

Supplementary Information for manuscript:
***A View-Engage-Predict Framework for Enhancing
Brain-Behavior Mapping with Naturalistic Movie-
Watching fMRI***

Yunrui Zhang¹, Emily S. Finn², Mert R. Sabuncu^{1,3,4}, Amy Kuceyeski^{3,5,*}

¹School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA

²Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA

³Department of Radiology, Weill Cornell Medicine, New York, NY, USA

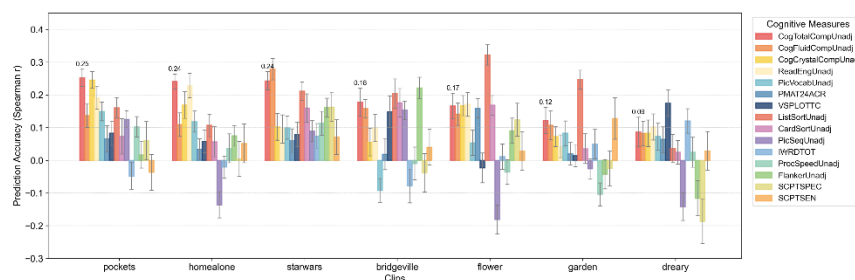
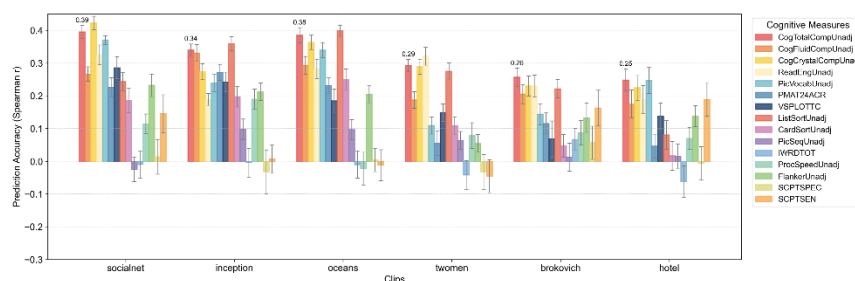
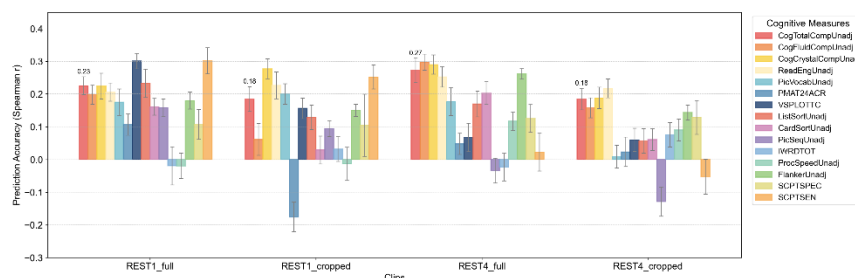
⁴Cornell Tech, New York, NY, USA

⁵Department of Computational Biology, Cornell University, Ithaca, NY, USA

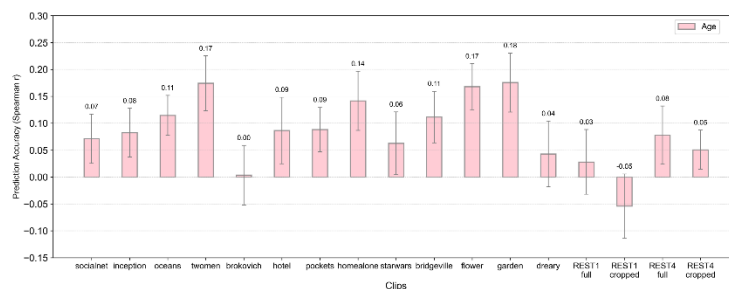
*To whom correspondence should be addressed; E-mail: amk2012@med.cornell.edu

Supplementary Figures

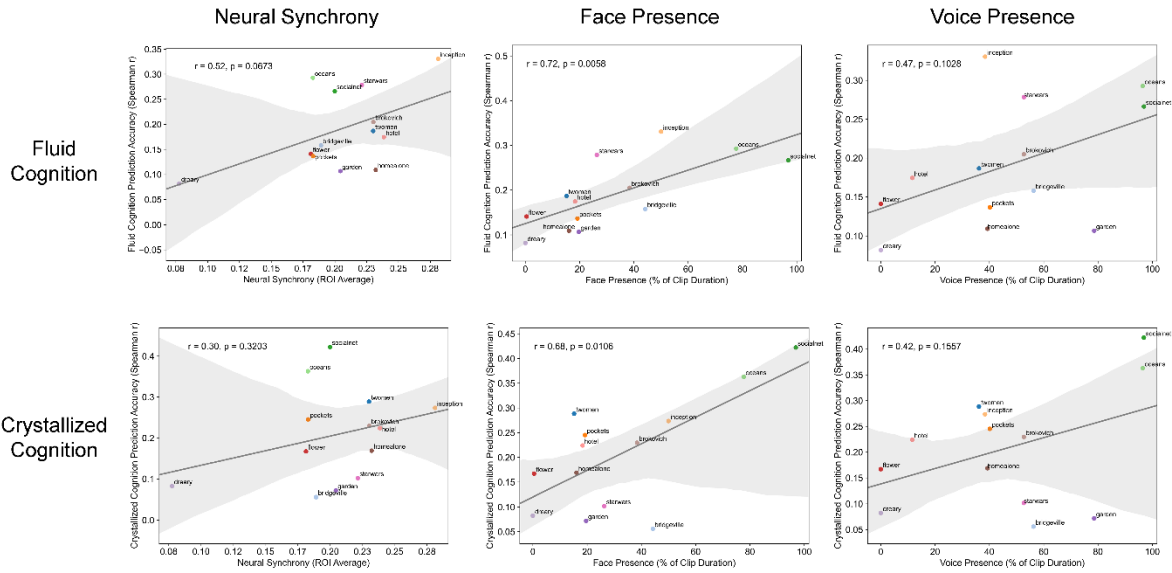
All cognition scores:



