Act Natural: Functional Connectivity from Naturalistic Stimuli fMRI Outperforms Resting-State in Predicting Brain Activity

Shachar Gal[1,2], Yael Coldham[1,2], Michal Bernstein-Eliav[1], Ido Tavor[1,2,3]

[1]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

[2]Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

[3]Strauss Center for Computational Neuroimaging, Tel Aviv University, Tel Aviv, Israel

Corresponding Author:

Dr. Ido Tavor; ORCID iD https://orcid.org/0000-0002-9117-4449
Department of Anatomy and Anthropology
Sackler Faculty of Medicine
Tel Aviv University
Ramat Aviv 6997801
Israel
idotavor@tauex.tau.ac.il
Tel: +972-73-7304420
Fax: +972-73-7304421

Abstract

The search for an 'ideal' approach to investigate the functional connections in the human brain is an ongoing challenge for the neuroscience community. While resting-state functional magnetic resonance imaging (fMRI) has been widely used to study individual functional connectivity patterns, recent work has highlighted the benefits of collecting functional connectivity data while participants are exposed to naturalistic stimuli, such as watching a movie or listening to a story. For example, functional connectivity data collected during movie-watching were shown to predict cognitive and emotional scores more accurately than resting-state-derived functional connectivity. We have previously reported a tight link between resting-state functional connectivity and task-derived neural activity, such that the former successfully predicts the latter. In the current work we use data from the Human Connectome Project to demonstrate that naturalistic-stimulus-derived functional connectivity predicts task-induced brain activation maps more accurately than resting-state-derived functional connectivity. We then show that activation maps predicted using naturalistic stimuli are better predictors of individual intelligence scores than activation maps predicted using resting-state. We additionally examine the influence of naturalistic-stimulus type on prediction accuracy. Our findings emphasize the potential of naturalistic stimuli as a promising alternative to resting-state fMRI for connectome-based predictive modelling of individual brain activity and cognitive traits.

Keywords

Naturalistic stimuli, Functional connectivity, Predictive modelling, Individual differences, Resting-state fMRI, Task-fMRI

1. Introduction

The field of cognitive neuroscience is ultimately aimed at understanding how our brains function in a real-life environment. How do we integrate the dynamic flow of multiple stimuli in a complex, multi-modal experience such as watching a play, taking a test, or meeting a friend? These real-life situations are quite challenging to simulate in laboratory settings. Therefore, in order to investigate the neural mechanisms that operate in such naturalistic experiences, researchers have developed various experimental paradigms and tasks to serve as proxies for the complex scenarios we face 'in the wild'. These tasks usually involve relatively simple stimuli that are delivered in a strictly controlled temporal order (Sonkusare, Breakspear, & Guo, 2019). They enable researchers to isolate behavioral constructs or mental processes and to localize specific cognitive functions to defined brain regions.

While these laboratory-constrained experimental tasks have offered valuable insights into the function of the human brain, their ecological validity is debatable. In daily experiences, we rarely perform isolated functions such as recognizing a face or memorizing a series of numbers. More often, plenty of cognitive and emotional processes take place simultaneously in a dynamic, multi-stimuli and context-dependent experience. Therefore, an 'ideal' experimental paradigm would allow researchers to extract brain mechanisms of discrete mental processes from 'holistic', real-life situations in which participants are placed. In other words, replacing rather simplified experimental tasks with a richer, multi-stimulus design that is more similar to a naturalistic experience, while still being able to uncover the neural correlates of the cognitive processes involved.

A promising candidate for baring information on task-related brain activity across cognitive domains is functional connectivity, which is considered to represent an intrinsic functional organization of the human brain (Biswal, Zerrin Yetkin, Haughton, & Hyde, 1995; Greicius, Krasnow, Reiss, & Menon, 2003; van den Heuvel & Hulshoff Pol, 2010). Several recent studies have provided evidence linking functional connectivity patterns to task-induced brain activity (Cole, Ito, Bassett, & Schultz, 2016; Parker Jones, Voets, Adcock, Stacey, & Jbabdi, 2017; Tavor et al., 2016; Tik et al., 2021). These studies have demonstrated that distinctive functional connectivity patterns are extremely successful predictors of task-evoked brain activity, in both healthy participants (Tavor et al., 2016) as well as psychiatric (Tik et al., 2021) and neurological patients (Parker Jones et al., 2017). Functional connectivity has also been reported to successfully predict individual scores of intelligence and cognition (Dhamala, Jamison, Jaywant, Dennis, & Kuceyeski, 2021; Dubois, Galdi, Paul, & Adolphs, 2018; Finn et al., 2015;

Rosenberg et al., 2016; Song et al., 2008) as well as personality traits (Hsu, Rosenberg, Scheinost, Constable, & Chun, 2018; Nostro et al., 2018).

Functional connectivity data are commonly collected using resting-state functional magnetic resonance imaging (rs-fMRI), by measuring the temporal correlations between spontaneous neural activation of different brain regions (van den Heuvel & Hulshoff Pol, 2010). Participants are scanned at rest, i.e., while not performing any specific task but are merely instructed to avoid falling asleep. The rs-fMRI data are therefore susceptible to influences of mood variations, level of concentration and tiredness (Christoff, Gordon, Smallwood, Smith, & Schooler, 2009; Duncan & Northoff, 2013; Finn et al., 2017; Harrison et al., 2008; Morcom & Fletcher, 2007; Shirer, Ryali, Rykhlevskaia, Menon, & Greicius, 2012), and are prone to excessive head motion (Eickhoff, Milham, & Vanderwal, 2020; Vanderwal, Kelly, Eilbott, Mayes, & Castellanos, 2015).

An alternative approach to studying intrinsic functional connectivity patterns is by measuring "spontaneous" brain activity while participants are exposed to a naturalistic stimulus (Sonkusare et al., 2019). A naturalistic stimulus, such as listening to a story or watching a movie, is a complex and dynamic stimulus which is used as an ecologically valid condition to measure neural activity (Eickhoff et al., 2020). While previous work has demonstrated a synchronized neural response across participants during movie-watching (Hasson, Malach, & Heeger, 2010; Hasson, Nir, Levy, Fuhrmann, & Malach, 2004), recent studies claim that the remaining individual variance in brain activity is more reflective of inherent, stable, connectivity features (Finn & Bandettini, 2021; Finn et al., 2017). Furthermore, a widely reported advantage of naturalistic stimuli is an increase in participants' compliance in the form of decreased head motion compared to rs-fMRI (Vanderwal, Eilbott, & Castellanos, 2019; Vanderwal et al., 2017, 2015), as well as increased wakefulness throughout the MRI scan (Sonkusare et al., 2019; Vanderwal et al., 2017, 2015).

