# A topography-based predictive framework for naturalistic viewing fMRI

Xuan Li [a,b,*], Patrick Friedrich [a,b], Kaustubh R. Patil [a,b], Simon B. Eickhoff [a,b], Susanne Weis [a,b]

[a] Institute of Neuroscience and Medicine, Brain and Behaviour (INM-7), Research Centre Jülich, Jülich 52428, Germany
[b] Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf 40225, Germany

A B S T R A C T

Functional magnetic resonance imaging (fMRI) during naturalistic viewing (NV) provides exciting opportunities for studying brain functions in more ecologically valid settings. Understanding individual differences in brain functions during NV and their behavioural relevance has recently become an important goal. However, methods specifically designed for this purpose remain limited. Here, we propose a topography-based predictive framework (TOPF) to fill this methodological gap. TOPF identifies individual-specific evoked activity topographies in a data-driven manner and examines their behavioural relevance using a machine learning-based predictive framework. We validate TOPF on both NV and task-based fMRI data from multiple conditions. Our results show that TOPF effectively and stably captures individual differences in evoked brain activity and successfully predicts phenotypes across cognition, emotion and personality on unseen subjects from their activity topographies. Moreover, TOPF compares favourably with functional connectivity-based approaches in prediction performance, with the identified predictive brain regions being neurobiologically interpretable. Crucially, we highlight the importance of examining individual evoked brain activity topographies in advancing our understanding of the brain-behaviour relationship. We believe that the TOPF approach provides a simple but powerful tool for understanding brain-behaviour relationships on an individual level with a strong potential for clinical applications.

## 1. Introduction

Functional magnetic resonance imaging (fMRI) during naturalistic viewing (NV), i.e., movie-watching, has recently emerged as a useful tool for studying brain function. NV fMRI uses naturalistic stimuli (e.g., movie clips) in the scanner to approximate real-life situations, overcoming the limited ecological validity of conventional task-based paradigms (Sonkusare et al., 2019). Such naturalistic stimuli also elicit complex cognitive processes that may not be observable when using simplified conventional tasks, such as hierarchical memory in processing unfolding stories (Hasson et al., 2015). Furthermore, NV settings improve subject compliance and engagement, overcoming the unsystematic noise caused by unconstrained brain states in resting-state (RS) settings (Vanderwal et al., 2019). While the neural response during NV tends to synchronise across participants due to exposure to the same stimulus (Hasson et al., 2004), recent studies have found that it still preserves substantial individual differences (Vanderwal et al., 2017; Finn et al., 2020). These advantages make NV fMRI promising for exploring individual differences (Dubois and Adolphs, 2016), particularly in higher-order brain functions and clinical applications (Eickhoff et al., 2020).

NV fMRI signals can be assumed to contain three components: stimulus-evoked activity, spontaneous brain activity and noise, among which the stimulus-evoked activity is generally of primary interest. As the focus of fMRI studies is gradually shifting from groups to individuals, a fundamental question regarding the stimulus-evoked activity is: why do different individual brains produce different responses to the same naturalistic stimulus? One practical way to interpret individual differences in brain measures is to link them to individual behavioural measures (Dubois and Adolphs, 2016). Recent studies have found that individual differences in neural response to the same naturalistic stimulus are associated with individual differences in not only behavioural measures related to the presented stimulus but also intrinsic personal traits (Campbell et al., 2015; Di & Biswal, 2022; Finn et al., 2020; Gruskin et al., 2020). Methodologically, these studies commonly interpret individual differences in stimulus-evoked activity by correlating them with individual behavioural measures for each voxel or region of interest (ROI) separately. However, this approach does not sufficiently uncover the brain-behaviour relationship, because it typically only examines the behavioural relevance of a single voxel or ROI at each time, rather than that of the spatially distributed activity patterns across the whole brain. Furthermore, it does not ensure the generalisability of the learned brain-behaviour relationship to unseen data. Therefore, a novel

computational framework for understanding individual differences in neural response during NV fMRI in relation to behaviour is needed.

A predictive framework may be better suited for investigating the brain-behaviour relationship compared to correlational approaches (Dubois and Adolphs, 2016), as it will allow making predictions for individual subjects from novel data, which has great practical utility particularly in clinical settings. It will also provide important insights into the neural substrates of behaviours and biomarkers of diseases (Rosenberg et al., 2018; Woo et al., 2017). More importantly, execution of complex brain functions often requires participation of multiple brain regions. Examining distributed activity patterns facilitates not only the understanding of brain functional organisation but also mapping between brain function and behaviour (Cohen et al., 2017; Eickhoff et al., 2018; Gonzalez-Castillo et al., 2012). A predictive framework can open the door to a wealth of advanced machine learning models, which will enable investigation of the behavioural relevance of distributed brain activity patterns and identification of more complex brain-behaviour relationships.

Another important consideration for NV fMRI data analysis is separating stimulus-evoked activity from other signal components (Simony & Chang, 2020). Traditional general linear model (GLM)-based approaches have been used to identify the stimulus-evoked activity based on models of relevant features of presented stimulus (e.g., movie annotations) (Lahnakoski et al., 2012). These models can be difficult to obtain due to the dynamic and multimodal nature of naturalistic stimuli. By contrast, data-driven approaches extract the stimulus-evoked signals without the need of explicit descriptions of the stimulus. They leverage the time-locked nature of the stimulus across subjects, typically using the temporal variance shared across different subjects to reflect the stimulus-evoked activity (Nastase et al., 2019). Applications of data-driven approaches have greatly advanced our understanding of brain function in NV settings.

One commonly used data-driven approach is the intersubject correlation (ISC) method (Hasson et al., 2004; Nastase et al., 2019), which quantifies the shared variance across subjects by the correlation of their fMRI time series. However, in individual differences studies, applications of ISC often focus on the relative similarity of brain activity between subjects, rather than provide a measure of brain activity of individuals. While more sophisticated data-driven approaches, such as the shared response model (SRM) (Chen et al., 2015) and hyperalignment (Haxby et al., 2011), can extract stimulus-evoked signals in individuals, they often target fine-grained activity patterns within certain brain regions and have high memory and computational demands. On the other hand, approaches that have been widely used for characterising individual brain activity patterns on RS fMRI data, such as functional connectivity (FC) and independent component analysis (ICA), typically concern the interactions among ROIs or voxels rather than the stimulus-evoked activity.

In this study, we propose a novel approach for understanding individual differences and brain-behaviour relationships on NV fMRI data, called topography-based predictive framework (TOPF). TOPF consists of two components: (i) identifying individual evoked activity topographies across all ROIs in a data-driven manner, and (ii) examining their relationship with individual behaviour via machine learning-based prediction. Specifically, TOPF applies a principal component analysis (PCA) to each ROI separately to identify stimulus-evoked activity time courses shared across subjects by principal components (PCs). Subject-wise PC loadings thereby reflect the expression levels of these shared time courses specific to each subject. The activity topography for each subject is then characterised by the pattern of these PC loadings across all ROIs, termed individual-specific topography (Fig. 1). These topographies intuitively delineate the unique patterns of how strongly each subject's brain activity follows the shared stimulus-evoked activity across the whole brain. A machine learning-based predictive framework is then employed to predict individual behavioural phenotypes based on these topographies.

Capitalising on fMRI data of multiple NV and task-based paradigms from the human connectome project (HCP) (Van Essen et al., 2013), we first show that TOPF can effectively and stably identify stimulus-evoked responses and capture meaningful individual differences therein in a data-driven way. Next, we show that TOPF successfully predicts phenotypes across cognition, emotion and personality on unseen subjects based on their activity topographies. Additionally, we show that prediction performance of TOPF is similar to and mostly better than that of three popular FC-based approaches. Finally, we localise brain regions that are most predictive of the phenotypes and show that the prediction models learned by TOPF provide promising interpretability.

## 2. Results

We used two datasets from the HCP (dataset1, n = 100; dataset2, n = 179; see Tables S1 and S2 for an overview), covering fMRI data from three NV, three task and one RS conditions and eight behavioural phenotypes, to evaluate the TOPF approach. Detailed descriptions of the datasets and methodology of TOPF can be found in Section 4 "Materials and Methods". In this section, we introduce several key concepts in TOPF and then report the results of our analyses.

### 2.1. TOPF identifies shared responses and individual-specific topographies

The first step in TOPF is to identify individual stimulus-evoked activity patterns across ROIs. Broadly, this step is built on an assumption that the observed NV fMRI time series in each ROI contains a stimulus-evoked component, such that it is shared but expressed with different intensities across individuals (Di & Biswal, 2022; Finn et al., 2020). In this study, this time series component and its expression level for each subject are termed shared response and individual-specific expression (IE), respectively. The pattern of the IE values across ROIs for each subject is what we refer to as individual-specific topography. Such individual-specific topographies will be later used as features for phenotype prediction.

Aiming at individual topographies of a reasonable resolution rather than the fine-grained patterns across voxels, we parcellated the whole brain into 268 functionally defined ROIs (Shen et al., 2013) and computed the fMRI time series averaged over voxels within each ROI (Fig. 1A). This reduced the spatial dimensionality of fMRI data, thereby the computational load, and increased the signal-to-noise ratio. TOPF uses PCA to identify the shared response and IEs. For each ROI separately, the voxel-averaged fMRI time series of all subjects were subjected to a PCA after z-score normalisation, where dimensionality reduction was performed on the subject dimension rather than the temporal dimension of the fMRI data (Fig. 1B). The shared response and IEs for each ROI are represented by the detected PC time series and the subject-wise PC loadings, respectively (Fig. 1C). The individual-specific topography of each subject is operationalised as the pattern of that subject's PC loadings across all ROIs (Fig. 1D). As the first PC (PC1) captures the largest temporal variance shared between subjects across the duration of the scan, we assume that PC1 is most likely to reflect the shared stimulus-evoked activity. In the main text we will mainly focus on PC1-based results (see results for the later PCs in Supplementary Information). Results on how different choices of the number of PCs influence prediction performance are shown in Section 2.8.

### 2.2. TOPF effectively identifies task-evoked activity

We used dataset1, containing task-based fMRI data of 100 unrelated subjects, to evaluate the validity of TOPF in identifying stimulus-evoked activity. Two representative tasks of the seven HCP tasks (Barch et al., 2013) that tapped into two different levels of the hierarchy of brain function, i.e., the motor task and the social cognition (social) task, were used for the evaluation. Task-based fMRI data were used because the availability of the temporal structure of these conventional tasks permitted
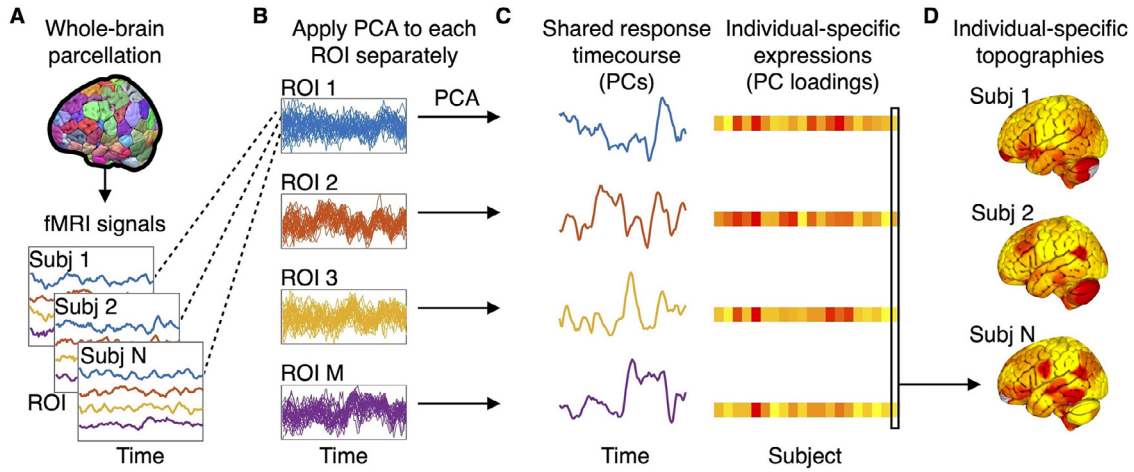
**Fig. 1.** Schematic of TOPF for identifying shared responses and individual-specific topographies. (*A*) For each subject, the whole brain is parcellated into distinct functionally defined ROIs and the fMRI BOLD time series averaged across voxels within each ROI is extracted. (*B*) For each ROI, the extracted fMRI time series is z-score standardised, collected across all subjects and then subjected to a PCA. (*C*) The resultant PCs and PC loadings of each ROI represent the shared response time series and individual-specific expressions (IEs), respectively. Here, we only plot the first PC (PC1) that explains the largest amount of variance for each of these example ROIs for illustration. (*D*) The pattern of the PC loadings of each subject across all ROIs is defined as an individual-specific topography (e.g., marked by the black box).

the application of GLM, which is expected to accurately identify task-evoked activations, thus providing a "ground truth" to evaluate TOPF (Pajula et al., 2012). For each task, a PCA was applied to the time series across the whole task of each ROI separately and the resulting PC1 time series was considered as a shared response for the respective ROI.