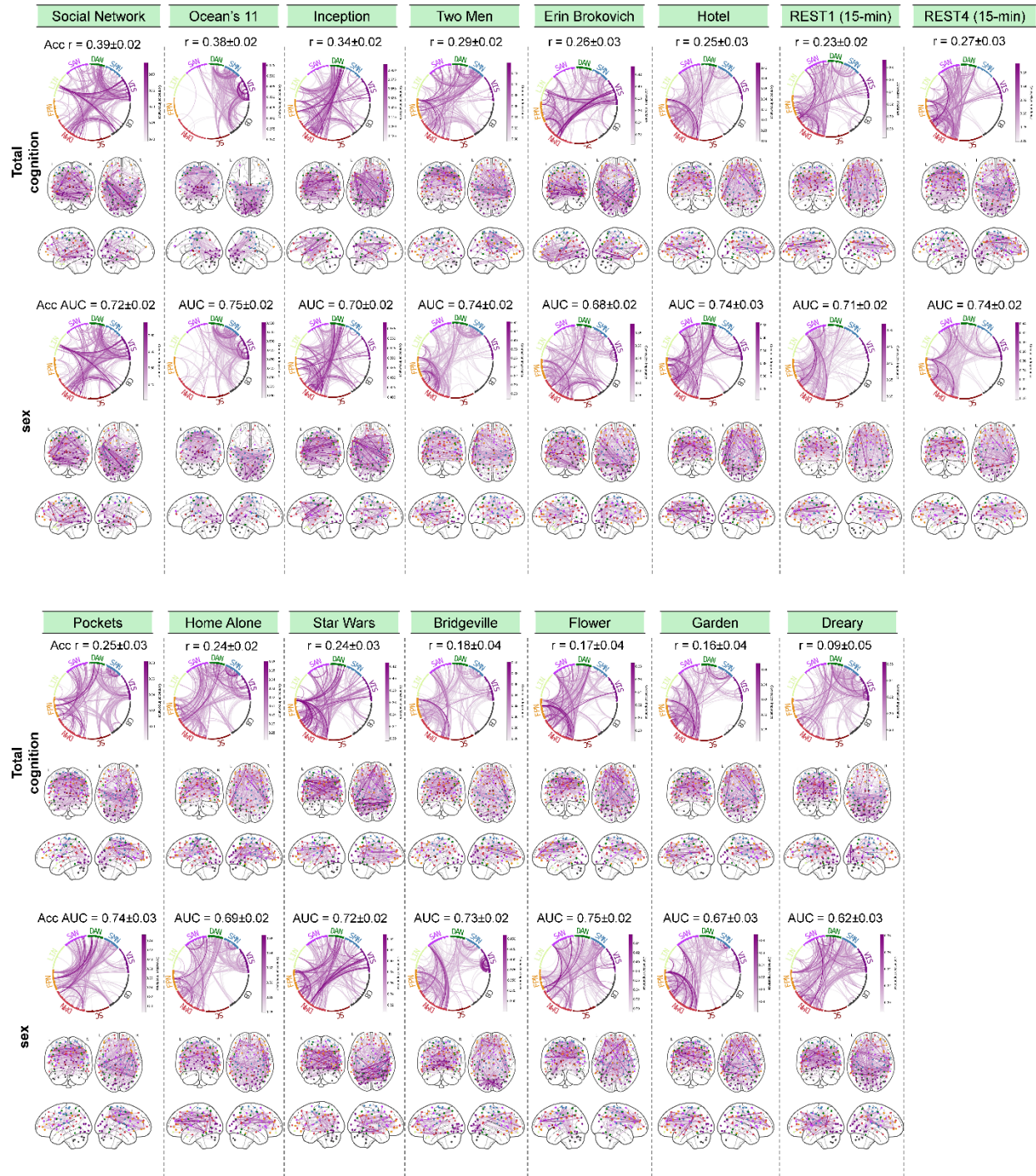
Chronological age:



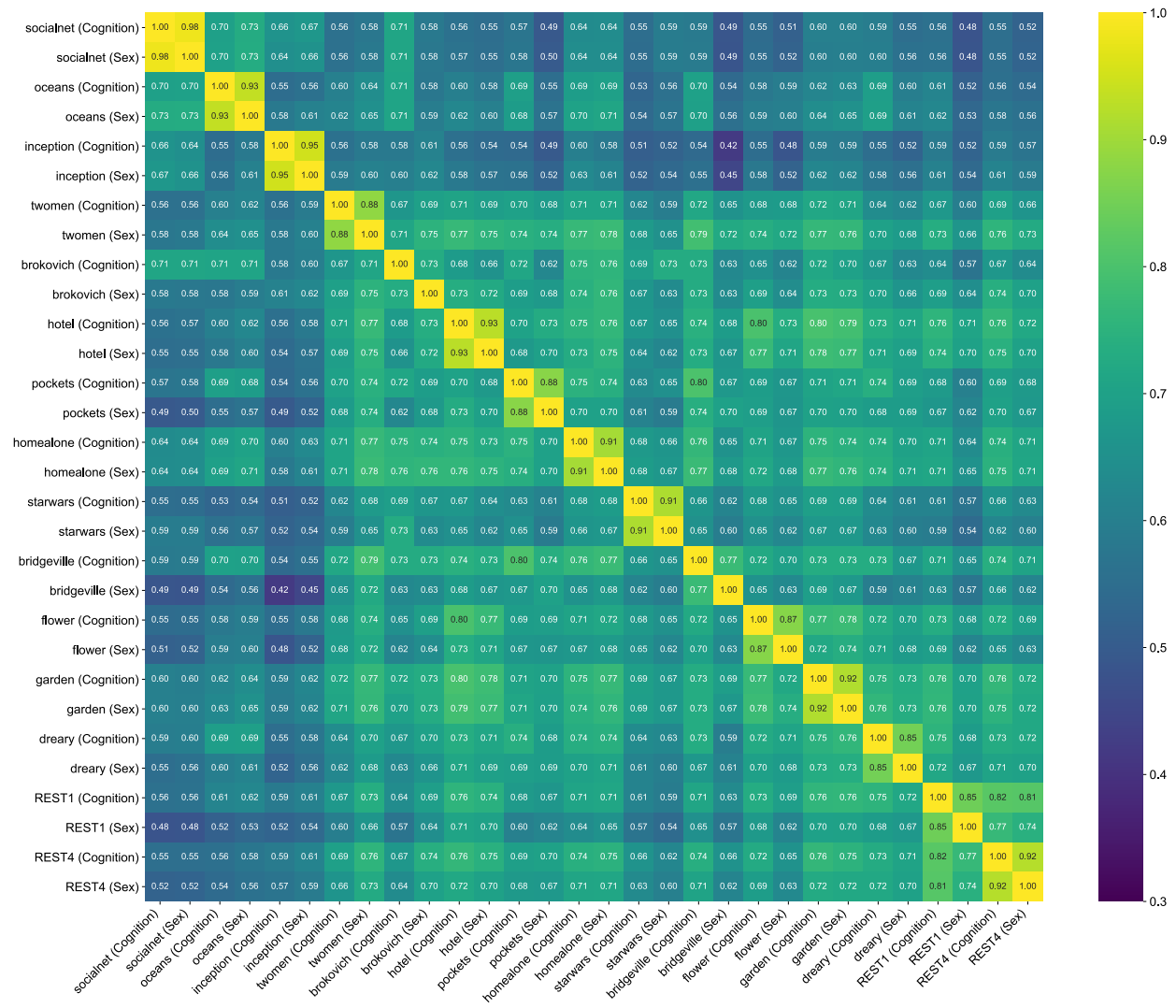
Supplementary Figure 1: Supplement of Fig.2. Prediction accuracies for all cognitive measures from the NIH Toolbox and chronological age, computed for each movie clip from the HCP 7T release. Here, we show results of both the first 164 seconds of the rest sessions and the 15-min full rest sessions.