Naturalistic stimuli derived data are increasingly being used to extract functional connectivity features (Finn et al., 2020; Vanderwal et al., 2019) and seem to outperform rs-fMRI derived data in studying various neural and behavioral traits. For example, the use of naturalistic stimuli has been shown to enhance the detection of unique individual functional connectivity patterns compared to rs-fMRI, possibly due to increased arousal levels during movie-watching (Vanderwal et al., 2017). Functional connectivity data derived from movie-watching has also outperformed data collected at rest in predicting cognitive and emotional scores (Finn & Bandettini, 2021). Notably, the appropriate content to be used in naturalistic stimuli is

debatable (Grall & Finn, 2021), with different studies employing different types of audiovisual content such as abstract videoclips as opposed to scenes from familiar films (Vanderwal et al., 2015).

In light of previous evidence tightly linking functional connectivity patterns to task-induced brain activity, as well as evidence demonstrating the advantages of naturalistic stimuli compared to rs-fMRI, in the current work we examine the predictability of task-induced brain activity from patterns of brain connectivity while participants watch a naturalistic stimulus. We hypothesize that brain-connectivity data derived from movie-watching fMRI will be successful at predicting individual task-evoked neural activity across a range of cognitive paradigms. Moreover, we expect prediction of activation maps from naturalistic-stimulus data to outperform prediction from rs-fMRI data, and to successfully detect individually unique activation patterns. We additionally compare the accuracy of task-activation maps predicted from different types of naturalistic stimuli (abstract vs. familiar). Finally, we demonstrate the predictability of cognitive scores from the predicted task-induced activation maps derived from functional connectivity data collected using a naturalistic stimulus, compared to functional connectivity data collected using rs-fMRI.

2. Methods

2.1. Participants

The data used in this study were provided by the Human Connectome Project (HCP; Van Essen et al., 2013). While the HCP dataset includes over 1000 participants, only a subset of 184 participants were scanned during naturalistic stimuli viewing. Naturalistic stimuli fMRI scans were performed on a 7 Tesla Siemens Magnetom. Since this work focuses on naturalistic stimuli, we used this subset of participants for whom naturalistic viewing data were available. Besides movie-watching scans, these participants were also scanned at rest on both 7T and 3T scanners, and during task performance on a 3T scanner. Participants that did not complete all movie-watching, resting-state and task fMRI scans were excluded from this study, resulting in a final dataset of n = 159 (mean age = 29.4, standard deviation = 3.25, 95 females). The HCP data includes multiple sibships of monozygotic and dizygotic twins as well as non-twin siblings. These family relations were accounted for in our analyses (see below).

2.2. Functional MRI data

2.2.1. 7T data

The 7T dataset includes four resting-state scans and four movie-watching scans for each participant, acquired over four scanning sessions: Session #1 included two movie-watching scans and a resting-state scan; Sessions #2 and #3 included a resting-state scan only; Session #4 included two movie-watching scans and a resting-state scan. Therefore, the first two movie-watching scans were acquired in the same session as the first resting-state scan, while the third and fourth movie-watching scans were acquired in the same session as the fourth resting-state scan (see Table 1).

7T data were acquired using a gradient-echo-planar imaging (EPI) sequence with a TR of 1 second and a spatial resolution of 1.6 mm^3. The phase encoding direction alternated between posterior-to-anterior (PA; Rest1, Movie2, Movie3) and anterior-to-posterior (AP; Rest4, Movie1, Movie4).

Each resting-state scan included 900 timepoints (15min long). Movie-watching scans varied slightly in duration between 901-921 timepoints, according to the clips being viewed during each scan. During movie-watching scans, participants were presented with a series of 4 or 5 audiovisual clips, separated by 20 seconds of rest. The first and third scans included clips from fiction and documentary independent (obscure) films, while the second and fourth scans

included clips from Hollywood (familiar) films. Importantly, there was a dissociation between the phase-encoding direction of the scan and the type of movie presented in the scan (Table 1).

In order to retain the subsets of the data in which brain activity was affected by movie-watching, we excluded the last 17 seconds of each rest interval, keeping the first three seconds to account for the delay in the hemodynamic response. Following this exclusion procedure, the shortest scan consisted of 808 timepoints (Movie3). We then excluded timepoints from the end of each scan, both movie and rest, to keep a fixed number of 808 timepoints in all scans. Notably, while timepoints were excluded from movie-watching scans in an interleaved manner (i.e., from each no-movie interval), in the resting-state scans all timepoints were removed from the end of the scan. While this difference in timepoint-removal may have introduced a bias in favor of the resting-state data, as it made these data more continuous than the movie-watching data, it is noteworthy that such potential bias is in the opposite direction of our hypothesis.

Movie run	Scanning session	Place in session	Stimuli type	Phase encoding
1	1	1	Independent (obscure)	AP
2	1	2	Hollywood (familiar)	PA
3	4	1	Independent (obscure)	PA
4	4	2	Hollywood (familiar)	AP

Table 1. Description of the movie-watching scans included in the 7T dataset.

2.2.2. 3T data

All functional 3T data were acquired with a TR of 0.72 seconds and a spatial resolution of 2 mm^3. For the task-fMRI data, we used the z-score statistical maps provided by the HCP (see Barch et al., 2013) for a representative contrast from each task (see Table 2). The Motor task was not included in the analysis due to the low inter-subject variability in its activation maps, resulting in low specificity of the predicted maps. For the resting-state fMRI, we used all four scans available and trimmed them similarly to the 7T resting-state scans, to include a total of 808 timepoints per scan.

Task	Contrast number	Contrast name
Emotion	02	Faces-Shapes
Gambling	01	Punish
Language	03	Math-Story
Relational	02	Rel
Social	01	Random
Working Memory	09	2BK

Table 2. A list of the six fMRI-tasks included in this study. An exemplary contrast from each task in the HCP dataset, except the motor task, was used. See Barch et al., 2013 for more details regarding the tasks and their statistical analysis.

2.3. Preprocessing

We used the minimally preprocessed fMRI data provided by the HCP (Glasser et al., 2013). These data underwent motion and distortion corrections and nonlinear alignment to the MNI template space. Furthermore, the data were denoised using FMRIB's ICA-based Xnoiseifier (FIX) (Griffanti et al., 2014; Salimi-Khorshidi et al., 2014), and then resampled and "projected" onto a surface representation of 91,282 vertices ("grayordinates") in standard space. Data were aligned and registered using Multimodal Surface Matching (MSMAll) (Robinson et al., 2018).

