistency ($\sim 50\%$ of trials) than

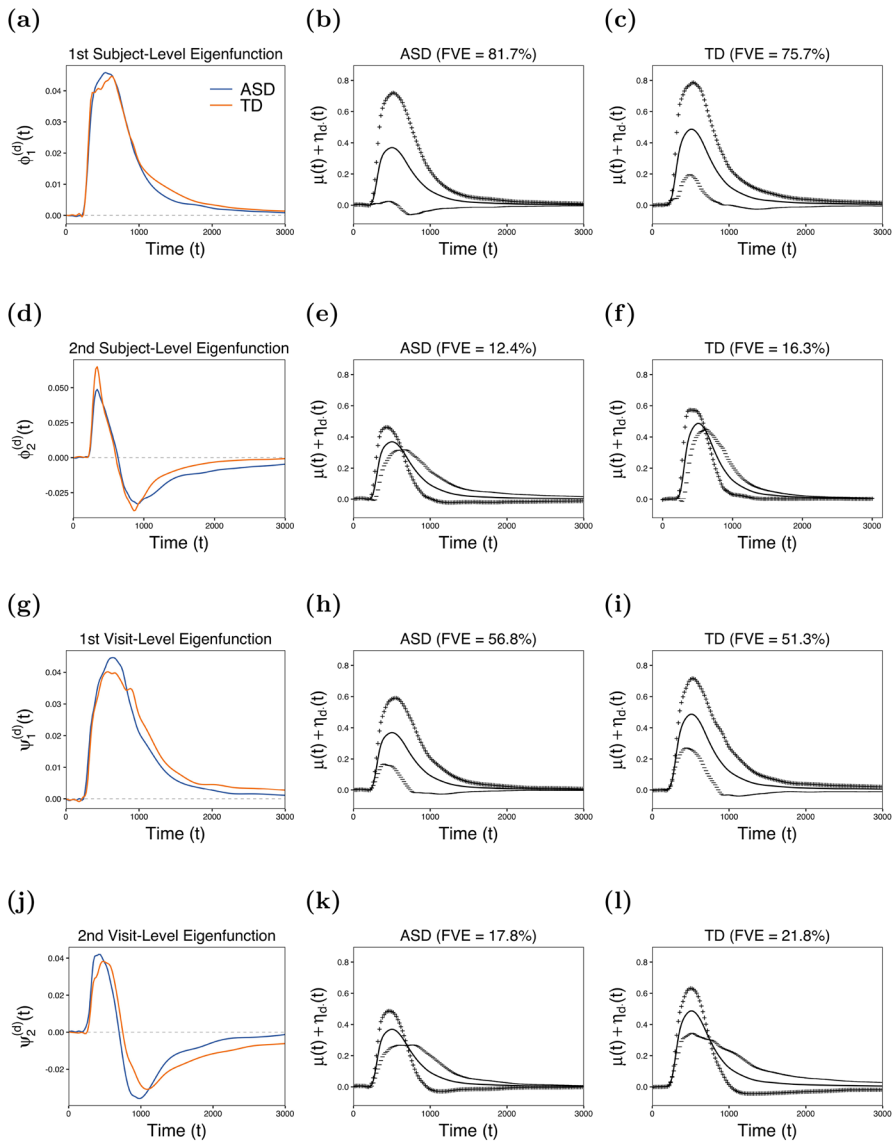


Fig. 7 **a, d, g, h** The first two leading subject- and visit-level eigenfunctions, $\phi_k(t), k = 1, 2$ (**a, d**), $\psi_k(t), k = 1, 2$ (**g, h**) estimated from MFPCA. **b, c, e, f** Group-specific mean estimates plus and minus two times the amount of variation along the leading two subject-level eigenfunctions, $\mu(t) + \eta_d(t) \pm 2\sqrt{\lambda_1}\phi_1(t)$ (**b, c**), and $\mu(t) + \eta_d(t) \pm 2\sqrt{\lambda_2}\phi_2(t)$ (**e, f**). **h, i, k, l** Group-specific mean estimates plus and minus two times the amount of variation along the leading two visit-level eigenfunctions, $\mu(t) + \eta_d(t) \pm 2\sqrt{v_1}\psi_1(t)$ (**h, i**), and $\mu(t) + \eta_d(t) \pm 2\sqrt{v_2}\psi_2(t)$ (**k, l**)