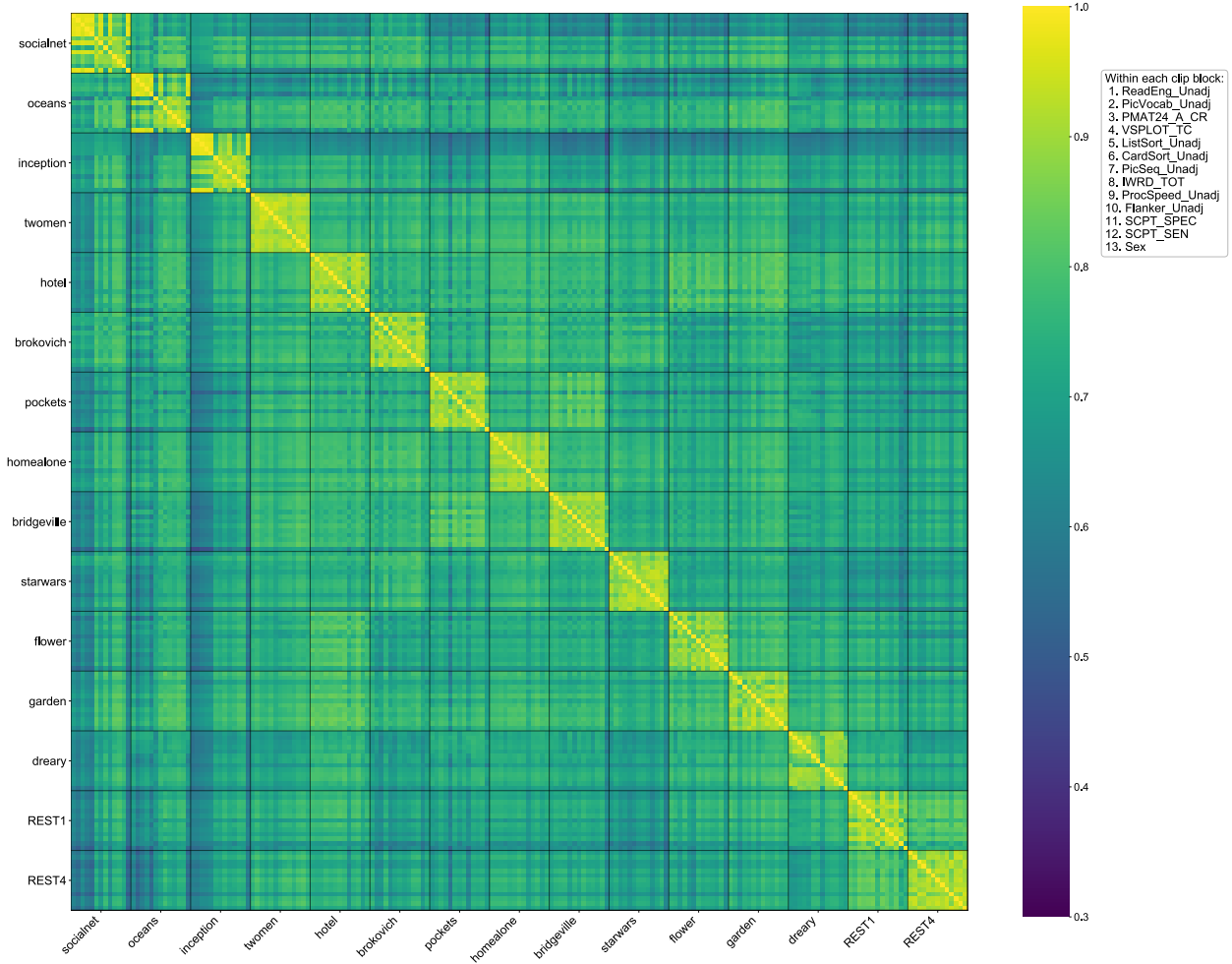
Supplementary Figure 2: Supplement of Fig.2. Correlation between movie clip FC-based prediction accuracy for fluid/crystallized cognition and neural synchrony, face and voice presence duration.



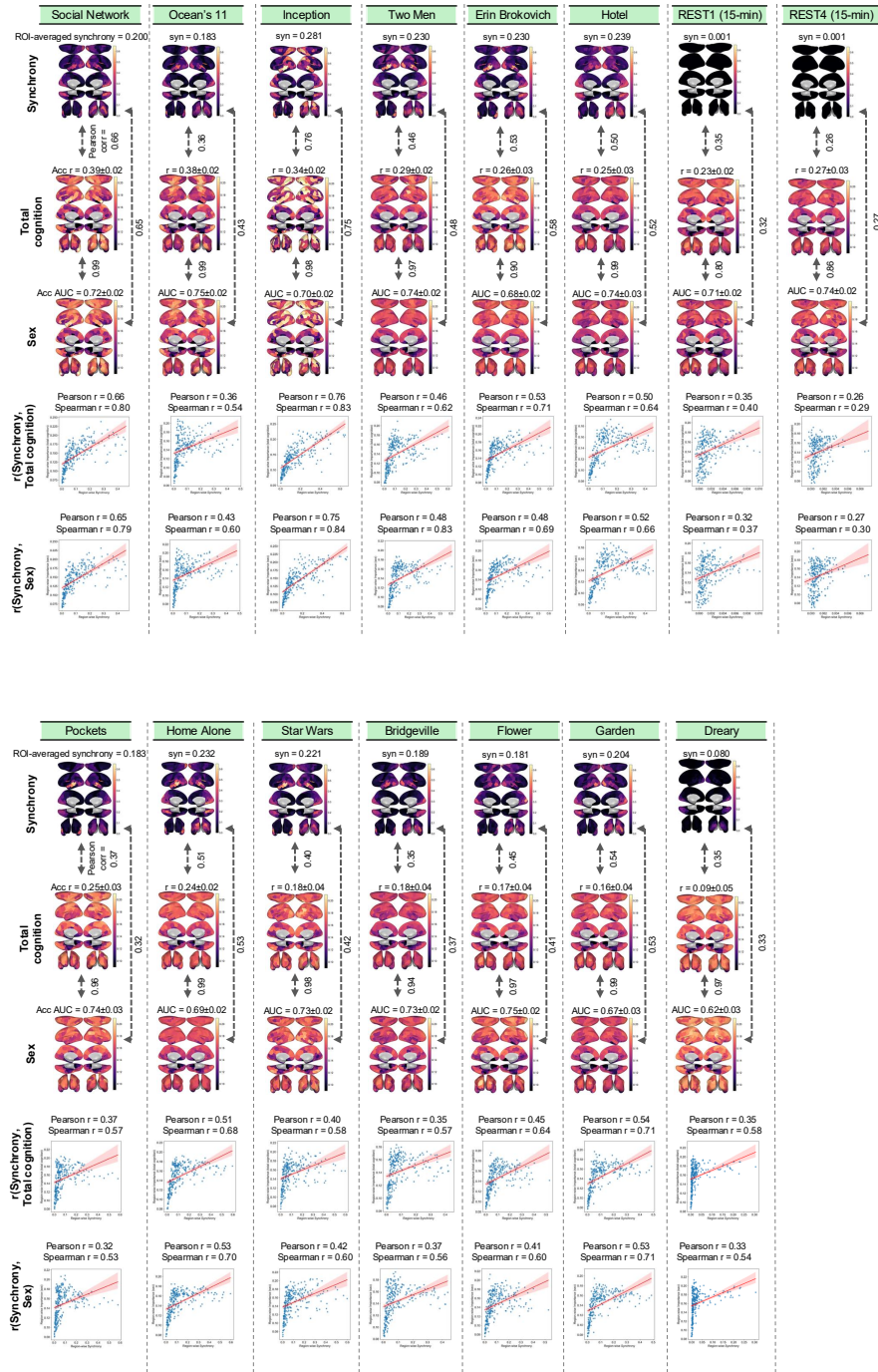
Supplementary Figure 3: Supplement of Fig.3. For each clip and resting-state session, chord and corresponding glass brain diagrams visualize the most important connections—specifically, those within the top 1% of average importance scores—for both total cognition and sex prediction.