2.4. Predicting task-activation maps from task-free fMRI data

This work employs a previously developed method to predict task-induced brain activation maps from scans acquired while no explicit task is performed (Gal et al., 2021; Tavor et al., 2016; Tik et al., 2021). Here, we compared activation maps predicted from movie-watching scans, 7T resting-state scans and 3T resting-state scans. To create these maps (hereafter referred to as *connTask* maps), a set of 45 independent features was first extracted from each input type (i.e., movie-watching, 7T resting-state and 3T resting-state) separately. This was done using an iterative principal component analysis (PCA; Smith et al., 2014), followed by a group independent component analysis (ICA; Beckmann, DeLuca, Devlin, & Smith, 2005) to yield a group-wise set of components. We then performed a dual regression (Beckmann, Mackay, Filippini, & Smith, 2009), resulting in 45 spatially independent cortical individual maps per participant, each representing a brain network. These maps were then used as seeds in a weighted seed-to-vertex connectivity analysis. The resulting connectivity maps were used as the final features for model fitting and prediction. An elaborated explanation of the feature extraction procedure can be found elsewhere (Tavor et al., 2016).

These feature maps were then fed into a general linear model to predict task activation maps of a variety of tasks (Table 2). For each task, a 10-fold cross validation routine was employed to create a predicted *connTask* map, such that in each of the 10 iterations, 9/10 of the data were used as the training-set and 1/10 as the test-set. Taking into account the family structure in the HCP data, we made sure that in each iteration, members of the same family were kept in one group (either training or test) to prevent data leakage resulting from the heritability of connectivity patterns (Colclough et al., 2017).

2.5. Comparing the accuracy of task-activation maps predicted from different inputs

We tested whether *connTask* maps accuracy was higher when predicted from movie-watching data than from resting-state data. The accuracy of a *connTask* map was defined as the Pearson's

correlation coefficient between the predicted and actual activation maps. For each task-contrast, we performed a paired t-test comparing the accuracy of the maps derived from movie-watching with that of the maps derived from 3T or 7T resting-state data. The resulting p values were corrected for 12 (6 contrasts X 2 input types) comparisons using Bonferroni correction for multiple comparisons. We also tested whether prediction accuracy differed between 3T and 7T resting-state data, using a paired t-test for each task-contrast, corrected for six comparisons.

Next, we explored the effect of audiovisual content type in the movie-watching scans on the accuracy of the *connTask* maps predicted from these scans. For this purpose, we generated *connTask* maps using data from each scanning session separately. For each participant and each scanning session, we defined a measurement of prediction improvement as the difference between the accuracies of the movie-watching-derived and the resting-state-derived *connTask* maps, using scans from the same session (i.e., movie scans 1&2 were compared with resting-state scan 1; movie scans 3&4 were compared with resting-state scan 4). We then grouped these improvement measures according to the movie-watching audiovisual content type (Obscure [1,3] or Familiar [2,4]), and performed a paired t-test comparing the two content types. Resulting p values were corrected for six comparisons using Bonferroni correction for multiple comparisons.

Finally, we fitted a mixed linear model using 'lme4' package in R (Bates, Mächler, Bolker, & Walker, 2015) that explains the improvement in *connTask* maps accuracy by the audiovisual content (*stimuli_type*), with a random intercept across task contrasts, as well as a random intercept and slope across participants:

$$accuracy_improvement \sim 1 + stimuli_type + (1|contrast) + (1 + stimuli_type|participant)$$

2.6. Relating the extent of improvement in accuracy to the level of brain activity

Improvement in prediction accuracy could arise from reduced predicted activity in regions where original activity was low (i.e., decreasing the "false positives"), or from increased predicted activity in regions where original activity was high (i.e., increasing the "true positives"). To examine what drives the effect of improved accuracy of movie-watching-derived task-activation maps relative to those derived from resting-state, we created maps that depict the mean activity and maps that depict the mean improvement of prediction by the movie-watching data, for 100 different cortical areas. For this purpose we used a 100-node parcellation based on the division of the human cortex to seven distinct connectivity networks (Schaefer et al., 2018; Yeo et al., 2011).

Prediction improvement was calculated for each participant and each session as the difference between the accuracies of the movie-watching-derived and the resting-state-derived *connTask* maps, using scans from the same session. We then averaged prediction improvement values across participants and sessions to create the parcellated maps. We created such parcellated maps for each task-contrast and examined the correlations between prediction improvement and mean activity in each task separately.

To further examine the relationship between prediction improvement and task-activity beyond the level of a single task, we used a mixed linear model (Bates et al., 2015) that explains the improvement in *connTask* accuracy in each parcel by the mean activity in this parcel in the original task-activation map, with a random intercept and slope across task contrasts and across parcels:

$$accuracy_improvement \sim 1 + mean_activity + (1 + mean_activity|contrast) + (1 + mean_activity|parcel)$$

Notably, the degrees of freedom in this model are limited by the number of task-contrasts we used. In this case, only six task-contrasts do not allow sufficient data for fitting a reliable model. Therefore, we used eight more task-contrasts from the HCP dataset, as in previous work on *connTask* maps (Gal et al., 2021).

2.7. Prediction of individual traits using *connTask* maps

While it has been previously shown that connectomes derived from movie-watching data outperformed those derived from resting-state data in the prediction of individual traits (Finn & Bandettini, 2021), here we tested whether a similar effect also exists when using *connTask* maps rather than the traditional connectome.

We used the resting-state and movie-watching-derived *connTask* maps to predict individual scores of intelligence. Intelligence scores were constructed by applying a factor analysis to 10 cognitive measures reported in the HCP data (Dubois et al., 2018). Prediction was performed using the Basis Brain Set pipeline (BBS) (Sripada et al., 2019; Sripada, Angstadt, Rutherford, Taxali, & Shedden, 2020), which has been previously found effective for predicting individual traits from *connTask* maps (Gal et al., 2021).

