Supplementary material

In Supplementary Tables S1~S3, values indicate Spearman's ρ between predicted and observed d' within the internal validation sample in both dynamic functional connectivity (DFC) and combined (static + dynamic) FC models. * : $p < .05$; ** : $p < .01$; ** : $p < .001$.

Supplementary Table S1: Internal validation results with high-pass filtering

A high-pass Butterworth filter with cutoff frequency = $\frac{1}{w}$ was applied to fMRI timecourses before the dynamic DFC and combined FC matrices with corresponding window sizes were generated.

Training data	**Rest**		**Task**	
Testing data	**Rest**	**Task**	**Rest**	**Task**
Dynamic model	0.26	0.57 **	0.33 *	0.81 ***
Dynamic model (with high-pass filtering)	0.22	0.55 **	0.34 *	0.70 ***
Combined model	0.45 *	0.77 ***	0.54 **	0.86 ***
Combined model (with high-pass filtering in DFC connectome construction)	0.45 **	0.76 ***	0.54 *	0.86 ***

Supplementary Table S2: Internal validation results without global signal regression

In this control analysis, global signal regression (GSR) was omitted during fMRI preprocessing.

Training data	**Rest**		**Task**	
Testing data	**Rest**	**Task**	**Rest**	**Task**
Static model	0.44 **	0.73 ***	0.51 **	0.81 ***
Static model (without GSR)	0.03	0.32 *	0.33 *	0.76 ***
Dynamic model	0.26	0.57 **	0.33 *	0.81 ***
Dynamic model (without GSR)	0.08	0.42 **	0.29	0.62 ***
Combined model	0.45 *	0.77 ***	0.54 **	0.86 ***
Combined model (without GSR)	0.12	0.42 *	0.39 *	0.77 ***

Supplementary Table S3: Internal validation results with BOLD variability as model features

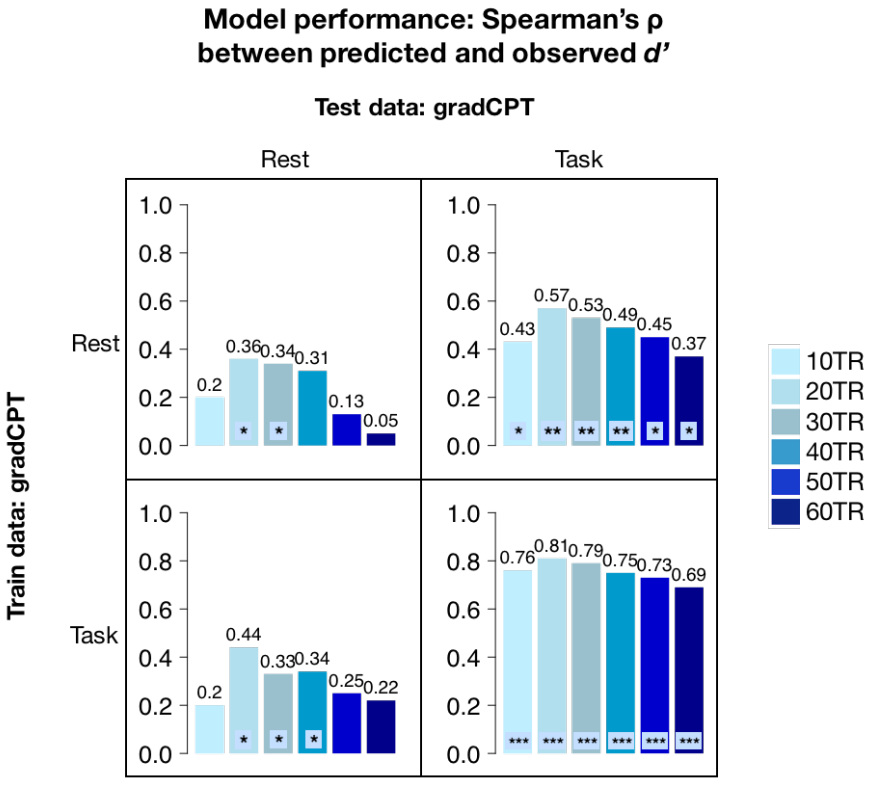
BOLD variability was calculated for each node using the 268 nodal timecourses under rest and task and used as PLSR model features in the downstream prediction pipeline.

Training data	**Rest**		**Task**	
Testing data	**Rest**	**Task**	**Rest**	**Task**
Static model	0.44 **	0.73 ***	0.51 **	0.81 ***
Dynamic model	0.26	0.57 **	0.33 *	0.81 ***
BOLD variability model	0.01	0.44 *	0.43 *	0.31
Combined SFC + DFC model	0.45 *	0.77 ***	0.54 **	0.86 ***
Combined SFC + BOLD variability model	0.27	0.46 *	0.46 **	0.63 ***