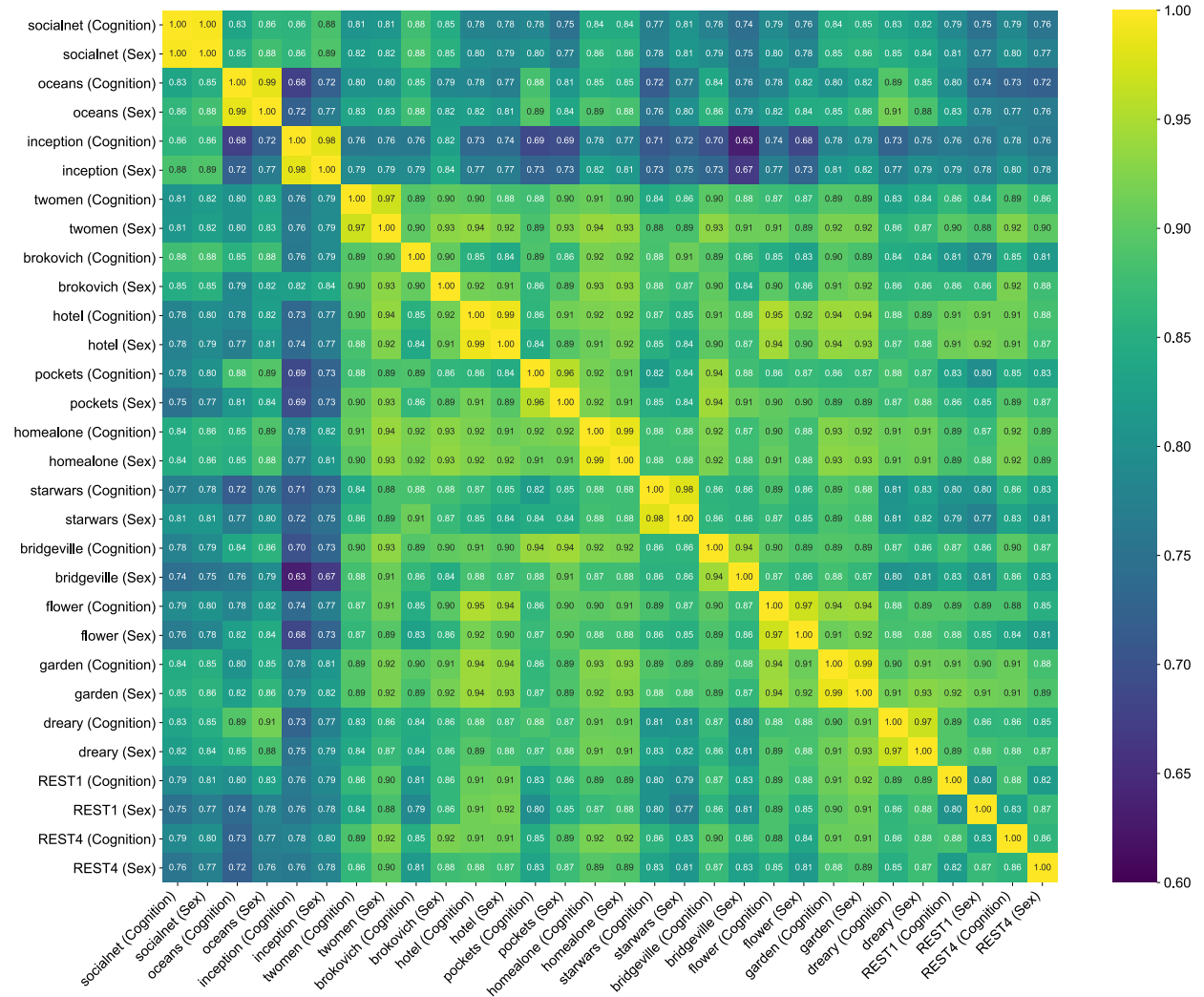
Supplementary Figure 4: Supplement of Fig.3b. Correlation between feature maps (edge-wise). Pairwise Pearson correlations between edge-wise importance maps across all clips/resting-state sessions and both outcomes (cognition and sex).



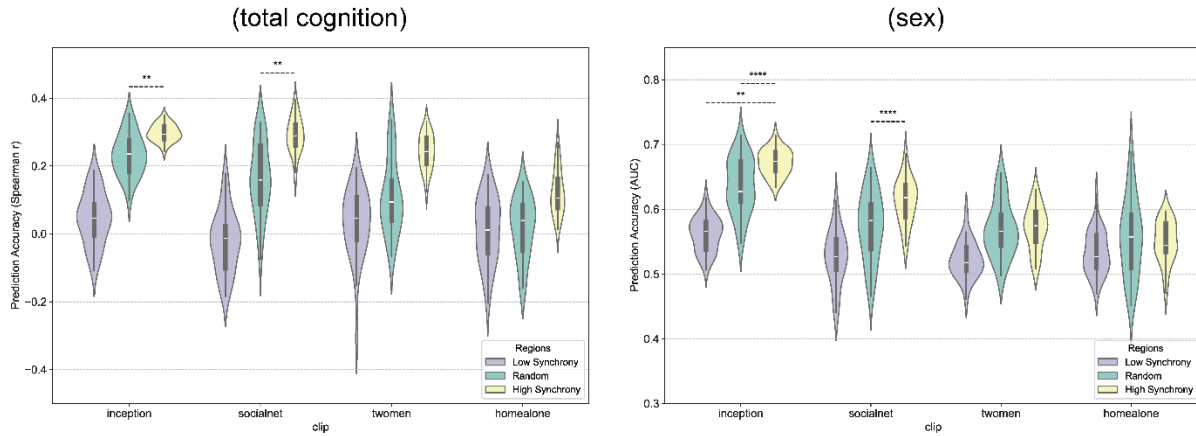
Supplementary Figure 5: Supplement of Fig.3. Correlation matrix of connection-level prediction importance maps across all movie clips and behavioral outcomes (12 NIH Toolbox cognition scores and sex). Each block corresponds to a movie clip (labeled on the axes), and within each block the rows/columns represent the different behavioral predictions (see legend for order). For some movie clips, the importance maps for sex and some cognition scores are less correlated than others (e.g. for *Social Network*, *Ocean's 11* and *Inception*, the Dimensional Change Card Sort Test score shows lower correlation to other cognitive scores). However, the higher correlation values near the diagonal show that importance maps tend to be more similar within a given clip for the varied cognition and sex predictions as compared to the same cognitive score/sex across different clips.



