

# View, engage, predict: enhancing brain-behavior mapping with naturalistic movie-watching fMRI

Yunrui Zhang<sup>1</sup>, Emily S. Finn<sup>2</sup>, Mert R. Sabuncu<sup>1,3,4</sup>, and Amy Kuceyeski<sup>3,5,\*</sup>

<sup>1</sup>School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA

<sup>2</sup>Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA

<sup>3</sup>Department of Radiology, Weill Cornell Medicine, New York, NY, USA

<sup>4</sup>Cornell Tech, New York, NY, USA

<sup>5</sup>Department of Computational Biology, Cornell University, Ithaca, NY, USA

\*Corresponding author. [amk2012@med.cornell.edu](mailto:amk2012@med.cornell.edu)

## Abstract

Most brain-behavior mapping studies rely on resting-state functional connectivity (FC), but this approach has known accuracy limits and can be outperformed by movie-watching FC. Here, we present a novel deep neural network framework to predict cognitive scores and sex from FC during naturalistic movie viewing, and examine how movie content and its ability to synchronize brain activity across individuals relate to prediction performance. We show that FC from movie-watching generally outperforms resting-state FC - even when compared to five times more temporal data - with sensory and higher-order brain networks emerging as the most important for prediction. Using both static and sliding-window dynamic FC approaches, we find that higher cognitive prediction accuracy is significantly associated with greater inter-subject synchrony and the duration of human faces and voices in the movies; these effects were not found for sex prediction. This work underscores the promise of naturalistic movie viewing as a powerful tool for probing individual differences in the brain and revealing neural underpinnings of human behavior.

**Keywords:** Functional Connectivity, Deep Neural Networks, Movie-Watching fMRI, Neural Synchrony, Behavior Prediction, Dynamic Functional Connectivity

## 1. Introduction

Understanding brain regions and circuits that underlie individual differences in cognition and behavior is a fundamental goal in neuroscience. Functional MRI, which allows non-invasive recording of brain activity, has revealed that functional connectivity (FC)

patterns are uniquely identifiable across individuals, i.e. FC fingerprinting, and are associated with various behavioral traits [1] [2] [3] [4] [5]. Traditionally, FC fingerprinting and brain-behavior mapping has relied on resting-state fMRI, but naturalistic paradigms such as movie watching offer a more ecologically valid and engaging framework for that at times results in more accurate brain-behavior mappings [6] [7] [8] [9] [10]. Watching a movie induces rich and dynamic neural responses, characterized by synchronized activity across individuals and structured temporal fluctuations in functional connectivity [11] [12] [13] [14]. The use of movie stimuli provides a controlled yet immersive setting, capturing brain activity fluctuations that more closely resemble real-world cognitive processes compared to static resting-state paradigms, in which a person is told not to think about anything in particular.

Recent studies have used FC to predict behavior, i.e. cognitive scores, motor outcomes or diagnosis categories, or demographics like age or sex, but existing models and data remain limited in their predictive power. Most prior brain-behavior mapping studies have employed linear models like linear regression [1] [15]; however, these methods may be limited, as the mapping between functional connectivity (FC) and behavior may involve nonlinear components. The application of deep learning to FC-based behavior prediction has been an area of recent exploration, and may have the potential to uncover complex mappings that traditional models overlook [16] [17] [18] [19] [20] [21]. In this study, we introduce a novel deep neural network (DNN) framework for predicting cognition and sex from FC patterns derived from movie-watching fMRI data, leveraging machine learning techniques to enhance predictive performance and interpretability. One commonly cited drawback of DNNs is that they suffer from a lack of explainability, which is an important aspect of brain-behavior mapping. Being able to reliably identify the functional connections, regions or networks that are important for predicting a given task are of utmost importance for the field of network neuroscience [22] [4] [23] [24] [25]. Knowing the neural underpinnings of behavior or diagnoses would enable better understanding of how we may intervene to enhance performance or reduce symptoms.

Another largely unexplored aspect of FC-based behavior prediction using movie-watching data is its relationship with neural synchrony, or the level of temporal alignment across individuals during shared experiences [26] [27] [28] [29] [30]. Increased neural synchrony has been associated with higher levels of attentional engagement and cognitive processing [31] [32]. However, the connection between neural synchrony and brain-behavior prediction accuracy remains under explored.

Finally, prior research has largely focused on static functional connectivity, overlooking the temporal dynamics of FC during movie-watching. While static FC provides an aggregate measure of connectivity over an entire scan or movie clip, it cannot capture the moment-to-moment fluctuations that occur in response to evolving stimuli [33] [34]. Dynamic functional connectivity (dFC) can reveal transient connectivity states that may be more informative for predicting behavior [35] [36] [37] [38] [39] [40], and in the context

of brain-behavior mapping allow finer scale analysis of the movie features that may be driving prediction accuracy.

In this study, we introduce a novel deep learning-based framework for predicting cognition and sex from functional connectivity during movie-watching and at rest. We compare the predictive performance of various movie clips with differing content, systematically analyzing the contributions of individual brain connections and regions to the predictions. Additionally, we explore the relationship between neural synchrony and brain-behavior prediction accuracy to test our hypothesis that higher neural synchrony generally correlates with better prediction performance. At a finer temporal scale, we segment the movie clips into sliding windows to capture the dynamic fluctuations in neural synchrony and prediction accuracy, further testing for correlations between synchrony and prediction performance. Finally, we investigate the visual, auditory, and semantic features of the movies, identifying key factors that correlate with predictive performance at both the level of the entire movie clip as well as the sliding windows within a clip. By integrating these approaches, our work provides novel insights into the relationship between functional brain connectivity, behavior, neural synchrony, and the features of naturalistic stimuli, offering a more comprehensive understanding of the brain and stimulus features that drive brain-behavior mapping using naturalistic paradigms.

## 2. Results

### 2.1 Brain-behavior mapping accuracy varies across movie clips

We applied a deep neural network (DNN) to the static FC extracted from different movie clips (and rest) to predict individual traits, including cognitive scores and sex, in a cohort of 176 participants (106 female, 70 male, aged 22-36) from the 7T substudy in the Human Connectome Project's Young Adult dataset [41]. Cross-validation was performed with 10-folds, ensuring family members did not span folds; this cross-validation was repeated 100 times to assess robustness. The prediction accuracy for cognitive scores was quantified using the Spearman correlation coefficient between true and predicted scores across participants in the test set, then averaged across the 100 iterations. To assess statistical significance, p-values were computed by randomly permuting the predicted scores 10,000 times in each iteration. We report the median p-value across all 100 iterations. For sex classification, prediction performance was evaluated using the area under the receiver operating characteristic curve (AUC) for the test set, also averaged over 100 iterations. Our results show wide variability in the predictive performance across the different movie clips. Most movie-based FCs outperformed the rest FC calculated using the same duration of fMRI data, whereas using the full 15 mins of resting FC gave accuracy that was higher than most movie clip FC models. However, the best performing movie clip FC models, each with only 3 mins of fMRI data, outperformed resting state FC models using 15 mins of data (Fig. 2a,b).

For total cognition, the highest prediction accuracies were observed for models based on movie-watching FC, including *Social Network* (Spearman  $r = 0.39 \pm 0.02$ , permutation-based  $p = 0$ ), *Ocean's 11* ( $r = 0.38 \pm 0.02, p = 0$ ), and *Inception* ( $r = 0.34 \pm 0.02, p = 0$ ). In contrast, the lowest predictive performance was associated with clips from the third session, including *Flower* ( $r = 0.17 \pm 0.04, p = 0.043$ ), *Garden* ( $r = 0.12 \pm 0.04, p = 0.078$ ), and *Dreary* ( $r = 0.09 \pm 0.05, p = 0.120$ ). The two rest sessions also demonstrate predictive power, with *REST1* ( $r = 0.18 \pm 0.04, p = 0.034$ ) corresponding to the same day as the first and second movie sessions, and *REST4* ( $r = 0.18 \pm 0.03, p = 0.033$ ) aligning with the third and fourth sessions (Fig. 2a). Note that all movie clips and rest sessions were cropped to 164 seconds from their onset to ensure comparability across conditions. The 15-min full rest sessions show better prediction accuracy than the cropped rest sessions (full *REST1*:  $r = 0.23 \pm 0.02, p = 0.009$ ; full *REST4*:  $r = 0.27 \pm 0.03, p = 0.002$ ), but this result is outperformed by the movie clips from *Social Network*, *Inception*, *Ocean's 11* and *Two Men*, which are 5 times shorter than the full rest sessions. The relative ranking of prediction accuracy for other cognitive measures, including crystallized cognition and fluid cognition, followed a similar trend to that of total cognition (Fig. 2a). However, there are slight variations in the prediction accuracy ranking for crystallized and fluid cognition; for example, while *Social Network* yielded the highest predictive accuracy for total cognition ( $r = 0.39 \pm 0.02, p = 0$ ) and crystallized cognition ( $r = 0.42 \pm 0.02, p = 0$ ), *Inception* exhibited the highest predictive accuracy for fluid cognition ( $r = 0.33 \pm 0.03, p = 0$ ). Results for all 12 individual NIH toolbox cognitive measures and chronological age are presented in Supplementary Figure 1; in the main text, we do not present the individual cognitive scores as they are noisier than the composites and we do not present age as the spread of age values is quite small in this population and thus the accuracy quite low.

For sex classification, the highest predictive accuracies were observed for *Ocean's 11* ( $AUC = 0.75 \pm 0.02$ ), *Flower* ( $0.75 \pm 0.02$ ), and *Pockets* ( $0.74 \pm 0.03$ ), whereas *Erin Brockovich* ( $0.68 \pm 0.02$ ), *Garden* ( $0.67 \pm 0.03$ ), and *Dreary* ( $0.62 \pm 0.03$ ) demonstrated the lowest predictive power (Fig. 2b). 164-second *REST1* ( $0.71 \pm 0.03$ ) and *REST4* ( $0.72 \pm 0.02$ ) demonstrated predictive accuracies that were either lower or higher than those of specific movie clips, as observed in cognition prediction. The 15-min full rest sessions show slightly better prediction accuracy (full *REST1*:  $AUC = 0.71 \pm 0.02$ , full *REST4*:  $0.74 \pm 0.02$ ), which is outperformed by shorter movie clips from *Ocean's 11*, *Flower*, *Pockets* and *Hotel*. Notably, while some movie clips consistently outperformed rest sessions in both cognition and sex prediction, the ranking of prediction accuracies for sex classification differed from that of total cognition.