We first evaluated the validity of the spatial pattern of evoked activity captured by TOPF. It is assumed that any variance shared across subjects comes from processing the same stimulus or performing the same task at the same time (Hasson et al., 2004; Nastase et al., 2019). Therefore, here we used the variance explained by PC1 to reflect group-level task-evoked "activation". A higher amount of variance explained by PC1 indicates a stronger synchronisation of brain activity across subjects evoked by the task (Di and Biswal, 2022). The map of the variance explained by PC1 across ROIs was compared against a group-level activation (z-score) map derived by GLM (Van Essen et al., 2013). The latter was generated by aggregating the activation maps (i.e., computing the maximum absolute values of the z-scores) across all experimental conditions. A high Pearson's correlation coefficient ($r$) between these two maps was achieved for both tasks (motor: $r = 0.72$; Fig. 2*A*; social: $r = 0.77$; Fig. 2*C*; both $p < e^{-10}$). At an individual subject level, the activity topographies derived by TOPF (i.e., subject-specific PC1 loadings across ROIs) also achieved a moderate correspondence with those derived by GLM (motor: $r = 0.50 \pm 0.15$, social: $r = 0.64 \pm 0.09$; Fig. S1*A* and *B*), with TOPF capturing greater individual differences (Fig. S1*C*). These results indicate that the activity topographies derived by TOPF effectively reflect the spatial patterns of task-evoked activations.

The validity of the shared response time series of each ROI was assessed by comparing it against a combination of models used in GLM using Pearson's correlation. The latter was constructed as the convolution of the timing of the relevant events with a canonical hemodynamic response function (HRF) aggregated across experimental conditions. For both tasks, a high correlation coefficient was observed in those strongly activated ROIs, such as the left premotor cortex (related to tongue movement) (Schubotz et al., 2010) for the motor task ($r = 0.82$; Fig. 2*B*) and the right temporoparietal junction (TPJ, related to social cognition) (Van Overwalle, 2009) for the social task ($r = 0.89$; Fig. 2*D*; see Fig. S2 for results across the whole brain). These results further demonstrate the ability of TOPF for capturing evoked brain activity. Besides, as we expected, PC1 better reflected task-evoked activity than PC2 and PC3 (Fig. S3).

### 2.3. Shared responses are stable across subsamples

In TOPF, we assume that a stable identification of the shared responses is essential for a meaningful characterisation of individual differences (i.e., individual-specific topographies). To evaluate how the shared responses detected by TOPF change with sample composition and sample size, we used NV fMRI data of 179 subjects in dataset2, which were acquired while watching three different movie clips, namely "Two Men" (Movie1), "Welcome to Bridgeville" (Movie2), and "Pockets" (Movie3). For each ROI and movie clip separately, we measured the stability of the derived PC1 across 100 different subsamples over a range of sample sizes ($n$) from 10 to 90 by using Pearson's correlation. To avoid possible bias from the family structure of this dataset (Van Essen et al., 2013), no subjects from the same family were included in the same subsample.

We quantified the overall stability as the mean stability across all ROIs (Fig. 3*A*). As sample size increased, the overall stability also increased and achieved a high value for all three movies with small variability across subsamples at $n = 90$ (Movie1: $0.90 \pm 0.01$; Movie2: $0.83 \pm 0.02$; Movie3: $0.84 \pm 0.02$). In particular, the stability of the PC1 time series achieved above $0.80$ at $n = 90$ for most ROIs in the sensory, frontal and parietal cortices (Fig. 3*B*; Fig. S4). The overall stability was much lower for PC2 and PC3 than for PC1 (Fig. S5).

### 2.4. Individual differences can be captured by individual topographies

As TOPF reflects individual differences by IEs of shared responses rather than directly captures individuals' idiosyncratic stimulus-evoked brain activity, it is necessary to evaluate whether such characterisation leaves enough room for individual variation. The individual-specific topographies derived by TOPF at $n = 90$ on the NV fMRI data stated above were used for this evaluation (see Fig. 4*A* for an illustrative example). For each pair of subjects, we computed their similarity in individual topographies by Pearson's correlation ($r$) within each subsample for each movie clip separately (Fig. 4*B*). On average, the between-subject similarity for the three movies achieved $r = 0.47 \pm 0.14$, $r = 0.41 \pm 0.17$, and $r = 0.38 \pm 0.18$, respectively. This result demonstrates that TOPF-derived individual topographies capture a considerable amount of individual differences. Additional analyses show that these individual differences mainly came from ROIs beyond the sensory cortex (Fig. S6).
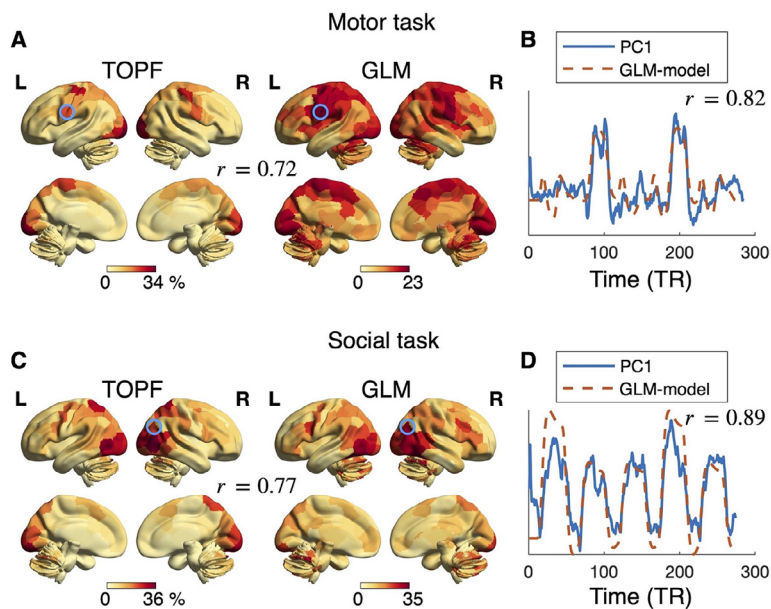
**Fig. 2.** TOPF identifies stimulus-evoked brain activity on task-based fMRI data. Correspondence between TOPF-derived topographies and GLM-derived activation maps, measured by Pearson's correlation coefficient (*r*), for the motor (*A*) and social (*C*) tasks separately at the group level. For TOPF, each value represents the amount of variance explained by the PC1 time series in the given ROI. For GLM, each value represents the activation strength (z-score) aggregated across all experimental conditions within each task (maximum of their absolute values) for the given ROI. The colour from yellow to red indicates the value from low to high. Correspondence (*r*) between the detected PC1 time series (blue) and the model used in GLM (red) for representative ROIs marked in circles in (*A*) and (*C*) for the motor (*B*) and social (*D*) tasks separately. For GLM, the model of each task is computed as the convolution of the HRF with the temporal structure of the given task aggregated over all experimental conditions. For TOPF, the results are computed over the time series across the whole scan of each task. The brain maps are visualised using BrainNet Viewer (Xia et al., 2013).
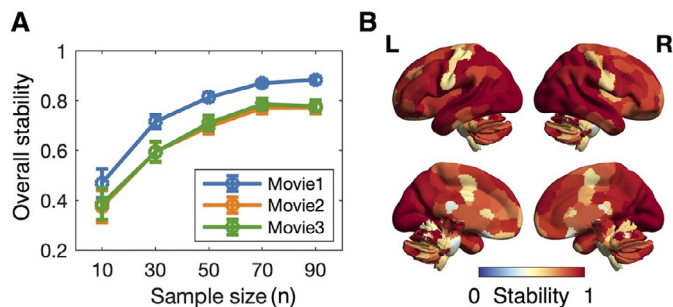


**Fig. 3.** TOPF stably identifies the shared response time series on NV fMRI data. (*A*) Stability of PC1 over a range of sample sizes (*n*) from 10 to 90. The overall stability for PC1 is computed as the stability (i.e., the mean absolute value of the Pearson's correlation coefficients of PC1 over all pairs of subsamples) averaged across all ROIs for each movie clip separately. The error bars depict the standard deviation across subsample pairs. (*B*) Stability map of the PC1 time series across the whole brain at *n* = 90. For each ROI, the stability is averaged across all three movie clips. The colour from blue to red indicates the stability from low to high.

On the other hand, to understand how individual topographies captured by TOPF change across movie clips, we computed the within-subject similarity for each subject and each pair of movie clips on each subsample separately. The mean within-subject similarity (*r*) across subjects and subsamples achieved approximately *r* = 0.50 for all movie-clip pairs (Fig. 4*C*). In particular, the pattern of the PC1 loadings over subjects remained relatively stable across movie clips (Spearman's correlation *r* = 0.44 − 0.75; Fig. S7) for the ROIs in the sensory cortex (Gao et al., 2020). These results suggest that these individual topographies may reflect stable personal traits to a certain degree.

### 2.5. Individual topographies predict individual phenotypes

The second step in TOPF is to link the identified individual-specific topographies to individual phenotypes using a machine learning-based predictive framework. We investigated a total of 8 phenotypes across cognition (fluid intelligence and working memory), personality (openness, agreeableness, conscientiousness, extraversion and neuroticism) and emotion (emotion recognition). For each phenotype, we used a single summary score as its measure (Table S2). For prediction of each
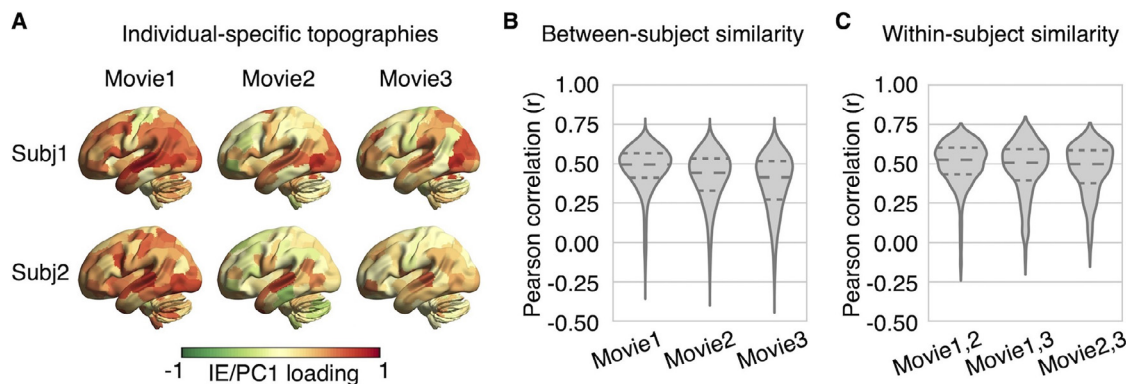


**Fig. 4.** TOPF captures individual differences by individual-specific topographies on NV fMRI data. (*A*) Individual-specific topographies of two representative subjects derived on a randomly selected subsample (*n* = 90) for each movie clip separately. The colour from green to red indicates the IE value (PC1 loading) from low to high. Each subject shows a unique spatial pattern of the IE values while watching the movie clips. (*B*) Distribution of the between-subject similarity (Pearson's correlation coefficient, *r*) in the individual-specific topographies over all pairs of subjects within each subsample (*n* = 90) for each movie clip separately. (*C*) Distribution of the within-subject similarity (*r*) in the individual-specific topography over all subjects within each subsample (*n* = 90) for each pair of movies separately. The three dashed lines inside each violin plot from top to bottom denote the third quartile, median and first quartile of the corresponding distribution, respectively.
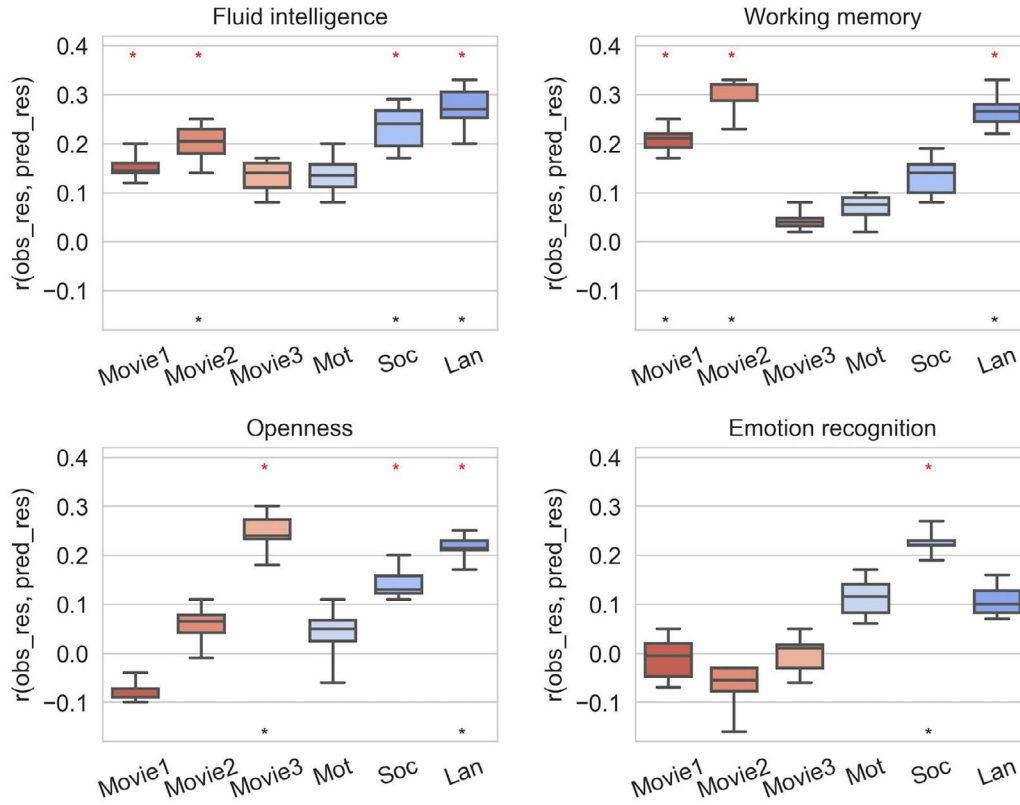
**Fig. 5.** Performance of TOPF for phenotype prediction. Prediction performance is computed by the Pearson's correlation coefficient ($r$) between the residuals of predicted (pred_res) and observed scores (obs_res) over all subjects ($n = 179$) after regressing out sex, age and head motion (RMS-FD). Each box represents results from 10 repetitions of 10-fold CVs of each fMRI paradigm-phenotype combination (box: middle bar, median; upper and lower bars, third and first quartiles; upper and lower whiskers: maximum and minimum). Significant predictions ($p < 0.05$, evaluated by permutation tests with 5000 iterations) before and after FDR correction are marked by red stars and black stars, respectively. NV and task paradigms are marked in red and blue, respectively. Only the phenotypes with significant predictions are shown here (see Tables S4 and S5 for the complete results). Mot: motor task; Soc: social task; Lan: language task.