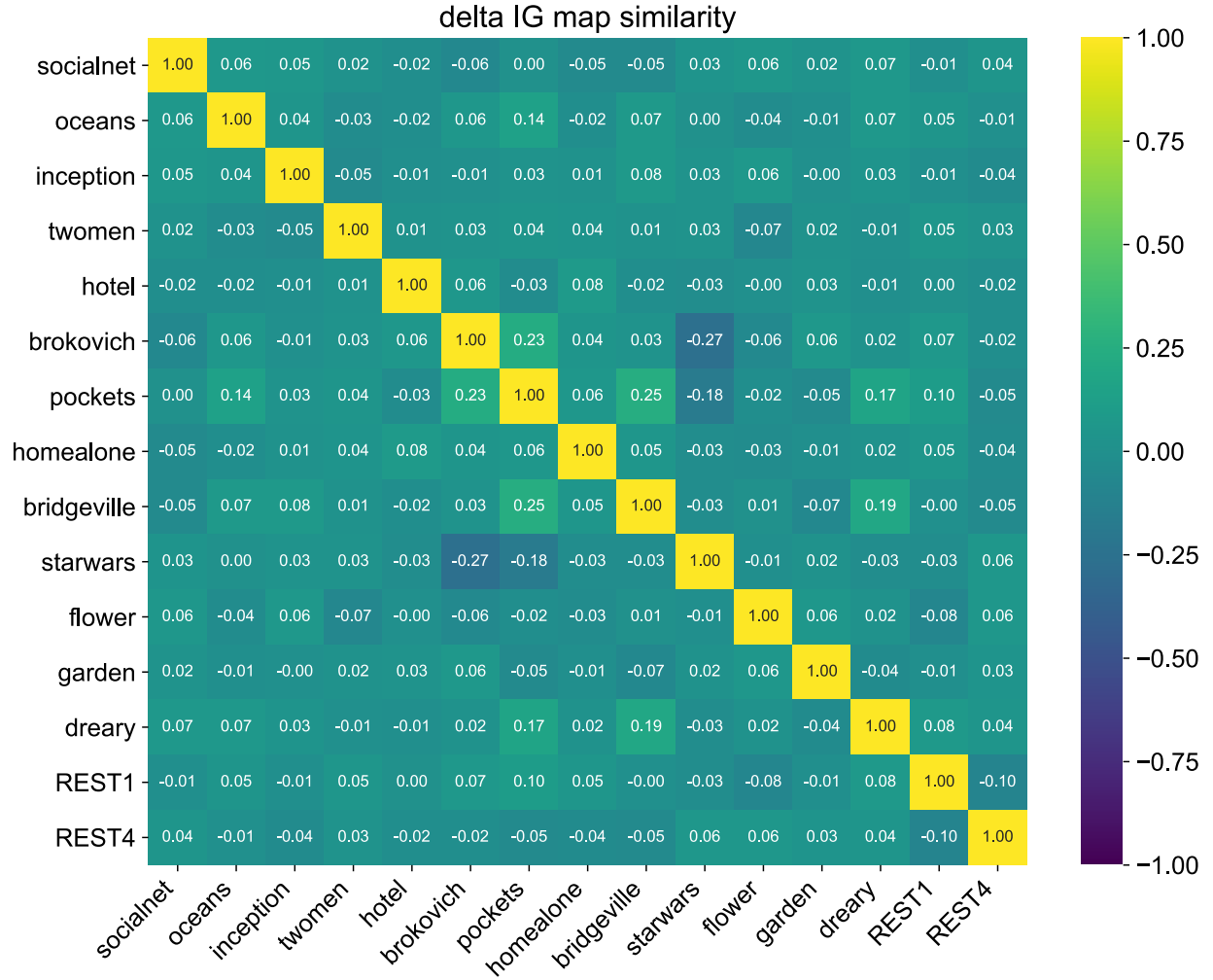
Supplementary Figure 6: Supplement of Fig.4. For each clip, the first row shows region-wise neural synchrony maps. The second and third rows display region-wise prediction importance maps for total cognition and sex. The importance of each region is calculated as the average of the importance scores of all connections connected to that region. The fourth and fifth rows show the Pearson correlation between region-wise neural synchrony and prediction importance maps for total cognition and sex.



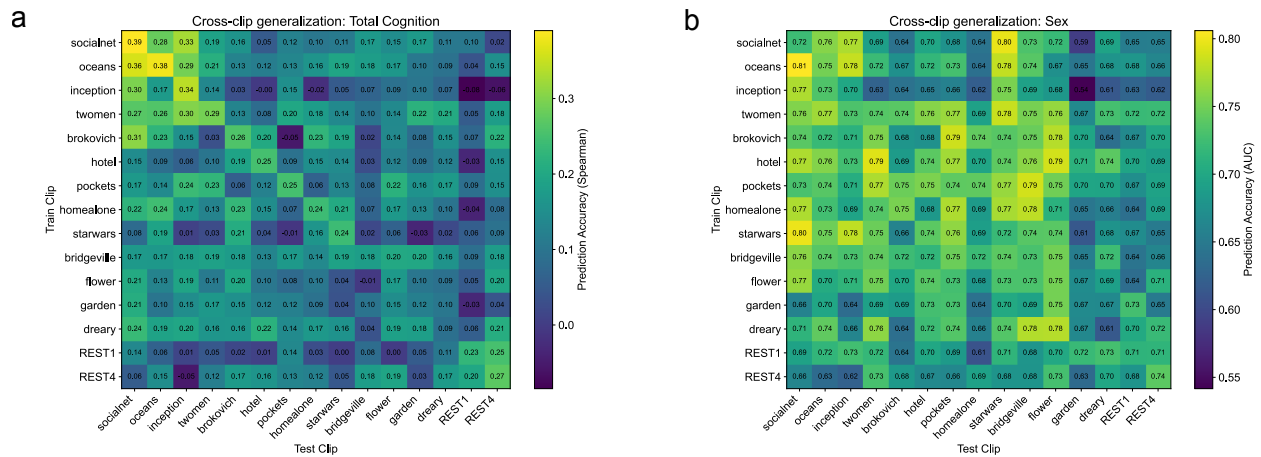
Supplementary Figure 7: Supplement of Fig.4b. Correlation between feature maps (region-wise). Pairwise Pearson correlations between region-wise importance maps across all clips/resting-state sessions and both outcomes (cognition and sex).



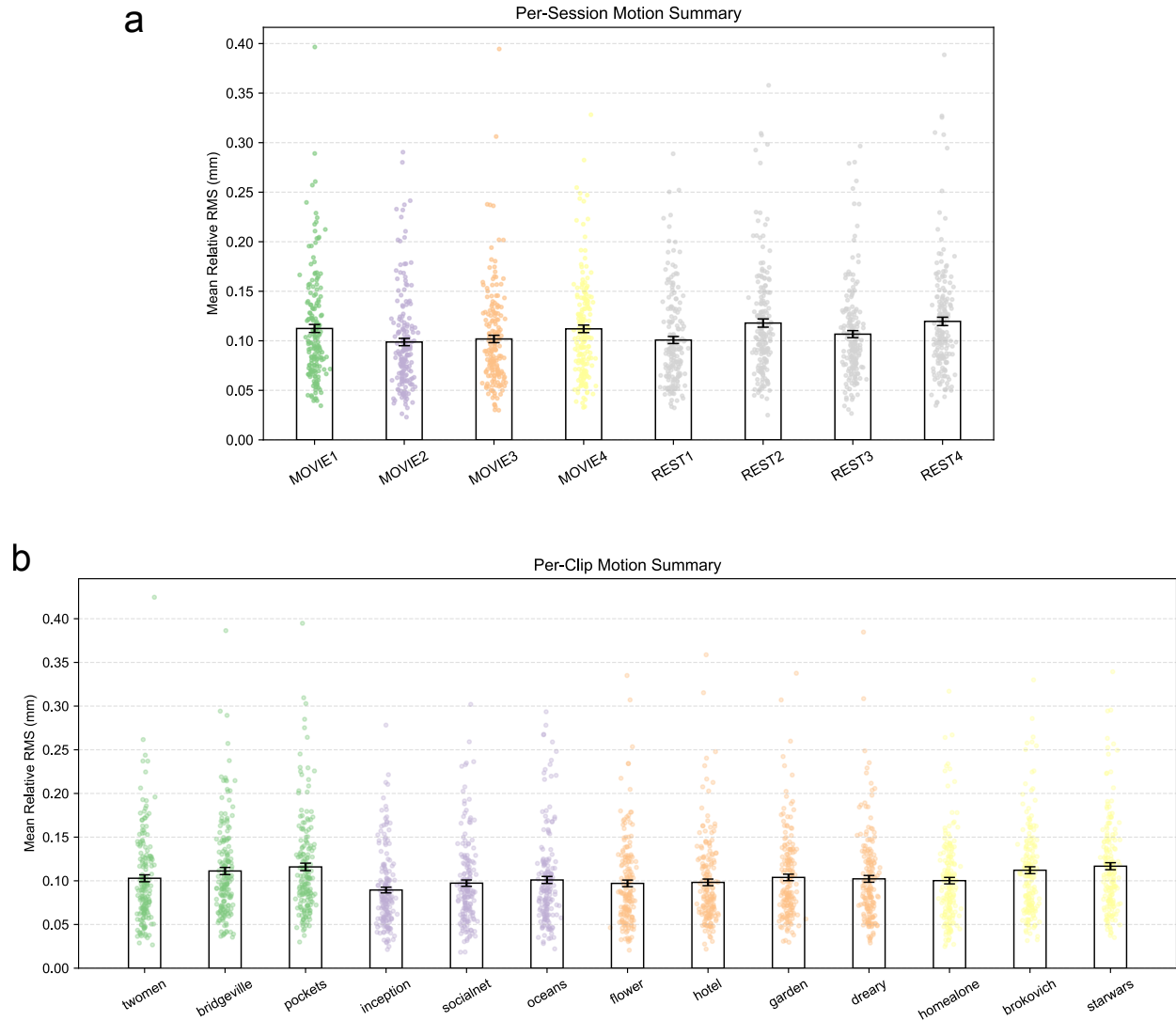
Supplementary Figure 8: Supplement of Fig.4. Prediction performance using an equal number of high-synchrony, low-synchrony, and randomly selected regions. For both cognition and sex prediction, models using FC of high-synchrony regions outperform models based on random regions' FC, which in turn outperform models based on FC from low-synchrony regions.



