Evaluating fMRI-Based Estimation of Eye Movements during Naturalistic Viewing

Jake Son[1,2], Lei Ai[1], Ting Xu[1], Stan Colcombe[3], Stephen LaConte[4], Jonathan Lisinski[4], Arno Klein[1,2], Cameron Craddock[1,3,5], Michael Milham[1,3]

[1]Center for the Developing Brain, Child Mind Institute, New York, New York
[2]MATTER Lab, Child Mind Institute, New York, New York
[3]Center for Biomedical Imaging and Neuromodulation, Nathan S. Kline Institute for Psychiatric Research, New York, New York
[4]Viriginia Tech Carilion Research Institute, Blacksburg, Virginia
[5]Department of Diagnostic Medicine, Dell Medical School, Austin, Texas

Corresponding Author:
Michael Peter Milham, MD, PhD
Phyllis Green and Randolph Cowen Scholar
Child Mind Institute
445 Park Avenue
New York, NY 10022
Michael.Milham@childmind.org

ABSTRACT

The collection of eye movement information during functional magnetic resonance imaging (fMRI) is a valuable, though commonly overlooked component of monitoring variations in attention and task compliance, particularly for naturalistic viewing paradigms (e.g., movies). Predictive eye estimation regression (PEER) is a previously developed support vector regression-based method for retrospectively estimating eye gaze from fMRI data that simply adds a 1.5-minute calibration scan to any protocol. Here, we provide a large-scale assessment of PEER for inferring eye fixations on a TR-by-TR basis during movie viewing using a subset of data (n=448) from the Child Mind Institute Healthy Brain Network Biobank. Consistent with prior work, we demonstrate the ability of PEER to provide accurate estimates of fixation location throughout the course of fMRI scans and we establish head motion as the primary determinant of model accuracy. Minimum data requirement analyses suggest model estimation can be carried out with less than half the data obtained in the 1.5-minute calibration scan. We demonstrate the ability to predict the movie an individual is watching (i.e., *Despicable Me*, *The Present*) based on the PEER time series. Out-of-scanner eye tracker-based measurements obtained during a repeat viewing of the movie *The Present* was used to further validate the time series obtained using PEER. Consistent with prior findings in the eye tracking literature, the fixation sequences showed a high consistency across participants, reducing the ability to identify an individual based on their fixation sequence. Finally, examination of neural activations associated with the PEER time series replicated prior findings regarding the neural correlates of eye movements. In summary, we demonstrate that PEER is an inexpensive, easy-to-use tool for researchers to determine eye fixations from naturalistic viewing data that overcomes the cost and burdens of in-scanner eye tracking.

INTRODUCTION

A classic challenge for functional magnetic resonance imaging (fMRI) is the identification of variations in attention and compliance with task demands during a given scan[1–4]. This is particularly true when participants are required to perform tasks that are tedious or tiresome (e.g., rest). In studies requiring active task performance during fMRI scans, these concerns can be addressed in part by analyzing performance measures (e.g., accuracy, reaction time)[5]. However, such solutions do not work for studies of passive states (e.g., resting state fMRI, naturalistic viewing fMRI), as there are no such responses to monitor. The resting state fMRI literature has struggled with this issue due to the increased likelihood of individuals falling asleep[6,7], driving most to require participants to remain with their eyes open during scans - a requirement that can be tracked through direct observation by a technician or via video recording. The requirements of conducting fMRI movie viewing experiments are greater, as one needs to know that an individual is actually paying attention to specific information on the screen.

Eye tracking devices provide an obvious solution to the additional challenges inherent to naturalistic viewing fMRI, as they can provide high fidelity fixation information from which the level of participant engagement and preferential viewing patterns can be readily inferred[4,8]. A growing number of systems are available for use in the MRI environment, meeting ferromagnetic requirements and overcoming logistical requirements regarding their positioning (e.g., long-range mounts)[9–11]. However, these devices remain outside the range of access for many due to either their costs or the added layers of complexity (e.g. operator training, synchronization and analysis of an additional data type) that can be dissuasive.

An alternative solution is predictive eye estimation regression (PEER)[12–15], an imaging-based method that uses machine learning algorithms (i.e., support vector regression [SVR][16]) to estimate the direction of gaze during each repetition (TR) in the fMRI time series based on voxel-wise data from the eyes. The feasibility of this approach is demonstrated in previous studies that used support vector regression to successfully predict fixation locations on a TR-by-TR basis. Prior works have validated this approach by comparing PEER-derived fixations with simultaneously acquired eye tracking[13], as well as with perimetry[15]. However, PEER has only been used in a few studies to date, which were small in size and tended to focus on neurologically normal adults[12–15].

Here, we provide a large-scale assessment of PEER for inferring eye fixations on a TR-by-TR basis during movie viewing using a subset of data (n=448) from the Child Mind Institute (CMI) Healthy Brain Network Biobank[17]. For each participant, 2-3 PEER scans were available, allowing us to address questions regarding the reproducibility of model accuracy, as well as to assess minimum data requirements. The inclusion of children of varying ages (5-21 years old) and ability to remain still during scan sessions allowed us to assess potential sources of artifact related to compliance[17–22]. In addition to PEER scans, two movie fMRI scans were available for most participants (*Despicable Me*, *The Present*), allowing us to look at both the consistency of findings across participants and the specificity of eye tracking patterns obtained for differing

movie stimuli. For one of the two movies, out-of-scanner eye-tracking device data were available for a subset of individuals, allowing us to assess comparability to gold standard results. This demonstration is a step toward deploying new methods of monitoring variations in task compliance that overcome the burdens of MRI compatible eye-tracking devices.

RESULTS

Model accuracy and data requirements. Consistent with prior implementations of the PEER method, support vector regression (C = 100; epsilon = 0.01) was used to generate a predictive model for eye fixation location based on the EPI signal measured in the eye orbit at each time point. Our primary assessment of the PEER method's predictive accuracy used the data from the first PEER scan (Scan1) in each participant's imaging session to train their SVR model (PEER1) and the second PEER scan to test model accuracy. We assessed the fit of the predicted fixation time series with the stimulus location time series using Pearson's correlation coefficient, which we found to be more robust to individual outlier predictions in the model-generated fixation time series than alternative measures (e.g., Euclidean distance - see supplementary figure 1). Accuracy scores ranged from .35 – .97 (.66 ± .31) and .26 – .90 (.58 ± .32) in the x- and y- directions (see Figure 1a). For our sample of 448 participants, a Pearson's r value of 0.3 or greater achieved statistical significance in a one-directional test (alpha = .05) after taking into account the number of simultaneous comparisons (i.e., the # of participants). Overall, this analysis confirmed that the EPI signal from the orbit contains enough information to reliably predict fixations from a calibration sequence. As a point of comparison with the literature focused on adults only, accuracy scores from prior works were .65 – .85 and .78 – .92 in the x- and y- directions[12,13].

Given the potential impact of changes in head position over the course of an imaging session, we tested for differences in prediction accuracy if the third scan were used for training rather than the first scan, as well as if the third scan were combined with the first scan for training. We compared the predictive accuracies of the PEER1, PEER3, and PEER1&3 models when estimating eye fixations from the second PEER scan. Paired t-tests found significantly high accuracy for PEER1 relative to the other models ($p < 0.01$ in all tests), except in the y-direction when compared against PEER1&3. These differences were not dramatic (see Figure 1b) but may be explained by increased head motion, since paired t-tests showed that mean framewise displacement (FD) was significantly greater for the third PEER scan than the first ($p < 0.01$). We used the PEER1 model to estimate eye fixations for the remainder of the work.