## 2.2 Key Brain Connections and Networks for Multiple Cognitive Measures and Sex Prediction Partially Overlap in Movie-Watching Sessions

We evaluated the contribution of specific brain connections and functional networks to cognition and sex prediction using the Integrated Gradients (IG) method. The importance of individual connections was quantified by calculating the absolute IG values derived from the deep neural network (DNN) model, averaged across all 176 subjects. We identified the connections with the top 1% average importance scores and mapped to their corresponding Yeo 7 functional networks [42], and visualized using chord diagrams (Fig. 3 and, alternatively, glass brain visualizations in Supplementary Figure 3). During movie-watching sessions, the most predictive connections involved interactions between sensory networks—including the visual (VIS) and auditory (within the somatomotor network, SMN)—and high-order networks such as the default mode network (DMN), frontoparietal network (FPN), dorsal attention network (DAN), and salience network (SAN). During the two rest sessions, DMN, FPN and SAN are the most important for behavior prediction.

Although the most predictive connections across movie clips often originate from the same functional networks, the clip-specific top connections vary considerably. For example, in *Inception*, the most predictive connections were primarily within high-order networks (DAN and DMN), whereas in *Ocean's 11*, they were more concentrated within the lower-order visual network. Notably, during both movie-watching and rest sessions, many connections important for cognition prediction also played a significant role in sex classification. The overlap between cognition and sex prediction was more pronounced within the same clip than across different clips. This indicates that perhaps the most important functional connections underlying behavior and demographic prediction may be primarily modulated by the features of the movie stimuli.

To obtain a more quantitative and comprehensive analysis of the association between varied movie clips' FC feature importances, we correlated the FC edge-wise feature importance maps for both total cognition and sex, see Fig. 3b. For the most part, the correlations ranged from moderate to high with *Ocean's 11*, *Inception* and *Social Network* clips appearing somewhat as outliers in both cognition and sex predictions. We performed the same analyses across all clips and all individual NIH toolbox cognition scores (plus sex) in Supplementary Figure 4. We observe the same overall pattern of edge-wise FC feature importances being more similar within the same movie clip for different cognitive scores or sex, compared to the same cognitive score or sex across different movie clips. It appears that the clip features are driving the importance of the FCs in predicting varied cognitive outcomes and sex.

## 2.3 Neural Synchrony and Functional Connectivity-based prediction accuracy are correlated

We observe that different movie clips elicit varying degrees of regional synchrony in the fMRI BOLD responses across subjects. Intuitively, movies that induce higher neural synchrony across the brain may be more engaging for most participants, as higher engagement likely means more time-locked, stimulus-driven brain activity. In addition, as neural activity increases in stimulus-driven regions during such movies, the signal-to-noise ratio (SNR) of the BOLD response and functional connectivity likely improves, potentially enhancing behavior prediction. However, very high levels of neural synchronization could mean reduced inter-individual variability, which may limit prediction capabilities across the population. This raises the question: how does neural synchrony impact brain-behavior prediction?

To investigate this, we first examined whether neural synchrony across all movie clips was correlated with prediction accuracy (Fig. 2c). For each brain region, neural synchrony was quantified using inter-subject correlation (ISC) of preprocessed fMRI BOLD signals across the duration of a movie clip or rest session, averaged over all subject pairs. The overall synchrony for a movie clip or rest session was then computed by averaging regional synchrony values across all 55 regions that had significantly synchronized activity in at least one movie clip. We found a positive, near significant correlation between neural synchrony and total cognition prediction accuracy (Pearson  $r = 0.52, p = 0.0667$ ), and a weaker, non-significant correlation between synchrony and sex prediction accuracy ( $r = 0.38, p = 0.1971$ ). Although the relationships did not reach statistical significance ( $p < 0.05$ ), the positive trend suggests that higher synchrony may contribute to improved total cognition prediction accuracy.

Regional synchrony maps showing the average inter-subject correlation (ISC) across all pairs of individuals for each movie clip and rest session are shown in Fig. 4b, top row. During most movie-watching sessions, the highest synchrony was observed in auditory, visual, and language-related regions. For example, peak synchrony reached around 0.65 in the auditory cortex during *Inception* and 0.48 in *Ocean's 11*. Generally, auditory region synchrony was higher than visual region synchrony, which has been shown in the past[43] and may be ascribed to more variability in gaze location during the movie watching than in what the person is hearing, the latter of which should be the identical across participants. As expected, neural synchrony during rest sessions was minimal across all regions.

Next, we show region-wise feature importance maps, computed by averaging the prediction importance scores of all 267 brain connections involving that region, for total cognition and sex prediction, see Fig. 4a (middle and bottom rows). We then assessed the correlation between regional synchrony and regional feature importance for each movie clip. The correlation between cognition and sex feature importance maps for models based on the FC from the same movie clip exceeded 0.90, highlighting a strong overlap in the predictive features for both outcomes. Additionally, the synchrony maps showed

consistent, moderate, positive correlations with both models' feature importance maps, indicating that regions exhibiting higher inter-subject synchrony also tended to have greater predictive relevance for cognition and sex. See scatter plots showing these correlations in Supplementary Figure 5. Notably, we also observed considerable similarity in regional synchrony and feature importance maps across different clips, with correlations ranging from 0.67-0.95, see in Fig. 4b.

To directly evaluate the role of high-synchrony regions in behavior prediction, we conducted targeted predictions using FC in subsets of brain regions with different levels of synchrony. As shown in Supplementary Figure 6, models based on FC in regions exhibiting high synchrony consistently outperformed models based on FC of the same number of randomly selected regions in multiple movie clips, which in turn outperformed models based on FC of low-synchrony regions. These trends were consistent for both cognition and sex prediction. This finding highlights the importance of high-synchrony regions in accurate brain-behavior models using movie-watching FC.

## 2.4 Sliding-window FC prediction accuracy, Neural Synchrony and Movie Features are correlated

At a finer temporal resolution, movie clips elicit rich dynamic fluctuations in brain activity as visual and auditory stimuli unfold over time. To capture these moment-to-moment changes, we segmented each movie clip into overlapping sliding windows (window size = 60s; step size = 1s) and tracked the temporal evolution of dynamic functional connectivity (dFC) prediction accuracy, neural synchrony, and high-level movie features within each window (Fig. 5a). Several clips exhibited significant positive correlations between dynamic synchrony and total cognition prediction accuracy, including Ocean's 11 ( $r = 0.48$ , permutation-based  $p = 0$ ), Inception ( $r = 0.37$ ,  $p = 0$ ), Erin Brockovich ( $r = 0.26$ ,  $p = 0$ ), Home Alone ( $r = 0.29$ ,  $p = 0$ ), Bridgeville ( $r = 0.31$ ,  $p = 0$ ), Flower ( $r = 0.33$ ,  $p = 0$ ), and Garden ( $r = 0.40$ ,  $p = 0$ ). Similarly, significant positive correlations between synchrony and sex prediction accuracy were observed in Two Men ( $r = 0.19$ ,  $p = 0.02$ ), Home Alone ( $r = 0.40$ ,  $p = 0$ ), Bridgeville ( $r = 0.48$ ,  $p = 0$ ), and Flower ( $r = 0.72$ ,  $p = 0$ ).

To gain a more comprehensive view across the different clips, we aggregated all sliding windows from all movie clips and recalculated the correlations. In this combined analysis, each point in Fig. 5b represents a sliding window, colored by its corresponding movie clip. Across all windows, neural synchrony was significantly positively correlated with both total cognition dFC accuracy (Pearson  $r = 0.37$ , permutation-based  $p = 0$ ) and, to a lesser extent, sex dFC accuracy ( $r = 0.22$ ,  $p = 0$ ). These findings are consistent with the clip-level correlations shown in Fig. 2c, reinforcing the observation that periods of heightened neural synchrony are generally associated with improved cognitive and (lesser so) sex prediction.

We further explored how high-level, sliding window stimulus features modulate dynamic FC prediction performance. Total cognition dFC accuracy was positively correlated

with the presence of human faces ( $r = 0.50$ ,  $p = 0$ ) and to a lesser extent voices ( $r = 0.30$ ,  $p = 0$ ). Here, face/voice presence was defined as the percentage of time points in which these features appeared cross each 60-second window. Sex prediction accuracy was not correlated with face presence ( $r = 0.13$ ,  $p = 0.22$ ) or voice presence ( $r = 0.03$ ,  $p = 0.78$ ). These results align with prior findings from static FC prediction, where the presence of human faces was associated with improved cognitive prediction accuracy [6]. Together, these results suggest that socially rich moments in naturalistic stimuli may enhance both neural synchrony and the predictive power of functional connectivity patterns in mapping to cognitive scores - but not in sex classifications.

To further investigate the semantic content underlying accurate dFC predictions, we analyzed frame-level WordNet features of the movie clips. Within each sliding window, we computed the frequency of WordNet features and weighted them by dFC prediction accuracy (for either cognition or sex). We then visualized the weighted feature distributions as word clouds (Fig. 5c), where word size reflects the weighted average frequency (weighted by prediction accuracy so more predictive windows are given larger weights) and averaged across all windows and clips. For total cognition, human- and social-related features emerged as most predictive, including terms such as “male,” “man,” “person,” “adult,” “talk,” and “look.” This further supports our finding that naturalistic stimuli with high social contents would elicit brain activity that has higher cognition predictive power. For sex prediction, the associated features were more abstract and general, including “entity,” “physical entity,” “object,” “living thing,” “artifact,” “whole,” and “causal agent” suggesting that sex-related neural patterns may be less tied to socially specific stimuli and more broadly distributed.