For each input type (i.e., resting-state or movie-watching) and each task, individual trait prediction from *connTask* maps was performed as follows: We used a 5-fold cross validation method to split the data while keeping all members of each sibship in either the training or the test dataset. In each iteration, dimensionality of the training set data was reduced using PCA to a predetermined number of components, k. We used k=75 based on previous work (Sripada et

al., 2019, 2020). These components were regressed against the individual data of each participant, test and training set alike, to yield an "expression score" for each component. These K "expression scores" were then used to fit a linear model that predicts the desired trait. For significance testing, this procedure was repeated 1000 times with 2/3 of the data used for training and 1/3 for testing, accounting for family structure. Differences between the accuracy of predictions based on the two input types were tested using a non-parametric, related samples test (Wilcoxon signed ranks test) (Demšar, 2006).

3. Results

3.1. Movie-watching-derived *connTask* maps are more accurate than resting-state-derived maps

Task-activation maps predicted from movie-watching data were significantly more accurate than those predicted from 7T resting-state data for all task-contrasts (Figure 1). In five out of six task-contrasts, no significant differences were observed between the accuracy of *connTask* maps derived from 3T and 7T resting-state data, demonstrating that the strength of magnetic field does not affect prediction accuracy, while *connTask* maps derived from movie-watching data outperformed both. In the Language task (Math vs. Story contrast), *connTask* maps derived from the 3T resting-state data were more accurate than those based on 7T resting-state data, and did not differ significantly from the movie-watching based maps.

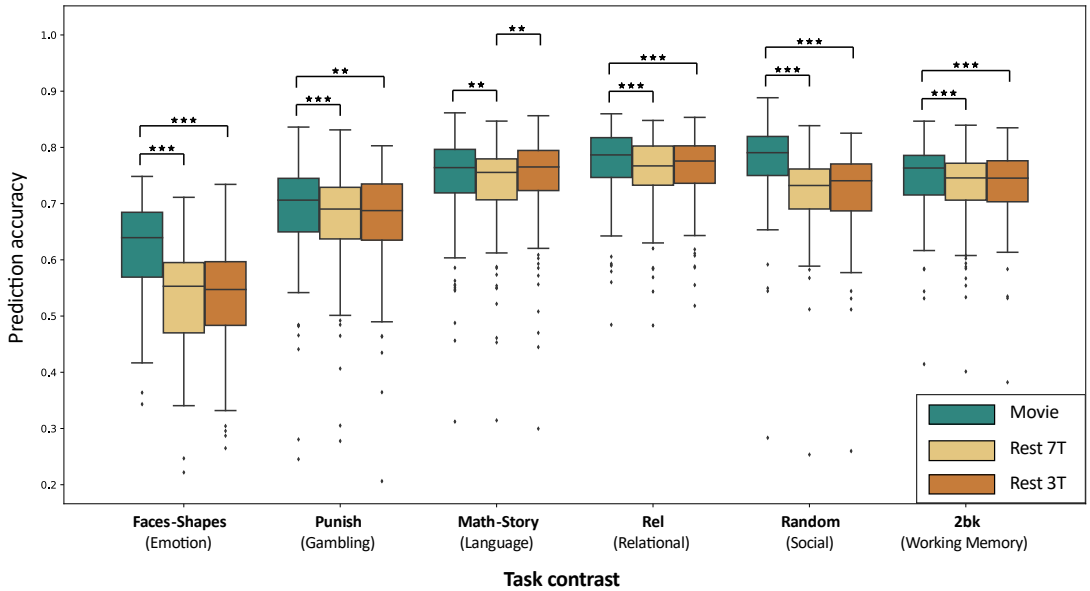

Figure 1. Movie-watching-derived *connTask* maps are more accurate than resting-state-derived maps. For each task-contrast and each input type (movie-watching, 3T resting-state and 7T resting-state), boxplots portray the distribution of prediction accuracies. Prediction accuracy for each participant was calculated as the Pearson correlation between *connTask* and actual task-induced activation maps. Differences in *connTask* accuracy between input types and within each task-contrast were tested using a paired t-test and corrected using Bonferroni correction for multiple comparisons. $^{**}p < 0.01$; $^{***}p < 0.001$

3.2. The type of naturalistic stimulus impacts the accuracy of the predicted task-activation maps

In four of the six task-contrasts, we found that *connTask* maps predicted from movie-watching data where participants viewed obscure naturalistic stimuli (e.g., excerpts from independent films) compared to familiar naturalistic stimuli (e.g., excerpts from Hollywood films) offered

a larger improvement in the prediction of activation patterns, relative to the prediction from resting-state data (Figure 2).

A mixed linear model that explains improvement in *connTask* accuracy by the input type while controlling for variance driven by the different tasks and participants further supported this effect ($p < 0.001$, $df=158$), suggesting that task-activation maps predicted from data of obscure naturalistic stimuli were more accurate than those predicted from familiar naturalistic stimuli.

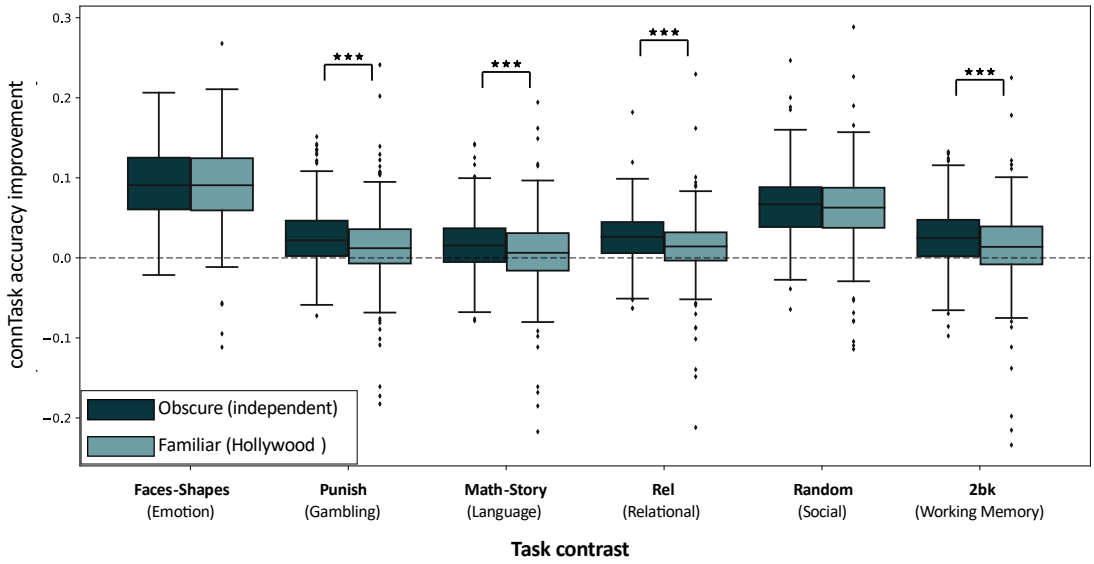

Figure 2. The type of naturalistic stimuli impacts the accuracy of subsequent *connTask* map. For each task-contrast and each type of audiovisual content, boxplots portray the distribution of prediction improvement. For each participant and each content type, we defined prediction improvement as the difference between the accuracies of the movie-watching-derived and the resting-state-derived *connTask* maps, using scans from the same session. Differences in accuracy improvement between input types and within each task-contrast were tested using a paired t-test and corrected using Bonferroni correction for multiple comparisons. ***$p < 0.001$