phenotype, we applied TOPF to fMRI data of 6 different paradigms from the same cohort of subjects ($n = 179$; dataset2) separately, including the three NV paradigms (Movies 1-3) and three conventional tasks of similar scan durations, namely the motor, social, and language tasks (Table S1). Prediction was performed by using a ridge linear regression model for its simplicity and robustness. The model was evaluated by a nested 10-fold cross-validation (CV) procedure with 10 repetitions. In each fold, the model was fitted on the training set by using the individual-specific topographies (i.e., the PC1 loadings of all ROIs) as features. The hyperparameter was optimised via an inner 5-fold CV. The fitted model was then tested by predicting the phenotypes of the test set. To avoid data leakage from the test set to the training set, features of the test set of each fold were derived based on the shared response learned on the training set, and subjects from the same family were ensured to stay either in the training or the test set. Prediction performance was assessed by the Pearson's correlation coefficient ($r$) between predicted and observed scores over all subjects after regressing out potential confounds (age, sex and head motion; Table S3) from both scores (Dinga et al., 2020). Head motion was measured by the relative root-mean-square framewise displacement (RMS-FD).

The prediction performance was significant for fluid intelligence, working memory, openness and emotion recognition (permutation-based $p < 0.05$, 5000 iterations; Fig. 5; see Tables S4 and S5 for results of all phenotypes). Our subsequent analyses will thus mainly focus on these four phenotypes. Notably, the prediction performance varied substantially across fMRI paradigms. For example, working memory and openness were best predicted by Movie2 ($r = 0.30 \pm 0.04$, permutation-based $p < 0.001$) and Movie3 ($r = 0.24 \pm 0.04$, $p = 0.002$), respectively, whereas fluid intelligence and emotion recognition were best predicted by the

language task ($r = 0.27 \pm 0.04$, $p = 0.001$) and social task ($r = 0.23 \pm 0.02$, $p = 0.002$), respectively. Similar results were obtained before confound removal (Table S6), using the leave-one-out CV (Table S7), and after we controlled for the scan length by truncating all fMRI data to 3 mins (Fig. S8).

### 2.6. Comparisons with FC-based predictions

To further validate the utility of TOPF for phenotype prediction, we compared TOPF with three commonly used prediction approaches in fMRI studies that employ different types of FC-based features. These include whole-brain connectome-based prediction (WConn), nodal connectivity strength-based prediction (NConn), and connectome-based predictive modelling (CPM) (Finn et al., 2015; Shen et al., 2017). To facilitate comparisons across methods, ridge regression models were applied for WConn and NConn. Performance of all the three approaches was evaluated via the same procedure as used by TOPF.

Overall, TOPF outperformed all the FC-based approaches across the four phenotypes for all the fMRI paradigms except Movie3 (social: TOPF vs. WConn: $p = 0.011$, TOPF vs. NConn: $p = 0.001$, TOPF vs. CPM: $p = 0.049$, language: TOPF vs. NConn: $p = 0.041$, others: $p > 0.05$, corrected resampled t-tests (Nadeau & Bengio, 2003); Fig. 6A). For individual phenotypes, TOPF exhibited better performance than the other approaches for predicting the fluid intelligence, openness and emotion recognition scores across all fMRI paradigms (openness: TOPF vs. WConn/NConn: both $p = 0.036$, others: $p > 0.05$; Fig. 6B). While the best approach varied across individual fMRI paradigms and phenotypes, TOPF achieved the best prediction performance for all the four phenotypes (Fig. S9). Moreover, performance of applying TOPF on NV and task data also out-

**Fig. 6.** Comparisons between TOPF and FC-based prediction approaches. Each method is evaluated via 10 repetitions of a 10-fold CV for each fMRI paradigm and phenotype separately. Results averaged (*A*) over the four phenotypes and (*B*) over the fMRI paradigms are shown. Prediction performance is measured as the Pearson's correlation coefficient (*r*) between the residuals of predicted (pred_res) and observed scores (obs_res) after regressing out sex, age and head motion (RMS-FD) over all subjects (*n* = 179). Boxes: upper and lower whiskers: maximum and minimum; bars within each box from top to bottom: third quartile, median and first quartile. Statistical significance between TOPF and each of the other approaches is examined by using a corrected resampled paired t-test (Nadeau & Bengio, 2003). *: $p < 0.05$.

performed that of applying the FC-based approaches on RS data (Fig. S10).

### 2.7. Phenotype-related brain regions

To understand the neurobiological interpretation behind the prediction models obtained by TOPF, we identified ROIs that were predictive of these phenotypes based on their permutation feature importance (Breiman, 2001). For each phenotype, we identified the predictive ROIs only for the fMRI paradigms on which the phenotype was significantly predicted. For fluid intelligence, we identified a broad range of predictive ROIs, spreading across the frontal, parietal, temporal and cerebellar cortices (Dubois et al., 2018a) (Fig. 7*A*). For working memory, the identified ROIs were mainly located in the prefrontal, parietal, medial temporal and cerebellar cortices (Rottschy et al., 2012) (Fig. 7*B*). For openness, the identified ROIs were mainly located in the left frontal lobe and cerebellum (Adelstein et al., 2011) (Fig. 7*C*). For emotion recognition, the identified ROIs were mainly located in the right hemisphere, covering the frontal cortex, TPJ and subcortical regions (Ruffman et al., 2008) (Fig. 7*D*). Furthermore, different fMRI paradigms exhibited distinct spatial patterns of predictive ROIs for the same phenotype (Jaccard similarity: 0.02 to 0.10; Fig. 7*E*).

### 2.8. Impact of different choices of the number of PCs on prediction performance

Our previous analyses of TOPF focused on using PC1 time series as shared responses, as we assume that PC1 best reflects the stimulus-evoked activity. Here, we further investigate the impact of different choices of PCs on prediction performance of TOPF. First, we evaluated the number of significant PCs for each ROI and each fMRI paradigm by using a conservative permutation test, where the null distribution was built with the amount of variance explained by PC1 from 10000 permutations (Di & Biswal, 2022; Kauppi et al., 2010); a PC is considered

to be meaningful if the variance explained by it significantly exceeds chance level. For all fMRI paradigms, most ROIs (237 - 266/268) exhibited a significant PC1 ($p < 0.05$; Fig. 8A), 2 - 40 ROIs had a significant PC2, and no ROIs had significant results for the later PCs. Therefore, our following analyses focused on PC1 and PC2.

The variance explained by PC1 varied remarkably across ROIs for all the fMRI paradigms (ranging from 1.6% to 57.0%; Fig. 8B), with the mean across ROIs achieving around 12% for NV and 8% for task data. ROIs for which PC1 explained larger amounts of variance were distributed in visual, auditory and sensory association cortices for NV paradigms and in task-evoked regions for task paradigms (Fig. S11). By contrast, PC2 explained around 4% and 3% of the total variance averaged across ROIs for NV and task data, respectively, with the explained variance of individual ROIs ranging from 1.6% to 8.4%.

For phenotype prediction, the average performance over all fMRI paradigms was better for PC1-based than for PC2-based TOPF for all the four phenotypes (Fig. 8C). As the amount of explained variance may also influence the utility of the PCs for prediction, we additionally conducted a feature-selection step with five different thresholds of explained variance (0%, 3%, 5%, 7%, 10%). The feature selection was conducted across all ROIs for PC1, PC2 and PC1+2 (using loadings of both PC1 and PC2 as features) separately, resulting in 15 different settings. Overall, prediction performance was improved after feature selection, and the best performance was achieved in most cases by using PC1- or PC1+2-based prediction (Fig. S12).

### 3. Discussion

NV settings provide promising opportunities for facilitating our understanding of individual differences in brain functioning, yet they present new challenges for fMRI data analysis. In this study, we propose a simple computational framework, TOPF, which allows us to characterise stimulus-evoked activity topographies in individuals and investigate their behavioural relevance using machine learning-based predic-
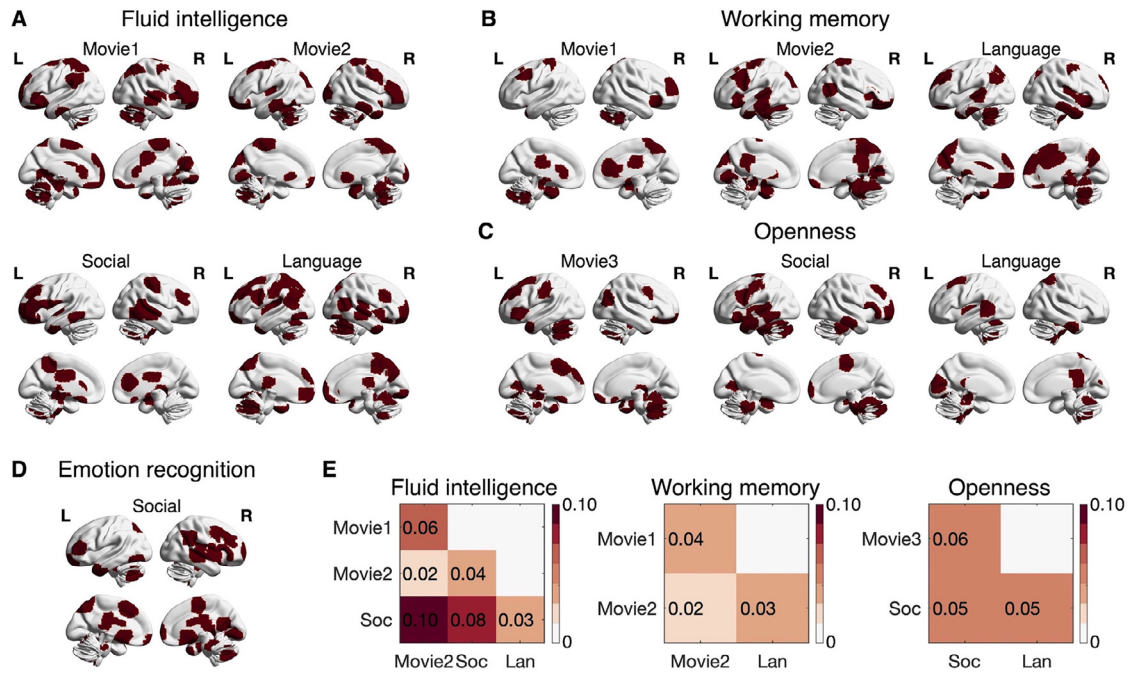
**Fig. 7.** Predictive ROIs for each phenotype and fMRI paradigm with significant predictions. (*A–D*) The predictive ROIs (marked in red) are identified for each fMRI paradigm with significant predictions ($p < 0.05$ in Fig. 5) for fluid intelligence (*A*), working memory (*B*), openness (*C*) and emotion recognition (*D*) separately. Only the ROIs for which the permutation feature importance over all CV folds and repetitions is significantly larger than zero (permutation-based, corrected $p < 0.05$, with 5000 permutations) are identified as predictive ROIs. (*E*) The similarity in the spatial patterns of the predictive ROIs is assessed by the Jaccard similarity for each pair of fMRI paradigms within each phenotype separately. No result is plotted for emotion recognition, as only one fMRI paradigm (the social task) achieves a significant result.
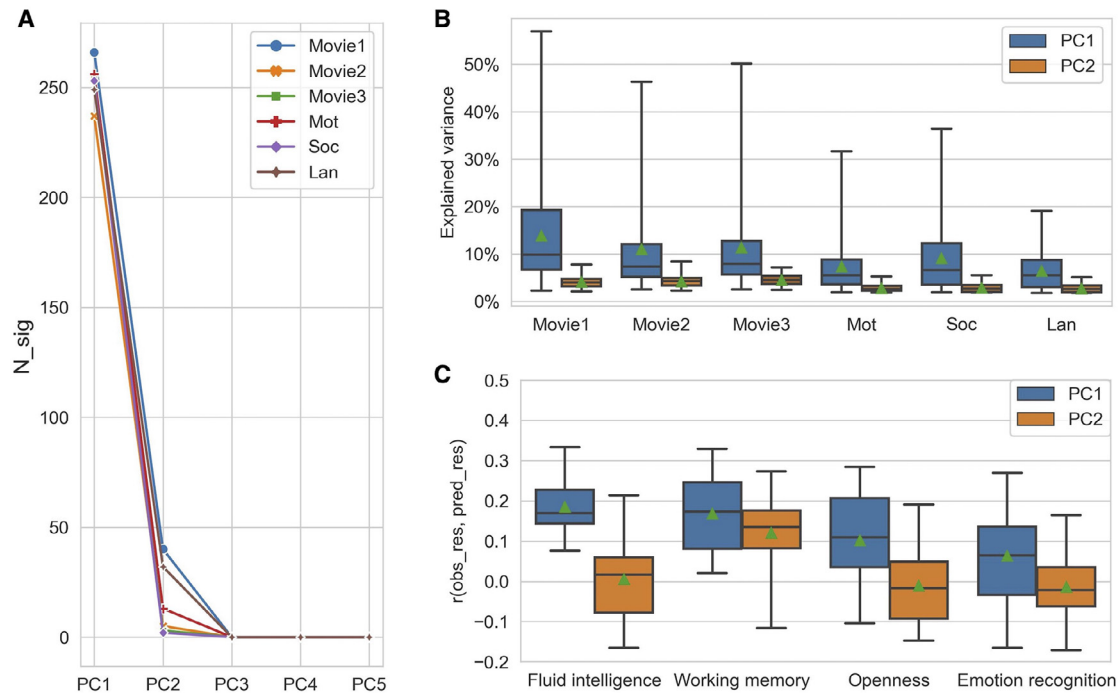