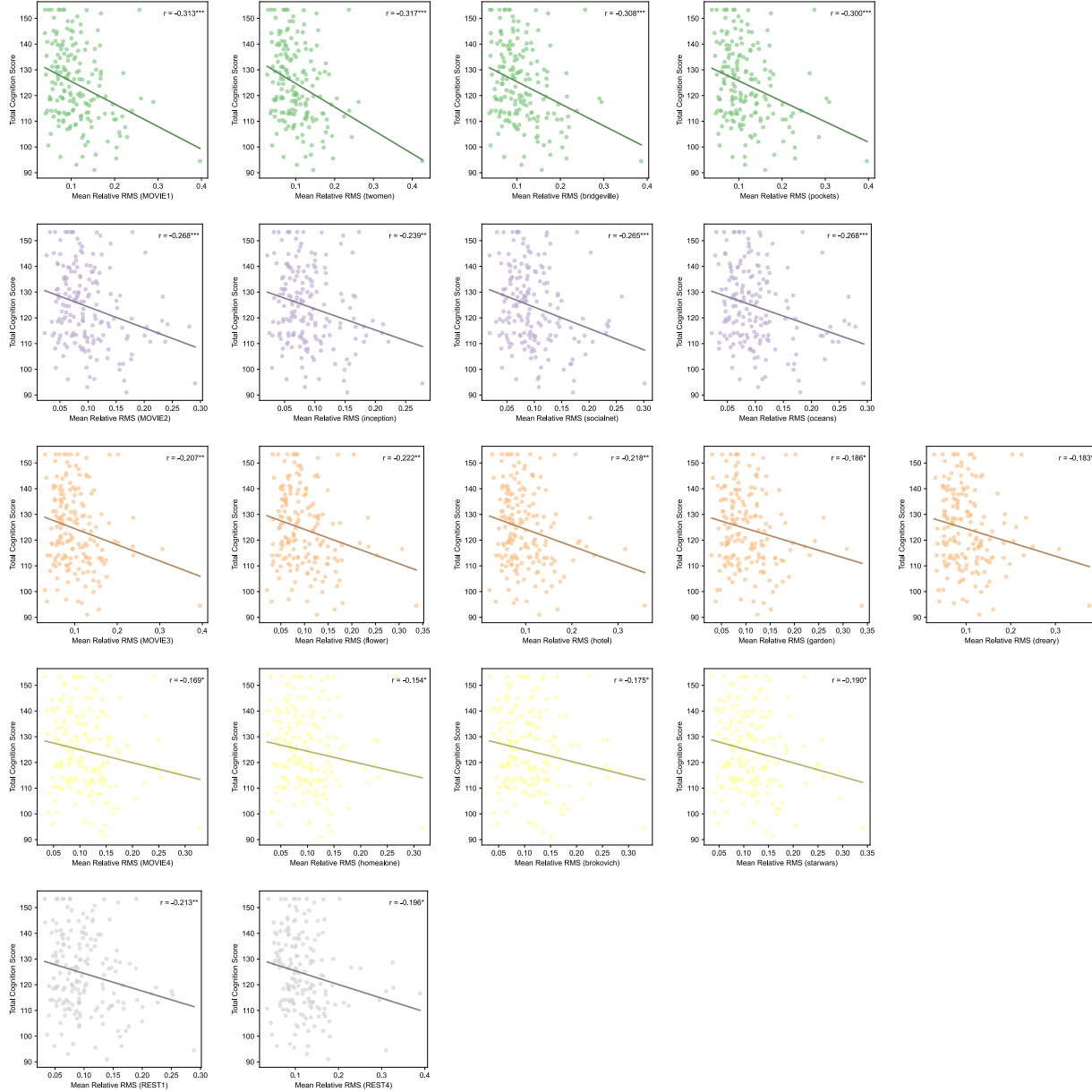
Supplementary Figure 9: Cross-clip similarity of outcome-specific importance differences (ΔIG). Pairwise correlations between edge-wise ΔIG maps ($\Delta IG^{(c)} = \Delta IG_{cognition}^{(c)} - \Delta IG_{sex}^{(c)}$) computed separately for each movie clip and resting-state session.



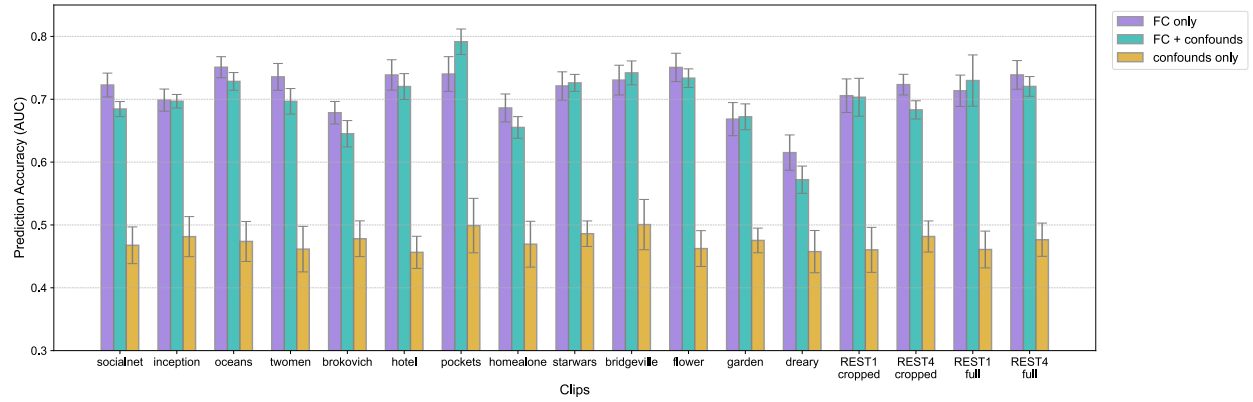
Supplementary Figure 10: Cross-clip generalization of FC-based prediction models. Cross-clip generalization matrices for (a) total cognition prediction (Spearman correlation) and (b) sex prediction (AUC). Each row corresponds to the clip used for training, and each column corresponds to the clip used for testing; values represent mean performance across cross-validation iterations.