3.3. Improvement in task-activity prediction from movie-watching data is larger in highly active brain areas

In five of the six task-contrasts, prediction improvement was significantly correlated with the absolute values of mean activation across parcels (Figure 3). This effect is corroborated by the results of the mixed linear model analysis, showing a significant relationship between mean activity and prediction improvement across task-contrasts and brain areas ($p = 0.042$, $df=14$).

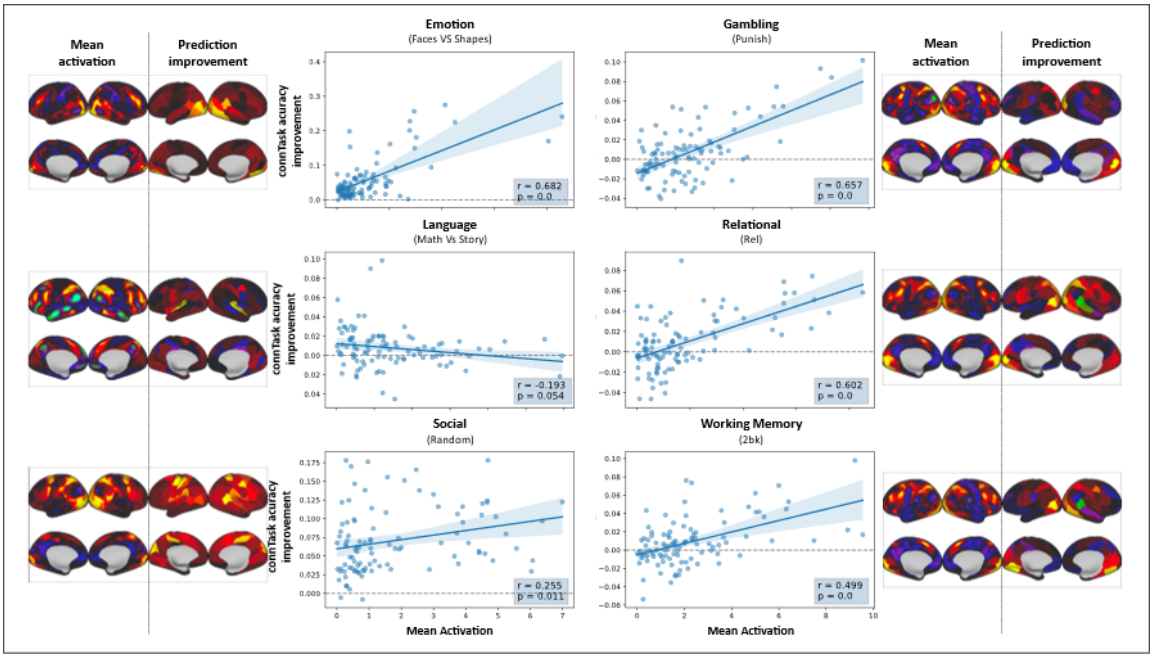

Figure 3. Improvement in task-activity prediction from movie-watching data is larger in highly active brain areas. Center: Scatter plots depicting the relationship between mean activation and prediction improvement in 100 distinct brain areas (Schaefer et al., 2018). Left and right: mean activation maps and parcellated prediction improvement maps for each contrast.

3.4. Prediction of individual traits from movie-watching-derived *connTask* maps is more accurate than from resting-state-derived maps

Individual scores of intelligence (G-scores) were successfully predicted from *connTask* maps derived from both movie-watching and resting-state data. All models predicted significantly better than chance, as confirmed by a permutation test with 5000 iterations ($p<0.05$, Bonferroni corrected for multiple comparisons). Models based on movie-watching-derived *connTask* maps were significantly more accurate than models based on resting-state-derived *connTask* maps ($p<0.0001$, Bonferroni corrected for multiple comparisons).