**Fig. 8.** Comparisons across different choices of PCs for TOPF. (*A*) Number of ROIs with significant PCs (N_sig) for each of the first five PCs. Significance of each PC of each ROI is evaluated separately via a conservative permutation test, where the null distribution is built with the amount of variance explained by PC1 of the given ROI from 10000 permutations (Di & Biswal, 2022). The p-value threshold here is set to 0.05. (*B*) Distribution of the amount of variance explained by PC1 and PC2 separately for individual fMRI paradigms. (*C*) Prediction performance of PC1- and PC2-based TOPF separately. Each condition is evaluated via 10 repetitions of 10-fold CVs for each fMRI paradigm and phenotype separately. Results shown are aggregated across all fMRI paradigms for each phenotype. Prediction performance is measured as the Pearson's correlation coefficient (*r*) between the residuals of predicted (pred_res) and observed scores (obs_res) after regressing out sex, age and head motion over all subjects ($n = 179$). Boxes: upper and lower whiskers: maximum and minimum; bars within each box from top to bottom: third quartile, median and first quartile; green triangles: means.

tive modelling. By validating TOPF on both NV and task fMRI data, we demonstrate that the individual-specific topographies conceptualised by TOPF provide a practical characterisation for individual activity patterns during NV. These topographies also predict multiple behavioural phenotypes and often outperform FC-based features. Additionally, the learned predictive models provide neurobiologically meaningful interpretations. TOPF is a generic framework that can be readily adapted for clinical applications, thus holding great potential for advancing basic and clinical neuroscience studies.

### 3.1. Data-driven perspective on stimulus-evoked activity

While classic ICA and FC-based approaches are useful for studying individual differences in the interactions among brain regions, individual differences in brain activity directly driven by the naturalistic stimulus, that is, the stimulus-evoked activity, might be of more interest for NV fMRI studies. On the other hand, the utility of GLM-based approaches to capture individual stimulus-evoked activity on NV fMRI data is limited by their requirement for accurate stimulus descriptions. Our TOPF approach avoids these issues by identifying stimulus-evoked activity in a data-driven way. It detects stimulus-evoked time courses shared across subjects for each ROI separately and characterises individual activity patterns by the individual-specific expression (termed IE) of these shared responses across ROIs (termed individual-specific topography).

Similar to GLM, TOPF also defines a (few) consistent response time course(s) across subjects (TOPF: PC time courses; GLM: predefined regressors) and uses them as a common basis to estimate individual task/stimulus-evoked activity (TOPF: PC loadings; GLM: regression coefficients). However, TOPF, as a data-driven approach, does not rely on prior knowledge about relevant features of the presented naturalistic stimulus for extracting stimulus-evoked signals as GLM does, thus highly suitable for dealing with fMRI data under NV and possibly other complex naturalistic conditions. Notably, whereas GLM picks up individual idiosyncrasies by using predefined models to fit individual fMRI signals, TOPF uses IEs of shared responses to reflect individual variation. Even so, TOPF still leaves enough room for preserving individual differences. Moreover, the data-driven nature may allow TOPF to capture unique individual differences that are not induced by explicitly predefined task designs (Bolt et al., 2018) and thus may not be detected by GLM. Future work could compare the prediction performance of TOPF and GLM-based methods to better understand their differences in the characterisation of individual differences.

TOPF is also closely related to several existing data-driven approaches for detecting stimulus-evoked activity. A common rationale behind these data-driven approaches is that any temporal variance shared across subjects can only originate in the processing of the same stimulus. For example, the commonly used ISC approach often computes the pairwise or leave-one-out (LOO) correlation (i.e., correlation between the fMRI time series of a subject and the time series averaged across the other subjects) to reflect the similarity in brain activity between subjects (Finn et al., 2020; Nastase et al., 2019). Although similar, PC1 loadings have been shown in a previous study to be more computationally efficient than the LOO correlations (Di & Biswal, 2022), thus allowing for more efficient integrations with machine learning models, especially when the number of subjects/ROIs is large. More importantly, TOPF aims to use an "absolute" expression of a group-level basis to reflect individual activity instead of providing a relative measure. While SRM (Chen et al., 2015) and hyperalignment (Haxby et al., 2011) approaches adopt similar logic, they often focus on more fine-grained individual functional topographies within certain brain regions rather than the whole-brain topographies. Tensor-ICA (Beckmann & Smith, 2005; Campbell et al., 2015), on the other hand, identifies a set of common stimulus-evoked spatial components throughout the whole brain and associated time courses, with individual differences being characterised for each component. By contrast, TOPF characterises individual varia-

tion by ROI, delineating individual topographies across the whole brain at a resolution suitable for machine learning-based predictions.

### 3.2. Individual differences in activity topographies evoked during NV

TOPF characterises individual variation by IEs of stimulus-evoked responses that are shared across subjects rather than directly captures individual idiosyncratic responses. Although the latter might be conceptually more desirable, we show that the IEs also captured considerable and meaningful individual differences. These IEs were more variable across subjects in frontoparietal, limbic, subcortical and default mode network regions than in sensory regions. These results partially align with previous studies observing higher intersubject variability in FC patterns of brain regions associated with cognitive control in both RS (Finn et al., 2015; Laumann et al., 2015; Mueller et al., 2013) and NV settings (Vanderwal et al., 2017). Moreover, for the sensory regions, although the intersubject variability was relatively small, patterns of the IEs across subjects were stable across movie clips and thus less sensitive to the specific movie content. This is in line with recent studies showing that the way the brain processes complex sensory and social information during NV may be an intrinsic characteristic of individuals (Gao et al., 2020; Lahnakoski et al., 2012) and associated with brain baseline functional organisation (Gruskin & Patel, 2022). We note that further work is needed to assess whether our findings here are generalisable across different movie types.

### 3.3. Comparisons across fMRI paradigms for TOPF-based phenotype prediction

NV fMRI paradigms performed better than the conventional task paradigms for predicting working memory and openness. This result is in line with previous studies discovering significant relationships between individual differences in brain responses during NV and individual differences in personality traits and working memory capability by using univariate analyses (Finn et al., 2018; Finn et al., 2020). The better performance of these NV paradigms suggests that watching these movies may evoke stronger individual differences relevant to these phenotypes than performing those strictly-controlled tasks. Such advantages of NV settings may arise from improved subject compliance and engagement, reduced head motion during the scan (Vanderwal et al., 2017), as well as presence of brain states and cognitive processes that are uniquely evoked when processing complex naturalistic stimuli (Finn et al., 2017; Hasson et al., 2015; Van der Meer et al., 2020).

Different movies showed substantially different prediction performance, with the observation remaining the same after we controlled for scan (movie) length. When movie length was shortened, the prediction performance slightly degraded in general but in certain cases was even improved. These results suggest that movie length may not be a key factor that determines the utility of a movie for phenotype prediction. The differences in prediction performance across movies may actually result from various factors rather than purely movie length. For instance, movies with richer social content tend to evoke greater inter-subject variability in interpretations of the movies, thus potentially reflecting stronger behaviourally relevant individual differences (Finn et al., 2018; Gruskin et al., 2020). Commercial movies can evoke more reliable responses across subjects than real-life, unedited movies by using professional filmmaking techniques to enhance audiences' engagement (Hasson et al., 2010). Furthermore, familiarity with the movies (e.g., whether subjects have seen the movie before) may also affect subjects' interpretations of the movie (Jääskeläinen et al., 2008), and thus be important to consider when understanding the evoked individual differences in neural responses. Future work is needed to better understand which factors and how they influence prediction performance of behaviour in NV settings by using more movies with different features.

NV fMRI paradigms did not always achieve the best performance for predicting various phenotypes. For fluid intelligence and emotion recog-

nition scores, the language and social tasks performed best, respectively. One distinct feature of these two tasks is that their experimental designs and the brain function they target are closely related to the respective phenotype they predict. Brain activations of language comprehension and maths processing in the HCP language task have shown significant genetic correlations with fluid intelligence (Le Guen et al., 2018). Emotion recognition has been acknowledged as a core aspect of social cognition (Gallese et al., 2004) and the two brain functions have shown overlapping activations in an fMRI study (Mier et al., 2010). These findings suggest that the choice of fMRI paradigms for phenotype prediction should be made individually for each specific research question.

### 3.4. Comparisons across methods for phenotype prediction

Motivated by the development of machine learning techniques, a growing body of work has recently applied machine learning predictive models on various brain measures to predict individual phenotypes (Dosenbach et al., 2010; Finn et al., 2015; Nostro et al., 2018; Weis et al., 2020). Most brain measures used for phenotype prediction in fMRI studies have so far been constructed based on FC. In this study, the overall prediction performance of TOPF across fMRI paradigms and phenotypes was similar to and in most cases better than that of the three popular FC-based prediction approaches (i.e., WConn, NConn and CPM), demonstrating the validity of TOPF. More importantly, this result shows that, beyond FC-based brain measures, individual differences in evoked responses are also useful for phenotype prediction. Whereas FC depicts individual differences in interactions between brain regions, TOPF directly captures individual differences in the activity of each brain region, and thus they may reflect different aspects of brain function. In fact, previous studies have shown that individual phenotypes can be predicted from individual differences in brain activations on task fMRI data (Greene et al., 2020; Sripada et al., 2020), which are computationally different but conceptually similar to the individual differences captured by TOPF.

For predicting fluid intelligence and openness, TOPF on average outperformed the FC-based methods. Many previous studies have reliably predicted fluid intelligence and openness from resting state fMRI data (Smith et al., 2013; Dubois et al., 2018b). Recent evidence has shown that task- or movie-induced brain states can further improve prediction of these phenotypes, by amplifying or bringing out relevant individual differences that are unique to the evoked brain states (Finn and Bandettini, 2021; Greene et al., 2018)). These findings may suggest that TOPF better captured individual differences related to these phenotypes by having a higher level of focus on the evoked brain signals compared to the FC-based methods. For emotion recognition scores, neither TOPF nor the FC-based methods achieved a satisfying prediction performance. This result is consistent with previous findings that emotional traits are in general harder to predict than cognitive traits (Finn & Bandettini, 2021, Kong et al., 2019). One explanation could be that individual differences in emotional traits are not effectively reflected by brain measures in activity and FC. Alternatively, this result may stem from limitations in reliability and validity of the phenotypic measures (Tiego and Fornito, 2022). For predicting working memory scores, whether TOPF outperformed the other methods largely depended on the fMRI paradigms. Previous studies have shown that different tasks or naturalistic stimuli can elicit highly distinct individual differences and brain states (Finn et al., 2017; Hasson et al., 2008a). For example, watching certain movie clips has been shown to elicit a unique hierarchical organisation of working memory (Hasson et al., 2015). On the other hand, activity- and FC-based features may capture unique and complementary information in individual differences that are useful in different situations (Di & Biswal, 2019; Tsvetanov et al., 2018). Further investigation is needed to better understand the differences across fMRI paradigms and between FC- and activity-based features. Future work could also consider combining FC- and activity-based features to further improve prediction performance (Greene et al., 2020).

Computationally, TOPF uses a remarkably smaller number of features ($n$ ROIs) than approaches using whole-brain connectomes as features ($n(n-1)/2$ connections). This feature of TOPF greatly eases the problem of overfitting and reduces the number of observations (i.e., subjects) needed for a meaningful prediction, alleviating the critical problem of relatively small sample sizes of the current public NV fMRI datasets. Furthermore, different from FC-based approaches which are often applied to fMRI data of individual subjects, TOPF captures individual differences via an inter-subject approach. The latter is particularly useful for analysing NV fMRI data, which separates stimulus-driven responses based on their synchronisation across subjects (Hasson et al., 2004; Nastase et al., 2019; Simony & Chang, 2020). In TOPF, we use PCA to detect shared stimulus-evoked responses and their loadings onto subjects' fMRI signals to reflect individual differences, which have shown potential for phenotype prediction in a recent study (Di and Biswal, 2022). It is worth noting that from a computational perspective, TOPF uses PCA to perform a dimensionality reduction on the subject dimension. This distinguishes our study from the majority of fMRI studies which use PCA to reduce the dimensions of spatial or temporal features of fMRI data.