the ASD group ($\sim 35\%$ of trials) and compared to T1, subjects display longer initial gaze at the social object at subsequent visits (T2 and T3). Subjects, especially in the ASD group, also seem to display higher consistency of the target behavior at T2 and

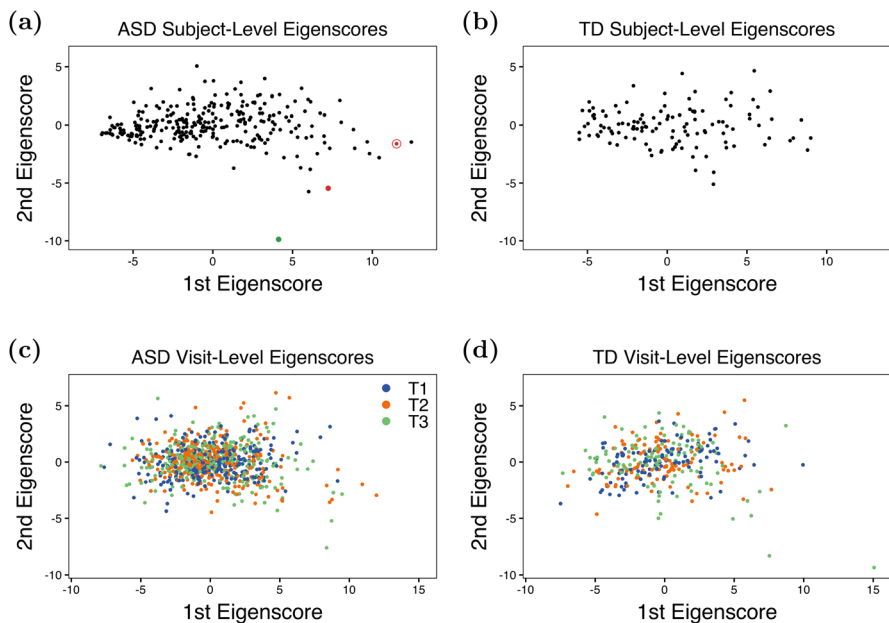


Fig. 8 **a** Estimated first two leading subject-level eigenscores for the ASD group. Three ASD subjects with either the first or the second leading eigenscore outside 2 standard deviations from the mean, also identified in the baseline analysis, are highlighted in green and red. **b** Estimated first two leading subject-level eigenscores for the TD group. **c, d** Estimated first two leading visit-level eigenscores for the ASD and TD groups, respectively (Color figure online)

T3, compared to T1. The bootstrap tests outlined in Sect. 5.2 facilitate formal inference on the estimated mean differences as will be reported below.

The functional ICC, ρ_d , is estimated to be 0.58 and 0.50 in the ASD and TD groups, respectively, suggesting that the total variation is split almost equally between within- and between-subject variation. In other words, there is approximately equal variation across subjects as there is across visits within subjects in both groups. The leading two eigencomponents are retained in decompositions at the subject level, based on the scree plots given in Figs. S1(c) and S1(d). Figure 7a and d display the estimated first two leading subject-level eigenfunctions which explain 81.7% and 12.4% of the between-subject variation in the ASD and 75.7% and 16.3% of the between-subject variation in the TD groups, respectively. The estimated eigenfunctions signal similar modes of variation across the two diagnostic groups. The first leading subject-level eigenfunctions capture the variation in consistency of the initial gaze to the social object across trials, while the second leading subject-level eigenfunctions signal variation in the duration of the initial gaze, similar to findings of Sect. 4.3. This can also be seen in Fig. 7b, c, e, f, which display group- and visit-specific mean trajectories plus and minus two times the amount of variation along each eigenfunction. While a positive leading subject-level eigenscore suggests higher consistency of the initial gaze at the social object across trials, a

negative second leading subject-level eigenscore suggests a longer initial gaze at the social object.