Impact of Head Motion and Age on Model Accuracy. We next considered the impact of factors that we thought may compromise model accuracy. First, we considered head motion, a common source of artifact in image-based analyses[18,23,24]. Then, we looked at age and IQ, two factors that may affect an individual's ability to comply with task instructions. The impact of head motion on model accuracy was readily discernible by visual inspection of a heatmap of the data for all participants when sorted by motion. Specifically, in Figure 2a, we plotted each participant's predicted fixation sequence (for the second calibration scan) in rows, with the participants being sorted from top to bottom in ascending order based on their mean FD in the training scan (i.e, Scan1); color is used to represent the predicted fixation location relative to the center of the screen and the calibration stimulus locations are depicted on top and bottom.

The consistency of fixation sequences is made obvious by distinct vertical bands observed at different time points, which indicate that an identical fixation location is being predicted for most of the individuals; these bands decrease among those participants in the lower portion of the map, where mean FD is greatest. Findings by visual inspection were similar when sorting participants based on age, but were notably less apparent when sorting participants based on mean DVARS (D referring to the temporal derivative of timecourses, VARS meaning root mean square variance over voxels), an indirect measure of head motion, as well as other potential factors that may impact compliance (i.e., full scale IQ [FSIQ])[25]. We used multiple regression to statistically test for associations between model accuracy and variables of interest, finding that mean FD and age are statistically significant predictors of model accuracy ($p < .01$), but not IQ or DVARS (see Figure 2b). As expected, predictive accuracy is negatively correlated with head motion and positively correlated with age among those in our sample (i.e., 5.0-21.0 years old).

Impact of Global Signal Regression and Volume Censoring on Model Accuracy. Given the deleterious effects of head motion on model accuracy, we explored the ability of two methods that are commonly used to ameliorate its impact on fMRI analyses — global signal regression (GSR) and volume censoring (framewise displacement > 0.2)[23,25,26]. Both of these methods appeared to significantly decrease model accuracy when applied to the training scan according to paired t-tests ($p < .01$), though the differences were relatively minimal in size (see Figure 2c). Limiting the GSR analysis to only those participants with high motion (i.e., mean FD > 0.2) did not increase its utility; instead, no difference in model accuracy was found with respect to whether or not GSR was employed with these datasets.

Minimum Data Requirements. Although limited in number, past implementations of PEER have differed with respect to the number of target locations for training (i.e., 9 vs. 25)[13,14]. As such, we examined the impact of the number of locations included in training on PEER model accuracy by creating random subsets of our data that systematically varied the number of training points included (i.e., from 1 to 25; 50 random subsets per number of locations). We found that PEER model accuracy appeared to asymptote with as few as 8 calibration targets; minor increases in accuracy were noted with additional training points. This is consistent with our finding that addition of a second PEER scan during training did not result in substantial increases in model performance.

Eye movement patterns during movie viewing. We next applied PEER to the movie fMRI scans. Specifically, the PEER1 model was used to predict fixation sequences for two scans during which video clips were viewed — *Despicable Me* and *The Present*. Figures 3a and 3b depict the heatmap of predicted time series for the two movie scans. In each of the heatmaps, distinct vertical bands were noted, indicating consistency in fixation location across participants at many of the time points, similar to what was observed for the calibration sequences. This is unsurprising since visually and emotionally salient scenes throughout the movie are expected to automatically capture a person's attention (e.g. a human face in the center of the screen), while those that are less important to the narrative will be less consistent in viewing patterns across participants[27–29].

As would be expected, the fixation pattern observed for each of the two movies was unique. To demonstrate this point, we first calculated pairwise correlations between the fixation time series generated for each of the participants for the two movies, allowing us to compare within vs. between movie relationships. As depicted in Figure 3c, correlations were significantly higher between two time series from different participants watching the same movie than correlations between different participants watching different movies. To quantify the level of discriminability between movies based on the predicted series, we trained a linear SVM using data from half of the available participants and tested on the remaining half. The model performed well in distinguishing which of the two movies were being watched based on the PEER-derived fixations, with AUROC curve values of .902 and .935 in the x- and y- directions respectively. Assessment of the confusion matrices indicated that there was no class imbalance for samples that were incorrectly classified. We looked for phenotypic variables that may distinguish the participants whose fixation sequences were classified accurately from those that were misclassified — however, no such features were identified. Thus, we establish that eye fixation information from movie viewing is reliably encoded in the fixation time series from PEER, allowing us to predict which movie is viewed with a high degree of accuracy.

Eye Tracker Validation. We next worked to validate the fixation sequences obtained for *The Present* by comparing them with those obtained using an eye tracker in the same participants on a repeat viewing of *The Present* (outside of the scanner on a different date). First, we visually inspected the consistency in the predicted fixation sequences from PEER with those from eye tracking by overlaying the sequences from both modalities as a heatmap. In Figure 4, we see that there is a moderate level of consistency in viewing patterns across modalities, indicated by consistency along columns (i.e. eye gaze for each time point), which hints at the congruence between fixation information from PEER and eye-tracking devices. To quantify this relationship, we compared the median fixation series (calculated across participants) for each modality with correlation values of .85 and .81 in the x- and y- directions. Overall, we conclude that PEER captures relevant information about the gaze location on a TR-by-TR basis.

Reliability of individual-specific eye movement patterns. We first tested whether the correlation between the PEER and eye-tracker based fixation sequences were greater when collected from the same individual than different - no difference was detected. Next, we more directly quantified the reliability of fixation locations at each point in the sequence using intraclass correlation coefficient (ICC) as well as that of the entire fixation sequence using I2C2[30] (a multivariate extension of ICC). Both approaches yielded poor reliability (i.e., < 0.3). *The Present* was specifically selected by the Healthy Brain Network for its simple and clear scenes of high emotional valence[17], which are of interest for imaging and voice sample analyses (recordings of participants retelling the story are obtained outside the scanner). However, there may be limited differences in between-participant fixation sequences since viewing is driven by salient stimuli that draw attention[29]. It is also possible that more precise measures derived from an eye tracker may carry meaningful individual-specific variation, though this is beyond the scope of the present work.

Neural Correlates of Eye Movement. Finally, we took the opportunity to examine the neural correlates of eye movements, as indexed by changes in fixation location during each of the scans (i.e., calibration, *Despicable Me*, *The Present*). While the low temporal resolution of PEER clearly limits the ability to identify individual saccades from one another, or from sustained smooth pursuits, we expected that it should afford a gross perspective of neural activity associated with eye movements. Each participant's fixation sequence was first converted into a change time series by calculating the euclidean distance from one time point to the next. Convolution with the hemodynamic response was then used to identify the neural bases of eye movements; to minimize potential head motion-related saccades, we modeled 24 motion-related parameters[31] and the PEER time series itself. We observed a similar pattern of activations across the three scan types, with activations being most robust for the *Despicable Me* scans, likely reflecting the longer scan duration (see Figure 5). We replicated patterns of activation in regions within distributed brain networks known to be associated with eye movement[32,33]. Specifically, there was activation in Brodmann Areas 6, 8, and 17, which contain the premotor cortex, frontal eye fields, and primary visual cortex.

DISCUSSION

This present work demonstrates the ability of the EPI-based PEER technique to accurately capture eye gaze patterns in movie viewing using a simple 1.5-minute calibration scan. Such information is a prerequisite for identifying variations in compliance with task demands during functional MRI scans, particularly in naturalistic viewing paradigms where there are no observable behaviors. We found that the eye orbit contains enough signal without the brain to reliably predict fixation sequences from both calibration scans and video clips. In addition, the PEER method's predictive accuracy increased with the number of calibration targets, but stabilized after 8 targets. Not surprisingly, head motion appeared to be the primary determinant of prediction accuracy, whether during the training scan or the scan to be predicted. Eye-fixation patterns were found to be highly distinct for each movie and consistent across participants, allowing for relatively easy identification of the movie being viewed based on PEER alone. Eye tracking data obtained from a repeat viewing of the clip *The Present* outside the scanner was used to validate the results from PEER. Consistent with prior eye tracking studies, we found that consistency of the eye tracking patterns in movie viewing[28] observed across individuals limited the ability to reliably detect individual-specific variations in eye-fixation patterns. Finally, we found that the neural correlates of eye movement identified through PEER mirror those found in literature. Thus, PEER is a cost-efficient, easy to set up method of retrospectively determining eye-gaze patterns from fMRI scans.