Moreover, there are other inter-subject approaches that could be used for studying individual differences in brain activity during NV from different perspectives. For example, inter-subject functional correlation (Simony et al., 2016) can capture stimulus-driven FC profiles of individuals, while SRM (Chen et al., 2015) and hyperalignment (Haxby et al., 2011) can offer a more fine-grained representation of individual idiosyncrasies in brain functional topographies. A recent study has shown that functionally hyperaligned, fine-grained FC profiles can remarkably improve prediction performance for behaviour on RS and task fMRI data (Feilong et al., 2021). Future work may also integrate these inter-subject approaches into a machine learning framework as a preprocessing or feature extraction step for phenotype prediction on NV fMRI data.

### 3.5. Different choices of number of PCs for TOPF-based phenotype prediction

In general, PC1 outperformed PC2-based TOPF for predicting the four phenotypes. This result fits our expectation that PC1 better reflects shared brain activity across subjects and provides more reliable grounds for characterising meaningful individual differences than the later PCs. Furthermore, the performance of TOPF was improved overall when applying a threshold to preserve only the PCs for which the captured variance exceeded a certain amount. This result further supports our assumption that PCs capturing larger amounts of variance tend to reflect more stable and meaningful individual differences by their loadings. However, we also observed that combining PC1 and PC2-based features can sometimes improve prediction performance. This result suggests that different PCs may reflect different aspects of a cognitive process, e.g., different consistent responses of different subject groups, in particular for clinical and ageing populations (Byrge et al., 2015; Campbell et al., 2015; Di & Biswal, 2022). Therefore, including multiple PCs may provide a more comprehensive characterisation of individual differences across the whole sample and thus benefit behavioural prediction.

### 3.6. Interpretations of identified phenotype-related brain regions

In addition to good prediction performance, we also expected TOPF to provide good interpretability so that it can help understand the brain-behaviour relationship. In general, the predictive brain regions we identified in this study showed overlaps with previous literature. For example, the ROIs identified for fluid intelligence and working memory were mainly located in brain regions supporting cognitive functions, e.g., frontal and parietal cortices (Jung & Haier, 2007), showing overlaps with findings of previous meta-analytic studies (Rottschy et al., 2012).

The association between the left frontal lobe and the openness personality trait has been reported by previous studies using conventional task fMRI data and brain structural data (DeYoung, 2010; Vartanian et al., 2018). The TPJ identified for emotion recognition has been recognised as a key region for social and emotion processing (Van Overwalle, 2009). These results demonstrate that TOPF holds great potential for facilitating our understanding of the neurobiological bases of behaviour. Furthermore, the patterns of predictive ROIs identified for TOPF varied largely across fMRI paradigms, indicating that different tasks and movies may elicit unique neural processes related to the same behaviour or phenotype (Geerligs et al., 2015). We expect that future studies on the important question of whether and how brain function under naturalistic conditions differs from that measured in strictly controlled task paradigms can be facilitated by using our approach.

### 3.7. Limitations and future directions

We note several limitations of this study as well as possible future directions. First, as our interest in this study focuses on individual differences in evoked brain activity, we only included the first two PCs that captured the largest amount of variance in our analyses for TOPF. Previous studies have shown that a low-dimensional representation of brain activity and dynamics can reflect meaningful individual differences, although such representation may capture only a small part of the total variance (Misra et al., 2021; Shine et al., 2019). These findings are similar to our result in this study where the first two PCs on average captured less than 20% the total variance. However, it may still be interesting for future studies on NV fMRI to explore the remaining variance, as it contains a large part of individual idiosyncratic information that could potentially be useful for prediction of phenotypes beyond cognition. Besides, the orthogonality of these PCs may complicate their interpretations. Other methods, such as non-negative matrix factorisation (Lee & Seung, 1999), might be used to detect the shared responses in future work for better interpretability.

Second, although we showed that TOPF can achieve significant performance on both naturalistic paradigms and conventional tasks, we note that these two are essentially different from each other. While naturalistic settings use a common, continuous stimulus for all subjects during the whole scan, tasks often use a block or event-related design. For block-designed tasks, the exact stimulus content and event onset timing within each block may not always be consistent across subjects. However, our results in this study demonstrate that TOPF is not interested in specific events, but rather the performance of the task across the full block. Further investigation is needed to better understand the temporal variance shared across subjects and individual differences captured by TOPF as well as other inter-subject approaches. Future work could also seek to improve temporal alignment across subjects for better performance of inter-subject approaches (Joshi et al., 2018). Besides, whether these data-driven approaches are suitable for analysing event-related designs remains to be tested.

Third, the performance of phenotype prediction achieved up to r = 0.30 in our study, while larger r values have been reported by previous studies. Such a difference may be a consequence of various factors, such as the relatively small size of our sample, whether and how confounding variables are controlled for, whether and how data leakage is dealt with (e.g., controlling for family structure of the HCP samples) and the complexity of machine learning models (e.g., linear regression vs. deep neural networks). Moreover, we used all ROIs as features for prediction in our main analyses. It is possible that the variance captured in some ROIs may not reflect the cognitive processes relevant to the phenotype of interest, but rather reflect, for example, the processing of low-level features of movie stimuli, such as brightness and audio power (Finn & Bandettini, 2021; Hasson et al., 2008b). Future work could further evaluate TOPF (e.g., by using other datasets) and optimise predictive models (e.g., by preserving only phenotype-related features or using more advanced models). We also note that we chose to use simple mod-

els with minimal machine learning steps in this study because our main aim here is to draw attention especially from NV fMRI studies to the utility of individual differences in evoked brain activity for phenotype prediction.

Fourth, we characterised individual differences in brain activity during watching each whole movie clip. Actually, brain activity during NV is highly dynamic and such neural dynamics may also underlie cognitive processing and behaviour (Betzel et al., 2020; Van der Meer et al., 2020). Future work could extend TOPF to capturing individual differences in a finer temporal scale to pursue a better understanding of the brain-behaviour relationships and better prediction performance (Yang et al., 2020).

Finally, whole-brain parcellations play an important role in TOPF because they can reduce the spatial dimensionality of fMRI data in a biologically meaningful way (Eickhoff et al., 2018) and ease the computational load for subsequent analyses. As the best-suited parcellations may vary across different brain states and for answering different research questions (Salehi et al., 2020), how different parcellation schemes influence the performance of TOPF needs further investigation.

### 3.8. Conclusions

In sum, the TOPF approach presented here provides a simple and intuitive tool for studying individual differences in evoked brain activity and their behavioural relevance. Essentially, TOPF highlights the value of investigating whole-brain evoked activity topographies and applying machine learning tools for understanding brain-behaviour relationships on NV fMRI data. Although in this study we only test TOPF on healthy participants, this principled and flexible approach should be readily adapted for clinical applications. We envision that TOPF will provide a powerful tool for not only predicting symptom severity or clinical outcomes but also identifying potential biomarkers in clinical neuroscience studies.

## 4. Materials and Methods

### 4.1. Datasets

All participants used in this study were from the HCP S1200 release (Van Essen et al., 2013). Informed consent was obtained from all participants and data acquisition was approved by the Washington University institutional review board. An overview of all data used in this study can be found in Tables S1 and S2. Detailed acquisition protocols and study designs can be found elsewhere (Barch et al., 2013; Van Essen et al., 2013).

Dataset1 was from the HCP "100 Unrelated Subjects" subset (n = 100; age range: 22-36 years; mean age = 29.11±3.68 years; 54 females/46 males) of the full HCP dataset. This subset was selected because it contained only unrelated subjects, which straightforwardly avoided the possible bias from the genetic relatedness in the full HCP dataset. All subjects were scanned on a 3T Siemens scanner (TR = 720 ms, TE = 33.1 ms, resolution = 2.0 mm$^3$). Each subject was required to complete seven different tasks during the scan in two fMRI sessions, and each task was performed in two runs with different phase-encoding directions. Among these tasks, we chose the motor task and the social cognition (social) task as two representative tasks because they tapped into two different levels of the brain functional hierarchy: whereas the motor task mainly involved simple movements of body parts, the social task involved one higher-order cognitive function of the brain, i.e., the theory of mind ability. For both tasks, we limited our analysis to the data from the first run (with right-to-left phase encoding). The scan duration was 3'34" (284 TRs) for the motor task and 3'27" (274 TRs) for the social task.

Dataset2 was from the HCP 7T subset (n = 184; age range: 22-36 years; mean age = 29.43±3.35 years; 112 females/72 males). Subjects in dataset2 contained twins and siblings from 93 unique families. This

subset included all available subjects in the full HCP dataset that had NV fMRI data. Each subject underwent four NV fMRI runs in two sessions on a 7T Siemens scanner (TR = 1000 ms, TE = 22.2 ms, resolution = 1.6 mm³). Each run lasted approximately 15 mins, where subjects were presented with a sequence of different short movie clips (4-5 clips) interleaved with 20 s rest blocks. Detailed descriptions of the movie clips have been published elsewhere (Finn & Bandettini, 2021). In this study, we limited our analysis to the NV fMRI data from the first run because the movie stimuli used in the other three runs had slight differences in timing across subjects, which may potentially affect the results. Moreover, only the movie clips over 3 mins from the first run were analysed, resulting in three movie clips: "Two Men" (Movie1), "Welcome to Bridgeville" (Movie2), and "Pockets" (Movie3). Their scan durations were 4'04" (244 TRs), 03'41" (221 TRs) and 03'08" (189 TRs), respectively.

Apart from the NV fMRI data, task-based and RS fMRI data of all subjects in dataset2 were also included in our analysis for comparison. All task-based fMRI images were acquired on a 3T Siemens scanner with the same protocols as described above for dataset1. Only the data of the motor (3'34"), social (3'27"), and language tasks (3'57") from the first run (with right-to-left phase encoding) were included. The RS fMRI images were acquired before the NV fMRI run we used in this study in the same session on a 7T Siemens scanner with the same imaging parameters as described above. The scan duration of the RS fMRI run was 15 mins (900 TRs).

A total of 8 different behavioural phenotypes, covering fluid intelligence, working memory, personality and emotion recognition were studied here. Each phenotype was measured by a single summary score for each subject (fluid intelligence: "PMAT24_A_CR", working memory: "ListSort_Unadj", openness: "NEOFAC_O", agreeableness: "NEOFAC_A", conscientiousness: "NEOFAC_C", extraversion: "NEOFAC_E", neuroticism: "NEOFAC_N", and emotion recognition: "ER40_CR"). Descriptive statistics of these 8 measures and their respective behavioural tests are provided in Table S2. Detailed descriptions of the tests to measure these phenotypes can be found elsewhere (Barch et al., 2013). Five subjects were excluded due to the lack of the complete fMRI and behavioural data mentioned above, resulting in a total of 179 subjects (108 females/71 males) for dataset2.

### 4.2. fMRI data preprocessing

All 3T task-based fMRI images were preprocessed with the HCP minimal preprocessing pipeline (Glasser et al., 2013), which includes gradient unwarping, motion correction, spatial normalisation to the Montreal Neurological Institute (MNI) space and intensity normalisation. We further preprocessed these images by regressing out Friston's 24 head motion parameters (Friston et al., 1996), as well as the mean time series of white matter and cerebrospinal fluid and the linear trend, using the Data Processing and Analysis for Brain Imaging (DPABI) toolbox (Yan et al., 2016) (http://rfmri.org/dpabi) in Matlab R2019a. All 7T NV and RS fMRI images were preprocessed with the standard HCP pipelines (Glasser et al., 2013), including correction for distortion and motion, registration to the MNI space, high-pass filtering, removal of 24 motion parameters and FIX-denoising (Salimi-Khorshidi et al., 2014). The first 10 volumes of the NV fMRI data were discarded for obtaining stable signals for each movie clip separately. For all the NV, task-based and RS fMRI images, the whole brain was divided into 268 ROIs using a functionally defined parcellation (Shen et al., 2013) that has been widely applied to the HCP dataset. The parcellation was resampled to match the spatial resolutions of the corresponding fMRI images. Within each ROI, the mean time series over all voxels was extracted and z-score normalised (i.e., zero-mean with unit-variance) for each subject. Head motion was measured by the relative root-mean-square framewise displacement (RMS-FD; "Movement_RelativeRMS_mean.txt") for each subject and each fMRI paradigm. The RMS-FD values of all subjects from

all these fMRI paradigms were less than 0.5 mm, and thus no subject was further excluded from the analysis.

### 4.3. TOPF: Identification of shared responses and individual-specific topographies

The first step of TOPF is to delineate individual stimulus-evoked brain activity patterns across ROIs from NV fMRI data in a data-driven manner. For NV fMRI signals, it is often assumed that, in each brain region, for a given subject $i$, the observed fMRI time series $x_i$ consists of three time series components (Nastase et al., 2019):

$$x_i(t) = c(t) + id_i(t) + \varepsilon_i(t) \tag{1}$$

where $c$ is a stimulus-evoked component that is shared across all subjects (denoted as shared response), $id_i$ is also stimulus-evoked but unique to each individual (denoted as idiosyncratic response), $\varepsilon_i$ is the residual representing the other signal components, and $t$ is a specific time point. However, it is usually difficult to identify the $id_i$s and later link them to other known aspects of the individuals (e.g., behavioural performance or personal traits). To represent individual differences in these stimulus-evoked responses, recent studies have modified the formulation as follows (Finn et al., 2020; Di & Biswal, 2022):

$$x_i(t) = c(t) + \beta_i id(t) + \varepsilon_i(t). \tag{2}$$

This formulation is based on an assumption that there is some consistent response, $id$, across subjects and it is expressed differently across individuals by $\beta_i$. Note that $id$ and $c$ in this equation is interchangeable as they both represent some consistent response across subjects. Actually, it is possible that there are multiple consistent responses underlying the brain activity across subjects. Therefore, in this study, we further modify the formulation as:

$$x_i(t) = \beta_{i1}c_1(t) + \beta_{i2}c_2(t) + \ldots + \beta_{ik}c_k(t) + \varepsilon_i(t), \tag{3}$$

where each $c_j$ is a consistent response across subjects, which we refer to as a shared response. Each scaler $\beta_{ij}$ represents the individual-specific expression (IE) level of $c_j$ for subject $i$. That means, we define a set of common bases ($c_j$) across subjects to reflect the stimulus-evoked brain activity and summarise individual differences therein as the IE values of each $c_j$, i.e., $\beta_{ij}$.