### 3. Discussion

In this study, we demonstrate the potential of deep neural networks (DNNs) to predict behavioral traits from the brain’s functional connectivity (FC) during movie watching, using data from the HCP 7T substudy. We show that while certain movie clips induce FC that has higher predictive power in terms of cognitive scores and sex than resting-state, others do not. We identify that movie clips with higher across-subject synchrony of brain activity also have better prediction performance for cognition, but not for sex classification. On a dynamic, finer temporal scale, we show that neural synchrony and prediction performance are significantly correlated over time for both cognition and, to a lesser extent, sex. Additionally, cognitive prediction accuracy is positively influenced by the presence of human faces and voices, while sex classification accuracy was not. Finally, we demonstrate that semantic content related to human and social stimuli enhances cognition prediction performance and content related to non-human objects enhances sex classification performance. We find that connections within and between sensory networks (VIS, SMN) and higher-order networks (DMN, FPN, SAN, DAN) are key for predicting both cogni-

tion and sex, with the most salient connections for both cognition and sex being similar 298  
but varying across different movie stimuli and rest. In conclusion, our findings highlight 299  
the importance of visual, auditory, and higher-order network connections in behavior pre- 300  
diction under naturalistic stimuli, underscore the role of neural synchrony in prediction 301  
accuracy, and suggest that human- and social-related movie content is especially effective 302  
in stimulating cognition-related brain activity. 303

The proposed deep neural network (DNN) for brain-behavior mapping introduces non- 304  
linearity and complex feature spaces, effectively capturing intricate relationships between 305  
functional connectivity (FC) and behavioral traits. We highlight that movie-watching 306  
FC yields a wide range of prediction accuracies for cognition and sex, depending on each 307  
clip's ability to elicit neural synchrony and the clip's content. Most clips outperform 308  
resting-state FCs of the same duration, and some even outperform rest sessions that 309  
are 5 times longer. While the model's prediction results demonstrate its effectiveness, 310  
there are some limitations. Compared to previous studies on cognition prediction from 311  
movie-watching connectivity using linear models (connectome-based predictive modeling, 312  
CPM)[6], our results are similar overall, but the prediction accuracy is slightly higher 313  
for highly predictive clips (e.g. *Social Network*) and lower for less predictive clips (e.g. 314  
*Dreary*). We also note that some studies can achieve higher sex classification accuracy 315  
from a training set with more subjects and longer movie-watching [44]. This discrepancy 316  
may be due to overfitting, which can arise from the high complexity of the DNN, the 317  
high-dimensional input FC vectors, and the relatively small number of training subjects. 318  
To mitigate overfitting, we employed several strategies in the model design, applied strong 319  
regularization techniques, and reduced dimensionality via PCA. While the PCA+MLP 320  
framework performed better than other machine learning approaches (e.g., XGBoost, con- 321  
volutional neural networks) we have tried, future research could explore more advanced 322  
non-linear approaches including topography-based networks [16], convolutional-recurrent 323  
models [19] and graph neural networks [45] [46]. 324

Our findings also highlight the previously unexplored relationship between inter-subject 325  
neural synchrony during movie watching and FC prediction performance. There is a po- 326  
tential tradeoff in inducing neural synchrony - on one hand, higher synchrony improves 327  
the signal-to-noise ratio (SNR) of BOLD activity and corresponding FC; on the other 328  
hand, higher synchrony means less inter-subject variability, which may negatively impact 329  
behavior prediction. However, our overall finding of a positive correlation between neural 330  
synchrony and cognition prediction performance (and to a lesser extent sex classification) 331  
supports the conclusion that the primary benefit of synchrony lies in improving SNR of 332  
FCs - at least for this level of synchrony. While previous studies have pointed out that 333  
periods of high inter-subject FC similarity tend to align across viewers and are linked to 334  
the brain's modular architecture [33], we here directly examine how similarity in BOLD 335  
responses affect brain-behavior mapping. Periods of higher synchrony tend to provide 336  
more accurate predictions, and brain regions of high synchrony tend to be more inform- 337

ative in terms of cognitive abilities and sex classification. These synchronizing moments 338  
are sometimes referred to as "events", though in the context of co-fluctuations in z-scored 339  
BOLD activity across brain connections [29]. This finding is especially relevant for fu- 340  
ture fMRI experiment design, suggesting that stimuli that evoke higher synchrony may 341  
be preferable. It would be of interest to try and maximize synchrony to test if accuracy 342  
in predicting individual metrics would hit a ceiling - it is our opinion that such level of 343  
synchrony was not inducible due to inherent inter-individual variability in brain activity 344  
responses to the same stimuli. 345

At both static and dynamic scales, we found that the percentage of movie frames with 346  
human faces and voices was positively correlated with prediction accuracy for cognition, 347  
but not for sex. Additionally, we found that predictions of fluid and crystallized cognition 348  
are strongly associated with face presence (more so fluid), and to a lesser extent, voice 349  
presence, as shown in Supplementary Figure 2. These findings suggest that movies rich 350  
in social content engage brain networks relevant to cognition, but not to sex. This result 351  
aligns with previous findings at both the clip and dynamic levels: at the clip level, studies 352  
have shown that the duration of face presence is positively correlated with prediction ac- 353  
curacy during movie watching [6]; at the dynamic level, movie watching has been shown 354  
to elicit co-fluctuations in both core and extended face-processing regions, which in turn 355  
correlate with face recognition scores [47]. The stronger correlation between cognitive 356  
prediction and face presence than voice presence is also consistent with previous findings 357  
that face-selective brain regions show more robust and distributed identity-related neural 358  
patterns than voice-selective regions [48]. On the other hand, objects not related to hu- 359  
mans - both living and non-living, were related to better sex classification. These findings 360  
have practical implications for designing future naturalistic viewing experiments — the 361  
content of the movie may be varied to obtain more accurate brain-behavior mappings, 362  
but this accuracy appears to depend on the specific outcome one is trying to predict. 363

Our implementation of DNNs in the brain-behavior mapping allowed us to employ the 364  
Integrated Gradients (IG) method to identify the most predictive connections on an in- 365  
dividual subject basis. The gradient-based approach measures feature importance at the 366  
connection level, while the interpolation method enhances robustness. Our way to assess 367  
general feature importance across the population was to average the individual-level fea- 368  
ture importances from the IG framework, however, one could also look at variability of the 369  
feature importances across individuals or subgroups within the overall group. We observe 370  
that the top FC features important for predicting cognition and sex are more consistent 371  
across movie clips than across the different outcomes that are being predicted, but that, 372  
generally speaking, the entire vector of feature importances were still moderately to highly 373  
correlated across movies (Fig. 3, 4, Supplementary Figure 4). The most important con- 374  
nections across the movie-watching FC mapping included connections within and between 375  
sensory networks (VIS, SMN) and higher-order networks (DMN, FPN, SAN, DAN). This 376  
finding is consistent with prior work using support vector machine models [25], which 377

identified the FPN, DMN, and dorsal attention network (DAN) as key contributors to 378  
predicting cognitive states during movie-watching. It also partly agrees with work using 379  
convolutional models and deconvolution-based importance mapping [44], which showed 380  
that the default mode network (DMN) has the highest predictive power for sex classific- 381  
ation, while the frontoparietal network (FPN) is most predictive of fluid and crystallized 382  
intelligence. 383

Importantly, our results reveal a high level of similarity between cognition and sex 384  
prediction importance maps within a clip compared to across a clip for the same outcome. 385  
Looking at the individual NIH Toolbox cognition test scores, we see a similar pattern of 386  
moderate levels of agreement across all movies (and rest) but tighter correlations within 387  
a movie for different cognitive scores than across a movie for the same cognitive score. 388  
This indicates that different movies may modulate somewhat different brain connections 389  
which could cause some differences in the feature importance maps. Interestingly, we see 390  
that the highest performing movie FC (Oceans 11, Social Network and Inception) feature 391  
importances are more similar to one another and tended to be more different from the 392  
other movie FC feature importance maps. Finally, the movie FC feature importances at 393  
both the regional and pairwise level were correlated with the FC feature importances from 394  
the resting-state model to an extent that was similar to the feature correlations across 395  
different movies. This may suggest that the neural substrates important for predicting 396  
to sex and cognition are similar across movies and rest, but that there is some variability 397  
induced by the specific movie being watched which likely depends on the across-subject 398  
synchrony of various brain regions and the resulting change in SNR of FC. 399

The analysis of movie-watching functional connectivity-based brain-behavior mapping 400  
is crucial for understanding the unique value of this paradigm in understanding the neural 401  
circuits underlying behavior traits or demographics. We find that two key elements correl- 402  
ated with movie-watching FC brain-behavior mapping accuracy are across-subject neural 403  
synchrony and movie features; we find that the movie features driving accuracy may de- 404  
pend on the metric we are trying to predict. Specifically, we find that movies with human- 405  
and social-related content are most effective for cognition and non-human object elements 406  
drive sex classification. In summary, these results highlight the importance of naturalistic 407  
viewing in fMRI neuroimaging and brain-behavior mapping, offering rich insights into 408  
brain mechanisms, and points us to further exploration of the mapping between movie 409  
features, neural synchrony, and accuracy in brain-behavior mapping using naturalistic 410  
paradigms. 411

## 4. Methods 412

### 4.1 Data 413

#### 4.1.1 Dataset, Participants, and Behavioral Measures 414

In this study, we use functional magnetic resonance imaging (fMRI) and behavioral data 415  
from the Human Connectome Project (HCP) 7T movie-watching dataset [41]. The dataset 416  
comprises n=184 healthy adult participants (ages 22 to 35), who underwent scanning 417  
using a Siemens Magnetom 7T scanner equipped with an optimized multiband echo- 418  
planar imaging (EPI) sequence ( $TR = 1000ms$ ,  $TE = 22.2ms$ ,  $voxelsize = 1.6mm^3$ ). We 419  
use a subset of n=176 subjects with complete data for all resting state and movie fMRI 420  
runs as well as behavioral data needed in this study. 421

In this study, we focus on predicting two items: biological sex (106 female, 70 male) 422  
and cognitive performance, both of which have been associated with functional connectiv- 423  
ity patterns. We use the total cognition score which is the average of crystallized cognition 424  
and fluid cognition, which are derived from the NIH toolbox cognition tests including Pic- 425  
ture Vocabulary, Reading Recognition, Dimensional Change Card Sort, Flanker Inhibitory 426  
Control and Attention, Picture Sequence Memory, List Sorting, and Pattern Comparison 427  
tests [49]. For completeness, we included the analysis for individual cognitive measures 428  
in Supplementary Information, but generally speaking the individual cognitive scores are 429  
noisier than the composite scores. 430

#### 4.1.2 fMRI Acquisition and Preprocessing 431

Resting-state fMRI (rfMRI) data were collected in four separate runs, each lasting ap- 432  
proximately 16 minutes, at the beginning of each of the four 7T imaging sessions. During 433  
these runs, participants were instructed to keep their eyes open and maintain a relaxed 434  
fixation on a bright cross-hair projected against a dark background in a dimly lit scanning 435  
environment. For this study, we used data from the first (*REST1*) and fourth (*REST4*) 436  
resting-state sessions, as these were the only ones conducted on the same day as the 437  
movie-watching fMRI sessions. 438

Following the resting-state scans in the first and fourth imaging sessions, participants 439  
underwent two additional fMRI runs in which they viewed a total of 15 movie clips 440  
designed to engage cognitive and emotional processing. These clips, ranging from 1 to 4.3 441  
minutes in length, consisted of both independent short films and excerpts from commercial 442  
Hollywood movies. The individual clips were concatenated into 12- to 14-minute-long 443  
.mp4 files, with a common Vimeo repeat validation clip included to assess test-retest 444  
reliability [50]. 445

We excluded the test-retest clip, as well as another clip that was too short compared to 446  
the others. This resulted in a final set of 13 movie clips for our analysis. The movies were 447  
displayed using VisuaStim Digital MRI-compatible goggles to ensure consistent visual 448

stimulation. We used brain activity data from all movie clips and resting-state sessions 449  
between the 10th and 174th seconds from the beginning of the scan (except where we also 450  
include all 15 mins of rsfMRI data for comparison), excluding the initial unstable period 451  
and aligning the end of the window with the shortest clip duration to ensure comparability 452  
of brain activity across all clips. 453