At the visit level, the leading three eigencomponents are retained in the decompositions using the scree-plots (Fig. S1e and f). Figure 7g and j display the first two leading eigenfunctions which explain 56.8% and 17.8% of the within-subject variation in the ASD and 51.3% and 21.8% of the within-subject variation in the TD group, respectively. While the first visit-level eigenfunctions signal variability in both consistency and duration of the initial gaze at the social object, the second visit-level eigenfunction signals mostly variability at the duration, with positive first visit-level eigenscores indicating both a higher consistency and longer gaze (Fig. 7h, i) and a negative second visit-level eigenscores indicating longer gaze (Fig. 7k, l). The third leading visit-level eigenfunctions, explaining 7.9% and 10% of within-subject variation in ASD and TD group, respectively, signal variability in location of the initial gaze at the social object in trial time, with positive third visit-level eigenscores indicating an earlier gaze (Fig. S4). Note that in summary, while most of the variation across subjects is explained by differences in consistency of the initial gaze at the social object, variation within subjects across visits is explained by both consistency and initial gaze duration differences. Figure 8a and b display the two leading subject-level eigenscores in both groups, where similar to analysis with FPCA, we observe that subjects with negative first leading subject-level eigenscores tend to vary less in their second leading subject-level eigenscores, suggesting that subjects with low consistency of the initial gaze at the social object have less variation in their length of gaze, as expected. The three ASD subjects highlighted in green and red for having high consistency or longer duration of the initial gaze at the social object in Fig. 5a are also highlighted in Fig. 8a. While the two subjects highlighted in red have higher consistency, the subject in green is still identified as having the longest gaze duration in the initial gaze at the social object. Figure 8c and d displays the two leading visit-level eigenscores in both groups, where the variation is similar across the two groups and the three visits.

Finally, utilizing the estimated MFPCA components and the parametric bootstrap tests outlined in Sect. 5.2, the diagnosis group mean differences observed in Fig. 6c are found to be significant (p value < 0.001) analogous to the results reported in Sect. 4.3, where TD subjects are significantly more consistent in their initial gaze at the social object across trials. The longer initial gaze duration at the social object in the latter visits, observed in Fig. 6d, is also found to be significant (p value < 0.001 for overall visit difference). Whether the detected group mean difference stays the same across the three visits is also tested using the third bootstrap test for group difference across visits. The null hypothesis that the group difference stays the same across the visits is rejected (p value = 0.035), implying that the diagnosis group differences vary across three visits. Although the TD group displays higher consistency of the initial gaze at the social object at all three visits, the consistency difference between group seems to be smaller at the latter visits (T2, T3) (Fig. 6b), especially at T3. (Note that the mean latency group difference was also found to be slightly smaller at T3 using the LMMs of Sect. 2.2).

6 Discussion

Traditional analysis of ET data collapse information across trial time through the use of data summaries such as total ROI viewed, average looking time at or latency to ROI. Viewing profiles are proposed to retain common trends in gaze across trial time in ET experiments. The constructed functional outcomes either obtained from one visit (cross-sectional) or collected at multiple visits (collected longitudinally) are decomposed via FPCA. FPCA decomposes the total variation in the data into between-subject and within-subject variation, where dominant modes of variation (i.e. eigenfunctions) are identified for between- and within-subject variation. Projection of the data onto the dominant modes of variation yields representation of the high-dimensional functional data in lower dimensions through eigenscores. This lower dimensional representation not only allows for easy interpretation and comparison of mean and variation trends in the data (possibly across different diagnostic groups), but it also facilitates formal inference via bootstrap. Subject-specific eigenscores obtained from FPCA can further be used in assessing associations between