In this work, we applied PCA to identify the shared responses $c_j$ and their corresponding IE values of each subject $\beta_{ij}$ for each ROI separately. Specifically, for each ROI, we first constructed a data matrix of preprocessed fMRI time series across subjects, $X = [x_1, x_2, \ldots, x_n] \in R^{d \times n}$, where $n$ is the number of subjects, $d$ is the number of time points, and $x_i$ is the z-score normalised time series for subject $i$. A PCA was then performed on $X$, where the subjects were treated as variables and the time points as samples. The data matrix $X$ of each ROI can be written as:

$$X = C W^T, \tag{4}$$

where each column, $c_j$, of $C = [c_1, c_2, \ldots, c_p] \in R^{d \times p}$, denotes the $j$-th PC scores across time, each column, $w_j$, of $W = [w_1, w_2, \ldots, w_p] \in R^{n \times p}$, denotes the weights of the $j$-th PC across subjects, and $p$ denotes the total number of PCs. These PCs are ordered according to the amount of variance they explain and $c_1$ explains the largest amount of variance. The time series of subject $i$ ($x_i$) can thus be represented as:

$$x_i = w_{i1}c_1 + w_{i2}c_2 + \ldots + w_{ip}c_p, \tag{5}$$

where $w_{ij}$ represents the weight of subject $i$ in PC $j$. We assume that the larger amount of variance a PC captures, the more likely it reflects a meaningful shared response. Therefore, we considered only the first few PCs that explain the largest amount of variance as shared responses for subsequent analyses. The IE value of PC $j$ in subject $i$, $\beta_{ij}$, was operationalised as the loading of $c_j$ onto $x_i$, which is equivalent to the PC weight $w_{ij}$ multiplied by the square root of the eigenvalue of $c_j$. This procedure was repeated for each ROI separately. Finally, for each subject, we collected the PC loadings of that subject across all ROIs to reflect

a brain activity pattern specific to that subject, a construct we refer to as individual-specific topography. PCA was implemented in Python using the "*sklearn.decomposition.PCA*" function in the scikit-learn package (Pedregosa et al., 2011) (https://scikit-learn.org/stable/).

### 4.4. Validation of TOPF for identifying task-evoked activity

The validity of TOPF in capturing stimulus-evoked activity was assessed on dataset1, which contained fMRI data of two conventional tasks with known timing of individual events. We note that as TOPF is a fully data-driven approach, it does not rely on information of the event structure of the tasks; instead, TOPF derives the topography and the temporal profile of shared responses across the whole scan of each task.

For each task, we applied a PCA to each ROI separately. We constructed a group-level topography across ROIs by using the variance explained by the PC1 time series to reflect group-level "activation" evoked by the task. This topography was then compared against a group-level, task-evoked activation (z-score) map derived by GLM by Pearson's correlation. The GLM-derived map was provided by the HCP and downloaded from NeuroVault (Gorgolewski et al., 2015) (https://neurovault.org). We then conducted similar analyses at an individual-subject level. For each subject, the TOPF-derived individual-specific topography (i.e., PC1 loadings across all ROIs) was compared to the GLM-derived activation map of the given subject (available from the HCP) by Pearson's correlation. The amount of captured individual differences was quantified for each approach by between-subject correlations. Note that the GLM-derived activation maps were initially generated for each experimental condition separately (e.g., left-hand movement or tongue movement), whereas the TOPF-identified topographies reflected the overall brain activity across all experimental conditions within a task. Therefore, before each comparison, we aggregated the GLM-derived activation maps across all experimental conditions by computing the maximum of the absolute activation values for each voxel in FSL (Jenkinson et al., 2012) (http://fsl.fmrib.ox.ac.uk/fsl) and then averaged over all voxels within each ROI.

We next evaluated how well the detected PC1 time series reflected stimulus-evoked activity temporally for each ROI separately. Specifically, for each experimental condition $i$, we modelled a canonical response ($\hat{c}_i$) by convolving the timing of the relevant events with the canonical HRF as implemented by the "*spm_hrf.m*" function of the Statistical Parametric Mapping (SPM) software (Penny et al., 2011) with default settings in Matlab R2019a. For each ROI, we then combined these canonical responses across all experimental conditions with the GLM-derived activations of the given ROI in these conditions as weights:

$$y^j(t) = \sum_i z_i^j \cdot \hat{c}_i(t) \tag{6}$$

where $y^j(t)$ denotes the combined model of ROI $j$ at time point $t$ and $z_i^j$ denotes the GLM-derived activation of ROI $j$ in task condition $i$. Correspondence between the detected PC1 time series and the combined model $y^j$ was measured by Pearson's correlation for each ROI

### 4.5. Stability analysis of shared responses

The stability of the TOPF-derived shared response time series across different subsamples was evaluated on the NV fMRI data in dataset2. For each movie clip, we varied the sample size (number of subjects, $n$) of the input fMRI data matrix from 10 to 90 with a step size of 20 and created 100 different subsamples for each sample size. Only unrelated subjects were included in the same subsample to avoid possible bias from the family structure of the dataset. The stability was first evaluated for each ROI separately at each value of $n$ by computing the mean Pearson's correlation coefficient of the derived PC1 time series across all pairs of subsamples. Given that the derived PC1s may be sign-flipped for some subsamples, we used the absolute correlation value to reflect the

similarity of any two PC time series. The stability was then averaged over all ROIs to reflect the overall stability across the whole brain.

### 4.6. Evaluation of individual differences captured by individual-specific topographies

The ability of TOPF to capture individual differences was assessed by the between-subject similarity of individual-specific topographies (i.e., PC1 loadings across ROIs). The topography was derived by TOPF for each subject on each subsample (as described previously) of the NV fMRI data at $n = 90$. The between-subject similarity in these topographies was computed by using Pearson's correlation for each pair of subjects within each subsample for each movie clip separately. Intersubject variability in brain activity was also quantified for each ROI separately as the standard deviation of its PC1 weights. Note that here we used PC1 weights for a fair comparison across different ROIs instead of PC1 loadings. Likewise, the within-subject similarity was computed by Pearson's correlation coefficient of the topographies between each pair of movie clips for each subject within each subsample separately. The cross-movie stability of the pattern of the PC1 loadings across all subjects was computed by using Spearman's correlation for each ROI and each pair of movie clips on each subsample separately.

### 4.7. TOPF: Phenotype prediction

The second step of TOPF is to investigate the behavioural relevance of the identified individual-specific topographies under a machine learning-based predictive framework. fMRI data from three NV paradigms and three tasks in dataset2 were used to predict eight different phenotypes separately to evaluate the predictive framework. In this study, we used a ridge linear regression model for the prediction. We chose this particular model because it is simple and has been shown to achieve robust performance in various neuroimaging studies (Cui & Gong, 2018; Tian & Zalesky, 2021). We note that other prediction models could also be used within TOPF in general.

Prediction models were trained in a 10-fold cross-validation (CV) scheme. Specifically, all subjects ($n = 179$) were randomly divided into 10 groups of roughly the same size. To account for the family structure of dataset2, subjects from the same family were ensured to be all included in either the training set or the test set for each fold. In each fold, one group was held out for testing, with subjects from the other groups being used for training. Each group was used for testing once and only once. In each fold, the training set was used to train a prediction model, where for each subject all IE values (PC1 loadings) in the individual-specific topography were used as features, resulting in a $1 \times 268$ feature vector. Each feature was then z-score normalised.

We note that to avoid information leakage from the test to the training set, test subjects were totally separated from training subjects during both the feature extraction and phenotype prediction phases. Instead of adding test subjects to the training sample and redoing PCA, the computation of features of test subjects was separate from that of training subjects. Specifically, for each test subject, each feature was computed as the Pearson's correlation coefficient between that subject's fMRI time series and the PC1 time series which was previously derived on the training set of a given ROI. We note that in PCA, a PC loading of a variable (i.e., a training subject in our case) is mathematically equivalent to the Pearson's correlation coefficient between this original variable and the new variable (i.e., the PC) (Kanti et al., 1979). Therefore, the features of both training and test subjects are computed by comparing to the same group-level templates (i.e., the PCs learned on the training sample), thus ensuring their comparability. For each training set, the optimal regularisation parameter $\lambda$ of ridge regression was determined by an inner 5-fold CV from 12 values [$2^{-5}, 2^{-4}, ..., 2^6$]. The whole pipeline was implemented in Python using the scikit-learn and julearn packages (https://juaml.github.io/julearn/main/index.html). For each

fMRI paradigm-phenotype combination, the above procedure was repeated 10 times with different groupings of subjects.

For each computation, prediction performance was measured by the Pearson's correlation coefficient between the predicted and the observed phenotypic scores across all subjects and all 10 folds. To control for the influence of head motion (measured by the RMS-FD values), age and sex, we regressed out these confounds from both the observed and predicted scores and computed the correlation of the two residuals. The correlation was then averaged over the 10 repetitions. Permutation tests with 5000 iterations were used to assess whether the obtained result was significantly higher than the chance level (Dinga et al., 2020). For each iteration, features were randomly shuffled across subjects with the original orders of the confounds and observed scores being kept. A correlation coefficient between predicted and observed scores was computed on the permuted data for each iteration and collected across all iterations to construct the null distribution. A p-value was then obtained by computing the proportion of iterations for which the true correlation was higher than that on the non-permuted data. The prediction performance of TOPF was further evaluated (i) without confound removal and (ii) using leave-one-out CV. Additionally, to rule out the impact of scan length, we further evaluated the prediction performance of TOPF after truncating all fMRI data to the same length. All NV fMRI data were truncated to the length of the shortest movie (Movie3), roughly 3 mins (179 TRs), by removing the last few TRs of each data. Similarly, all task fMRI data were also truncated to the length of 3 mins (250 TRs). Prediction performance was recalculated on the truncated data using the same procedure as described above.

### 4.8. Comparisons with FC-based prediction approaches

Three commonly used FC-based prediction approaches, including the whole-brain connectome-based prediction (WConn), nodal connectivity strength-based prediction (NConn), and connectome-based predictive modelling (CPM) (Shen et al., 2017), were compared with TOPF to further evaluate the utility of TOPF for phenotype prediction. Apart from the six fMRI paradigms (three NV paradigms and three tasks), we further included RS fMRI data of all the subjects in dataset2 (as described above). For each subject and each fMRI paradigm, we computed the FC as the Pearson's correlation coefficient of the fMRI time series for each pair of ROIs separately, resulting in a total of eight $268 \times 268$ connectivity matrices for each subject. Each correlation coefficient was then Fisher z-transformed. For WConn, we extracted the upper triangle of the connectivity matrix without the diagonal as features, yielding a total of 35778 features (connections) for each subject and each fMRI paradigm. For NConn, we computed the node strength for each ROI separately, which is defined as the sum of the absolute values of the connectivities between the given ROI and all the other ROIs, resulting in a total of 268 features (ROIs) for each subject and each fMRI paradigm. The WConn and NConn features were then used to predict fluid intelligence, working memory, openness and emotion recognition on each fMRI paradigm separately by repeating the same prediction pipeline as used in TOPF. CPM uses a linear regression model and applies an additional feature selection step in each fold, where only the features (i.e., connections) that are correlated with the phenotype to be predicted are preserved. The selected features are then grouped into two groups according to whether they are positively and negatively correlated with the phenotype and the sum over all features within each group is then used as a feature for further steps (Shen et al., 2017). We used a correlation threshold of $|r| = 0.2$ for feature selection as used in a previous study (Finn & Bandettini, 2021). Code for implementing CPM was also adapted from that study (Finn & Bandettini, 2021). Prediction performance of CPM was also evaluated via 10 repetitions of a 10-fold CV. Comparisons between methods were conducted by using corrected resampled paired t-tests over scores of all CV folds and repetitions (Nadeau & Bengio, 2003).

### 4.9. Identification of phenotype-related brain regions

Permutation feature importance (Breiman, 2001) was computed for each ROI in each fMRI paradigm-phenotype combination with significant predictions separately. Specifically, for each successful combination, we collected the models fitted on the training sets across all CV folds and all repetitions, resulting in a total of 100 models. For each ROI and each model, we randomly shuffled the observations of the given ROI in the test set and recomputed the prediction performance. After repeating the procedure 100 times, the importance of each ROI in the given model was then quantified as the decrease in prediction performance (i.e., the difference between the true performance and the performance derived on the permuted test set), averaged across the 100 iterations. This analysis was implemented in Python using the "*permutation_importance*" function in the scikit-learn package. Finally, for each successful fMRI paradigm-phenotype combination, a right-tailed, one-sample permutation test (Manly, 2007) with 5000 iterations was used to identify the ROIs for which the importance values over all the models were significantly larger than zero (corrected $p < 0.05$). The correction for multiple comparisons was performed against a null distribution of the maximum value across all ROIs in each iteration (Nichols & Holmes, 2002). We note that in this study we used the permutation feature importance instead of the feature weights (i.e., regression coefficients) to identify the predictive features because the latter are known to have certain limitations in their interpretability (Haufe et al., 2014; Tian & Zalesky, 2021). The similarity of the spatial patterns of the identified predictive ROIs between fMRI paradigms was computed by the Jaccard similarity index for each pair of fMRI paradigms with significant predictions for each phenotype separately.