The specific movie clips analyzed in this study were as follows: 454

- MOVIE1 (first session, first day): Two Men, Welcome to Bridgeville, Pockets 455
- MOVIE2 (second session, first day): Inception, The Social Network, Ocean's 11 456
- MOVIE3 (first session, second day): Flower, Hotel, Garden, Dreary 457
- MOVIE4 (second session, second day): Home Alone, Erin Brockovich, Star Wars 458

Clips in MOVIE1 and MOVIE3 are freely available independent films under Creative 459  
Commons licensing, and clips in MOVIE2 and MOVIE4 are from Hollywood movies as 460  
prepared and published by Cutting et al. 2012 [51]. 461

Standard preprocessing of fMRI data, including motion correction, distortion correc- 462  
tion, high-pass filtering, and nonlinear alignment to MNI template space [52] was used in 463  
the FIX-denoised volume-space data from the HCP 7T release. We also applied global 464  
signal regression (GSR) by removing the global mean time series signal from each voxel's 465  
BOLD time series, thereby reducing potential confounds arising from widespread non- 466  
neuronal noise, including physiological artifacts and scanner-related fluctuations[53]. 467

### 4.1.3 Static and Dynamic Functional Connectivity Construction 468

**Static functional connectivity(sFC):** To quantify static functional connectivity (sFC), 469  
we employed the Shen 268-region brain atlas [54]. For each region in the atlas, we com- 470  
puted the mean BOLD time series by averaging voxel-wise preprocessed signals within 471  
that region for each subject, run, and time frame. We then computed pairwise Pearson 472  
correlation coefficients between all 268 regions, yielding a  $268 \times 268$  functional connectiv- 473  
ity matrix for each movie clip or rest run per subject. We then extracted the upper 474  
triangular portion of each matrix (excluding the diagonal) and vectorized it into a 35,778- 475  
dimensional feature vector [6]. 476

**Dynamic functional connectivity(dFC):** To capture temporal fluctuations in func- 477  
tional connectivity during movie-watching, we adopted a sliding-window approach for dy- 478  
namic functional connectivity (dFC). Each movie clip was segmented into overlapping 479  
time windows of 60 seconds, with a step size of 1 second, resulting in 104 sliding win- 480  
dows for each movie clip. Within each window, the mean BOLD signals were computed, 481  
and Pearson correlation was used to generate a  $268 \times 268$  FC matrix for each window, 482  
resulting in 163 sliding-window dFC matrices for each clip. 483

## 4.2 Deep Neural Network(DNN) Prediction Model 484

We propose a novel deep neural network (DNN)-based behavior prediction model designed 485 to capture the complex relationship between dynamic/static functional connectivity and 486 behavioral and demographic traits (Fig. 1). The model takes as input a set of  $n = 176$  487 functional connectivity vectors, each of size (35,778), representing the upper triangular 488 elements of the  $268 \times 268$  FC matrix. The output corresponds to either cognition score (a 489 continuous variable regression task) or biological sex (a binary classification task). Given 490 the relatively small number of subjects, we employed a 10-fold cross-validation approach 491 to ensure robust model evaluation. The 176 subjects were randomly divided into 10 492 groups, while making sure that individuals from the same family were always assigned 493 to the same group. This step prevents data leakage and reduces the risk of inflated 494 performance estimates due to genetic or familial similarities in brain connectivity. In 495 each cross-validation iteration, the model was trained on 9 groups and tested on the 496 remaining 1 group, ensuring that every subject was included in the test set exactly once. 497 Predicted values for all subjects were collected across the 10 folds and then correlated with 498 the true values to evaluate model performance. The division into 10 folds was performed 499 100 times to assess uncertainty due to varied groupings of individuals into train and test. 500

The prediction model consists of two primary components: 501

### (1) Principal Component Analysis (PCA) Block: 502

To reduce dimensionality and minimize noise, we apply PCA to the input FC vectors 503 across all training subjects, retaining the first 100 principal components. This results in a 504 (#training subjects, 100) feature matrix, capturing generally around 84% of the variance. 505

### (2) Multi-Layer Perceptron (MLP) Block: 506

- The PCA-reduced data is fed into a three-layer MLP, each layer containing 500 507 nodes with ReLU activation. 508
- To reduce overfitting, we employ dropout (rate = 0.5) in each hidden layer and 509 apply L2 weight decay (0.01) as regularization. 510
- The model is optimized using the Adam optimizer with a learning rate of 0.001. 511
- The output layer is: 512
  - Linear for cognition prediction (regression task). 513
  - Sigmoid for sex classification (binary classification task). 514
- The loss function is: 515
  - Mean Squared Error (MSE) loss for cognition prediction. 516
  - Binary Cross-Entropy Loss for sex prediction. 517

While overfitting remains a challenge due to the limited sample size and the high- 518  
dimensional, noisy nature of fMRI-based FC measures [55] [56] [57] [58], the model’s 519  
hyperparameters were carefully tuned to strike a balance between learning the complex 520  
FC-behavior mapping and avoiding excessive overfitting. The model is developed using 521  
Pytorch on a Cuda GPU environment. To evaluate the model’s predictive performance, we 522  
computed Spearman’s correlation coefficient ( $r$ ) between the predicted and true behavior 523  
scores for cognition. For sex classification, we assessed the model using the Area Under 524  
the Curve (AUC) of the receiver operating characteristic (ROC) curve. 525

### 4.3 Statistical Analysis of Predictive Connections and Networks 526

To quantify the predictive contribution of individual brain connections to the predictions, 527  
we employed Integrated Gradients (IG), a technique designed to estimate feature im- 528  
portance in deep neural networks [59] [60]. IG uses backward propagation to compute 529  
gradients of the model’s output with respect to the input, and improves robustness by 530  
integrating these gradients across a path between a baseline and the actual input. 531

In our DNN behavior prediction model, the IG procedure is as follows: 532

- PCA transformation: First, we perform Principal Component Analysis (PCA) on 533  
the training functional connectivity (FC) vectors, resulting in a (100,1) principal 534  
component(PC) vector for each training subject. 535
- Baseline Definition: The baseline reference input is defined as the average of each 536  
of the PC vectors across all training subjects (i.e. one vector per PC). 537
- Interpolation: A series of intermediate vectors are generated by linearly interpolat-  
ing between the baseline and the actual input principle component (PC) vector of  
a subject. Specifically, for an input PC vector  $\mathbf{x}$  and baseline  $\bar{\mathbf{x}}$ , we compute the  
 $k$ th step:

$$\mathbf{x}^{(k)} = \bar{\mathbf{x}} + \frac{k}{m}(\mathbf{x} - \bar{\mathbf{x}})$$

where  $m = 50$  is the number of interpolation steps. 538

- Gradient Computation: For each interpolated vector  $\mathbf{x}^{(k)}$ , we compute the gradient 539  
of the DNN’s output (predicted cognition score or sex classification probability) 540  
with respect to its  $i$ th feature  $\mathbf{x}_i^{(k)}$ ,  $i = 0, \dots, 99$ . 541
- Integration Along the Path: The importance of each feature  $i$  of the PC vector is  
estimated by integrating these gradients over all interpolation steps:

$$\begin{aligned} IG(x_i) &= \frac{1}{m}(x_i - \bar{x}_i) \sum_{k=1}^m \frac{\partial F(\mathbf{x}^{(k)})}{\partial x_i} \\ &= \frac{1}{m}(x_i - \bar{x}_i) \sum_{k=1}^m \frac{\partial F(\bar{\mathbf{x}} + \frac{k}{m}(\mathbf{x} - \bar{\mathbf{x}}))}{\partial x_i}, i = 0, \dots, 99 \end{aligned}$$

where  $F(x)$  is the DNN's prediction function, and  $\frac{\partial F}{\partial x_i}$  represents the gradient of the model output with respect to the  $i$ th entry of the input PC vector, computed by backward propagation.

- PCA inverse: The gradient is multiplied by the inverse of PCA loadings to project back to the gradients of the whole set of 35,778 brain connections.
- Feature Importance Assignment: For each subject, we obtain a set of IG values for each functional connection (i.e., each of the 35,778 connections). The absolute values of the IG for each connection are averaged across all subjects to compute its importance score, reflecting its contribution to behavior prediction.

At the regional level, we computed the average of the absolute values of the importance scores for the 267 connections associated with each region, which was used as its relative importance. At the functional network level, we mapped the most important connections to their corresponding Yeo canonical functional networks [42] and analyzed which functional networks contained the most predictive connections.

## 4.4 Neural Synchrony and its Relationship to Prediction Performance

### 4.4.1 Neural Synchrony

During naturalistic viewing, people tend to exhibit higher neural synchrony when watching engaging and cognitively stimulating movies, whereas less engaging content often results in more variable neural responses across individuals [26] [27] [28] [29] [30]. To quantify neural synchrony, we calculated the inter-subject Pearson correlation of regional fMRI BOLD signals over time, reflecting shared neural responses across participants. This resulted in 268 region-specific synchrony values per subject pair. We then averaged these values across all subject pairs for each region to obtain regional population-level synchrony.

For each movie clip, we identified significantly synchronized regions based on the Pearson correlation of BOLD signals across all subject pairs. To correct for the large number of pairwise comparisons ( $176 \times 175/2$  total pairs), we applied False Discovery Rate (FDR) correction using the Benjamini-Hochberg procedure [61]. Regions with adjusted p-values below  $\alpha = 0.05$  were considered significantly synchronized. We took the union of significant regions across clips (collected into the set called ROI), and then averaged these to obtain a whole-brain synchrony measure; this procedure excludes regions with non-significant synchrony across all movie clips.