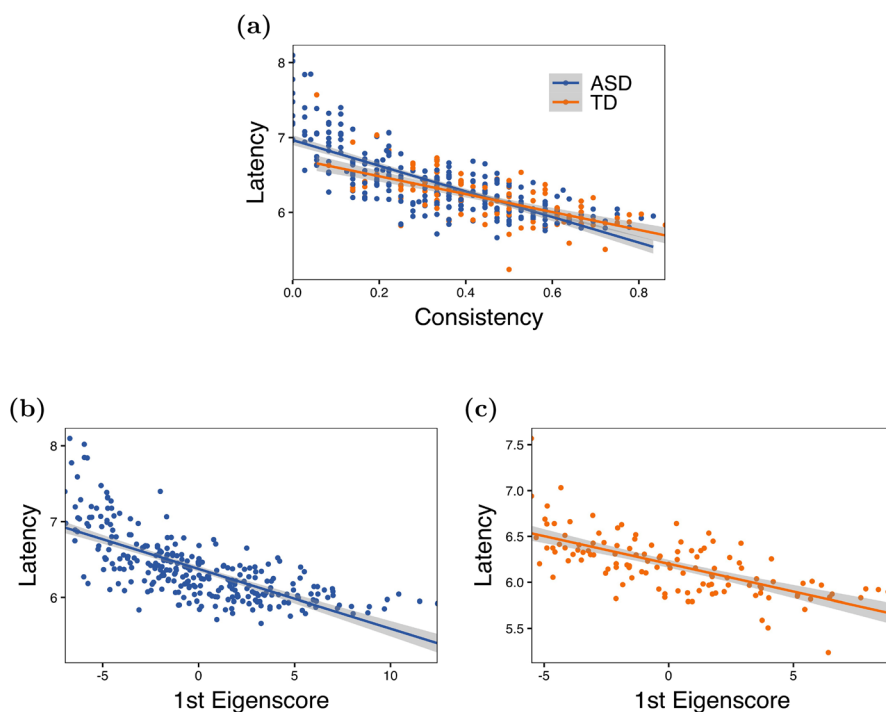


Fig. 9 **a** Significant negative correlations are detected between average consistency of the initial gaze being at the social object and the average logarithm of latency to social object in the entire data ($r = -0.758$, p value < 0.001) as well as the ASD ($r = -0.768$, p value < 0.001) and TD groups ($r = -0.678$, p value < 0.001) **(b), (c)** Significant negative correlations are also detected between the leading subject-level eigenscores (positive scores signaling higher consistency) in both ASD and TD groups (ASD: $r = -0.706$, p value < 0.001 , TD: $r = -0.678$, p value < 0.001)

the uncovered trends in ET data with behavioral measures. The proposed approach, to the best of our knowledge, is the first functional model proposed in literature for the analysis of ET data.

In applications to VE data, the proposed functional approach leads to new insights on trends across trial time that are lost in traditional analysis of the data. More specifically, an initial gaze at the social object by 500 ms is identified as a common trend in gaze across trial time in both diagnostic groups. In addition, there are significant group differences detected in consistency of the initial gaze detected across trials, where the ASD group displays significantly lower consistency compared to the TD group. The consistency difference across groups explains the average difference observed for the latency to social object detected between groups through traditional analysis. Figure 9a displays significant negative correlations between average consistency of the initial gaze at the social object (i.e., average viewing profile values at 500 ms across trials and visits) and the average logarithm of latency to social object in the entire data ($r = -0.758$, p value < 0.001) as well as the ASD ($r = -0.768$, p value < 0.001) and TD groups ($r = -0.678$, p value < 0.001), signaling the association between lower consistency and higher than average latency to social object. Significant negative correlations are also observed between the leading subject-level eigenscores (positive scores signaling higher consistency) in both ASD and TD groups (Fig. 9b and c). Figures S5–S8 display viewing profiles for the target behaviors of the initial gaze being at the social object (outcome of main analysis) and initial social gaze for ASD subjects with the lowest and the highest average consistency. While the subjects with high consistency display their initial social gaze earlier on in trial time, those with lower consistency of the target behavior don't seem to display a prioritization of the gaze at the social object in trial time (gaze times at the social object distributed uniformly across trial time). In summary, the ASD group displays an initial gaze being at the social object in a lower percentage of trials, where in more than half of the trials ASD subjects gaze at the social object later in trial time (if at all), increasing average latency to social object across trials in the ASD group.

