Research Article: Theory/New Concepts | *Cognition and Behavior*

Decoding task-specific cognitive states with slow, directed functional networks in the human brain

https://doi.org/10.1523/ENEURO.0512-19.2019

Cite as: eNeuro 2020; 10.1523/ENEURO.0512-19.2019

Received: 4 December 2019
Accepted: 12 December 2019

1. Manuscript Title:

Decoding task-specific cognitive states with slow, directed functional networks in the human brain

2. Abbreviated Title:

Decoding task states with directed fMRI networks

3. Authors:

Shagun Ajmera[1], Hritik Jain[1], Mali Sundaresan[1] and Devarajan Sridharan[1,2]

Affiliations:

[1]Centre for Neuroscience and [2]Department of Computer Science and Automation,
Indian Institute of Science,
C. V. Raman Avenue,
Bangalore 560012, India

4. Author Contributions:

Designed research - DS

Performed research - SA, MS, HJ

Contributed unpublished reagents/analytic tools - MS, DS

Analyzed data - SA, MS, HJ

Wrote the paper - DS

5. Corresponding author:

Devarajan Sridharan
Assistant Professor and
Wellcome-Trust DBT India Alliance Fellow,
Centre for Neuroscience, Rm F104
Indian Institute of Science,
C. V. Raman Avenue,
Bangalore 560012, India

Phone: +91 802 293 3434

E-mail: sridhar@iisc.ac.in

6. Number of Figures

5 Main figures, 5 Extended Data figures

7. Number of Tables

1 Main (statistical) table, 8 Extended Data Tables

8. Number of Multimedia

No multimedia

9. Number of words for Abstract

250 words

10. Number of words for Significance Statement

119 words

11. Number of words for Introduction

761 words

12. Number of words for Discussion

1890 words

13. Acknowledgements

The authors would like to thank Lionel Barrett and Catie Chang for their comments on a preliminary version of this manuscript, and Govindan Rangarajan and Arshed Nabeel for helpful discussions. This research was funded by a Wellcome Trust-Department of Biotechnology India Alliance Intermediate fellowship, a Science and Engineering Research Board Early Career award, a Pratiksha Trust Young Investigator award, a Department of Biotechnology-Indian Institute of Science Partnership Program grant, a Sonata Software foundation grant and a Tata Trusts grant (all to DS).

14. Conflict of Interest

Authors report no conflict of interest.

15. Funding sources

Wellcome Trust-Department of Biotechnology India Alliance Intermediate fellowship; Science and Engineering Research Board Early Career award; Pratiksha Trust Young Investigator award; Department of Biotechnology-Indian Institute of Science Partnership Program grant; Sonata Software foundation grant; Tata Trusts grant

Keywords:

functional connectivity; Granger causality; partial correlations; support vector machines; emergent dynamics; cognitive score prediction

Abstract

Flexible functional interactions among brain regions mediate critical cognitive functions. Such interactions can be measured using functional magnetic resonance imaging (fMRI) data either with instantaneous (zero-lag) or lag-based (time-lagged) functional connectivity. Because the fMRI hemodynamic response is slow, and sampled at a timescale (seconds) several orders of magnitude slower than the underlying neural dynamics (milliseconds), simulation studies have shown that lag-based fMRI functional connectivity, measured with approaches like Granger-Geweke causality (GC), provides spurious and unreliable estimates of underlying neural interactions. Experimental verification of this claim is challenging because neural ground truth connectivity is, often, unavailable concurrently with fMRI recordings. Here we demonstrate that, despite these widely-held caveats, GC networks estimated from fMRI recordings contain useful information for classifying task specific cognitive states. We estimated instantaneous and lag-based GC functional connectivity networks using fMRI data from 1000 participants (Human Connectome Project database). A linear classifier, trained on either instantaneous or lag-based GC, reliably discriminated among seven different task and resting brain states, with over 80% cross-validation accuracy. With network simulations, we demonstrate that instantaneous and lag-based GC exploited interactions at fast and slow timescales, respectively, to achieve robust classification. With human fMRI data, instantaneous and lag-based GC identified complementary, task-core networks. Finally, variations in GC connectivity explained inter-individual variations in a variety of cognitive scores. Our findings show that instantaneous and lag-based methods reveal complementary aspects of functional connectivity in the brain, and suggest that slow, directed functional interactions, estimated with fMRI, may provide useful markers of behaviorally relevant cognitive states.