- Pairwise neural synchrony for a pair of subjects  $(i, j)$ , for a region  $v$ :

$$r_{i,j;v} = \text{Pearsonr}(BOLD_{i;v}, BOLD_{j;v})$$

- Region-wise synchrony for a region  $v$  for all subject pairs:

$$r_v = \frac{1}{n(n-1)/2} \sum_{i=1}^n \sum_{j=i+1}^n r_{i,j;v}, n = 176$$

- Global, whole brain synchrony for a pair of subjects  $(i, j)$ :

$$r_{i,j} = \frac{1}{\#ROI} \sum_{k=1}^{\#ROI} r_{i,j;v_k}, v_k \in \{ROI\}$$

#### 4.4.2 Relationship Between Neural Synchrony and Behavior Prediction Performance 574 575

To investigate the relationship between neural synchrony and behavior prediction, we analyzed both whole-brain and regional levels. At the whole-brain level, we computed an overall synchrony score for each movie clip by averaging synchrony across all regions of interest (ROIs). We then correlated this score with the prediction accuracy for cognition and sex across all movie clips to assess whether a significant relationship exists. 576  
577  
578  
579  
580

At the regional level, we examined the correlation between the region-wise synchrony map and the region-wise prediction importance map. The prediction importance map was calculated by averaging the importance scores of all 267 functional connections linked to each brain region. A positive correlation would indicate that regions with higher synchrony are also more predictive. To verify the robustness of the correlation, we used both Pearson and Spearman correlation metrics. This dual approach helps ensure the correlation is not driven solely by low-synchrony regions, where predictive power might still vary despite minimal synchrony. 581  
582  
583  
584  
585  
586  
587  
588

To further evaluate the impact of high- and low-synchrony brain regions on behavior prediction, we conducted targeted predictions using subsets of brain regions exhibiting distinct levels of synchrony. For each movie clip, we identified high-synchrony regions as those with inter-subject correlation values satisfying FDR-corrected  $p < 0.05$ . We then selected an equal number of randomly chosen regions and an equal number of low-synchrony regions (ranked in descending order of p-values). Using these subsets, we trained separate DNN prediction models while maintaining the same model architecture and compared their prediction accuracies. This approach allowed us to determine whether high-synchrony regions contribute more strongly to behavior prediction compared to randomly selected or low-synchrony regions. 589  
590  
591  
592  
593  
594  
595  
596  
597  
598

## 4.5 Sliding-Window Dynamic Functional Connectivity, Neural Synchrony and Movie Features

### 4.5.1 Sliding-Window Dynamic Functional Connectivity

To capture temporal dynamics, we employed a sliding-window approach that segments each 164-second movie clip into overlapping 60-second windows with a 1-second step size. This resulted in 104 windows per movie clip, with each window defined by its midpoint (e.g., the 30th second corresponds to the window from 0 to 60 seconds).

We trained separate DNN models to predict cognitive score and sex for each 60-second sliding window dynamic functional connectivity (dFC), ensuring that the model performance reflected the predictive power of the FC for that specific time window under the given naturalistic stimulus. While the structure of the DNN model and cross-validation method were identical to those used for static FC, the key difference was that the input FC vector was based on shorter temporal segments. This approach allowed us to generate an accuracy curve over time during the movie clip, which we later compared with fluctuations in neural synchrony and movie features.

### 4.5.2 Sliding-Window Neural Synchrony

We use the same set of overlapping 60-second sliding windows and compute the inter-subject correlation of fMRI BOLD time series within the window to generate a dynamic sequence of whole-brain neural synchrony levels. Similar to static neural synchrony, we took the average of inter-subject correlation across regions of interest (ROIs) to obtain the whole-brain synchrony level for each window. We can thus obtain the time fluctuation of synchrony level among the subjects. At the movie clip level, we computed the Pearson correlation between neural synchrony and dynamic FC prediction accuracies across the 104 windows for each clip. A strong positive correlation would indicate a high degree of co-fluctuation and similar temporal patterns between synchrony and FC prediction performance over time. At a more aggregate level, we combined the sliding windows from all 13 movie clips and computed the same Pearson correlation between neural synchrony and dynamic FC prediction accuracies across the  $13 \times 104$  windows. This approach provides much more data points compared to static FC, allowing for a more robust analysis and helping to further validate the relationship between synchrony and behavioral prediction performance.

### 4.5.3 Sliding-Window Movie Features

We also investigated whether specific high-level features of the movie, including the presence of human faces and voices, influenced the accuracy of behavior prediction. To quantify these features, we calculated the percentage of time each feature was present within each 60-second sliding window, defined as the proportion of TRs (time points)

where at least one human face or one person talking was detected. The presence of hu- 635  
man faces in each frame was identified using the open-source Python package *pliers* [62]. 636  
Human voice presence in each frame is identified using the open-source python package 637  
*webrtcvad* (<https://github.com/wiseman/py-webrtcvad>). Note that the TR for fMRI 638  
scanning in the HCP-7T dataset is 1 second, meaning the "frame" here is also 1 second to 639  
align with the dataset's temporal resolution. We then examined the correlations between 640  
these features and the sliding-window dFC (dynamic functional connectivity) prediction 641  
accuracy to explore how visual and auditory elements shape dFCs predictive power. 642

Additionally, we used WordNet features to analyze the semantic content driving high 643  
prediction accuracy for cognition and sex. The HCP-7T dataset includes frame-wise word 644  
descriptions that label the content of each frame, as specified by the HCP movie labeling 645  
pipeline [63] [64]. For each sliding window, we counted the occurrences of each word, then 646  
weighted these occurrences by the behavior prediction accuracy for that window. The 647  
average of these weighted occurrences was computed across all 13x104 sliding windows. 648  
This produced a weighted average for each word in the movie labeling dictionary, reflecting 649  
the words most influential for behavior prediction which we visualized using word clouds. 650

## 5. Figures

651

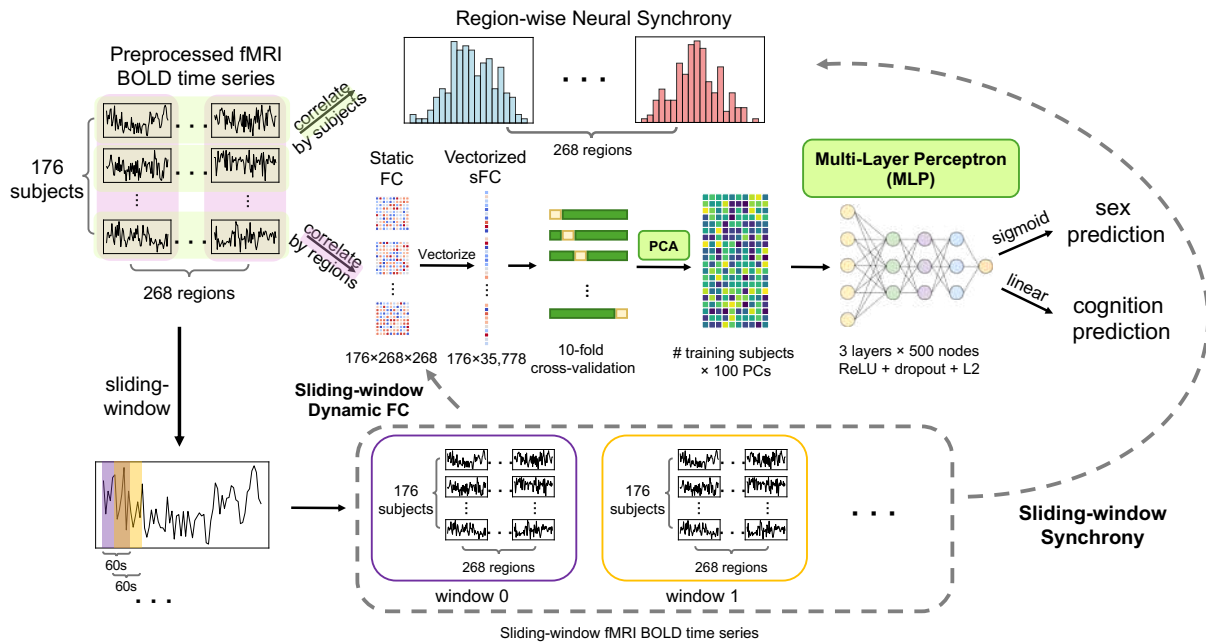
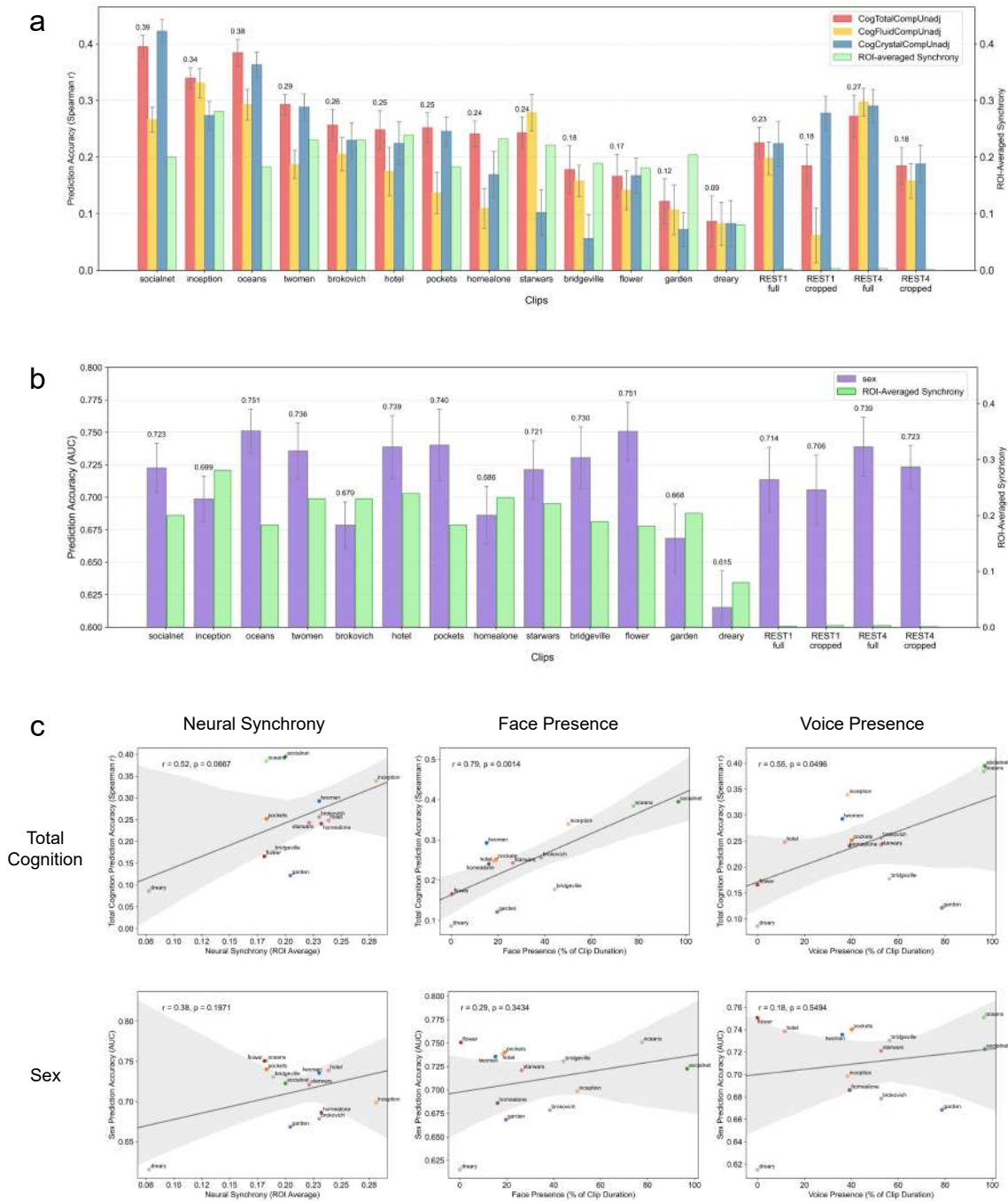
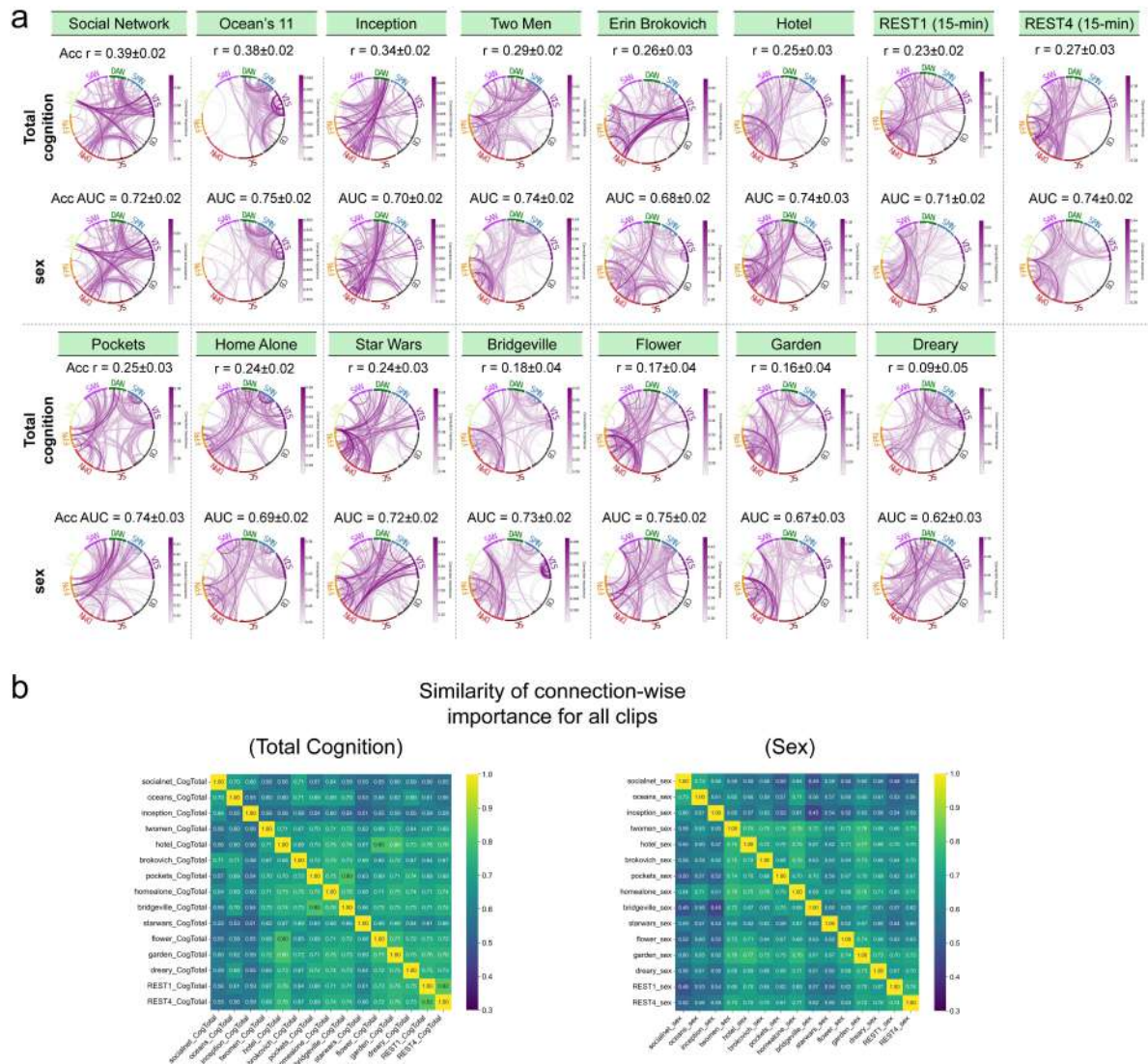