The findings portray a larger inter-trial variability of the target behavior in the ASD group, which connects to the wider emphasis in ASD research in recent years highlighting the importance of the study of intra-subject variability. Recent literature suggests that instead of a general failure to respond to certain tasks, there is greater intra-subject trial to trial variability across behavioral, fMRI, and EEG responses among ASD participants [33]. Intra-subject variability (across trials) is typically ignored in ASD studies, which instead focus on mean trends. Capturing such inter-trial variability in eye tracking paradigms contributes to the increasing evidence that suggests response variability as a possible endophenotype of the disorder.

In addition, the two leading subject-level eigenscores, where higher scores signal higher consistency and shorter gaze duration at the social object, respectively, have also been correlated with demographic and behavioral measures, including age, FSIQ, and scaled NEPSY score in both diagnostic groups and ADOS (only in ASD group). The leading subject-level eigenscore is significantly correlated with the scaled NEPSY score for face memory (ASD: $r = 0.211$, p value < 0.001 , TD: $r = 0.310$, p value < 0.001), where a better face memory is associated with a higher consistency of

the initial gaze at the social object (Fig. S9). The second leading eigenscore, signaling gaze duration at the social object is found to be significantly correlated with age in both diagnosis groups (ASD: $r = 0.149$, p value = 0.013, TD: $r = 0.265$, p value = 0.004), where older children tend to have shorter initial gaze duration at the social object (see Fig. S10c and d). This result is consistent with results in the traditional LMM approach controlling for age as well as findings in previous literature where fixation duration has been shown to decrease with age in children during reading or viewing tasks [34, 35]. No significant correlations with the leading subject-level eigenscores and FSIQ and ADOS are detected. The detected significant associations with behavioral measures further point to the fact that functional approach proposed leads to interpretable data summaries through FPCA.

While motivated by the VE application, the construction of viewing profiles can be useful in other ET paradigms as well, where there is interest in studying common trends across trial time. Consider the aforementioned visual world paradigm (VWP) by Allopenna et al. [5], for example, where participants hear an instruction to click on an object (e.g., beaker), while presented with several objects including the target (beaker), a competitor with similar pronunciation (beetle), and other irrelevant distractors (speaker and carriage). A common trend across trial time is an eye movement to beetle when only the first two segments have been heard, followed by a gaze at beaker 100 ms later. Two viewing profiles, with target behaviors of gaze at target and gaze at competitor, can be created in this example to portray trends across trial time. Consistency of looking at the target and competitor can then be compared throughout trial time, with the use of the two viewing profiles. In more complex experimental settings involving viewing of non-static stimuli, such as dynamic social videos presenting social interaction scenes, viewing profiles can be constructed corresponding to the target behavior of gaze at certain ROIs such as social interactions or faces.

Note that the smoothness level of the proposed functional outcomes depends on the number of trials averaged, as averaging of more trials will lead to smoother viewing profiles. A presmoothing step may be included in applications where viewing profiles are obtained using a lower number of trials. Another extension of the proposed methodology could consider incorporation of a curve registration step before the FPCA decompositions. By implementing suitable transformations on time, curve registration manages to align salient common features in a group of functional observations and analyze the temporal variation of these features [36–38]. In the VE application, data suggest that subjects initiate the initial gaze at the social object at about the same trial time and we do not observe much variation in the location of this behavior that would require a separate registration step. However, a preregistration step before the FPCA decomposition in other ET applications may be useful where there may be significant location variation in dominant features. Finally, covariate adjustments can be included in FPCA modeling similar to LMM analysis. While adjustments to the mean involve smoothing across the covariate domain, adjustments to the covariance can be made by allowing both the eigenvalues and eigenfunctions to change as a function of the covariate observed, or by assuming the eigenfunctions to be constant across the covariate dimension but allowing the corresponding eigenvalues and principal component scores to be covariate-dependent

[39, 40]. For covariate adjustments in multi-dimensional FPCA, we refer readers to Scheffler et al. [41].