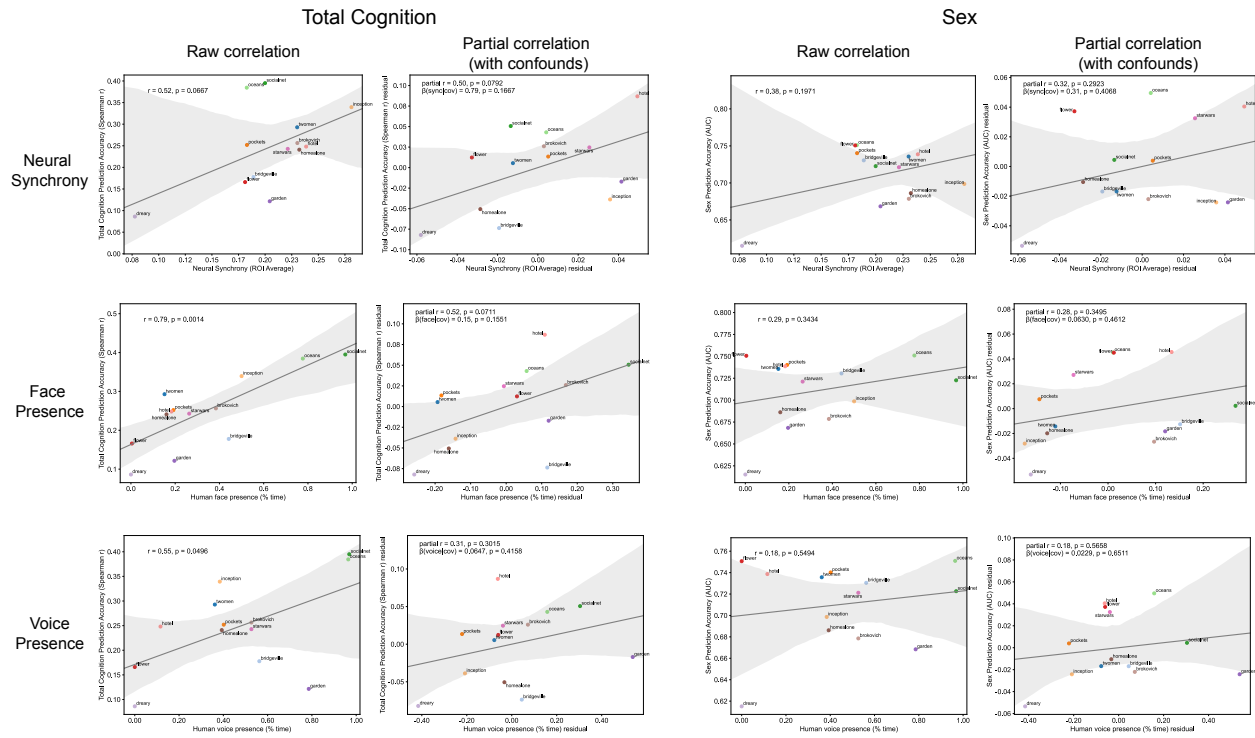
Supplementary Figure 11: Motion report. Mean and between-subject standard deviation of relative root-mean-square (RMS) framewise displacement for each resting-state session (a) and individual movie clips (b), computed as the subject-level mean relative RMS motion across the corresponding time interval.



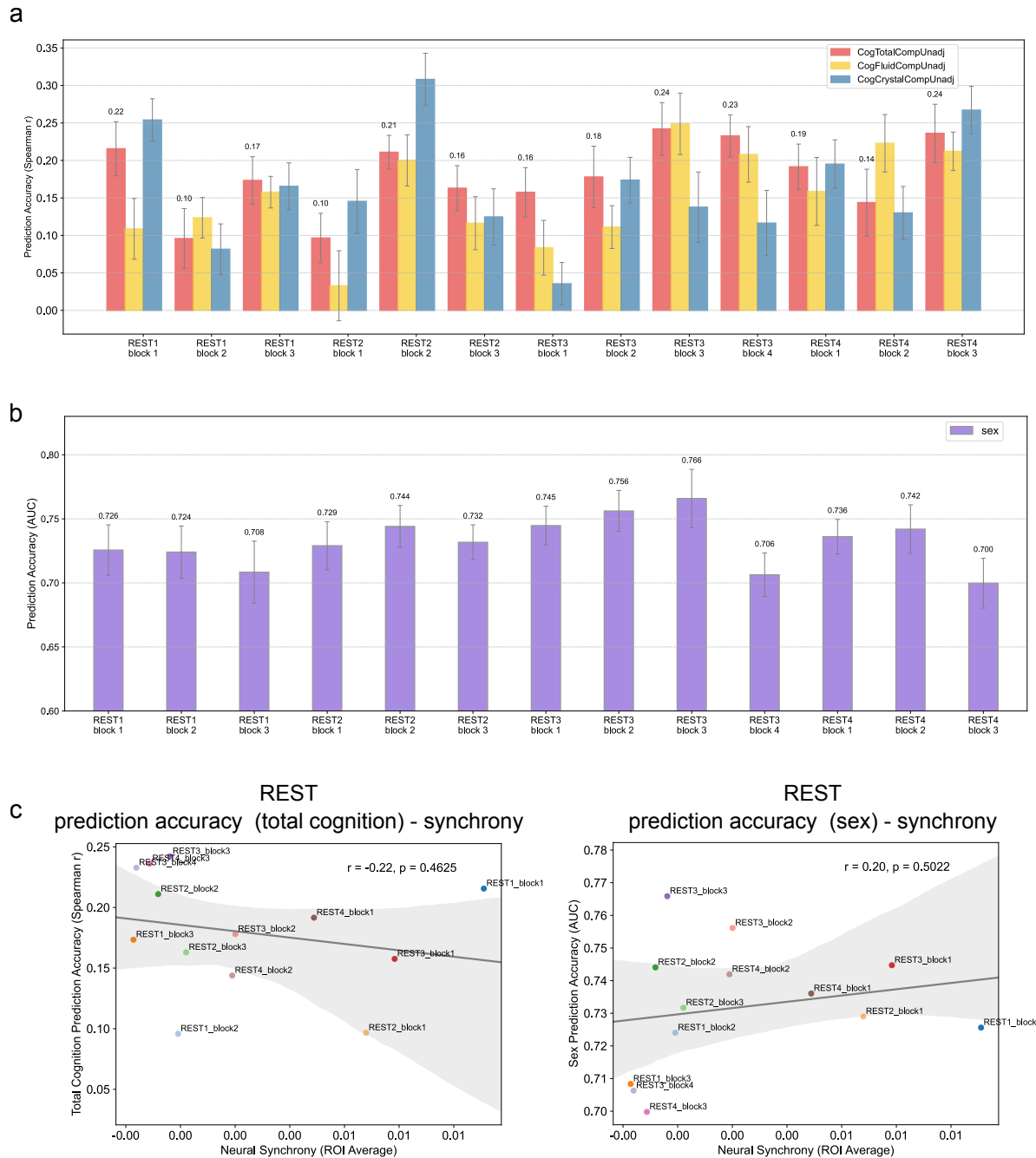
Supplementary Figure 12: Clip- and session- level correlation between head motion and cognition. Correlation between mean relative RMS head motion and total cognition score across subjects, shown for each movie clip, movie session and resting-state session. Each panel shows a scatter plot of subject-level mean relative RMS motion (x-axis) versus total cognition score (y-axis). Pearson's correlation coefficient (r) is reported in each panel. Head motion exhibits a consistent negative association with cognition across all movie clips and resting-state runs, with comparable effect sizes across conditions, indicating that motion represents a global confound rather than a clip-specific factor.