PEER is not intended to compete with the capabilities of modern eye trackers, most of which sample at a minimum of 60 Hz and contain additional information beyond eye fixations (e.g., pupillometry)[34]. PEER is a lightweight solution to one of the most basic confounds present in fMRI studies. As naturalistic viewing paradigms gain popularity and find more broad usage, there will be greater demand for methods that establish the validity of findings obtained in the absence of eye fixation data. The ease with which PEER can be added to any scan protocol, requiring only 1.5 minutes of data and no additional equipment or expertise, will make it appealing for many — particularly those pursuing large-scale studies. One area where PEER may have potential advantages over eye tracking is natural sleep imaging (e.g., infants, toddlers), as detection of eye movements is not dependent on the eyelid being open and the sampling frequency should be sufficient (there are typically 15.9 eye movements per minute)[35].

There were two key limitations in data quality for our PEER results. First and foremost is the deleterious impact of head motion. Consistent with observations from the resting state fMRI literature, we found some higher variability in the accuracy for model estimates from data with mean framewise displacement exceeding 0.2mm. Real-time motion detection systems in fMRI could be used to help establish data quality[36]. Second is compliance with the calibration scan. Similar to any eye tracking paradigm, failure to comply with the calibration will compromise detection accuracy. Not surprisingly, we found that model accuracy was predicted by age (after controlling for head motion) — a finding that likely reflects lower compliance with instructions in young children. One could potentially address the detection of such issues by adding a simple task into the calibration scan that requires a response (e.g., having a letter appear at each

calibration location and requiring the participant to identify it). Alternatively, integration of the calibration scan into a real-time fMRI could readily resolve the concern.

We found that despite consistency across participants, the eye movement patterns did not exhibit test-retest reliability, whether using a multivariate or univariate ICC framework. This may in part reflect the nature of the comparison afforded by the present work, which required the test-retest comparison to be between a PEER-based measurement of eye fixations and an eye-tracking-based measurement. We consider one of the primary challenges to be between-participant variation, as relatively consistent eye fixation patterns were detected across individuals, especially in dynamic scene viewing[28]. While studies that use summary statistics of viewing patterns (e.g. proportion of time fixating in a given region of interest) demonstrate higher intra-participant correlations, correlations are lower between the full fixation series for a given participant[28,37]. This suggests that the eye fixations detected with PEER are primarily driven by salient visual stimuli. The relatively high level of engagement that tends to be associated with the selected video clips may also be a factor. Our findings should not be taken to infer that individual variation is beyond the window of examination afforded by the more sophisticated measures obtainable from current eye tracking devices.

The exact reasons why PEER works remain unclear, although it likely involves the detection of variations in the MRI contrast between the vitreous and aqueous humors[38]. Looking forward, there is potential to create a generalizable model for PEER that will enable researchers to retrospectively determine fixations from fMRI data, even when calibration data are not available. In addition, there are potential optimizations, such as multiband imaging, which can increase sampling rates. We demonstrate that PEER is an inexpensive, easy-to-use method to retrospectively determine eye fixations from fMRI data, a step toward toward deploying new methods of monitoring variations in task compliance that overcome the burdens of MRI compatible eye-tracking devices.

METHODS

Participants. The Healthy Brain Network is a large-scale data collection effort focused on the generation of an open resource for studying pediatric mental health and learning disorders[17]. The ongoing data collection contains a range of phenotypes, spanning typical and atypical brain development. We included data from 480 participants (ages: 5.0-21.0; 10.3 ± 3.5) collected at the Rutgers University site, which had the largest number of complete imaging datasets available at the time of our analyses. As outlined in greater detail in the data descriptor publication for the Healthy Brain Network[17], approximately 80 percent of the participants in the sample have one or more diagnosable disorders according to the DSM-5. This dataset includes a high proportion of participants with Attention Deficit Hyperactivity Disorder (ADHD; ~50%) and children as young as age 5.0. Both of these participant types have a higher likelihood of head motion, allowing us to study its impact on PEER analyses[17–20].

Imaging Data. Data were collected using the Siemens 3T Tim Trio MRI scanner located at the Rutgers University Brain Imaging Center (RUBIC). For functional MRI scans, a multiband factor of 6 was employed to achieve a 2.4mm isotropic voxel size and TR = 800ms (TE = 30ms). During the imaging session, each participant completed a minimum of two PEER calibration scans (3 scans for n = 430, 2 scans for n = 50). Two movie viewing scans were included as well: *Despicable Me* [10 min clip, DVD version exact times 1:02:09 – 1:12:09] and *The Present* [~3.47 min; added November 23, 2016].

Predictive Eye Estimation Regression

PEER Scan Instructions. The participant is asked to fixate on a white dot that iterates through 25 different positions for 4 seconds each; the positions were selected to ensure coverage of all corners of the screen as well as the center (see supplementary figure 2). The PEER calibration scans were distributed throughout the imaging session such that they flank the other scan types (e.g. rest, movie viewing) and allow for a sampling of possible changes in image properties over time.

Image Processing. Consistent with prior work[12,13], a minimal image processing strategy was employed for the PEER scans. Using the Configurable Pipeline for the Analysis of Connectomes (C-PAC)[39], we performed the following steps: motion correction, image intensity normalization, temporal high-pass filtering (cutoff = 100s), and spatial filtering (FWHM at 6mm). The preprocessed functional data for each participant was then registered to the corresponding high-resolution anatomical T1 image using boundary-based registration via FLIRT[40,41]. The final fMRI data were registered to the MNI152[42] template space using ANTs[43].

Quality Assurance. Two researchers visually inspected the middle volume of each participant's PEER calibration scans for incomplete coverage of the orbit (i.e. missing eye signal), leading to the exclusion of 32 participants from analysis. At least two PEER scans were available for each participant (3 scans for n = 409, 2 scans for n = 39). Given that *The Present* was added to the imaging protocol later than *Despicable Me,* fewer participants had both scans. We inspected the movie scans and identified 427 *Despicable Me scans* and 360 scans of *The Present* with

complete coverage of eye signal. In the following experiments, we removed participants with low quality scans relevant to a given analysis.

Model Generation. We limited each fMRI scan to the region corresponding to the MNI152 eye-mask template[42]. Isolation of signal to the orbit was done for two reasons. First, prior works[12,13] suggest that signal from the eyes provides adequate information to predict eye fixations. Second, this would reduce the dimensionality of the dataset to accelerate model generation and fixation estimation using PEER.

At each voxel we: 1) mean centered and variance normalized (i.e., z-scored) the time series, and 2) averaged the consecutive time points associated with each stimulus presentation, reducing the time series to 25 points. The latter step can mitigate random noise, as well as the effects of subtle eye movements during fixation, given that fixation stability and saccadic eye movements vary across individuals[44–46]. For each participant, two separate support vector regression models were trained - one for x-direction fixation locations and one for y-direction fixation locations. In accord with prior works[12,13], PEER was used to predict the 25 positions using the voxel-wise time series data (i.e., a unique predictor was included for each voxel, with the following parameters: C=100, epsilon=0.01).

Estimating Model Accuracy. For each participant, the PEER-generated SVR model trained using Scan1 (PEER1) was used to predict eye fixations from their second calibration scan (Scan2). PEER1 model accuracy was assessed using Pearson's correlation coefficient between the model-generated fixation time series and the stimulus locations from the calibration sequence. Pediatric neuroimaging studies demonstrate a decline in scan quality over the course of an imaging session, which implies that the PEER scans (and the resulting SVR models) may differ in overall quality based on its timing in the imaging session[18–20,47–49]. We compared the models trained using Scan1, Scan3, or both scans in their ability to predict eye fixations from Scan2. To quantify differences in PEER scan quality, we conducted a paired t-test to compare Scan1 and Scan3 with respect to head motion (i.e. framewise displacement, DVARS)[23,25].