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12561-022-09354-6>.

Acknowledgements The authors would like to thank the associate editor and two anonymous referees who provided useful and detailed comments on a previous version of the manuscript.

Funding This work was supported by National Institute of Mental Health [R01 MH122428 (DS, DT) and U19 MH108206 (McPartland)].

Data Availability The ABC-CT data are publicly available at NIMH NDA (#2288) (https://nda.nih.gov/edit_collection.html?id=2288).

Declarations

Conflict of interest McPartland consults with Customer Value Partners, Bridgebio, Determined Health, and BlackThorn Therapeutics, has received research funding from Janssen Research and Development, serves on the Scientific Advisory Boards of Pastorius and Modern Clinics, and receives royalties from Guilford Press, Lambert, and Springer. Dawson is on the Scientific Advisory Boards of Janssen Research and Development, Akili Interactive, Inc, LabCorp, Inc, Roche Pharmaceutical Company, and Tris Pharma, and is a consultant to Apple, Gerson Lehrman Group, Guidepoint Global, Inc, and is CEO of DASIO, LLC. Dawson has stock interests in Neuvana, Inc. Dawson has the following patent No. 10,912,801 and patent applications: 62,757,234, 25,628,402, and 62,757,226. Dawson has developed technology, data, and/or products that have been licensed to Apple, Inc. and Cryocell, Inc. and Dawson and Duke University have benefited financially. Shic consults for Roche Pharmaceutical Company, Janssen Research and Development, BlackThorn Therapeutics, and BioStream Technologies.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. van Renswoude DR, Raijmakers ME, Koornneef A, Johnson SP, Hunnius S, Visser I (2018) Gaze-path: an eye-tracking analysis tool that accounts for individual differences and data quality. *Behav Res Methods* 50(2):834–852. <https://doi.org/10.3758/s13428-017-0909-3>
2. Shultz S, Klin A, Jones W (2018) Neonatal transitions in social behavior and their implications for autism. *Trends Cogn Sci* 22(5):452–469. <https://doi.org/10.1016/j.tics.2018.02.012>
3. Nayar K, Shic F, Winston M, Losh M (2021) A constellation of eye-tracking measures reveals social attention differences in ASD and the broad autism phenotype. Under review
4. McPartland JC, Bernier RA, Jeste SS, Dawson G, Nelson CA, Chawarska K et al (2020) The Autism Biomarkers Consortium for Clinical Trials (ABC-CT): scientific context, study design, and progress toward biomarker qualification. *Front Integr Neurosci* 14:16. <https://doi.org/10.1101/2019.12.18.19014548>
5. Allopenna PD, Magnuson JS, Tanenhaus MK (1998) Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *J Mem Lang* 38(4):419–439. <https://doi.org/10.1006/jmla.1997.2558>