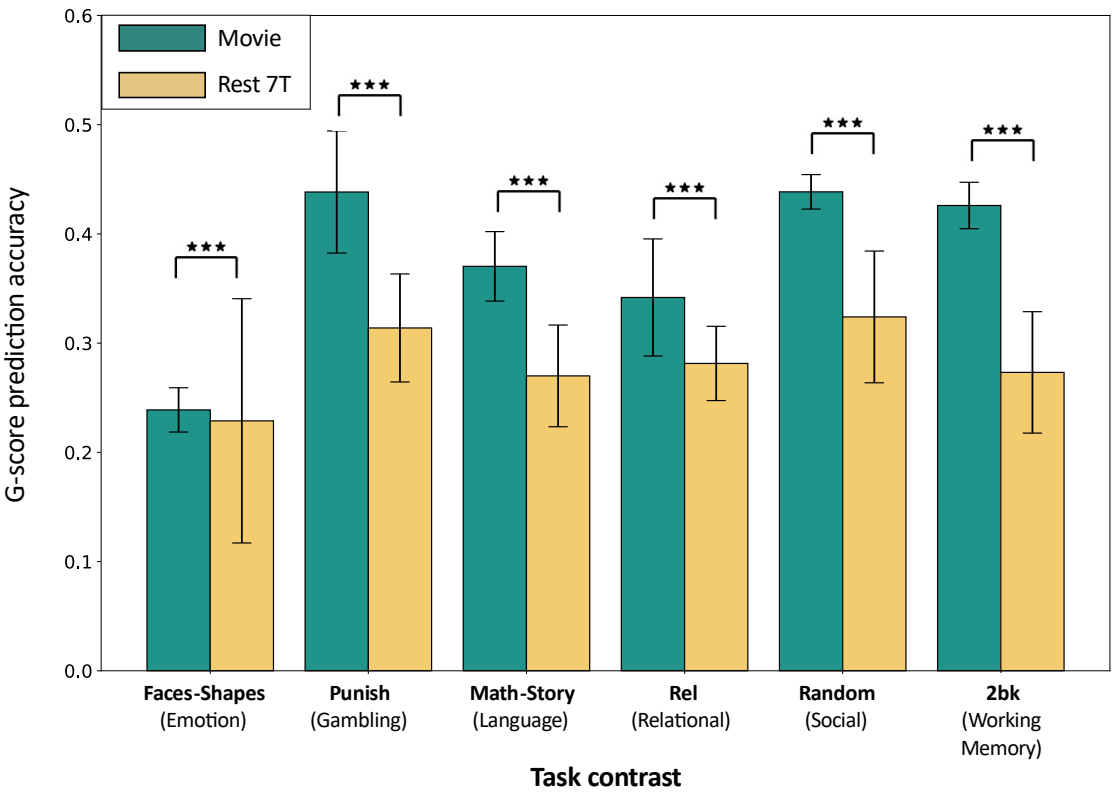

Figure 4. Prediction of individual traits is more accurate from movie-watching than resting-state-derived *connTask* maps. For each task-contrast and input type, bars height depicts the accuracy of G-scores prediction, calculated as the Pearson correlation between real and predicted G-scores. Error bars depict the standard error of prediction accuracy across the 5 folds of cross-validation. While all models predicted significantly more accurate than the null hypothesis ($p<0.05$, Bonferroni corrected for multiple comparisons), models based on movie-watching-derived *connTask* maps were significantly more accurate than models based on resting-state-derived *connTask* maps ($p<0.0001$, Bonferroni corrected for multiple comparisons).

4. Discussion

Functional connectivity patterns derived from data collected while participants were exposed to naturalistic stimuli outperformed connectivity patterns derived from resting-state fMRI in predicting inherent individual qualities, including task-induced brain activity and cognitive scores. While it has been previously demonstrated that data collected at rest can be used to successfully predict brain activity during task performance (Cole et al., 2016; Parker Jones et al., 2017; Tavor et al., 2016; Tik et al., 2021), here we demonstrate that data collected during movie-watching yield an even more accurate prediction of task-derived activation maps, across six different tasks: Emotion, Gambling, Language, Relational, Social and Working Memory (Table 2; Figure 1). Moreover, we found that connectivity patterns derived from data acquired while watching obscure videoclips provided higher accuracy of the predicted task-activation maps than familiar Hollywood scenes (Figure 2). Finally, task-activation maps derived from movie-watching outperformed those derived from rest in predicting individual cognitive scores (Figure 4).

Our results are in line with previous studies showing that movie-watching data outperform resting-state data in detecting unique functional connectivity patterns (Vanderwal et al., 2017) and that the prediction of cognitive and emotional scores is more accurate when based on movie-watching data than on data collected at rest (Finn & Bandettini, 2021). The superiority of movie-watching over rest may be attributed to the fact that movie-watching is a dynamic, ecologically valid stimulus, designed to resemble real-life experiences while measuring neural activity (Eickhoff et al., 2020). As such, movie-watching has been associated with increased participant compliance in the form of decreased head motion and sleepiness (Sonkusare et al., 2019; Vanderwal et al., 2017, 2015) and with less susceptibility to influences of volatile factors such as mood variations, level of concentration and tiredness (Christoff et al., 2009; Duncan & Northoff, 2013; Finn et al., 2017; Harrison et al., 2008; Morcom & Fletcher, 2007; Shirer et al., 2012).

In efforts to improve the prediction of human behavior and brain activity from functional connectivity, most research to date focuses on model enhancement. This includes exploring different machine-learning algorithms and testing different model parameters (Ngo, Khosla, Jamison, Kuceyeski, & Sabuncu, 2021; Zheng et al., 2021). Here, we take an alternative data-centric approach and focus on enhancing the features rather than the model. As such, by feeding the model with connectivity features derived from movie-watching fMRI rather than resting-state we were able to significantly improve model performance.

The ‘ideal’ type of naturalistic stimuli to be used for predictive models is an open question (Grall & Finn, 2021). Here, we found that obscure stimuli (independent films) offered a larger improvement in prediction of task-activity than familiar stimuli (Hollywood films) relative to rest. This result was consistent across the Language, Gambling, Relational, Social and Working Memory tasks, but not the Emotion and Social tasks (Figure 2). Notably, these two task-contrasts demonstrated the highest improvement in prediction accuracy to begin with (Figure 1), alluding to a possible ceiling effect.

While obscure videoclips outperformed familiar videoclips in the prediction of task-induced activation maps, previous studies have suggested that different types of naturalistic stimuli could be beneficial for different purposes. For example, in a recent work that used the HCP data in a connectome predictive modelling (CPM) paradigm, familiar videoclips with high social content provided better predictions of cognitive scores than obscure videoclips, with low or no social content (Finn & Bandettini, 2021). Notably, this result diverges from ours in the matter of which naturalistic stimulus yields superior predictions, emphasizing the variability offered by different stimuli in different prediction schemes. Moreover, different types of naturalistic stimuli were found to evoke different patterns of brain activation, with spatially limited and weaker synchronized activity across participants (Hasson et al., 2004) when watching an obscure videoclip compared to a familiar, highly social videoclip (*Ocean’s Eleven*) (Vanderwal et al. 2015). Taken together, these findings highlight the crucial impact of the choice of stimulus on the obtained results. Hence, future studies based on data collected while participants are exposed to naturalistic stimuli should thoroughly consider the type of stimulus to be used.

The improvement in prediction accuracy of task-activation maps reported here for movie-watching relative to resting-state data could be explained by either reduced predicted activity in regions where original activity was low (i.e., decreasing the “false positives”), or by increased predicted activity in regions where original activity was high (i.e., increasing the “true positives”). We revealed a positive correlation between the original task-induced activation levels and the improvement in activation prediction from movie-watching compared to resting-state data, suggesting that the improvement in prediction was due to enhanced prediction of actual activation patterns, rather than reduced prediction in areas of low actual activity. This result was consistent across the Emotion, Gambling, Relational, Social and Working Memory tasks, but not the Language task (Figure 3), which also showed the most moderate effect of improved prediction from movie than rest in the first place (Figure 1).

Previous research has shown that activity maps derived from resting-state data better predict cognitive scores than traditional functional connectivity properties (i.e., the resting-state connectome) (Gal et al., 2021). Here, we show that activity maps derived from data collected while viewing naturalistic stimuli outperform those that were derived from data collected at rest in the prediction of general intelligence scores (Figure 4). This finding may result from the fact that movie-watching data outperform resting-state data in the prediction of task-derived activation, however it may also arise from the former being a better predictor of intelligence than the latter, independently from task-activity prediction. The question of whether the increased accuracy is due to the better activity prediction or due to the suitability of movie-watching for predicting intelligence should be addressed in future studies exploring the uses and benefits of naturalistic stimuli. Regardless, this finding suggests that naturalistic stimuli designed to resemble complex, multi-modal real-life situations, may be more appropriate for the prediction of complex cognitive constructs.