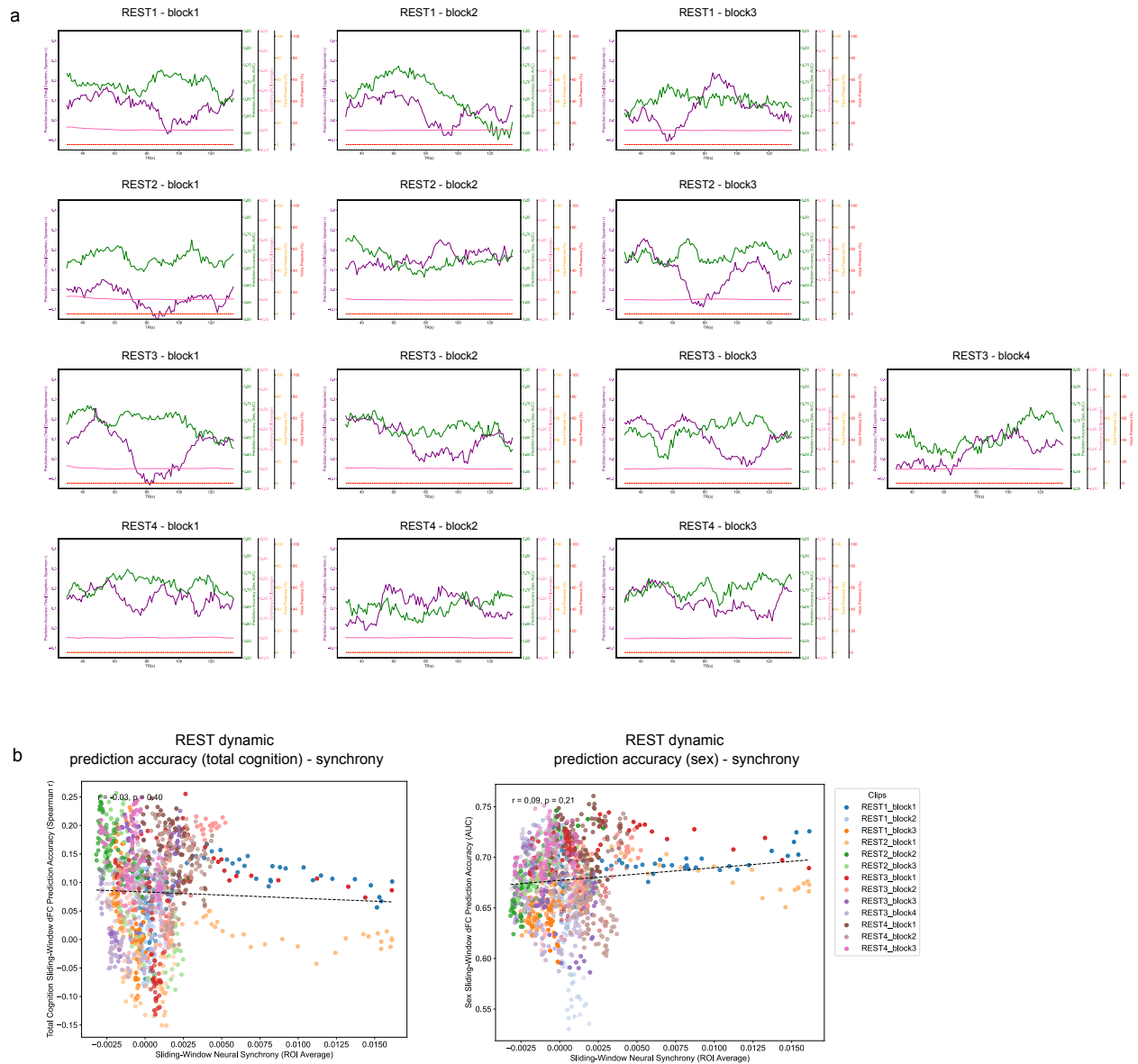
Supplementary Figure 13: Sex prediction performance (AUC) across movie clips and resting-state runs using three feature sets: FC only, confounds only (including head motion and time of day), and FC combined with confounds. Bars indicate mean cross-validated AUC across folds across 100 repetitions, with error bars showing standard deviation. Confound-only models perform at or near chance level ($AUC \approx 0.5$) across all clips, whereas FC-based models achieve substantially higher accuracy. Adding confounds to FC features does not improve prediction performance, indicating that sex prediction is primarily driven by functional connectivity patterns rather than head motion or session-related covariates.



Supplementary Figure 14: For each movie clip ($n=13$), we show the relationship between prediction accuracy (Total Cognition: Spearman r ; Sex: AUC) and three clip-level predictors: neural synchrony (top row), human face presence (middle row), and human voice presence (bottom row). Left panels show raw Pearson correlations across clips. Right panels show partial correlations after residualizing both predictor and accuracy with respect to clip-level covariates, including mean head motion, scanning day (Day 1 vs. Day 2), movie type (Hollywood vs. independent), and clip start time within the run. Each point corresponds to one movie clip; shaded regions indicate 95% confidence intervals of the fitted regression line. Correlation coefficients and associated p -values are reported in each panel.



Supplementary Figure 15: Prediction performance for 13 resting-state clips matched in duration (164 s) and temporal location to the movie clips in the naturalistic viewing sessions. (a) Prediction accuracy for total, fluid, and crystallized cognition, quantified using Spearman's rank correlation (r). (b) Prediction accuracy for sex classification, evaluated using AUC. (c) Relationship between resting-state prediction accuracy and inter-subject neural synchrony. No significant association was observed for either total cognition or sex. Bars show mean accuracy across 100 repetitions of 10-fold family-wise cross-validation, with error bars indicating standard deviation.



Supplementary Figure 16: Sliding-window dynamic FC prediction and neural synchrony during resting state. (a) Sliding-window dFC prediction accuracy for total cognition (purple) and sex (green), and neural synchrony (pink), shown across matched resting-state blocks (REST1–REST4). Resting-state scans were segmented into blocks with identical temporal locations and durations (164 s) as the movie clips, and sliding-window analyses were performed using the same window size (60 s) and step (1 s) as in Fig. 5. (b) Relationship between sliding-window neural synchrony and dFC prediction accuracy for total cognition (left) and sex (right) across all resting-state windows. Each point represents one sliding window, colored by rest block. No significant correlation was observed for either total cognition or sex.

Supplementary Tables

	REST1	REST2	REST3	REST4	MOVIE1	MOVIE2	MOVIE3	MOVIE4
Mean Relative RMS (mm)	0.101	0.118	0.107	0.120	0.112	0.099	0.102	0.112
Std Relative RMS (mm)	0.047	0.056	0.048	0.056	0.100	0.089	0.090	0.103

Supplementary Table 1: Motion report for 15-minute sessions. Mean and between-subject standard deviation of relative root-mean-square (RMS) framewise displacement for each 15-minute resting-state session and movie-watching session.

	twomen	bridgeville	pockets	inception	socialnet	oceans
Mean Relative RMS (mm)	0.103	0.111	0.116	0.089	0.097	0.101
Std Relative RMS (mm)	0.053	0.054	0.057	0.044	0.048	0.054

	flower	hotel	garden	dreary	homealone	brokovich	starwars
Mean Relative RMS (mm)	0.097	0.098	0.104	0.102	0.100	0.112	0.117
Std Relative RMS (mm)	0.050	0.050	0.049	0.052	0.049	0.053	0.054

Supplementary Table 2: Motion report for individual movie clips. Mean and between-subject standard deviation of relative root-mean-square (RMS) framewise displacement for individual movie clips, computed as the subject-level mean relative RMS motion across the corresponding time interval.