6. Ramsay JO, Silverman BW (2005) Functional data analysis. Springer, New York
7. Horváth L, Kokoszka P (2012) Inference for functional data with applications. Springer, New York
8. Müller HG (2005) Functional modeling and classification of longitudinal data. *Scand J Stat* 32(2):223–240. <https://doi.org/10.1111/j.1467-9469.2005.00429.x>
9. James GM, Hastie TJ, Sugar CA (2000) Principal component models for sparse functional data. *Biometrika* 87(3):587–602. <https://doi.org/10.1093/biomet/87.3.587>
10. Rice JA (2004) Functional and longitudinal data analysis: perspectives on smoothing. *Stat Sin* 14(3):631–647
11. Şentürk D, Müller HG (2010) Functional varying coefficient models for longitudinal data. *J Am Stat Assoc* 105(491):1256–1264. <https://doi.org/10.1198/jasa.2010.tm09228>
12. Shi M, Weiss RE, Taylor JM (1996) An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *J R Stat Soc Ser C* 45(2):151–163. <https://doi.org/10.2307/2986151>
13. Crainiceanu CM, Staicu AM, Di CZ (2009) Generalized multilevel functional regression. *J Am Stat Assoc* 104(488):1550–1561. <https://doi.org/10.1198/jasa.2009.tm08564>
14. Di CZ, Crainiceanu CM, Caffo BS, Punjabi NM (2009) Multilevel functional principal component analysis. *Ann Appl Stat* 3(1):458–488. <https://doi.org/10.1214/08-AOAS206>
15. Kundu MG, Harezlak J, Randolph TW (2016) Longitudinal functional models with structured penalties. *Stat Model* 16(2):114–139. <https://doi.org/10.1177/1471082X15626291>
16. Morris JS, Carroll RJ (2006) Wavelet-based functional mixed models. *J R Stat Soc Ser B* 68(2):179–199. <https://doi.org/10.1111/j.1467-9868.2006.00539.x>
17. Zippunnikov V, Caffo B, Yousem DM, Davatzikos C, Schwartz BS, Crainiceanu C (2011) Multilevel functional principal component analysis for high-dimensional data. *J Comput Graph Stat* 20(4):852–873. <https://doi.org/10.1198/jcgs.2011.10122>
18. Greven S, Crainiceanu C, Caffo B, Reich D (2011) Longitudinal functional principal component analysis. In: Recent advances in functional data analysis and related topics. Springer, pp 149–154
19. Chen K, Delicado Useros PF, Müller HG (2017) Modeling function-valued stochastic processes, with applications to fertility dynamics. *J R Stat Soc Ser B* 79(1):177–196. <https://doi.org/10.1111/rssb.12160>
20. Chen K, Müller HG (2012) Modeling repeated functional observations. *J Am Stat Assoc* 107(500):1599–1609. <https://doi.org/10.1080/01621459.2012.734196>
21. Park SY, Staicu AM (2015) Longitudinal functional data analysis. *Stat* 4(1):212–226. <https://doi.org/10.1002/sta4.89>
22. Mirman D, Dixon JA, Magnuson JS (2008) Statistical and computational models of the visual world paradigm: growth curves and individual differences. *J Mem Lang* 59(4):475–494. <https://doi.org/10.1016/j.jml.2007.11.006>
23. Del Bianco T, Mason L, Charman T, Tillman J, Loth E, Hayward H et al (2021) Temporal profiles of social attention are different across development in autistic and neurotypical people. *Biol Psychiatry* 6(8):813–824. <https://doi.org/10.1016/j.bpsc.2020.09.004>
24. Gliga T, Elsabbagh M, Andravizou A, Johnson M (2009) Faces attract infants' attention in complex displays. *Infancy* 14(5):550–562. <https://doi.org/10.1080/15250000903144199>
25. Tsang T, Bernier R, Dawson G, Jeste S, McPartland J, Nelson C, et al (2021) Attention allocation during exploration of visual arrays in ASD: results from the ABC-CT feasibility study. Under review
26. Xie M (2021) Attention allocation during visual exploration tasks in school-age children with ASD: Autism Biomarkers Consortium for Clinical Trials. Abstract presented at International Society for Autism Research Annual Meeting
27. Karhunen K (1946) Zur Spektraltheorie stochastischer prozesse. *Ann Acad Sci Fennicae AI* 1–37
28. Loeve M (1946) Fonctions aléatoires à décomposition orthogonale exponentielle. *La Revue Scientifique*, pp 159–162
29. Wang JL, Chiou JM, Müller HG (2016) Functional data analysis. *Annu Rev Stat Appl* 3:257–295. <https://doi.org/10.1146/annurev-statistics-041715-033624>
30. Goldsmith J, Scheipl F, Huang L, Wrobel J, Di C, Gellar J, et al (2021) refund: regression with Functional Data. R package version 0.1-24. <https://CRAN.R-project.org/package=refund>
31. Yao F, Müller HG, Wang JL (2005) Functional data analysis for sparse longitudinal data. *J Am Stat Assoc* 100(470):577–590. <https://doi.org/10.1198/016214504000001745>
32. Goldsmith J, Greven S, Crainiceanu C (2013) Corrected confidence bands for functional data using principal components. *Biometrics* 69(1):41–51. <https://doi.org/10.1111/j.1541-0420.2012.01808.x>