Factors Associated with PEER Prediction Quality.

Head Motion and Compliance. Head motion is one of the most consistent sources of artifacts in fMRI analyses[22,23,25]. We examined the impact of head motion on PEER accuracy using mean framewise displacement (a direct measure of head motion) and standardized DVARS (an indirect measure)[25]. To do so, we assessed the relationship between measures of motion and the predictive accuracy of PEER1 (ability to predict fixations from Scan2) via linear regression. Beyond head motion, compliance with PEER scan instructions (to fixate on the stimuli) can impact model accuracy. While we have no direct measure of this compliance, we did test for relationships between model accuracy and participant variables that varied considerably across participants and we hypothesized may impact compliance; in particular, age and full-scale IQ (FSIQ).

PEER Scan Image Preprocessing. We assessed model accuracy of PEER after implementing global signal regression (GSR), a method to remove non-neuronal contributions to fMRI data (e.g. head motion). Though the neuroimaging community has not reached a consensus on the use and interpretation of GSR, it has been shown to remove global changes to BOLD signal, caused for instance by respiration or head motion[23,26,50–53]. Given the increased likelihood of motion artifacts in the HBN dataset[17], which includes participants at various points of maturation (ages 5-21) with typical and atypical brain development (e.g. ADHD), we implemented GSR on Scan1 data prior to model training. The preprocessed data was used to train the PEER model, which was applied to Scan2 to measure fit of the predicted time series with known calibration targets. We repeated this analysis with volume censoring, using a framewise displacement threshold of 0.2mm on data from Scan1 prior to model training and estimation. Paired t-tests were used to compare the predictive accuracy between the original and preprocessed models.

Minimum Data Requirements. To establish minimum data requirements for accurate PEER model estimation, we systematically varied the number of calibration points used in model generation. Specifically, we randomly sampled N training points (N: 2-25) from the calibration scan and used the corresponding brain images to train PEER (50 random samples were generated per N to estimate confidence intervals). Consistent with our prior analyses, predictive accuracy for each PEER model was determined by comparing the predicted fixation time series with the known calibration locations. The composition of each random sample varied with respect to the combination of training points. Thus, this analysis specifically examines the number of training points required on average to adequately train the PEER model and not the spatial arrangement of the points that optimizes model performance.

Validation

Identifying a Movie Based on Eye Movements. We evaluated the ability of the PEER method to capture fixation sequence uniquely associated with a given movie stimulus. In order to accomplish this, we first applied each participant's PEER1 model to their corresponding movie scans (*The Present*, *Despicable Me*), thereby producing a participant-specific fixation sequence for each movie. We then used an unpaired t-test to compare the level of correlation observed between differing participants' time series when watching the same movie versus a different movie (given that fMRI scans of DM contained 750 volumes while scans for TP contained 250, only the first 250 volumes from DM were used in this analysis). To further quantify the level of discriminability between the fixation time series of the two movies, we trained a binary SVM classifier to predict which movie the individual was seeing based on a given fixation time series. The linear SVM classifier (C=100) was trained using half of the available participant datasets and tested on the remaining half of the participants. The results were assessed using a confusion matrix and ROC curves.

Out-of-Scanner Eye Tracker Measurement. For 248 participants included in this work, a second viewing of *The Present* was added outside the scanner at a later session in the study. Eye tracking data were obtained during this viewing using an infrared video-based eye tracker (iView-X Red-m, SensoMotoric Instruments [SMI] GmbH)[17], allowing us to compare

PEER-derived eye fixations and those from the current gold standard. Eye tracking data were collected at the Staten Island or Manhattan site (sampling rate: 60 and 120 Hz, respectively). Similar to the design of the MRI data collection protocol, clips of *The Present* were shown at the end of the EEG and eye tracking collection protocol; thus, participants with poor eye tracker calibration or participants who were unable to complete the protocol were missing data for *The Present*. Of the remaining participants, those missing more than 10% of raw samples from eye-tracking or with moderate to high levels of head motion in MRI data for *The Present* (defined by mean FD > 0.2mm) were removed from analysis; this left 116 participants in the dataset. To match the sampling rate (TR) of the MRI scanner, the raw data from eye-tracking was segmented into 800 ms windows. In each window, we calculated the median of the raw samples as an estimate of multiple eye movements that were detected by the eye tracker. The median fixation time series from PEER was compared to that of eye tracking, using Pearson's r to assess the similarity between the fixation sequences detected by each method.

Identifying a Participant Based on Eye Movements. We examined the similarity of the fixation time series from eye tracking and PEER predictions for *The Present* to assess the reproducibility of participant fixation patterns. *The Present* was selected specifically for its emotional content and easy-to-understand narrative; as such, we would expect viewing patterns to be congruent across scenes of high valence. However, these patterns may be identifiable based on subtle viewing pattern differences. Prior works have demonstrated reproducibility using summary statistics of eye movement (e.g. proportion of time fixating on mouth vs. eyes); however, we examined the whole time series, which has been shown to vary in viewing consistency by stimulus and age[28,37,54]. We examined the intra- and inter-individual variability in viewing patterns by calculating Pearson's r between all pairs of PEER and eye tracking fixation time series to assess the feasibility of identification. Then, we computed the correlation for each participant's PEER-estimated and eye tracking fixation time series when compared to the median fixation time series. Using these measures, we ran a univariate ICC analysis for the whole scan and for individual time points. We also completed a multivariate extension of ICC named the Image Intra-Class Correlation Coefficient (I2C2[30]) with 500 permutations to estimate the null distribution.

Neural Correlates of Eye Fixation Sequences. We took the opportunity to test the feasibility of using the PEER-derived fixations from movie viewing fMRI data to characterize the neural correlates associated with eye fixations. Based on prior work, we expected to observe activations in the frontal eye fields, ventral intraparietal sulcus and early visual processing areas[32,33,55,56]; deactivations were expected in the default mode network during movie viewing[57]. To model neural activity associated with fixations identified by PEER, we first computed the magnitude of change in fixation location from one time point to the next in each predicted PEER time series using Euclidean distance. The resulting stimulus function was convolved with the double-gamma hemodynamic response function with added temporal derivatives, which was then used to model voxel-wise activity in response to movie viewing. To minimize the impact of prediction errors, a hyperbolic tangent squashing function was used to identify and reduce spikes in the eye movement vector prior to the convolution. Twenty-four motion parameters

were regressed out from the model and FSL FEAT was used for all individual-level analyses. To assess group activation, we used the FSL FLAME mixed effect model with the following variables as nuisance covariates: sex, mean framewise displacement, age, model accuracy in the x- and y- directions and full scale IQ. Multiple comparison correction was carried out using Gaussian Random Field theory, as implemented in FSL ($Z > 3.1$, $p < 0.05$, corrected).

FIGURE LEGENDS

Figure 1 | PEER accuracy when estimating fixations from calibration scan 2. (a) Predictive accuracy for PEER when trained on calibration scan 1 (PEER1) in the x- and y- directions. (b) Differences in predictive accuracy based on the calibration scans used for SVR model training.