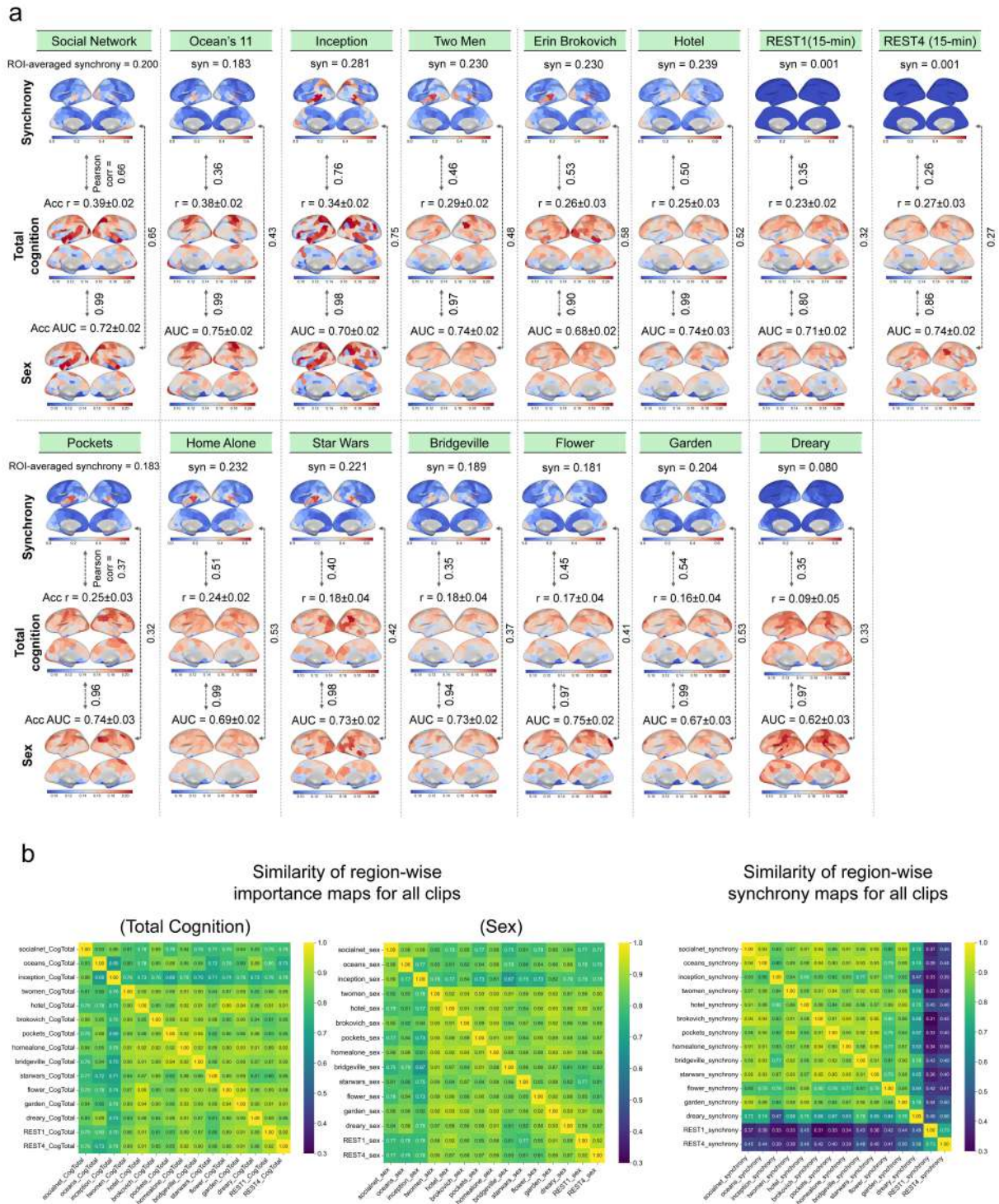
Figure 1: **Workflow of the study.** Static functional connectivity (sFC) is computed by correlating fMRI time series across pairs of brain regions, followed by vectorization and PCA to reduce dimensionality. Neural synchrony is defined by across-subject correlations of the same brain regions' fMRI time series and assessed by the mean/std of distribution across pairs of subjects. In parallel, a sliding-window approach is applied to compute dynamic functional connectivity (dFC) and sliding-window neural synchrony, capturing temporal fluctuations in connectivity patterns. The extracted features from both static and dynamic FC are fed into a 3-layer multi-layer perceptron (MLP) model with 500 nodes per layer for classification (sex prediction) and regression (cognition prediction).