Several limitations of the current work should be noted. First, the HCP dataset includes movie-watching scans acquired in 7T, resting-state scans acquired in 3T as well as 7T, and task-fMRI scans acquired in 3T. As such, the magnetic field strength differed between the data we used as the predictor (movie-watching) and the data we aimed to predict (task-fMRI). To examine whether the strength of the magnetic field may have introduced a bias to our analysis, we compared the accuracy of activation maps predicted from 7T rest data and 3T rest data. For five out of the six task-contrasts, no significant difference was found in the accuracy of prediction based on 7T rest data and 3T rest data. We therefore conclude that the use of data acquired at 7T to predict activation maps could not have biased our results in favor of movie-watching over resting-state data.

Second, this study was performed on 159 HCP participants who have completed all movie-watching, resting-state, and task fMRI scans. While it has been suggested that associations between neural and behavioral measures only stabilize and become reproducible with sample sizes of $n \gtrapprox 2{,}000$ (Marek et al., 2020), a recent work that examined the required sample size for the prediction of task-induced activity from functional connectivity has shown that a training set of 150-200 participants should be sufficient for accurate predictions (Cohen, Chen, Jones, Niu, & Wang, 2020).

5. Conclusions

The findings reported in this work highlight the benefits of using functional connectivity derived from naturalistic-stimuli rather than resting-state fMRI for the prediction of individual task-induced brain activity and cognitive traits. Our results also emphasize the influence of the type of naturalistic stimulus (e.g., obscure vs. familiar videoclips) on prediction accuracy, suggesting that further research is required in order to standardize and determine the 'optimal' naturalistic stimulus, if exists. We conclude that naturalistic stimuli such as movie-watching, that are designed to resemble complex, real-life experience, are a promising approach for unravelling the intrinsic functional organization of the human brain.

Declaration of competing interests

The authors have no conflict of interests to disclose.

CRediT authorship contribution statement

Shachar Gal: Conceptualization, Software, Formal analysis, Visualization, Writing - Original Draft. **Yael Coldham**: Writing - Original Draft, **Michal Bernstein-Eliav**: Writing - Review & Editing. **Ido Tavor:** Conceptualization, Software, Supervision, Writing - Review & Editing

Acknowledgments

Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

The authors acknowledge with thanks the support of the Israel Science Foundation (ISF grant no. 1603/18) and the National Institute of Psychobiology in Israel (grant 232-19-20.).

Data and materials availability

All the data included in this study are publicly available in the Human Connectome Project database at https://db.humanconnectome.org

References

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., … WU-Minn HCP Consortium. (2013). Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage*, *80*, 169–189. https://doi.org/10.1016/j.neuroimage.2013.05.033

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/JSS.V067.I01

Beckmann, C. F., DeLuca, M., Devlin, J. T., & Smith, S. M. (2005). Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1457), 1001–1013. https://doi.org/10.1098/rstb.2005.1634

Beckmann, C., Mackay, C., Filippini, N., & Smith, S. (2009). Group comparison of resting-state FMRI data using multi-subject ICA and dual regression. *NeuroImage*, *47*, S148. https://doi.org/10.1016/s1053-8119(09)71511-3

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine*, *34*(4), 537–541. https://doi.org/10.1002/mrm.1910340409

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(21), 8719–8724. https://doi.org/10.1073/pnas.0900234106

Cohen, A. D., Chen, Z., Jones, O. P., Niu, C., & Wang, Y. (2020). Regression-based machine-learning approaches to predict task activation using resting-state fMRI. *Human Brain Mapping*, *41*(3), 815. https://doi.org/10.1002/HBM.24841

Colclough, G. L., Smith, S. M., Nichols, T. E., Winkler, A. M., Sotiropoulos, S. N., Glasser, M. F., … Woolrich, M. W. (2017). The heritability of multi-modal connectivity in human brain activity. *ELife*, *6*. https://doi.org/10.7554/eLife.20178

Cole, M. W., Ito, T., Bassett, D. S., & Schultz, D. H. (2016). Activity flow over resting-state networks shapes cognitive task activations. *Nature Neuroscience*, *19*(12), 1718–1726. https://doi.org/10.1038/nn.4406

Demšar, J. (2006). *Statistical Comparisons of Classifiers over Multiple Data Sets*. *Journal of Machine Learning Research* (Vol. 7).

Dhamala, E., Jamison, K. W., Jaywant, A., Dennis, S., & Kuceyeski, A. (2021). Distinct functional and structural connections predict crystallised and fluid cognition in healthy adults. *Human*

Brain Mapping, *42*(10), hbm.25420. https://doi.org/10.1002/hbm.25420

Dubois, J., Galdi, P., Paul, L. K., & Adolphs, R. (2018). A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1756). https://doi.org/10.1098/rstb.2017.0284

Duncan, N. W., & Northoff, G. (2013). Overview of potential procedural and participant-related confounds for neuroimaging of the resting state. *Journal of Psychiatry & Neuroscience : JPN*, *38*(2), 84–96. https://doi.org/10.1503/jpn.120059

Eickhoff, S. B., Milham, M., & Vanderwal, T. (2020). Towards clinical applications of movie fMRI. *NeuroImage*, *217*. https://doi.org/10.1016/j.neuroimage.2020.116860

Finn, E. S., & Bandettini, P. A. (2021). Movie-watching outperforms rest for functional connectivity-based prediction of behavior. *NeuroImage*, *235*, 117963. https://doi.org/10.1016/j.neuroimage.2021.117963

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. https://doi.org/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. https://doi.org/10.1016/J.NEUROIMAGE.2017.03.064

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., … Constable, R. T. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, *18*(11), 1664–1671. https://doi.org/10.1038/nn.4135

Gal, S., Tik, N., Bernstein-Eliav, M., & Tavor, I. (2021). Predicting Individual Traits from Unperformed Tasks, 1–26.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., … WU-Minn HCP Consortium. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, *80*, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127

Grall, C., & Finn, E. S. (2021). The 'Naturalistic' Fallacy: Leveraging the Power of Media to Drive Cognition. *Psyarxiv*.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National*