Significance statement

Functional MRI is a leading, non-invasive technique for mapping functionally connected networks in the human brain. The fMRI hemodynamic response is slow, noisy and sampled far more slowly (seconds) than the timescale of neuronal spikes (milliseconds). fMRI data is, therefore, considered unsuitable for mapping directed, time-lagged functional connectivity among brain regions. Here, we apply machine learning to fMRI data from 1000 human participants and show that directed connectivity, estimated with Granger-Geweke Causality from fMRI data, accurately predicts task-specific cognitive states, and individual subjects' behavioral scores. Moreover, directed connectivity robustly identifies network configurations that may be challenging to identify with conventional, correlation-based approaches. Directed functional connectivity, as measured with fMRI, may be relevant for a complete understanding of brain function.

Introduction

Mapping functional coupling among brain regions is, key to mapping brain function and for understanding how the brain produces behavior (Fox et al., 2005). Human fMRI studies have commonly investigated such functional coupling with correlation-based measures, including the Pearson correlation coefficient (Vincent et al., 2008; Buckner et al., 2009) and partial correlations between pairs of brain regions (Marrelec et al., 2006; Ryali et al., 2012). Correlation-based measures characterize “instantaneous” functional interactions among brain regions that occur at timescales faster than the sampling rate of the measurement (Barnett and Seth, 2017). In contrast, comparatively few studies, have characterized functional connectivity with lag-based measures (Sridharan et al., 2008; Ryali et al., 2011).

Measures of linear dependence and feedback, based on Granger-Geweke causality (GC; Geweke, 1982, 1984) represent a powerful approach for estimating both instantaneous and lag-based functional connectivity. These measures are firmly grounded in information theory and statistical inferential frameworks (Geweke, 1982, 1984; Seth et al., 2015). GC measures have been widely applied to estimate functional connectivity in recordings of brain activity made with electroencephalography (EEG; Dhamala et al., 2008), magnetoencephalography (MEG; Ding and Wang, 2014) and electrocorticography (ECoG; Bastos et al., 2015). However, the application of GC measures to brain recordings made with functional magnetic resonance imaging (fMRI) has provoked significant controversy (Chang et al., 2008; Smith et al., 2011; Friston et al., 2013; Wen et al., 2013). Because the hemodynamic response is produced and sampled at a timescale (seconds) several orders of magnitude slower than the underlying neural processes (milliseconds), previous studies have argued that lag-based measures, particularly lag-based GC, produce spurious and unreliable estimates of functional connectivity, when applied to fMRI data (fMRI-GC; Lin et al., 2009; Smith et al., 2011; Seth et al., 2013; Solo et al., 2018).

Three primary confounds have been identified with inferring connectivity with fMRI-GC. First, systematic differences in hemodynamic lags across regions could yield spurious directionality of GC connections (Chang et al., 2008; Friston, 2009; Smith et al., 2011). Second, in simulations, measurement noise added to the signal during fMRI acquisition significantly degrades GC functional connectivity estimates (Nolte et al., 2008; Smith et al., 2012; Seth et al., 2013). Finally, downsampling recordings to the typical fMRI sampling rate (seconds), three orders of magnitude slower than the timescale of neural spiking (milliseconds), effectively eliminates all traces of functional connectivity inferred by GC (Seth et al., 2013).

The controversy regarding the application of GC to fMRI data continues to date. On the one hand, claims regarding the efficacy of GC estimates are primarily based on simulations (Seth et al., 2015; Solo, 2016), and are only as valid as the underlying model of neural activity and hemodynamic responses. Because the precise mechanism by which neural responses generate hemodynamic responses is an active area of research, strong conclusions cannot be drawn based on fMRI simulations alone. On the other hand, establishing ground-truth validity for fMRI functional connectivity requires invasive neurophysiological recordings across many brain regions, concurrently during fMRI scans, a challenging enterprise. For example, David et al. (2008) addressed this technical challenge, and showed that, in a rodent model, fMRI-GC functional connectivity estimates matched connectivity estimates from intracerebal EEG only when confounding hemodynamic effects were explicitly removed from the former.