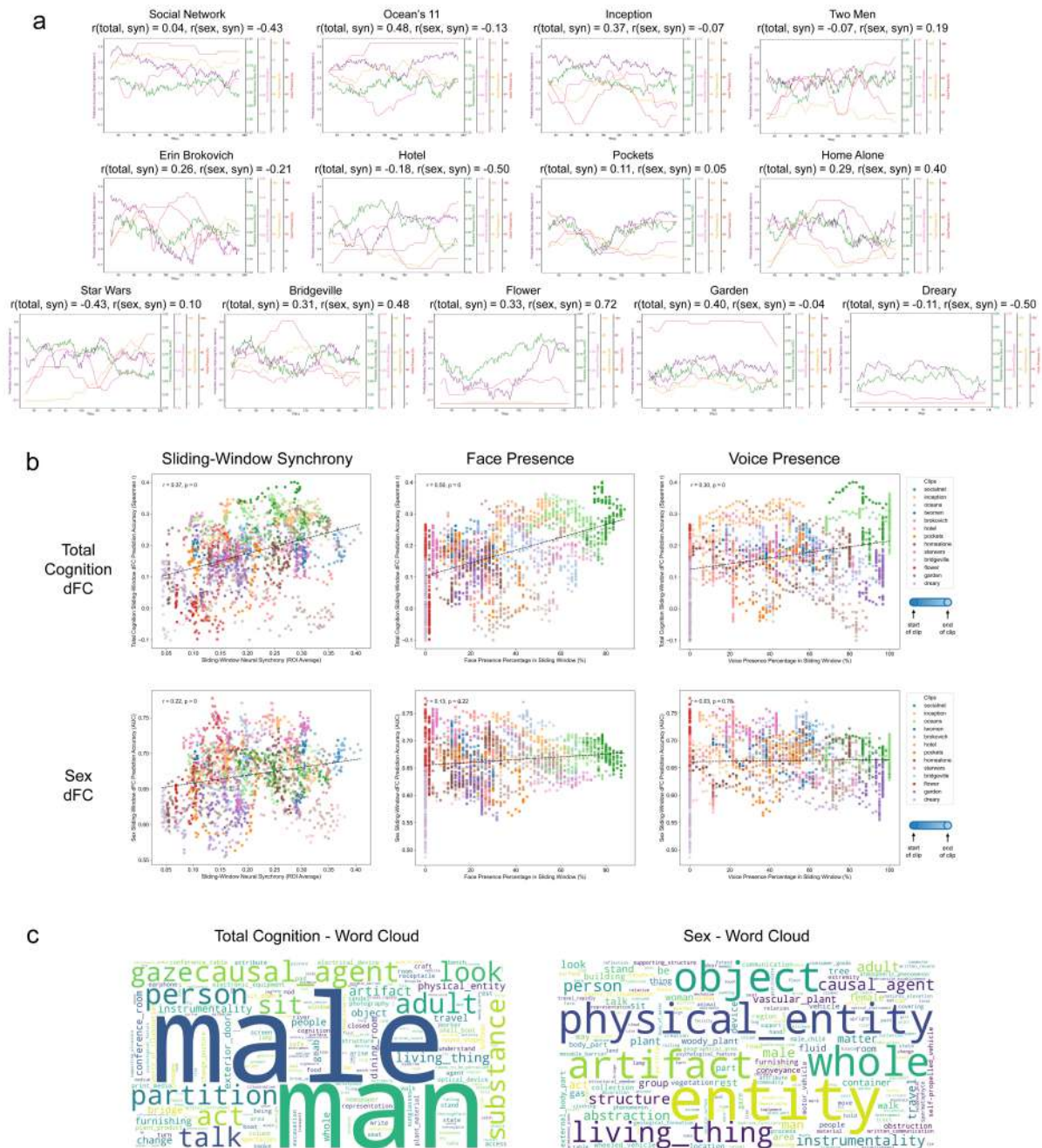
**Figure 2: Brain-behavior mapping model performance.** (a) Prediction accuracies for total, fluid, and crystallized composite cognition scores across all movie clips and resting-state sessions (both full and cropped), evaluated using Spearman's rank correlation coefficient ( $r$ ). The colored bar height is the average and the grey lines are the standard deviation of the accuracy values over the 100 cross-validation iterations. All sessions were cropped to 164 seconds from onset to ensure same input length across sessions, except for the full resting-state fMRI scan that consists of 15 mins of data. Regional average neural synchrony, over 55 brain regions with significant synchrony in at least one movie clip, averaged across all pairs of subjects are shown in green. (b) Prediction accuracies for sex classification across all movie clips and resting-state sessions, assessed using area under the curve (AUC); bar height is the average value and the grey lines are the standard deviation of the accuracies over the 100 iterations. Regional average neural synchrony over all pairs of subjects are shown in green. (c) Across movie clips, neural synchrony is positively and significantly correlated with total cognition prediction accuracy ( $r = 0.52, p = 0.0667$ ) and positively but not significantly related to sex classification accuracy ( $r = 0.38, p = 0.1971$ ). Total cognition prediction accuracy is significantly correlated with the percentage of clip duration containing human faces ( $r = 0.79, p = 0.0014$ ) and voices ( $r = 0.55, p = 0.0496$ ). Sex prediction accuracy also shows positive, though weak and non-significant, correlations with face presence ( $r = 0.29, p = 0.3434$ ) and voice ( $r = 0.18, p = 0.5494$ ) presence.



**Figure 3: Brain connections important to total cognition and sex predictions.** (a) For each clip and resting-state session, chord diagrams visualize the most important connections—specifically, those within the top 1% of average importance scores—for both total cognition and sex prediction models. The most important connections appear to vary across clips and remain consistent for total cognition and sex within a clip. These critical connections are predominantly located in sensory networks (VIS, SMN) and high-order networks (DMN, FPN, SAN, DAN), though the specific top connections vary across different movie clips. (b) Correlation of edge-wise FC importance across all clips, for total cognition prediction and sex prediction. Edge-wise FC importance similarity is assessed by correlating the importance scores of all pairwise FCs across two fMRI sessions.



**Figure 4: Neural synchrony and static FC feature importances.** (a) For each clip, the first row shows regional neural synchrony maps (average across synchronized regions listed at the top of the panel). The second and third rows display region-wise prediction importance maps, i.e. the average of the importance scores of all edge-wise FCs to that region, for total cognition and sex, with the average prediction accuracy at the top of the panel. For all clips, neural synchrony and feature importance maps are positively correlated, indicating that regions with higher synchrony also tend to play a key role in behavior prediction. (b) Similarity of region-wise importance maps across all clips, for total cognition prediction, sex prediction and neural synchrony respectively. Consistent with results from edge-level (see Fig. 3(b)), the region-level prediction importance is positively correlated across all clips. Neural synchrony is also consistent across clips, with the exception of clips showing particularly low synchrony, such as *Dreary*, and the two rest sessions.



**Figure 5: Sliding-window dynamic FC prediction, neural synchrony, and movie features co-fluctuate over time.** (a) Sliding-window dynamic FC prediction accuracies for total cognition (purple) and sex (green), neural synchrony (pink), and percent of human face/voice presence (yellow/red, respectively) for all movie clips. (b) Correlations between sliding-window dynamic functional connectivity (dFC) prediction accuracies and sliding-window neural synchrony, face presence, and voice presence are computed across all sliding windows in all movie clips. Each point represents a sliding window, each color a corresponding movie clip. Point transparency indicates temporal position within the clip, ranging from 100% opacity at the beginning to 50% at the end. Sliding-window synchrony shows a positive correlation with both total cognition dFC prediction ( $r = 0.37, p = 0$ ) and sex dFC prediction ( $r = 0.22, p = 0$ ), with  $r$ -values calculated by Pearson correlation and  $p$ -values estimated using a 10,000-iteration permutation test. Total cognition dFC prediction is also positively correlated with face ( $r = 0.50, p = 0$ ) and voice ( $r = 0.30, p = 0$ ) presence percentage within sliding-windows. Sex dFC prediction was non-significantly positively correlated with face presence ( $r = 0.13, p = 0.22$ ) but not correlated with voice presence ( $r = 0.03, p = 0.78$ ). (c) The average frequency of WordNet features within each sliding window, weighted by that window's dFC prediction accuracy for total cognition (left panel) or sex (right panel). The size of the words represents the weighted average count; larger words correspond to a feature that is present during more predictive sliding windows.

## 6. Data Availability 652

The HCP 7T Dataset is publicly available at <https://db.humanconnectome.org/>. Pre-processed shen268 atlas movie-watching BOLD time series data can be found at: [https://github.com/esfinn/movie\\_cpm](https://github.com/esfinn/movie_cpm). 653  
654  
655

## 7. Code Availability 656

The code used in this study, along with usage instructions, is publicly available at: [https://github.com/Kyrrego/MovieFC\\_Behav\\_Mapping.git](https://github.com/Kyrrego/MovieFC_Behav_Mapping.git). 657  
658

## 8. Acknowledgements 659

This work was supported by the Ann S. Bowers foundation through the Ann S. Bowers Women's Brain Health Initiative. 660  
661

## 9. Conflicts of Interest 662

The authors declare no conflict of interest. 663

## 10. Author Contributions 664

A.K. and M.S. conceived the experiments and interpreted the results. Y.Z. conducted the experiments, analyzed, and interpreted the results. E.F. interpreted the results. Y.Z. and A.K. wrote the manuscript. All authors reviewed the manuscript. 665  
666  
667

## 11. References 668

### References 669

1. Finn, E. S. *et al.* Functional Connectome Fingerprinting: Identifying Individuals Using Patterns of Brain Connectivity. *Nature Neuroscience* **18**, 1664–1671 (2015). 670  
671
2. Shen, X. *et al.* Using Connectome-Based Predictive Modeling to Predict Individual Behavior from Brain Connectivity. *Nature Protocols* **12**, 506–518 (2017). 672  
673
3. Miranda-Dominguez, O. *et al.* Connectotyping: Model Based Fingerprinting of the Functional Connectome. *PLOS ONE* **9**, e111048 (2014). 674  
675
4. Rosenberg, M. D. *et al.* A Neuromarker of Sustained Attention from Whole-Brain Functional Connectivity. *Nature Neuroscience* **19**, 165–171 (2016). 676  
677

5. Kong, R. *et al.* Spatial Topography of Individual-Specific Cortical Networks Predicts Human Cognition, Personality, and Emotion. *Cerebral Cortex* **29**, 2533–2551 (2019).
6. Finn, E. S. & Bandettini, P. A. Movie-Watching Outperforms Rest for Functional Connectivity-Based Prediction of Behavior. *NeuroImage* **235**, 117963 (2021).
7. Vanderwal, T. *et al.* Individual Differences in Functional Connectivity during Naturalistic Viewing Conditions. *NeuroImage* **157**, 521–530 (2017).
8. Gal, S., Coldham, Y., Tik, N., Bernstein-Eliav, M. & Tavor, I. Act Natural: Functional Connectivity from Naturalistic Stimuli fMRI Outperforms Resting-State in Predicting Brain Activity. *NeuroImage* **258**, 119359 (2022).
9. Vanderwal, T., Eilbott, J. & Castellanos, F. X. Movies in the Magnet: Naturalistic Paradigms in Developmental Functional Neuroimaging. *Developmental Cognitive Neuroscience* **36**, 100600 (2019).
10. Kröll, J.-P. *et al.* Naturalistic Viewing Increases Individual Identifiability Based on Connectivity within Functional Brain Networks. *NeuroImage* **273**, 120083 (2023).
11. Zamani Esfahlani, F. *et al.* High-Amplitude Cofluctuations in Cortical Activity Drive Functional Connectivity. *Proceedings of the National Academy of Sciences* **117**, 28393–28401 (2020).
12. Bolton, T. A., Jochaut, D., Giraud, A.-L. & Van De Ville, D. Brain Dynamics in ASD during Movie-Watching Show Idiosyncratic Functional Integration and Segregation. *Human Brain Mapping* **39**, 2391–2404 (2018).
13. Raz, G. *et al.* Functional Connectivity Dynamics during Film Viewing Reveal Common Networks for Different Emotional Experiences. *Cognitive, Affective, & Behavioral Neuroscience* **16**, 709–723 (2016).
14. Simony, E. *et al.* Dynamic Reconfiguration of the Default Mode Network during Narrative Comprehension. *Nature Communications* **7**, 12141 (2016).
15. Tian, L., Ye, M., Chen, C., Cao, X. & Shen, T. Consistency of Functional Connectivity across Different Movies. *NeuroImage* **233**, 117926 (2021).
16. Li, X., Friedrich, P., Patil, K. R., Eickhoff, S. B. & Weis, S. A Topography-Based Predictive Framework for Naturalistic Viewing fMRI. *NeuroImage* **277**, 120245 (2023).
17. Hüpen, P. *et al.* Functional Brain Network of Trait Impulsivity: Whole-brain Functional Connectivity Predicts Self-Reported Impulsivity. *Human Brain Mapping* **45**, e70059 (2024).
18. Peng, S. *et al.* Heterogenous Brain Activations across Individuals Localize to a Common Network. *Communications Biology* **7**, 1–11 (2024).
19. Almeida de Souza, E., Vieira, B. H. & Salmon, C. E. G. Individual Cognitive Traits Can Be Predicted from Task-Based Dynamic Functional Connectivity with a Deep Convolutional-Recurrent Model. *Cerebral Cortex* **34**, bhae412 (2024).