33. David N, Schneider TR, Peiker I, Al-Jawahiri R, Engel AK, Milne E (2016) Variability of cortical oscillation patterns: a possible endophenotype in autism spectrum disorders. *Neurosci Biobehav Rev* 71:590–600. <https://doi.org/10.1016/j.neubiorev.2016.09.031>
34. Seassau M, Bucci MP (2013) Reading and visual search: a developmental study in normal children. *PLoS ONE* 8(7):e70261. <https://doi.org/10.1371/journal.pone.0070261>
35. Helo A, Pannasch S, Sirri L, Rämä P (2014) The maturation of eye movement behavior: scene viewing characteristics in children and adults. *Vis Res* 103:83–91. <https://doi.org/10.1016/j.visres.2014.08.006>
36. Wang K, Gasser T (1997) Alignment of curves by dynamic time warping. *Ann Stat* 25(3):1251–1276
37. Ramsay JO, Li X (1998) Curve registration. *J R Stat Soc Ser B* 60(2):351–363
38. Liu X, Müller HG (2004) Functional convex averaging and synchronization for time-warped random curves. *J Am Stat Assoc* 99(467):687–699
39. Cardot H (2007) Conditional functional principal components analysis. *Scand J Stat* 34(2):317–335
40. Jiang CR, Wang JL (2009) Covariate adjusted functional principal component analysis for longitudinal data. *Ann Stat* 38:1194–1226. <https://doi.org/10.1214/09-AOS742>
41. Scheffler AW, Dickinson A, DiStefano C, Jeste S, Şentürk D (2021) Covariate-adjusted hybrid principal components analysis for region-referenced functional EEG data. *Stat Interface* (in press)

Authors and Affiliations

Mingfei Dong¹ · Donatello Telesca¹ · Catherine Sugar^{1,2} · Frederick Shic^{3,4} · Adam Naples⁵ · Scott P. Johnson⁶ · Beibin Li^{3,7} · Adham Atyabi^{3,8} · Minhang Xie³ · Sara J. Webb^{3,9} · Shafali Jeste¹⁰ · Susan Faja¹¹ · April R. Levin¹² · Geraldine Dawson¹³ · James C. McPartland⁵ · Damla Şentürk¹ · the Autism Biomarkers Consortium for Clinical Trials

¹ Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA, USA

² Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, Los Angeles, CA, USA

³ Center for Child Health, Behavior and Development, Seattle Children's Research Institute, Seattle, WA, USA

⁴ Department of Pediatrics, School of Medicine, University of Washington, Seattle, WA, USA

⁵ Child Study Center, School of Medicine, Yale University, New Haven, CT, USA

⁶ Department of Psychology, University of California, Los Angeles, CA, USA

⁷ Department of Computer Science and Engineering, University of Washington, Seattle, WA, USA

⁸ Department of Computer Science, University of Colorado, Colorado Springs, CO, USA

⁹ Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA

¹⁰ Children's Hospital Los Angeles, Keck School of Medicine, University of South California, Los Angeles, CA, USA

¹¹ Laboratory of Cognitive Neuroscience, Division of Developmental Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

¹² Department of Neurology, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA

¹³ Department of Psychiatry and Behavioral Sciences, Duke University, Durham, NC, USA