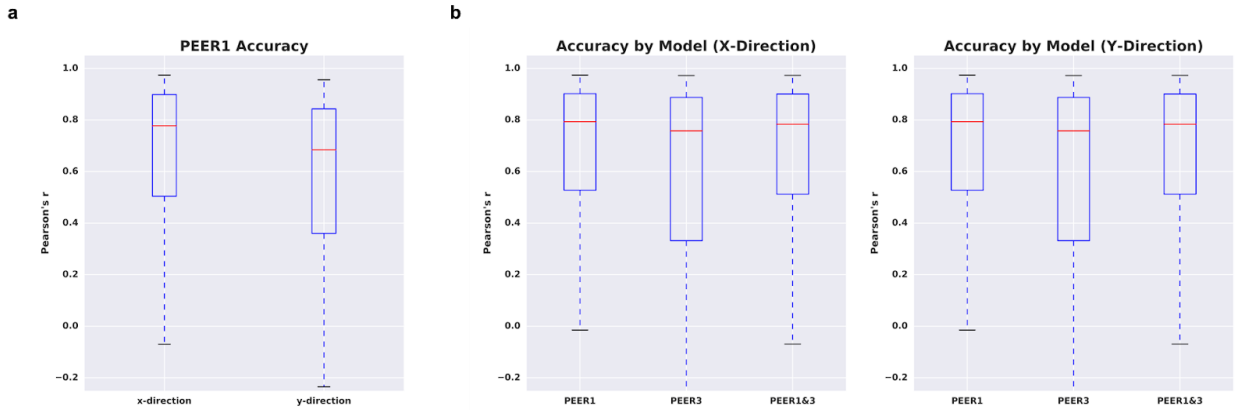


Figure 2 | Impact of phenotype and preprocessing on PEER accuracy. (a) Heatmap of predicted fixation time series from calibration scan 2, sorted by increasing motion (left) or decreasing age (right). Colorbar represents distance (pixels) of prediction from the center of the screen. Stimulus locations from the calibration scan are placed above and below participant predictions for reference. (b) Locally weighted scatterplot smoothing (Lowess) regression between PEER accuracy and mean framewise displacement or age. (c) Impact of global signal regression and volume censoring on PEER accuracy.

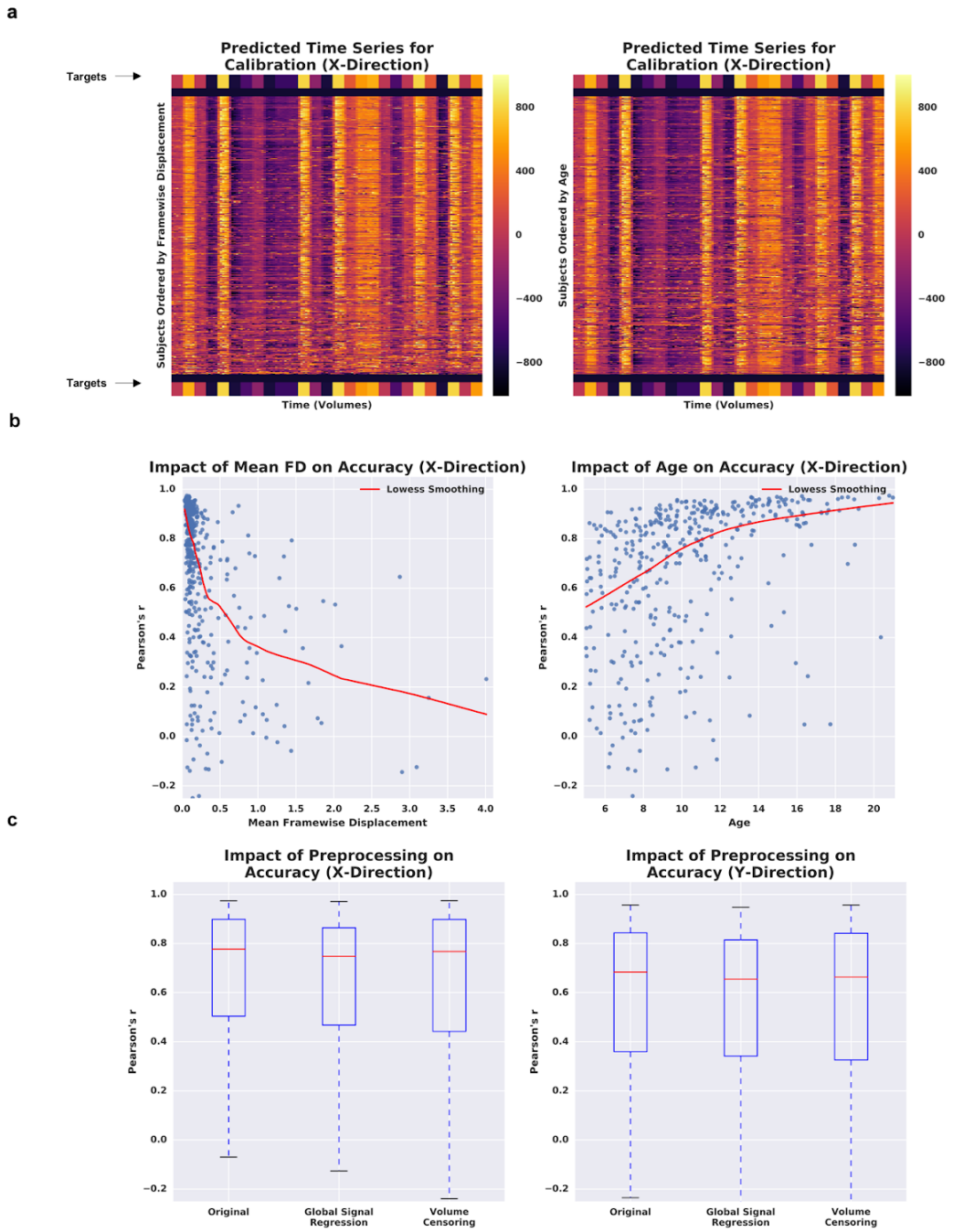


Figure 3 | PEER applied to movie viewing fMRI. (a) Heatmap of predicted fixations for *The Present*. The median fixation time series is placed above and below participant predictions. (b) Heatmap of predicted fixations for *Despicable Me*. (c) Pairwise correlations for all predicted fixation time series from both movies.