Here, we seek to examine the empirical relevance of fMRI-GC functional connectivity networks in human subjects for identifying task specific cognitive states, and for predicting behavior, by applying machine learning (Arbabshirani et al., 2017) to fMRI-GC networks. We estimated instantaneous and lag-based GC connectivity with fMRI data drawn from 1000 human subjects, recorded under seven different task conditions and in the resting state (Human Connectome Project database; Glasser et al., 2013). We trained a linear classifier, based on GC connectivity features, to discriminate among the different task and resting conditions, and assessed classifier accuracy with cross validation. Instantaneous and lag-based fMRI GC connectivity could decode task-specific cognitive states with superlative accuracies. Next, with simulations, we show that slow interactions at the timescale of seconds emerge in networks with sparse, random connectivity (Ganguli et al., 2008), despite individual neurons operating at fast, millisecond timescales. We further show that such interactions can be recovered with GC sampled at slow fMRI timescales, providing a putative explanation for the success of GC with classifying task states (Sundaresan et al., 2017). Finally, we demonstrate that GC connectivity features can be used as predictors (Aiken et al., 2003; Liégeois et al., 2019) to explain inter-individual variations in behavioral scores across a variety of cognitive tests. In summary, fMRI-GC may be relevant for understanding slow, emergent and behaviorally relevant functional interactions in the human brain.

Materials and Methods

Ethics statement.

The scanning protocol for the Human Connectome Project was approved by the Human Research Protection Office at Washington University at St. Louis' (IRB # 201204036). Only de-identified, publicly released data were used in this study. Secondary data analysis procedures were approved by the Institute Human Ethics Committee at Indian Institute of Science, Bangalore.

Data and code availability statement.

Data used in the study is available in the public domain at the Human Connectome Project database (https://db.humanconnectome.org/). Data sharing permissions can be found at the HCP website. The code required to replicate results described in the paper was developed at the Indian Institute of Science, Bangalore, India, and is freely available online at https://figshare.com/s/9d9131a6780fc8197cf1.

fMRI data, parcellation and time-series extraction.

We analyzed minimally preprocessed brain scans of 1000 subjects, drawn from the Human Connectome Project (HCP) database (S1200 release; age range: 22-35 years; 527 females); fMRI acquisition and preprocessing details are described elsewhere (Van Essen et al., 2012; Glasser et al., 2013). Briefly, in this preprocessing pipeline, subject’s data is firstly aligned to MNI space, volumes are segmented based on predefined subcortical parcels, and white matter and pial (cortical) surfaces are registered to the respective surface atlas. This is followed by gradient distortion correction, motion correction, image distortion correction, spline resampling, intensity normalization and brain masking. Next, cortical and subcortical grey matter voxels are mapped onto standard cortical surface vertices and subcortical parcels, respectively. Extended Data (ED) Figure 1-3 shows the identifiers of the subjects from whom data were analyzed. Data were analyzed from resting state and seven other task conditions (ED Figure 1-1): Emotion processing, Gambling, Language, Motor, Relational processing, Social cognition and Working memory; in most figures, these tasks are referred to with their initial letters. fMRI scans for the relational task were not available for 9/1000 subjects; therefore, we analyzed a total of 7991 scans across all tasks and subjects.

We employed five different brain parcellations based one anatomical atlas and four functional atlases (ED Figure 1-4). For the tasks versus resting-state classification based on GC connectivity (first section of Results), all 5 parcellations were used. Based on the classification performance in this analysis, we picked the three parcellations with the highest accuracies (90 node and 14 network parcellations of Shirer et al., 2012 and 96 network parcellation of Thomas Yeo et al., 2011) and these were used for the pairwise classification analysis of each task versus the other as well as the n-way task classification analyses. Analysis with averaging GC features across subjects (Fig. 1D) was performed with a 90 node parcellation (Shirer et al., 2012). Classification analyses with data purged of instantaneous correlations and unweighted digraph representations (second section of Results) were performed with the Shirer et al (2012) 14 network parcellations. Analyses involving identifying task-generic and task-discriminative networks, as well as behavioral score predictions, based on GC features (last section of the Results) were performed with the Shirer et al (2012) 14 network parcellation. Voxel time series were extracted using Matlab and SPM 8 (Penny et al., 2007) , and regional and network time series were computed by averaging the time series across all voxels in the respective region or network.

We employed parcellations with fewer, more coarse-grained regions, rather than fine-grained parcellations because Granger Causality estimates were more reliable when the number of regions was fewer than the number of timepoints. Both task and resting scans were of sufficient duration (~200-300 volumes) to permit robust GC estimation. Finally, we noticed that in some parcellations, there were overlapping voxels between some of the regions. To avoid mixing of signals, we assigned each overlapping voxel to the region to whose centroid it was closest, based on Euclidean distance.

Estimating functional connectivity with GC.