20. Gao, J. *et al.* Prediction of Cognitive Scores by Movie-Watching FMRI Connectivity and Eye Movement via Spectral Graph Convolutions in 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI) (2022), 1–5.
21. Kawahara, J. *et al.* BrainNetCNN: Convolutional Neural Networks for Brain Networks; towards Predicting Neurodevelopment. *NeuroImage* **146**, 1038–1049 (2017).
22. Greene, A. S., Gao, S., Scheinost, D. & Constable, R. T. Task-Induced Brain State Manipulation Improves Prediction of Individual Traits. *Nature Communications* **9**, 2807 (2018).
23. Yoo, K. *et al.* Multivariate Approaches Improve the Reliability and Validity of Functional Connectivity and Prediction of Individual Behaviors. *NeuroImage* **197**, 212–223 (2019).
24. Smith, S. M. *et al.* A Positive-Negative Mode of Population Covariation Links Brain Connectivity, Demographics and Behavior. *Nature Neuroscience* **18**, 1565–1567 (2015).
25. Sanchez-Alonso, S., Rosenberg, M. D. & Aslin, R. N. Functional Connectivity Patterns Predict Naturalistic Viewing versus Rest across Development. *NeuroImage* **229**, 117630 (2021).
26. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G. & Malach, R. Intersubject Synchronization of Cortical Activity during Natural Vision. *Science* **303**, 1634–1640 (2004).
27. Dziura, S. L. *et al.* Effects of Social and Emotional Context on Neural Activation and Synchrony during Movie Viewing. *Human Brain Mapping* **42**, 6053–6069 (2021).
28. Nguyen, M., Vanderwal, T. & Hasson, U. Shared Understanding of Narratives Is Correlated with Shared Neural Responses. *NeuroImage* **184**, 161–170 (2019).
29. Tanner, J. C. *et al.* Synchronous High-Amplitude Co-Fluctuations of Functional Brain Networks during Movie-Watching. *Imaging Neuroscience* **1**, 1–21 (2023).
30. Lyons, K., Stevenson, R., Owen, A. & Stojanoski, B. Examining the Relationship between Measures of Autistic Traits and Neural Synchrony during Movies in Children with and without Autism. *NeuroImage: Clinical* **28**, 102477 (2020).
31. Ohad, T. & Yeshurun, Y. Neural Synchronization as a Function of Engagement with the Narrative. *NeuroImage* **276**, 120215 (2023).
32. Song, H., Finn, E. S. & Rosenberg, M. D. Neural Signatures of Attentional Engagement during Narratives and Its Consequences for Event Memory. *Proceedings of the National Academy of Sciences* **118**, e2021905118 (2021).
33. Betzel, R. F., Byrge, L., Esfahlani, F. Z. & Kennedy, D. P. Temporal Fluctuations in the Brain’s Modular Architecture during Movie-Watching. *NeuroImage* **213**, 116687 (2020).

34. Bolton, T. A. W., Jochaut, D., Giraud, A.-L. & Ville, D. V. D. Dynamic Inter-Subject Functional Connectivity Reveals Moment-to-Moment Brain Network Configurations Driven by Continuous or Communication Paradigms. *Journal of Visualized Experiments (JoVE)*, e59083 (2019).
35. Calhoun, V. D., Miller, R., Pearlson, G. & Adali, T. The Chronnectome: Time-varying Connectivity Networks as the next Frontier in fMRI Data Discovery. *Neuron* **84**, 262–274 (2014).
36. Shine, J. M., Koyejo, O. & Poldrack, R. A. Temporal Metastates Are Associated with Differential Patterns of Time-Resolved Connectivity, Network Topology, and Attention. *Proceedings of the National Academy of Sciences* **113**, 9888–9891 (2016).
37. Vidaurre, D., Smith, S. M. & Woolrich, M. W. Brain Network Dynamics Are Hierarchically Organized in Time. *Proceedings of the National Academy of Sciences* **114**, 12827–12832 (2017).
38. Liégeois, R. *et al.* Resting Brain Dynamics at Different Timescales Capture Distinct Aspects of Human Behavior. *Nature Communications* **10**, 2317 (2019).
39. Lurie, D. J. *et al.* Questions and Controversies in the Study of Time-Varying Functional Connectivity in Resting fMRI. *Network Neuroscience* **4**, 30–69 (2020).
40. Preti, M. G., Bolton, T. A. & Van De Ville, D. The Dynamic Functional Connectome: State-of-the-art and Perspectives. *NeuroImage* **160**, 41–54 (2017).
41. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: An Overview. *NeuroImage* **80**, 62–79 (2013).
42. Thomas Yeo, B. T. *et al.* The Organization of the Human Cerebral Cortex Estimated by Intrinsic Functional Connectivity. *Journal of Neurophysiology* **106**, 1125–1165 (2011).
43. Khosla, M., Ngo, G. H., Jamison, K., Kuceyeski, A. & Sabuncu, M. R. Cortical Response to Naturalistic Stimuli Is Largely Predictable with Deep Neural Networks. *Science Advances* **7**, eabe7547 (2021).
44. Fan, L., Su, J., Qin, J., Hu, D. & Shen, H. A Deep Network Model on Dynamic Functional Connectivity With Applications to Gender Classification and Intelligence Prediction. *Frontiers in Neuroscience* **14** (2020).
45. Ding, J.-E., Luo, D., Zilverstand, A., Kulkarni, K. & Liu, F. *NeuroTree: Hierarchical Functional Brain Pathway Decoding for Mental Health Disorders* 2025. arXiv: [2502.18786](https://arxiv.org/abs/2502.18786) [cs].
46. Kim, B.-H. & Ye, J. C. Understanding Graph Isomorphism Network for Rs-fMRI Functional Connectivity Analysis. *Frontiers in Neuroscience* **14** (2020).

47. Levakov, G., Sporns, O. & Avidan, G. Fine-scale dynamics of functional connectivity in the face-processing network during movie watching. *en. Cell Reports* **42**, 112585. ISSN: 22111247. <https://linkinghub.elsevier.com/retrieve/pii/S221112472300596X> (2025) (June 2023).
48. Lally, C., Lavan, N., Garrido, L., Tsantani, M. & McGettigan, C. Neural Representations of Naturalistic Person Identities While Watching a Feature Film. *Imaging Neuroscience* **1**, 1–19 (2023).
49. Gershon, R. C. *et al.* NIH Toolbox for Assessment of Neurological and Behavioral Function. *Neurology* **80**, S2–S6 (2013).
50. Vu, A. *et al.* High Resolution Whole Brain Diffusion Imaging at 7 T for the Human Connectome Project. *NeuroImage* **122**, 318–331 (2015).
51. Cutting, J. E., Brunick, K. L. & Candan, A. Perceiving Event Dynamics and Parsing Hollywood Films. *Journal of Experimental Psychology: Human Perception and Performance* **38**, 1476–1490 (2012).
52. Glasser, M. F. *et al.* The Minimal Preprocessing Pipelines for the Human Connectome Project. *NeuroImage* **80**, 105–124 (2013).
53. Li, J. *et al.* Global Signal Regression Strengthens Association between Resting-State Functional Connectivity and Behavior. *NeuroImage* **196**, 126–141 (2019).
54. Shen, X., Tokoglu, F., Papademetris, X. & Constable, R. Groupwise Whole-Brain Parcellation from Resting-State fMRI Data for Network Node Identification. *NeuroImage* **82**, 403–415 (2013).
55. Grady, C. L., Rieck, J. R., Nichol, D., Rodrigue, K. M. & Kennedy, K. M. Influence of Sample Size and Analytic Approach on Stability and Interpretation of Brain-Behavior Correlations in Task-Related fMRI Data. *Human Brain Mapping* **42**, 204–219 (2021).
56. Marek, S. *et al.* Reproducible Brain-Wide Association Studies Require Thousands of Individuals. *Nature* **603**, 654–660 (2022).
57. Helmer, M. *et al.* On the Stability of Canonical Correlation Analysis and Partial Least Squares with Application to Brain-Behavior Associations. *Communications Biology* **7**, 1–15 (2024).
58. Scheinost, D. *et al.* Ten Simple Rules for Predictive Modeling of Individual Differences in Neuroimaging. *NeuroImage* **193**, 35–45 (2019).
59. Sundararajan, M., Taly, A. & Yan, Q. *Axiomatic Attribution for Deep Networks* in *Proceedings of the 34th International Conference on Machine Learning* (PMLR, 2017), 3319–3328.
60. Wang, D. *et al.* Deep Neural Network Heatmaps Capture Alzheimer’s Disease Patterns Reported in a Large Meta-Analysis of Neuroimaging Studies. *NeuroImage* **269**, 119929 (2023).

61. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical 824  
and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: 825*  
*Series B (Methodological)* **57**, 289–300 (1995). 826
62. McNamara, Q., De La Vega, A. & Yarkoni, T. *Developing a Comprehensive Frame- 827*  
*work for Multimodal Feature Extraction in Proceedings of the 23rd ACM SIGKDD 828*  
*International Conference on Knowledge Discovery and Data Mining* (Association for 829  
Computing Machinery, New York, NY, USA, 2017), 1567–1574. ISBN: 978-1-4503- 830  
4887-4. 831
63. Huth, A. G., Nishimoto, S., Vu, A. T. & Gallant, J. L. A Continuous Semantic Space 832  
Describes the Representation of Thousands of Object and Action Categories across 833  
the Human Brain. *Neuron* **76**, 1210–1224 (2012). 834
64. Nishimoto, S. *et al.* Reconstructing Visual Experiences from Brain Activity Evoked 835  
by Natural Movies. *Current Biology* **21**, 1641–1646 (2011). 836

## Appendix 837

Please see supplementary information in additional files. 838