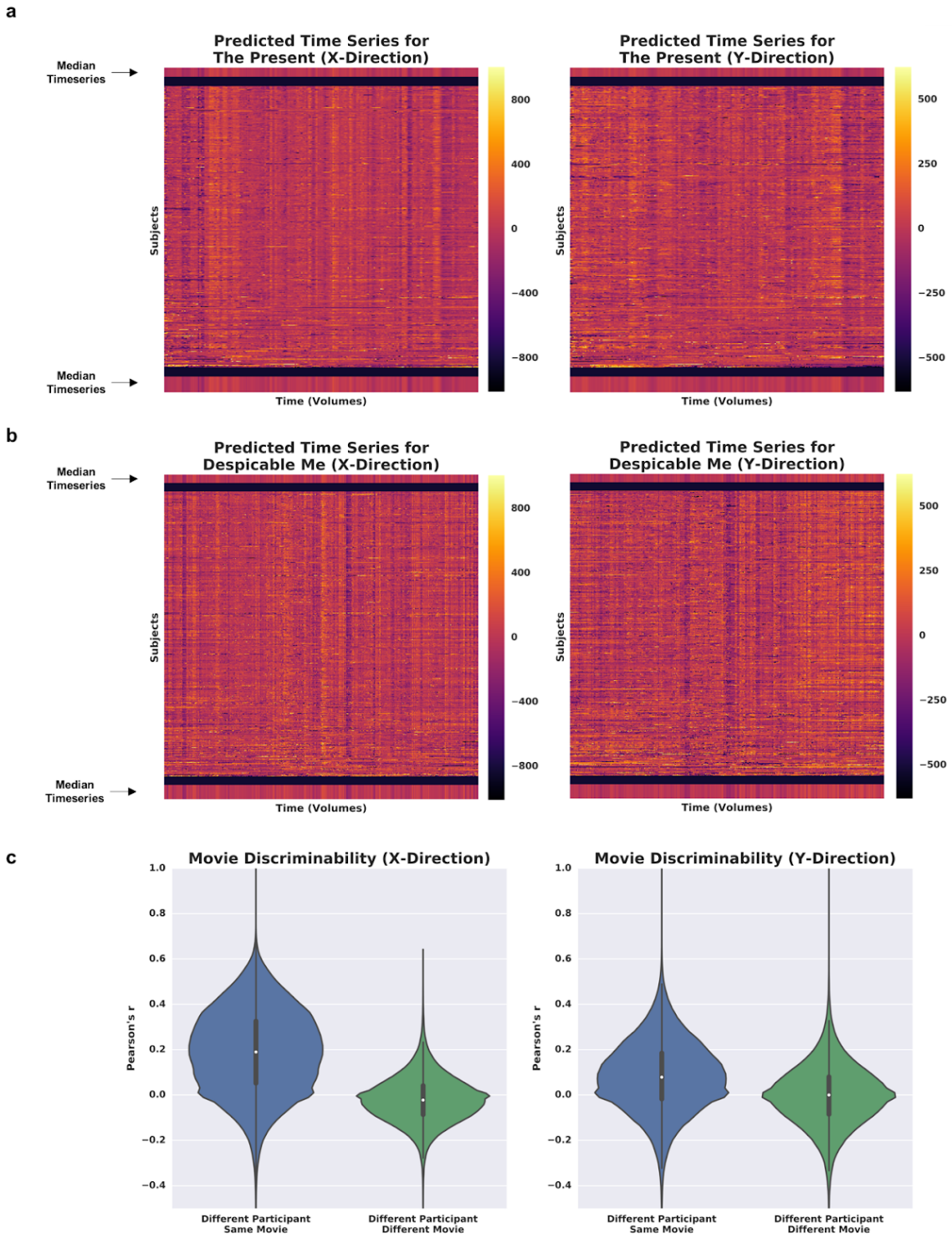


Figure 4 | Heatmap of PEER-derived fixation time series overlaid above eye-tracker measurements for *The Present*.

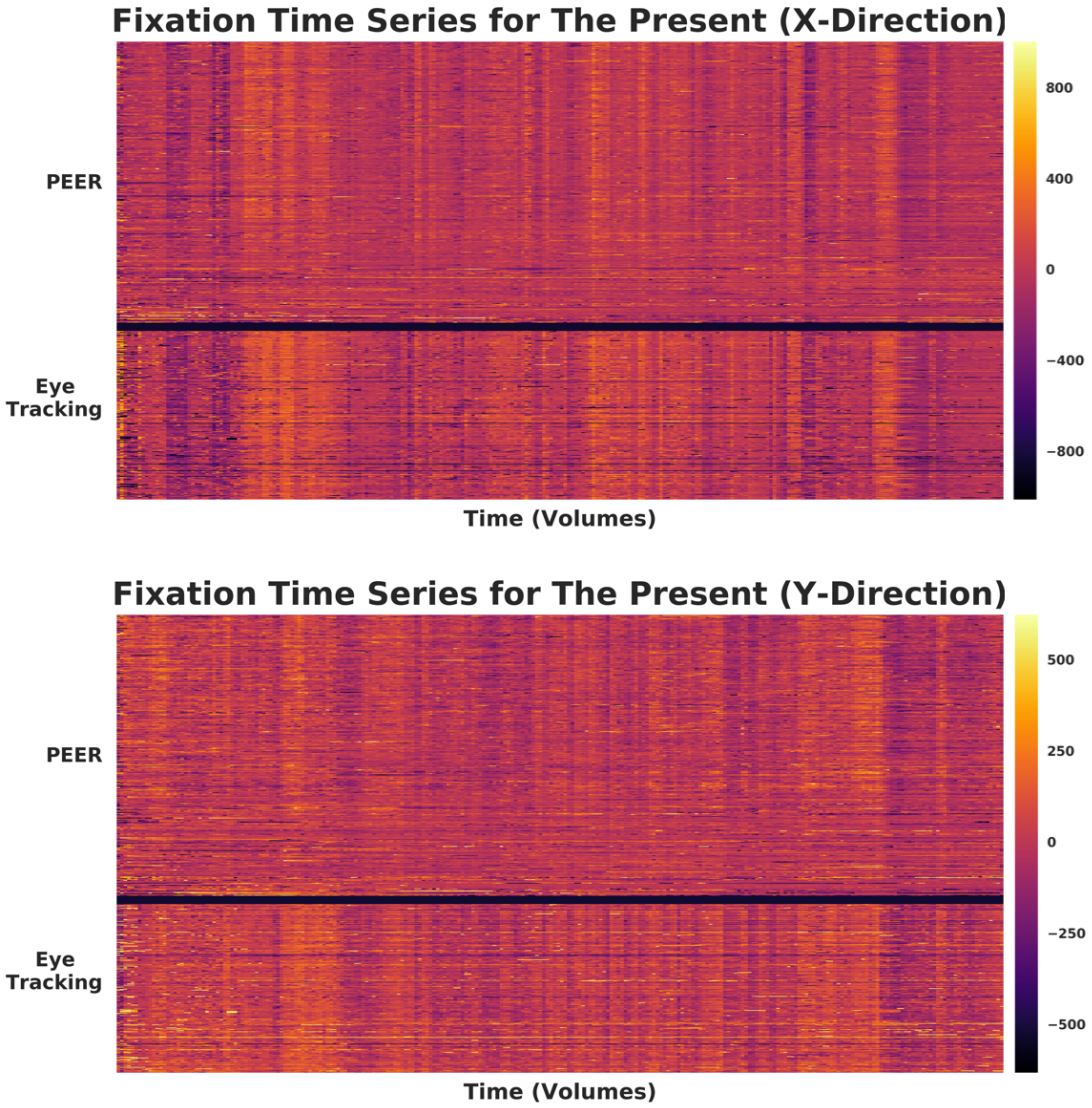
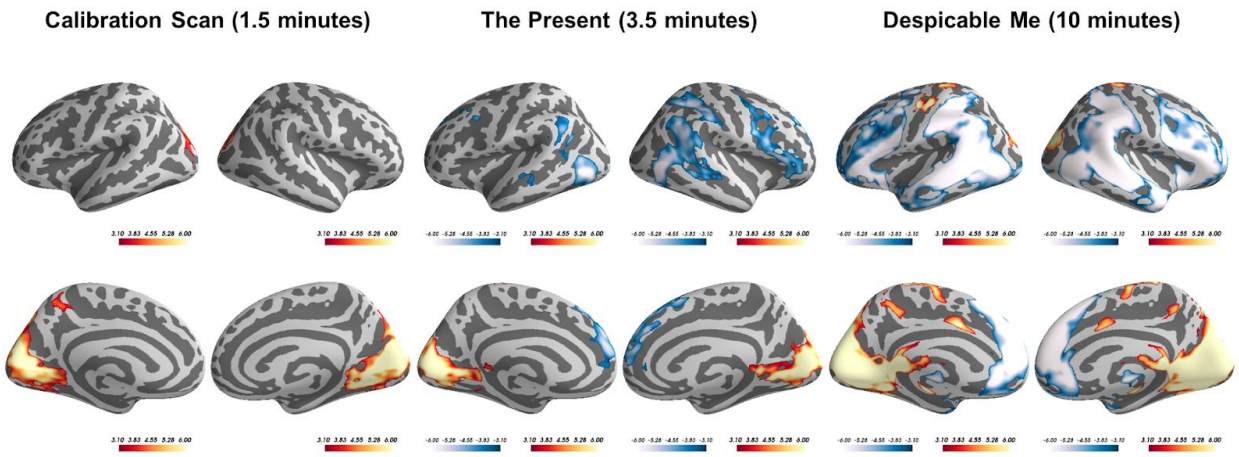
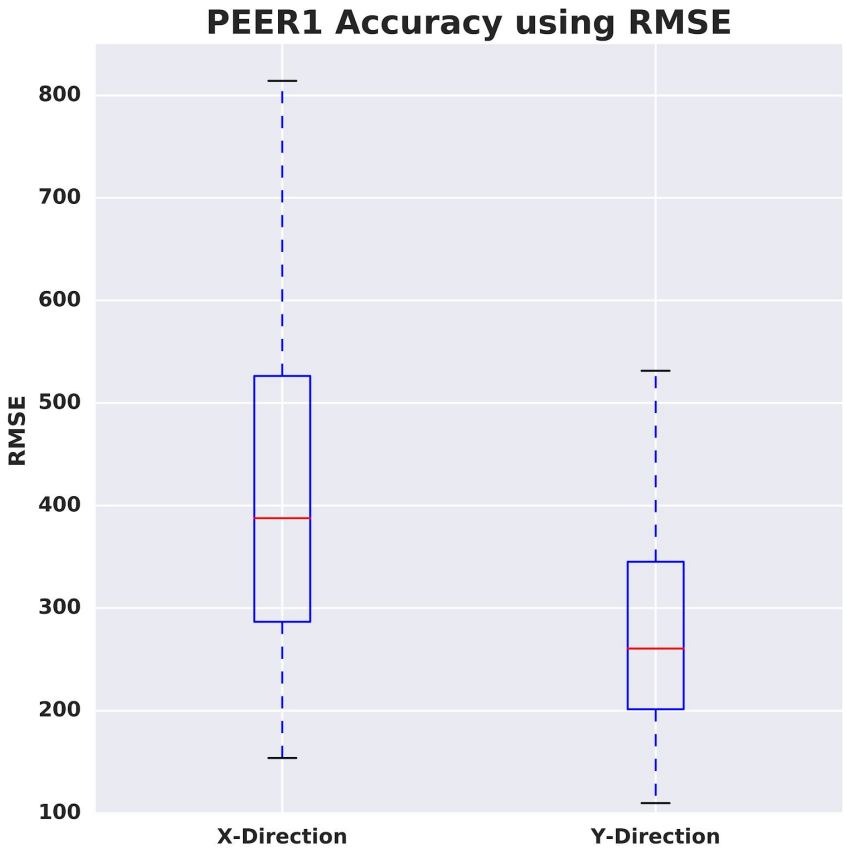