Supplementary analysis

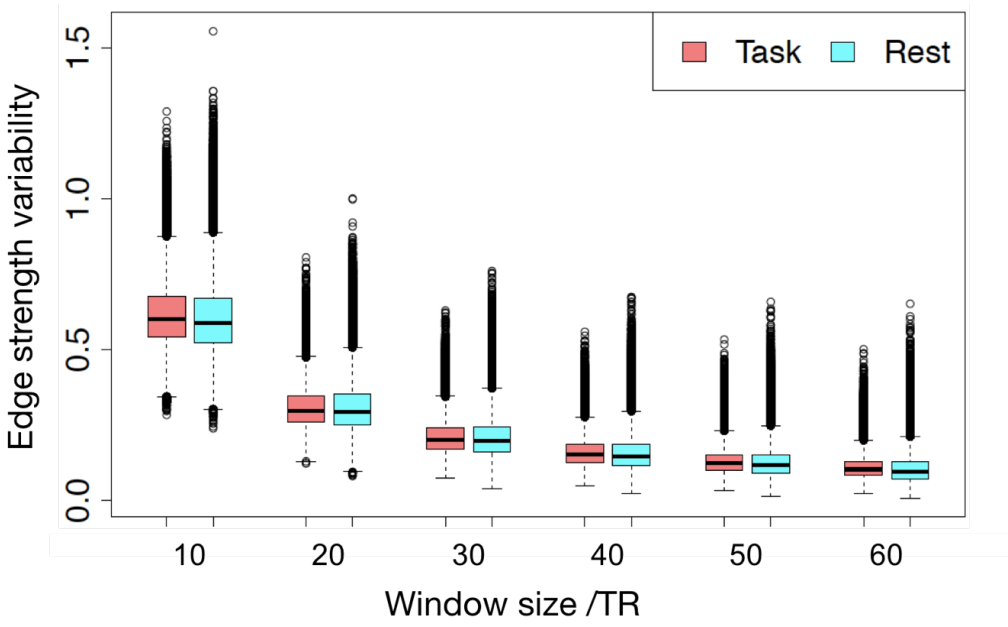
Effect of window size on dynamic functional connectivity model performance

To further investigate the relationship between window size and prediction performance, models were trained and tested on the internal validation (gradCPT) dataset in a LOOCV procedure, but with window sizes fixed across iterations at {10,20,30,40,50,60TR}. 20TR achieved the best performance across the 2x2 train-test condition combinations (Supplementary Figure **S1**). 10TR windows underperformed 20TR and 30TR windows in all cases, but between 20TR and 60TR, shorter windows consistently led to better performance. We observe that tested on task data, models achieved statistically significant performance across all window sizes. Moreover, across all train-test combinations, models from 20TR and 30TR windows achieved statistically significant performance. Based on this, results from longer windows (or at least 30TR windows) may be comparable to those from 20TR windows.



Supplementary Figure S1. *Influence of window length on DFC prediction performance, across rest- and task-based models.* DFC models were built using fixed window sizes; this analysis was performed on the gradCPT internal validation dataset. Values indicate Spearman's ρ between predicted and observed *d'*. * : $p < .05$; ** : $p < .01$; ** : $p < .001$.

Next, we compared the DFC connectomes generated on the {10,20,30,40,50,60TR} window sizes (Supplementary Figure **S2**). For each size and either condition (rest/task), matrix values were collapsed across all subjects. Shorter window sizes led to matrices with higher and more widely distributed edge variability values (for 10TR task, mean=0.601, range={0.000, 1.556}, sd=0.118; for 60TR task, mean=0.109, range={0.007,0.650}, sd=0.045). Thus, shorter windows captured more variability in transient FC patterns. This is reasonable, given that shorter windows receive input from fewer TR's, increasing the weight of each TR in computing transient FC while reducing the influence of more global FC patterns. For each window size, the distribution of edge variability values was similar across task and rest conditions, suggesting that this effect of window size is reliable within individuals across cognitive states. However, sliding-window DFC can still reflect various sources of non-stationary noise including sampling variability, head motion, and respiration/heart rate, which may also be relatively similar across conditions within an individual. The observed similarity of task and rest DFC connectomes could in part reflect persisting non-neuronal artefacts across conditions.



Supplementary Figure S2. *Distribution of edge strength variability values across window size (TR) and task vs rest conditions.* Matrix values were collapsed across all subjects. Bottom and top of boxes represent the 25% and 75% quartiles respectively, lines represents the median, and dots beyond the upper and lower whiskers represent individual data points.

Optimal number of PLS components in static, dynamic and combined CPM models

Supplementary Table S4: Median optimal number of components across LOOCV iterations, internal validation (range shown in brackets)

Training data	Rest		Task	
Testing data	Rest	Task	Rest	Task
Static model	**1** (1-1)	**1** (1-1)	**2** (1-6)	**3** (1-5)
Dynamic model	**1** (1-1)	**1** (1-1)	**3** (1-5)	**3** (1-6)
Combined model	**1** (1-1)	**1** (1-1)	**2** (1-3)	**1** (1-3)

Supplementary Table S4 shows the median of the optimal number of components across the 25 iterations of nested cross validation, for each PLSR model on the internal validation sample. This median number of components was subsequently applied to the external validation datasets for the respective training and testing conditions. PLSR models trained on rest data were optimally represented in a lower number of components than those trained on task data. Combined models did not require more PLS components to represent than static or dynamic matrices.

Additionally, we felt it important to demonstrate the consistency of PLSR components across LOOCV iterations. PLSR models were first fit with all possible number of components (1 to n-1). For each such number of components, Pearson's correlation between model coefficients in every pair of iterations was calculated, and the mean of these was taken across the pairs. Across static, dynamic, and combined FC models trained and tested on either rest or task data, and across all possible numbers of components, mean similarity exceeded 0.9, suggesting that PLSR components are highly robust to incremental changes in the training set features.