### 4.10. Selection of PCs for TOPF

The statistical significance of each PC time series derived by TOPF was evaluated by using a conservative permutation test (Di & Biswal, 2022). This analysis determined whether the amount of variance explained by each PC time series exceeded chance level. Specifically, for each ROI, we permuted the input fMRI data matrix $X$ by applying circular shifting. The fMRI time series of each subject was shifted with a random time interval so that the fMRI time series across subjects were mismatched in time while the autocorrelation structure of each fMRI time series was preserved (Kauppi et al., 2010). Next, a PCA was applied to the permuted data matrix and the procedure was repeated 10000 times. A null distribution was then constructed by collecting the variance explained by PC1 from the 10000 iterations. The p-value of each PC of the given ROI was determined as the proportion of iterations on which the explained variance was larger than the true variance explained by the PC.

In addition to using PC1 loadings, prediction performance of TOPF was further evaluated by using PC2 loadings (268 features) and the combination of PC1 and PC2 loadings (268*2 features; PC1+2) as features, together with five different thresholds of explained variance (0%, 3%,5%,7%,10%), resulting in a total of 15 different settings. In each CV fold, only the features for which the corresponding PC time series explained a larger amount of variance than the threshold were used in further steps for phenotype prediction. All the other procedures were the same as described in Section 4.7.

### Acknowledgments

McDonnell Center for Systems Neuroscience at Washington University. The authors thank the anonymous reviewers for their helpful feedback.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Credit authorship contribution statement

**Xuan Li:** Conceptualization, Data curation, Methodology, Formal analysis, Writing – original draft, Visualization. **Patrick Friedrich:** Software, Investigation, Writing – review & editing. **Kaustubh R. Patil:** Methodology, Software, Writing – review & editing. **Simon B. Eickhoff:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. **Susanne Weis:** Conceptualization, Methodology, Funding acquisition, Supervision, Writing – review & editing.

## Data availability

The HCP data used in this study (preprocessed task-based and NV fMRI data, behavioural data, and GLM-derived activation maps) are publicly available from the HCP website (https://db.humanconnectome.org/). Family structure information can be accessed after approval of the HCP Restricted Data Use Terms. The code to implement TOPF and subsequent analysis are available at https://github.com/xuanli-ac/TOPF.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2023.120245.

## References

Adelstein, J.S., Shehzad, Z., Mennes, M., DeYoung, C.G., Zuo, X.N., Kelly, C., Margulies, D.S., Bloomfield, A., Gray, J.R., Castellanos, F.X., 2011. Personality is reflected in the brain's intrinsic functional architecture. PLoS One 6, e27633. doi:10.1371/journal.pone.0027633.

Barch, D.M., Burgess, G.C., Harms, M.P., Petersen, S.E., Schlaggar, B.L., Corbetta, M., Glasser, M.F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J.M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A.Z., Van Essen, D.C., 2013. Function in the human connectome: Task-fMRI and individual differences in behavior. Neuroimage 80, 169–189. doi:10.1016/j.neuroimage.2013.05.033.

Beckmann, C.F., Smith, S.M., 2005. Tensorial extensions of independent component analysis for multisubject FMRI analysis. Neuroimage 25, 294–311. doi:10.1016/j.neuroimage.2004.10.043.

Betzel, R.F., Byrge, L., Esfahlani, F.Z., Kennedy, D.P., 2020. Temporal fluctuations in the brain's modular architecture during movie-watching. Neuroimage 213, 116687. doi:10.1016/j.neuroimage.2020.116687.

Bolt, T., Nomi, J.S., Vij, S.G., Chang, C., Uddin, L.Q., 2018. Inter-subject phase synchronization for exploratory analysis of task-fMRI. Neuroimage 176, 477–488. doi:10.1016/j.neuroimage.2018.04.015.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. doi:10.1023/A:1010933404324.

Byrge, L., Dubois, J., Tyszka, J.M., Adolphs, R., Kennedy, D.P., 2015. Idiosyncratic brain activation patterns are associated with poor social comprehension in autism. J. Neurosci. 35, 5837–5850. doi:10.1523/JNEUROSCI.5182-14.2015.

Campbell, K.L., Shafto, M.A., Wright, P., Tsvetanov, K.A., Geerligs, L., Cusack, R., Tyler, L.K., Brayne, C., Bullmore, E., Calder, A., Dalgleish, T., Duncan, J., Henson, R., Matthews, F., Marslen-Wilson, W., Rowe, J., Cheung, T., Davis, S., Kievit, R., McCarrey, A., Price, D., Taylor, J., Williams, N., Bates, L., Emery, T., Erzinçlioglu, S., Gadie, A., Gerbase, S., Georgieva, S., Hanley, C., Parkin, B., Troy, D., Allen, J., Amery, G., Amunts, L., Barcroft, A., Castle, A., Dias, C., Dowrick, J., Fair, M., Fisher, H., Goulding, A., Grewal, A., Hale, G., Hilton, A., Johnson, F., Johnston, P., Kavanagh-Williamson, T., Kwasniewska, M., McMinn, A., Norman, K., Penrose, J., Roby, F., Rowland, D., Sargeant, J., Squire, M., Stevens, B., Stoddart, A., Stone, C., Thompson, T., Yazlik, O., Dixon, M., Barnes, D., Hillman, J., Mitchell, J., Villi, L., 2015. Idiosyncratic responding during movie-watching predicted by age differences in attentional control. Neurobiol. Aging 36, 3045–3055. doi:10.1016/j.neurobiolaging.2015.07.028.

Chen, P.H., Chen, J., Yeshurun, Y., Hasson, U., Haxby, J.V., Ramadge, P.J., 2015. A reduced-dimension fMRI shared response model. In: Proceedings of the Advances in Neural Information Processing Systems, 28.

Cohen, J.D., Daw, N., Engelhardt, B., Hasson, U., Li, K., Niv, Y., Norman, K.A., Pillow, J., Ramadge, P.J., Turk-Browne, N.B., Willke, T.L., 2017. Computational approaches to fMRI analysis. Nat. Neurosci. 20, 304–313. doi:10.1038/nn.4499.

Cui, Z., Gong, G., 2018. The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. Neuroimage 178, 622–637. doi:10.1016/j.neuroimage.2018.06.001.

DeYoung, C.G., 2010. Personality neuroscience and the biology of traits. Soc. Person. Psychol. Compass 4, 1165–1180. doi:10.1111/j.1751-9004.2010.00327.x.

Di, X., Biswal, B.B., 2019. Toward task connectomics: examining whole-brain task modulated connectivity in different task domains. Cereb. Cortex 29, 1572–1583. doi:10.1093/cercor/bhy055.

Di, X., Biswal, B.B., 2022. Principal component analysis reveals multiple consistent responses to naturalistic stimuli in children and adults. Hum. Brain Mapp. 43, 3332–3345. doi:10.1002/hbm.25568.

Dinga, R., Schmaal, L., Schmaal, Penninx, B.W.J.H., Veltman, D.J., Marquand, A.F., 2020. Controlling for effects of confounding variables on machine learning predictions. bioRxiv doi:10.1101/2020.08.17.255034.

Dosenbach, N.U., Nardos, B., Cohen, A.L., Fair, D.A., Power, J.D., Church, J.A., Nelson, S.M., Wig, G.S., Vogel, A.C., Lessov-Schlaggar, C.N., 2010. Prediction of individual brain maturity using fMRI. Science 329, 1358–1361. doi:10.1126/science.1194144.

Dubois, J., Galdi, P., Paul, L.K., Adolphs, R., 2018a. A distributed brain network predicts general intelligence from resting-state human neuroimaging data. Philos. Trans. R. Soc. B Biol. Sci. 373. doi:10.1098/rstb.2017.0284.

Dubois, J., Adolphs, R., 2016. Building a science of individual differences from fMRI. Trends Cogn. Sci. 20, 425–443.

Dubois, J., Galdi, P., Han, Y., Paul, L.K., Adolphs, R., 2018b. Resting-state functional brain connectivity best predicts the personality dimension of openness to experience. Pers. Neurosci. 1, e6. doi:10.1017/pen.2018.8.

Eickhoff, S.B., Constable, R., Yeo, B.T., 2018. Topographic organization of the cerebral cortex and brain cartography. Neuroimage 170, 332–347. doi:10.1016/J.NEUROIMAGE.2017.02.018.

Eickhoff, S. B., Milham, M., Vanderwal, T., 2020. Towards clinical applications of movie fMRI. Neuroimage 217, 116860.

Feilong, M., Guntupalli, J.S., Haxby, J.V., 2021. The neural basis of intelligence in fine-grained cortical topographies. eLife 10, e64058. doi:10.7554/eLife.64058.

Finn, E.S., Bandettini, P.A., 2021. Movie-watching outperforms rest for functional connectivity-based prediction of behavior. Neuroimage 235, 117963. https://linkinghub.elsevier.com/retrieve/pii/S1053811921002408.

Finn, E.S., Corlett, P.R., Chen, G., Bandettini, P.A., Constable, R.T., 2018. Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. Nat. Commun. 9, 1–13. doi:10.1038/s41467-018-04387-2.

Finn, E.S., Glerean, E., Khojandi, A.Y., Nielson, D., Molfese, P.J., Handwerker, D.A., Bandettini, P.A., 2020. Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. Neuroimage 215, 116828. doi:10.1016/j.neuroimage.2020.116828.

Finn, E.S., Scheinost, D., Finn, D.M., Shen, X., Papademetris, X., Constable, R.T., 2017. Can brain state be manipulated to emphasize individual differences in functional connectivity?. Neuroimage 160, 140–151. doi:10.1016/j.neuroimage.2017.03.064.

Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, X., Constable, R.T., 2015. Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. Nat. Neurosci. 18, 1664–1671. doi:10.1038/nn.4135.

Friston, K.J., Williams, S., Howard, R., Frackowiak, R.S., Turner, R., 1996. Movement-related effects in fMRI time-series. Magn. Reson. Med. 35, 346–355. doi:10.1002/mrm.1910350312.

Gallese, V., Keysers, C., Rizzolatti, G., 2004. A unifying view of the basis of social cognition. Trends Cogn. Sci. 8, 396–403. doi:10.1016/j.tics.2004.07.002.

Gao, J., Chen, G., Wu, J., Wang, Y., Hu, Y., Xu, T., Zuo, X.N., Yang, Z., 2020. Reliability map of individual differences reflected in inter-subject correlation in naturalistic imaging. Neuroimage 223, 117277. doi:10.1016/J.NEUROIMAGE.2020.117277.

Geerligs, L., Rubinov, M., Tyler, L.K., Brayne, C., Bullmore, E.T., Calder, A.C., Cusack, R., Dalgleish, T., Duncan, J., Henson, R.N., Matthews, F.E., Marslen-Wilson, W.D., Rowe, J.B., Shafto, M.A., Campbell, K., Cheung, T., Davis, S., Geerligs, L., Kievit, R., McCarrey, A., Mustafa, A., Price, D., Samu, D., Taylor, J.R., Treder, M., Tsvetanov, K., van Belle, J., Williams, N., Bates, L., Emery, T., Erzinçlioglu, S., Gadie, A., Gerbase, S., Georgieva, S., Hanley, C., Parkin, B., Troy, D., Auer, T., Correia, M., Gao, L., Green, E., Henriques, R., Allen, J., Amery, G., Amunts, L., Barcroft, A., Castle, A., Dias, C., Dowrick, J., Fair, M., Fisher, H., Goulding, A., Grewal, A., Hale, G., Hilton, A., Johnson, F., Johnston, P., Kavanagh-Williamson, T., Kwasniewska, M., McMinn, A., Norman, K., Penrose, J., Roby, F., Rowland, D., Sargeant, J., Squire, M., Stevens, B., Stoddart, A., Stone, C., Thompson, T., Yazlik, O., Barnes, D., Dixon, M., Hillman, J., Mitchell, J., Villis, L., Henson, R.N., 2015. State and trait components of functional connectivity: individual differences vary with mental state. J. Neurosci. 35, 13949–13961. doi:10.1523/JNEUROSCI.1324-15.2015.

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Van Essen, D.C., Jenkinson, M., 2013. The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage 80, 105–124. doi:10.1016/j.neuroimage.2013.04.127.

Gorgolewski, K.J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S.S., Maumet, C., Sochat, V.V., Nichols, T.E., Poldrack, R.A., Poline, J.- B., 2015. NeuroVault. org: A web-based repository for collecting and sharing unthresholded statistical maps of the human brain. Front. Neuroinform. 9, 8. doi:10.3389/fninf.2015.00008.

Greene, A.S., Gao, S., Noble, S., Scheinost, D., Constable, R.T., 2020. How tasks change whole-brain functional organization to reveal brain-phenotype relationships. Cell Rep. 32, 108066. doi:10.1016/J.CELREP.2020.108066.