Figure 5 | Group activation maps generated using PEER-derived eye movements from the calibration scan, *The Present*, and *Despicable Me* ($Z > 3.1$, $p < 0.05$, corrected).

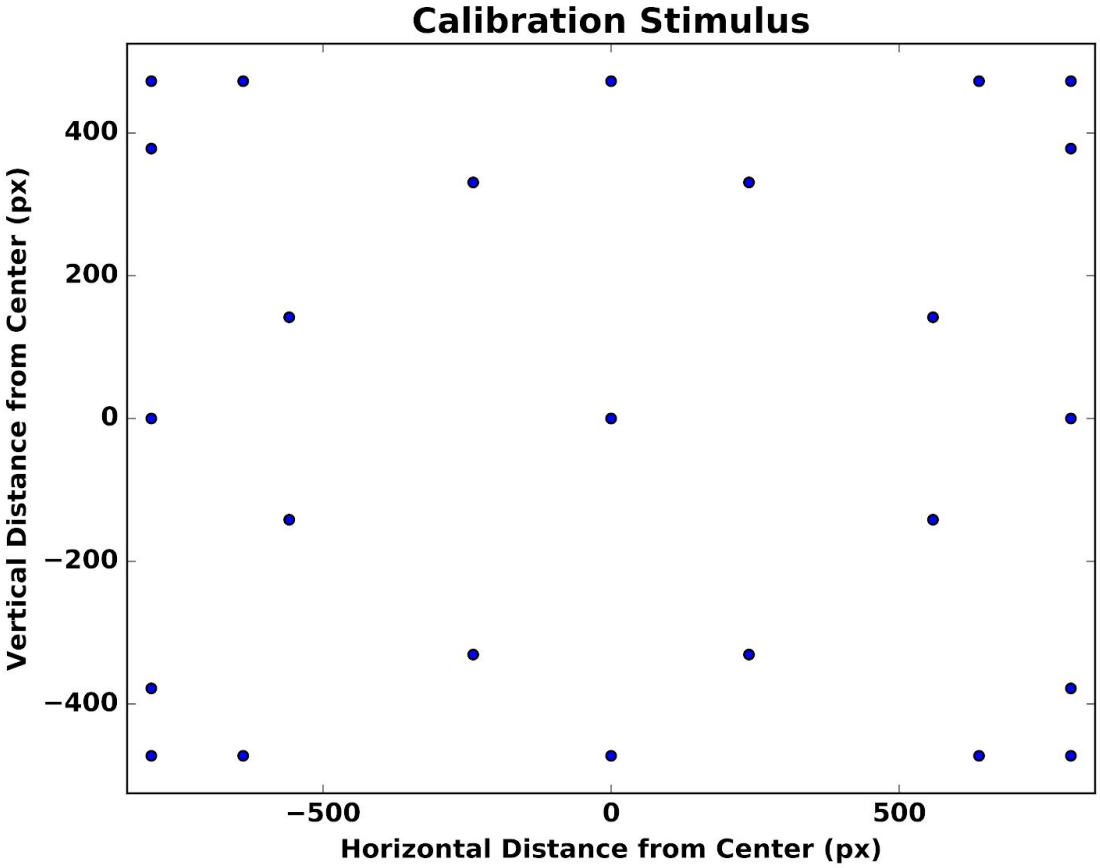


SUPPLEMENTARY FIGURES

Supplementary Figure 1 | Assessment of PEER1 accuracy when estimating fixations from calibration scan 2 using root mean square error (RMSE).



Supplementary Figure 2 | Stimulus location grid employed during the PEER calibration scans (one dot is displayed at a time; 4 seconds per dot).



ACKNOWLEDGEMENTS

The work presented here was primarily supported by gifts to the Child Mind Institute from Phyllis Green, Randolph Cowen, and Joseph Healey. We would like to thank the Healthy Brain Network participants and their families for taking the time to be in the study, as well as their willingness to have their data shared with the scientific community. We would also like to thank the many individuals who have provided financial support to the CMI Healthy Brain Network to make the creation and sharing of this resource possible. Stephen LaConte would like to thank Drs Chri Glielmi, Keith Heberlein, Scott Peltier, and Xiaoping Hu for helping to develop and test the PEER method. Jake Son would like to thank Youngwoo, Myungja, and Ickyoung for their love and support, and Drs. Perry, Swisher, Klein, and Milham for their patience and mentorship.

REFERENCES

1. Church, J. A., Petersen, S. E. & Schlaggar, B. L. The 'Task B problem' and other considerations in developmental functional neuroimaging. *Hum. Brain Mapp.* **31,** 852–862 (2010).
2. Price, C. J. & Friston, K. J. Scanning patients with tasks they can perform. *Hum. Brain Mapp.* **8,** 102–108 (1999).
3. Brown, T. T. *et al.* Developmental changes in human cerebral functional organization for word generation. *Cereb. Cortex* **15,** 275–290 (2005).
4. Murphy, K. & Garavan, H. Artifactual fMRI group and condition differences driven by performance confounds. *Neuroimage* **21,** 219–228 (2004).
5. Koten, J. W., Langner, R., Wood, G. & Willmes, K. Are reaction times obtained during fMRI scanning reliable and valid measures of behavior? *Exp. Brain Res.* **227,** 93–100 (2013).
6. Tagliazucchi, E. & Laufs, H. Decoding Wakefulness Levels from Typical fMRI Resting-State Data Reveals Reliable Drifts between Wakefulness and Sleep. *Neuron* **82,** 695–708 (2014).
7. Wang, J., Han, J., Nguyen, V. T., Guo, L. & Guo, C. C. Improving the Test-Retest Reliability of Resting State fMRI by Removing the Impact of Sleep. *Front. Neurosci.* **11,** 249 (2017).
8. Hansen, D. W. & Ji, Q. In the eye of the beholder: a survey of models for eyes and gaze. *IEEE Trans. Pattern Anal. Mach. Intell.* **32,** 478–500 (2010).
9. Kimmig, H., Greenlee, M. W., Huethe, F. & Mergner, T. MR-eyetracker: a new method for eye movement recording in functional magnetic resonance imaging. *Exp. Brain Res.* **126,** 443–449 (1999).
10. Kanowski, M., Rieger, J. W., Noesselt, T., Tempelmann, C. & Hinrichs, H. Endoscopic eye tracking system for fMRI. *J. Neurosci. Methods* **160,** 10–15 (2007).
11. Ferri, J., Schmidt, J., Hajcak, G. & Canli, T. Neural correlates of attentional deployment

within unpleasant pictures. *Neuroimage* **70,** 268–277 (2013).