Academy of Sciences of the United States of America, *100*(1), 253–258. https://doi.org/10.1073/pnas.0135058100

Griffanti, L., Salimi-Khorshidi, G., Beckmann, C. F., Auerbach, E. J., Douaud, G., Sexton, C. E., … Smith, S. M. (2014). ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage*, *95*, 232–247. https://doi.org/10.1016/j.neuroimage.2014.03.034

Harrison, B. J., Pujol, J., Ortiz, H., Fornito, A., Pantelis, C., & Yücel, M. (2008). Modulation of Brain Resting-State Networks by Sad Mood Induction. *PLoS ONE*, *3*(3), e1794. https://doi.org/10.1371/journal.pone.0001794

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences*, *14*(1), 40–48. https://doi.org/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*(5664), 1634–1640. https://doi.org/10.1126/science.1089506

Hsu, W. T., Rosenberg, M. D., Scheinost, D., Constable, R. T., & Chun, M. M. (2018). Resting-state functional connectivity predicts neuroticism and extraversion in novel individuals. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nsy002

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., … Dosenbach, N. U. F. (2020). Towards Reproducible Brain-Wide Association Studies. *BioRxiv*, 2020.08.21.257758. https://doi.org/10.1101/2020.08.21.257758

Morcom, A. M., & Fletcher, P. C. (2007). Does the brain have a baseline? Why we should be resisting a rest. *NeuroImage*, *37*(4), 1073–1082. https://doi.org/10.1016/J.NEUROIMAGE.2006.09.013

Ngo, G. H., Khosla, M., Jamison, K., Kuceyeski, A., & Sabuncu, M. R. (2021). Predicting Individual Task Contrasts From Resting-state Functional Connectivity using a Surface-based Convolutional Network. *BioRxiv*, 2021.04.19.440523. Retrieved from http://biorxiv.org/content/early/2021/04/20/2021.04.19.440523.abstract

Nostro, A. D., Müller, V. I., Varikuti, D. P., Pläschke, R. N., Hoffstaedter, F., Langner, R., … Eickhoff, S. B. (2018). Predicting personality from network-based resting-state functional connectivity. *Brain Structure and Function*, *223*(6), 2699–2719. https://doi.org/10.1007/s00429-018-1651-z

Parker Jones, O., Voets, N. L., Adcock, J. E., Stacey, R., & Jbabdi, S. (2017). Resting connectivity predicts task activation in pre-surgical populations. *NeuroImage. Clinical*, *13*, 378–385.

https://doi.org/10.1016/j.nicl.2016.12.028

Robinson, E. C., Garcia, K., Glasser, M. F., Chen, Z., Coalson, T. S., Makropoulos, A., … Rueckert, D. (2018). Multimodal surface matching with higher-order smoothness constraints. *NeuroImage*, *167*, 453–465. https://doi.org/10.1016/j.neuroimage.2017.10.037

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*, *19*(1), 165–171. https://doi.org/10.1038/nn.4179

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. https://doi.org/10.1016/j.neuroimage.2013.11.046

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., … Yeo, B. T. T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex (New York, N.Y. : 1991)*, *28*(9), 3095–3114. https://doi.org/10.1093/cercor/bhx179

Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V., & Greicius, M. D. (2012). Decoding Subject-Driven Cognitive States with Whole-Brain Connectivity Patterns. *Cerebral Cortex*, *22*(1), 158–165. https://doi.org/10.1093/cercor/bhr099

Smith, S. M., Hyvärinen, A., Varoquaux, G., Miller, K. L., & Beckmann, C. F. (2014). Group-PCA for very large fMRI datasets. *NeuroImage*, *101*, 738–749. https://doi.org/10.1016/j.neuroimage.2014.07.051

Song, M., Zhou, Y., Li, J., Liu, Y., Tian, L., Yu, C., & Jiang, T. (2008). Brain spontaneous functional connectivity and intelligence. *NeuroImage*, *41*(3), 1168–1176. https://doi.org/10.1016/J.NEUROIMAGE.2008.02.036

Sonkusare, S., Breakspear, M., & Guo, C. (2019). Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends in Cognitive Sciences*, *23*(8), 699–714. https://doi.org/10.1016/j.tics.2019.05.004

Sripada, C., Angstadt, M., Rutherford, S., Kessler, D., Kim, Y., Yee, M., & Levina, E. (2019). Basic Units of Inter-Individual Variation in Resting State Connectomes. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-018-38406-5

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. *Human Brain Mapping*, hbm.25007. https://doi.org/10.1002/hbm.25007

Tavor, I., Parker Jones, O., Mars, R. B., Smith, S. M., Behrens, T. E., & Jbabdi, S. (2016). Task-free MRI predicts individual differences in brain activity during task performance. *Science*, *352*(6282), 216–220. https://doi.org/10.1126/science.aad8127

Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., … Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Tik, N., Livny, A., Gal, S., Gigi, K., Tsarfaty, G., Weiser, M., & Tavor, I. (2021). Predicting individual variability in task-evoked brain activity in schizophrenia. *Human Brain Mapping*, hbm.25534. https://doi.org/10.1002/hbm.25534

van den Heuvel, M. P., & Hulshoff Pol, H. E. (2010). Exploring the brain network: a review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology : The Journal of the European College of Neuropsychopharmacology*, *20*(8), 519–534. https://doi.org/10.1016/j.euroneuro.2010.03.008

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., & Yacoub, E. (2013). The WU-Minn Human Connectome Project: An overview. https://doi.org/10.1016/j.neuroimage.2013.05.041

Vanderwal, T., Eilbott, J., & Castellanos, F. X. (2019, April). Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging. *Developmental Cognitive Neuroscience*. Elsevier Ltd. https://doi.org/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. https://doi.org/10.1016/j.neuroimage.2017.06.027

Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L. C., & Castellanos, F. X. (2015). Inscapes: A movie paradigm to improve compliance in functional magnetic resonance imaging. *NeuroImage*, *122*, 222–232. https://doi.org/10.1016/J.NEUROIMAGE.2015.07.069

Zheng, Y.-Q., Farahibozorg, S.-R., Gong, W., Rafipoor, H., Jbabdi, S., & Smith, S. (2021). Accurate predictions of individual differences in task-evoked brain activity from resting-state fMRI using a sparse ensemble learner. *BioRxiv*, 2021.08.19.456783. Retrieved from https://www.biorxiv.org/content/10.1101/2021.08.19.456783v1%0Ahttps://www.biorxiv.org/content/10.1101/2021.08.19.456783v1.abstract