Greene, A.S., Gao, S., Scheinost, D., et al., 2018. Task-induced brain state manipulation improves prediction of individual traits. Nat Commun 9, 2807. doi:10.1038/s41467-018-04920-3.

Gruskin, D.C., Patel, G.H., 2022. Brain connectivity at rest predicts individual differences in normative activity during movie watching. Neuroimage 253, 119100. doi:10.1016/j.neuroimage.2022.119100.

Gruskin, D.C., Rosenberg, M.D., Holmes, A.J., 2020. Relationships between depressive symptoms and brain responses during emotional movie viewing emerge in adolescence. Neuroimage 216, 116217. doi:10.1016/j.neuroimage.2019.116217.

Hasson, U., Chen, J., Honey, C.J., 2015. Hierarchical process memory: Memory as an integral component of information processing. Trends Cogn. Sci. 19, 304–313. doi:10.1016/j.tics.2015.04.006.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., Heeger, D.J., 2008a. Neurocinematics: The neuroscience of film. Projections 2, 1–26. doi:10.3167/proj.2008.020102.

Hasson, U., Malach, R., Heeger, D.J., 2010. Reliability of cortical activity during natural stimulation. Trends Cogn. Sci. 14, 40–48. doi:10.1016/j.tics.2009.10.011.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity during natural vision. Science 303, 1634–1640. doi:10.1126/science.1089506.

Hasson, U., Yang, E., Vallines, I., Heeger, D.J., Rubin, N., 2008b. A hierarchy of temporal receptive windows in human cortex. J. Neurosci. 28, 2539–2550. doi:10.1523/JNEUROSCI.5487-07.2008.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.D., Blankertz, B., Bießmann, F., 2014. On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 87, 96–110. doi:10.1016/j.neuroimage.2013.10.067.

Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., Hanke, M., Ramadge, P.J., 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron 72, 404–416. doi:10.1016/j.neuron.2011.08.026.

Jääskeläinen, I.P., Koskentalo, K., Balk, M.H., Autti, T., Kauramäki, J., Pomren, C., Sams, M., 2008. Inter-subject synchronization of prefrontal cortex hemodynamic activity during natural viewing. Open Neuroimaging J. 2, 14–19.

Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. Fsl. Neuroimage 62, 782–790. doi:10.1016/j.neuroimage.2011.09.015.

Joshi, A.A., Chong, M., Li, J., Choi, S., Leahy, R.M., 2018. Are you thinking what I'm thinking? Synchronization of resting fMRI time-series across subjects. Neuroimage 172, 740–752. doi:10.1016/j.neuroimage.2018.01.058.

Jung, R.E., Haier, R.J., 2007. The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. Behav. Brain Sci. 30, 135–154. doi:10.1017/S0140525X07001185.

Kanti, M., Kent, J., Bibby, J., 1979. Multivariate analysis. Academic Press.

Kauppi, J.P., Jääskeläinen, I.P., Sams, M., Tohka, J., 2010. Inter-subject correlation of brain hemodynamic responses during watching a movie: Localization in space and frequency. Front. Neuroinform. 4. doi:10.3389/fninf.2010.00005.

Kong, R., Li, J., Orban, C., Sabuncu, M.R., Liu, H., Schaefer, A., Schaefer, A., Sun, N., Zuo, X.N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T., 2019. Spatial topography of individual-specific cortical networks predicts human cognition, personality, and emotion. Cereb. Cortex 29, 2533–2551. doi:10.1093/cercor/bhy123.

Lahnakoski, J.M., Glerean, E., Salmi, J., Jääskeläinen, I.P., Sams, M., Hari, R., Nummenmaa, L., 2012. Naturalistic FMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. Front. Hum. Neurosci. 6, 233. doi:10.3389/fnhum.2012.00233.

Laumann, T.O., Gordon, E.M., Adeyemo, B., Snyder, A.Z., Joo, S.J., Chen, M.Y., Gilmore, A.W., McDermott, K.B., Nelson, S.M., Dosenbach, N.U.F., Schlaggar, B.L., Mumford, J.A., Poldrack, R.A., Petersen, S.E., 2015. Functional system and areal organization of a highly sampled individual human brain HHS public access. Neuron 87, 657–670. doi:10.1016/j.neuron.2015.06.037.

Lee, D.D., Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791.

Le Guen, Y., Amalric, M., Pinel, P., Pallier, C., Frouin, V., 2018. Shared genetic aetiology between cognitive performance and brain activations in language and math tasks. Sci. Rep. 8, 1–11. doi:10.1038/s41598-018-35665-0.

Manly, B.F.J., 2007. Randomization, Bootstrap and Monte Carlo Methods in Biology. Chapman and Hall.

Mier, D., Lis, S., Neuthe, K., Sauer, C., Esslinger, C., Gallhofer, B., Kirsch, P., 2010. The involvement of emotion recognition in affective theory of mind. Psychophysiology 47, 1028–1039. doi:10.1111/j.1469-8986.2010.01031.x.

Misra, J., Surampudi, S.G., Venkatesh, M., Limbachia, C., Jaja, J., Pessoa, L., 2021. Learning brain dynamics for decoding and predicting individual differences. PLoS Comput. Biol. 17, e1008943. doi:10.1371/journal.pcbi.1008943.

Mueller, S., Wang, D., Fox, M.D., Yeo, B.T., Sepulcre, J., Sabuncu, M.R., Shafee, R., Lu, J., Liu, H., 2013. Individual variability in functional connectivity architecture of the human brain. Neuron 77, 586–595. doi:10.1016/j.neuron.2012.12.028.

Nadeau, C., Bengio, Y., 2003. Inference for the generalization error. Mach. Learn. 52, 239–281. doi:10.1023/A:1024068626366.

Nastase, S.A., Gazzola, V., Hasson, U., Keysers, C., 2019. Measuring shared responses across subjects using intersubject correlation. Soc. Cogn. Affect. Neurosci. 14, 669–687. doi:10.1093/scan/nsz037.

Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: A primer with examples. Hum. Brain Mapp. 15, 1–25. doi:10.1002/hbm.1058.

Nostro, A.D., Müller, V.I., Varikuti, D.P., Pläschke, R.N., Hoffstaedter, F., Langner, R., Patil, K.R., Eickhoff, S.B., 2018. Predicting personality from network-based resting-state functional connectivity. Brain Structure and Function 223, 2699–2719. doi:10.1007/s00429-018-1651-z.

Pajula, J., Kauppi, J.P., Tohka, J., 2012. Inter-subject correlation in fMRI: Method validation against stimulus-model based analysis. PLoS One 7, 41196. doi:10.1371/journal.pone.0041196.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Penny, W.D., Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., 2011. Statistical Parametric Mapping: The Analysis of Functional Brain Images. Elsevier.

Rosenberg, M.D., Casey, B.J., Holmes, A.J., 2018. Prediction complements explanation in understanding the developing brain. Nat. Commun. 9, 1–13. doi:10.1038/s41467-018-02887-9.

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A.R., Schulz, J.B., Fox, P.T., Eickhoff, S.B., 2012. Modelling neural correlates of working memory: A coordinate-based meta-analysis. Neuroimage 60, 830–846. doi:10.1016/j.neuroimage.2011.11.050.

Ruffman, T., Henry, J.D., Livingstone, V., Phillips, L.H., 2008. A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. Neurosci. Biobehav. Rev. 32, 863–881. doi:10.1016/j.neubiorev.2008.01.001.

Salehi, M., Greene, A.S., Karbasi, A., Shen, X., Scheinost, D., Constable, R.T., 2020. There is no single functional atlas even for a single individual: Functional parcel definitions change with task. Neuroimage 208, 116366. doi:10.1016/J.NEUROIMAGE.2019.116366.

Salimi-Khorshidi, G., Douaud, G., Beckmann, C.F., Glasser, M.F., Griffanti, L., Smith, S.M., 2014. Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. Neuroimage 90, 449–468. doi:10.1016/j.neuroimage.2013.11.046.

Schubotz, R.I., Anwander, A., Knösche, T.R., von Cramon, D.Y., Tittgemeyer, M., 2010. Anatomical and functional parcellation of the human lateral premotor cortex. Neuroimage 50, 396–408. doi:10.1016/j.neuroimage.2009.12.069.

Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., Constable, R. T., 2017. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. Nature Protocols 12, 506–518. doi:10.1038/nprot.2016.178.

Shen, X., Tokoglu, F., Papademetris, X., Constable, R.T., 2013. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. Neuroimage 82, 403–415. doi:10.1016/j.neuroimage.2013.05.081.

Shine, J.M., Breakspear, M., Bell, P.T., Ehgoetz Martens, K., Shine, R., Koyejo, O., Sporns, O., Poldrack, R.A., 2019. Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. Nat. Neurosci. 22, 289–296. doi:10.1038/s41593-018-0312-0.

Simony, E., Chang, C., 2020. Analysis of stimulus-induced brain dynamics during naturalistic paradigms. Neuroimage 216, 116461. doi:10.1016/j.neuroimage.2019.116461.

Simony, E., Honey, C.J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., Hasson, U., 2016. Dynamic reconfiguration of the default mode network during narrative comprehension. Nat. Commun. 7, 12141. doi:10.1038/ncomms12141.

Smith, S.M., Vidaurre, D., Beckmann, C.F., Glasser, M.F., Jenkinson, M., Miller, K.L., Robinson, E.C., Salimi-Khorshidi, G., Woolrich, M.W., Barch, D.M., Uğurbil, K., Van Essen, D.C., 2013. Functional connectomics from resting-state fMRI. Trends Cogn. Sci. 17, 666–682.

Sonkusare, S., Breakspear, M., Guo, C., 2019. Naturalistic stimuli in neuroscience: critically acclaimed. Trends Cogn. Sci. 23, 699–714. doi:10.1016/j.tics.2019.05.004.

Sripada, C., Angstadt, M., Rutherford, S., Taxali, A., Shedden, K., 2020. Toward a "treadmill test" for cognition: Improved prediction of general cognitive ability from the task activated brain. Hum. Brain Mapp. 41, 3186–3197. doi:10.1002/HBM.25007.

Tian, Y., Zalesky, A., 2021. Machine learning prediction of cognition from functional connectivity: Are feature weights reliable? Neuroimage 245, 118648. doi:10.1016/j.neuroimage.2021.118648.

Tiego, J., Fornito, A., 2022. Putting behaviour back into brain-behaviour correlation analyses. doi:10.31219/osf.io/g84j2.

Tsvetanov, K.A., Ye, Z., Hughes, L., Samu, D., Treder, M.S., Wolpe, N., Tyler, L.K., Rowe, J.B., 2018. Activity and connectivity differences underlying inhibitory control across the adult life span. J. Neurosci. 38, 7887–7900. doi:10.1523/JNEUROSCI.2919-17.2018.

Van der Meer, J.N., Breakspear, M., Chang, L.J., Sonkusare, S., Cocchi, L., 2020. Movie viewing elicits rich and reliable brain state dynamics. Nat. Commun. 11, 1–14. doi:10.1038/s41467-020-18717-w.

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., 2013. The WU-Minn human connectome project: an overview. Neuroimage 80, 62–79. doi:10.1016/j.neuroimage.2013.05.041.

Van Overwalle, F., 2009. Social cognition and the brain: a meta-analysis. Hum. Brain Mapp. 30, 829–858. doi:10.1002/hbm.20547.

Vanderwal, T., Eilbott, J., Castellanos, F. X., 2019. Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging. Developmental cognitive neuroscience 36, 100600. doi:10.1016/j.dcn.2018.10.004.

Vanderwal, T., Eilbott, J., Finn, E.S., Craddock, R.C., Turnbull, A., Castellanos, F.X., 2017. Individual differences in functional connectivity during naturalistic viewing conditions. Neuroimage 157, 521–530. doi:10.1016/j.neuroimage.2017.06.027.

Vartanian, O., Wertz, C.J., Flores, R.A., Beatty, E.L., Smith, I., Blackler, K., Lam, Q., Jung, R.E., 2018. Structural correlates of openness and intellect: implications for the contribution of personality to creativity. Hum. Brain Mapp. 39, 2987–2996. doi:10.1002/hbm.24054.

Weis, S., Patil, K.R., Hoffstaedter, F., Nostro, A., Yeo, B.T.T., Eickhoff, S.B., 2020. Sex classification by resting state brain connectivity. Cereb. Cortex 30, 824–835. doi:10.1093/CERCOR/BHZ129.

Woo, C.W., Chang, L.J., Lindquist, M.A., Wager, T.D., 2017. Building better biomarkers: brain models in translational neuroimaging. Nat. Neurosci. 20, 365–377. doi:10.1038/nn.4478.

Xia, M., Wang, J., He, Y., 2013. BrainNet Viewer: A network visualization tool for human brain connectomics. PLoS One 8, e68910. doi:10.1371/journal.pone.0068910.

Yan, C.G., Wang, X.D., Zuo, X.N., Zang, Y.F., 2016. DPABI: Data processing & analysis for (resting-state) brain imaging. Neuroinformatics 14, 339–351. doi:10.1007/s12021-016-9299-4.

Yang, Z., Wu, J., Xu, L., Deng, Z., Tang, Y., Gao, J., Hu, Y., Zhang, Y., Qin, S., Li, C., Wang, J., 2020. Individualized psychiatric imaging based on inter-subject neural synchronization in movie watching. Neuroimage 216, 116227. doi:10.1016/J.NEUROIMAGE.2019.116227.