12. LaConte, S., Peltier, S., Heberlein, K. & Hu, X. Predictive eye estimation regression (PEER) for simultaneous eye tracking and fMRI. in *Proc. Intl. Soc. Magn. Reson. Med* **2808,** (2006).
13. LaConte, S., Glielmi, C., Heberlein, K. & Hu, X. Verifying visual fixation to improve fMRI with predictive eye estimation regression (PEER). in *Proc. Intl. Soc. Magn. Reson. Med* **3438,** (2007).
14. Sathian, K. *et al.* Dual pathways for haptic and visual perception of spatial and texture information. *Neuroimage* **57,** 462–475 (2011).
15. Adelore, T. A. Determining fixation stability of AMD patients using predictive eye estimation regression. Available at: https://smartech.gatech.edu/bitstream/handle/1853/26495/adelore_temilade_a_200812_mast.pdf.
16. Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. & Vapnik, V. Support Vector Regression Machines. in *Proceedings of the 9th International Conference on Neural Information Processing Systems* 155–161 (MIT Press, 1996).
17. Alexander, L. M. *et al.* An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Sci Data* **4,** 170181 (2017).
18. Kotsoni, E., Byrd, D. & Casey, B. J. Special considerations for functional magnetic resonance imaging of pediatric populations. *J. Magn. Reson. Imaging* **23,** 877–886 (2006).
19. Wilke, M., Holland, S. K., Myseros, J. S., Schmithorst, V. J. & Ball, W. S., Jr. Functional magnetic resonance imaging in pediatrics. *Neuropediatrics* **34,** 225–233 (2003).
20. Schlund, M. W. *et al.* Pediatric functional magnetic resonance neuroimaging: tactics for encouraging task compliance. *Behav. Brain Funct.* **7,** 10 (2011).
21. Gaillard, W. D., Grandin, C. B. & Xu, B. Developmental aspects of pediatric fMRI:

considerations for image acquisition, analysis, and interpretation. *Neuroimage* **13,** 239–249 (2001).

22. Satterthwaite, T. D. *et al.* Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. *Neuroimage* **60,** 623–632 (2012).
23. Power, J. D. *et al.* Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* **84,** 320–341 (2014).
24. Zuo, X.-N. *et al.* An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci Data* **1,** 140049 (2014).
25. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **59,** 2142–2154 (2012).
26. Murphy, K., Birn, R. M. & Bandettini, P. A. Resting-state fMRI confounds and cleanup. *Neuroimage* **80,** 349–359 (2013).
27. Hanke, M. *et al.* A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation. *Sci Data* **3,** 160092 (2016).
28. Dorr, M., Martinetz, T., Gegenfurtner, K. R. & Barth, E. Variability of eye movements when viewing dynamic natural scenes. *J. Vis.* **10,** 28 (2010).
29. Wang, H. X., Freeman, J., Merriam, E. P., Hasson, U. & Heeger, D. J. Temporal eye movement strategies during naturalistic viewing. *J. Vis.* **12,** 16 (2012).
30. Shou, H. *et al.* Quantifying the reliability of image replication studies: the image intraclass correlation coefficient (I2C2). *Cogn. Affect. Behav. Neurosci.* **13,** 714–724 (2013).
31. Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. & Turner, R. Movement-related effects in fMRI time-series. *Magn. Reson. Med.* **35,** 346–355 (1996).

32. Choi, W. & Henderson, J. M. Neural correlates of active vision: An fMRI comparison of natural reading and scene viewing. *Neuropsychologia* **75,** 109–118 (2015).

33. Marsman, J. B. C., Renken, R., Velichkovsky, B. M., Hooymans, J. M. M. & Cornelissen, F. W. Fixation based event-related fmri analysis: using eye fixations as events in functional magnetic resonance imaging to reveal cortical processing during the free exploration of visual images. *Hum. Brain Mapp.* **33,** 307–318 (2012).

34. Palinko, O., Kun, A. L., Shyrokov, A. & Heeman, P. Estimating Cognitive Load Using Remote Eye Tracking in a Driving Simulator. in *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* 141–144 (ACM, 2010).

35. Takahashi, K. & Atsumi, Y. Precise measurement of individual rapid eye movements in REM sleep of humans. *Sleep* **20,** 743–752 (1997).

36. Dosenbach, N. U. F. *et al.* Real-time motion analytics during brain MRI improve data quality and reduce costs. *Neuroimage* **161,** 80–93 (2017).

37. Rogers, S. L., Speelman, C. P., Guidetti, O. & Longmuir, M. Using dual eye tracking to uncover personal gaze patterns during social interaction. *Sci. Rep.* **8,** 4271 (2018).

38. Fanea, L. & Fagan, A. J. Review: magnetic resonance imaging techniques in ophthalmology. *Mol. Vis.* **18,** 2538–2560 (2012).

39. Cameron, C. *et al.* Towards Automated Analysis of Connectomes: The Configurable Pipeline for the Analysis of Connectomes (C-PAC). *Front. Neuroinform.* **7,** (2013).

40. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17,** 825–841 (2002).

41. Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* **5,** 143–156 (2001).

42. Grabner, G. *et al.* Symmetric Atlasing and Model Based Segmentation: An Application to the Hippocampus in Older Adults. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006* 58–66 (Springer Berlin Heidelberg, 2006).
43. Avants, B. B. *et al.* A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* **54,** 2033–2044 (2011).
44. Morales, M. U. *et al.* Reference Clinical Database for Fixation Stability Metrics in Normal Subjects Measured with the MAIA Microperimeter. *Transl. Vis. Sci. Technol.* **5,** 6 (2016).
45. Ditchburn, R. W. & Ginsborg, B. L. Involuntary eye movements during fixation. *J. Physiol.* **119,** 1–17 (1953).
46. Vingolo, E. M., Fragiotta, S. & Cutini, A. Eye-Fixation Behavior in Major Depressive Disorder and the Influence of Pharmacological Therapy: A Microperimetric Study. *Int. J. Ophthalmol.* (2014).
47. Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *International Joint Conference on Artificial Intelligence* (1995).
48. Refaeilzadeh, P., Tang, L. & Liu, H. Cross-Validation. in *Encyclopedia of Database Systems* (eds. Liu, L. & Özsu, M. T.) 1–7 (Springer New York, 2016).
49. Bradley, A. P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **30,** 1145–1159 (1997).
50. Birn, R. M., Diamond, J. B., Smith, M. A. & Bandettini, P. A. Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. *Neuroimage* **31,** 1536–1548 (2006).
51. Murphy, K. & Fox, M. D. Towards a consensus regarding global signal regression for resting state functional connectivity MRI. *Neuroimage* **154,** 169–173 (2017).
52. Murphy, K., Birn, R. M., Handwerker, D. A., Jones, T. B. & Bandettini, P. A. The impact of

global signal regression on resting state correlations: are anti-correlated networks introduced? *Neuroimage* **44,** 893–905 (2009).

53. Fox, M. D., Zhang, D., Snyder, A. Z. & Raichle, M. E. The global signal and observed anticorrelated resting state brain networks. *J. Neurophysiol.* **101,** 3270–3283 (2009).
54. Franchak, J. M., Heeger, D. J., Hasson, U. & Adolph, K. E. Free Viewing Gaze Behavior in Infants and Adults. *Infancy* **21,** 262–287 (2016).
55. Corbetta, M. & Shulman, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* **3,** 201–215 (2002).
56. Bruce, C. J. & Goldberg, M. E. Primate frontal eye fields. I. Single neurons discharging before saccades. *J. Neurophysiol.* **53,** 603–635 (1985).
57. Greicius, M. D., Krasnow, B., Reiss, A. L. & Menon, V. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 253–258 (2003).