We modeled instantaneous and lag-based functional connectivity between brain regions using conditional Granger-Geweke Causality (Geweke, 1984). The linear relationship between two multivariate signals **x** and **y** conditioned on a third multivariate signal **z** can be measured as the sum of linear feedback from **x** to **y**

($Fx\rightarrow y|z$), linear feedback from $\mathbf{y}$ to $\mathbf{x}$ ($Fy\rightarrow x|z$), and instantaneous linear feedback ($Fx\circ y|z$) (Geweke, 1984; Roebroeck et al., 2005). To quantify these linear relationships, we model the future of each time series in terms of their past values, using multivariate autoregressive (MVAR) modeling (Extended Data Mathematical Note, Section S1, equation 1). MVAR model order was determined with the Akaike Information Criterion (AIC) for each subject, and was typically 1. The MVAR model fit was used to estimate both an instantaneous connectivity matrix using iGC ($Fx\circ y|z$) and a lag-based connectivity matrix using dGC ($Fx\rightarrow y|z$). Details are provided in ED Mathematical Note, Section S1. Because the minimum number of scans across datasets (176) exceeded the number of nodes in all parcellations used (e.g. 90 nodes in the Shirer et al, 2012 parcellation), the GC estimation was well-posed.

Briefly, $Fx\rightarrow y|z$ is a measure of the improvement in the ability to predict the future values of $\mathbf{y}$ given the past values of $\mathbf{x}$, over and above what can be predicted from the past values of $\mathbf{z}$ and $\mathbf{y}$, itself (and vice versa for $Fy\rightarrow x|z$). $Fx\circ y|z$, on the other hand, measures the instantaneous influence between $\mathbf{x}$ and $\mathbf{y}$ conditioned on $\mathbf{z}$ (see ED Mathematical Note, Section S1). We refer to $Fx\circ y|z$, as *instantaneous* GC (iGC), and $Fx\rightarrow y|z$ and $Fy\rightarrow x|z$ as lag-based GC or *directed* GC (dGC), with the direction of the influence ($\mathbf{x}$ to $\mathbf{y}$ or vice versa) being indicated by the arrow. The "full" measure of linear dependence and feedback $Fx,y|z$ is given by: $Fx,y|z = Fx\rightarrow y|z + Fy\rightarrow x|z + Fx\circ y|z$. $Fx,y|z$ measures the complete conditional linear dependence between two time series. If, at a given instant, no aspect of one time series can be explained by a linear model containing all the values (past and present) of the other, $Fx,y|z$ will evaluate to zero (Roebroeck et al., 2005).

Classification with linear SVM based on GC connectivity.

The connection strengths of the estimated GC functional connectivity matrices were used as feature vectors with a linear classifier based on SVM for high dimensional predictor data. For a parcellation with n ROIs, the number of features for iGC-based classification was $n(n-1)/2$ (upper triangular portion of the symmetric $n\times n$ iGC matrix) and for dGC-based classification it was n^2-n (all entries of the $n\times n$ dGC matrix, excluding self-connections on the main diagonal).Based on these functional connectivity features, we asked if we could reliably distinguish each task condition from resting state (e.g. language versus resting) or each task condition from the other

For pairwise classification of resting state scans versus each task we used Matlab's `fitclinear` function, optimizing hyperparameters using a 5-fold approach: by estimating hyperparameters with five sets of 200 subjects in turn, and measuring classification accuracies with the remaining 800 subjects. Classification performance was assessed with leave-one-out and 10-fold cross-validation. We also assessed the significance of the classification accuracy with permutation testing (see Methods). In simulations, we observed that the magnitude of GC estimates varied based on the number of timepoints used in the estimation. To prevent this difference in number of timepoints from biasing classification performance, each scan was truncated to a common minimum number of time samples across the respective scans being classified (task, resting) before estimating GC. For each subject, GC connectivity was estimated independently for the two scan runs (left-to-right and right-to-left phase encoding runs), and averaged across the runs. Hyperparameters optimized included the regularization parameter, regularization method (ridge/lasso) and the learner (linear regression model, svm/logistic) using the `OptimizeHyperparameters` option to the `fitclinear` function. Hyperparameter optimization was performed only for task vs. rest classifications, but not for subject feature averaging, task vs. task, or N-way classification analyses.

For pairwise classification of each task versus the other, default hyperparameters were used in the `fitclinear` function and classification performance was assessed with leave-one-out cross-validation. For n-way classification, we used MATLAB's `fitcecoc` function, which is based on error-correcting output codes, and fits multiclass models for SVMs. Briefly, the function implemented a one-vs-all coding design, for which seven (number of classes in multiclass classification) binary learners were trained. For each binary learner, one class was assigned a positive label and the rest were assigned negative labels. This design exhausts all combinations of positive class assignments. Classification performance in n-way classification was assessed with leave-one-out cross-validation. For each classification analysis mentioned above, task scans were truncated to the common minimum number of time samples across each set of scans, before estimating GC.

Classification based on GC connectivity across sub-tasks and with sub-sampled data.

Tasks in the HCP data were run as a block design, alternating between various conditions (sub-tasks). We tested whether GC connectivity would be able to classify among sub-tasks within each task (ED Figure 1-2). fMRI time series corresponding to each sub task was obtained by concatenating blocks of fMRI task time series pertaining to the respective sub task; the temporal order across blocks was preserved while concatenating the data. We also ensured that data at the conjunction of two successive blocks, which represented non-contiguous time points, were not used for GC estimation. The two sub tasks to be classified were then truncated to have same number of time points. GC estimation and pair-wise classification across sub-tasks was performed with the procedure described in the previous section. The Shirer et al (2012) 14-network parcellation was used for these analyses. For the motor task, time series for the left and right finger movement blocks were combined into a “hand” movement sub-task, and left and right toe movement blocks were combined into a “foot” movement sub-task.

We also tested whether GC on fMRI data sampled at slower rates would suffice to classify among task and resting states. We obtained time series downsampled at 2x the original sampling interval by removing data at even numbered sample points, and retaining data at odd numbered sample points (k=1, 3, 5…). The even-sample point data were appended the end of odd-sample data series, thereby retaining the overall number of data points in the original time series. Again, we ensured that data at the conjunction of the odd- and even-sampled data series (last odd-sampled point and first even sampled point), which represented non-contiguous data points, were not used for GC estimation. Similarly, we obtained time series downsampled at 3x the original sampling interval by removing every third data point, starting with the second or third data point, and concatenating these timeseries to retain the overall number of data points in the original timeseries. As before, GC estimation and pair-wise classification was performed with the procedure described in the previous section

Permutation testing of classifier accuracies.

We performed permutation tests for evaluating the statistical significance of classifier performance, using the method outlined in (Ojala and Garriga, 2010). The test involved permuting task labels independently for each subject and computing a null distribution of 10-fold cross-validation accuracy. We employed 1000 surrogates and assessed significance of each empirically estimated 10-fold cross-validation accuracy values for dGC and iGC, based on the proportion of samples in the null distribution which were greater than the cross-validation accuracy estimated from the data. We conducted these analyses for the tasks versus resting state classifications, n-way task classification, classification analyses after purging instantaneous correlations and those based on digraph features, separately for the two metrics (dGC and iGC).

Testing for data stationarity and goodness of MVAR model fit.

Computing GC based on VAR modeling assumes that the timeseries represent a stationary process. Four different tests were performed to test whether the MVAR model provided a valid and adequate fit to the data (ED Figure 1-7). We performed these tests for parcellated time-series using scripts provided in the Multivariate Granger Causality (MVGC) toolbox (Barnett and Seth, 2014). First, we checked for the stability of the MVAR model fit by computing logarithm of the spectral radius using the *var_specrad()* function. A negative value was taken to indicate a stable fit. Second, we assessed consistency of the model fit, which quantifies what proportion of the correlation structure in data is accounted for by the VAR model, using the *consistency()* function. We adopted a threshold of 80% (or above) for both task and resting timeseries to consider the data to have passed the test for consistency (Barnett and Seth, 2014). Third, we evaluated the whiteness of residuals based on the Durbin-Watson test for absence of serial correlation of VAR residuals, using the *whiteness()* function. Values of the Durbin-Watson statistic less than 1 or greater than 3 signify a strong positive or negative correlation, respectively among the residuals (Barnett and Seth, 2014). Subjects for whom the Durbin-Watson statistic lay between 1 and 3 for more than 90% of the regional timeseries, for both task and resting state data, were considered to have passed the test. Fourth, we checked for stationarity based on the augmented Dicky-Fuller unit-root test (ADF), using the *mvgc_adf()* function. As in the previous case, subjects for whom the ADF test statistic was less than its critical value for more than 90% of the regional timeseries, for both task and resting state data, were considered to have passed the test.

Control for motion artifacts.

We checked whether systematic differences in motion artifacts could contribute to the superlative classification accuracies observed with GC. For this, we calculated Frame-wise Displacement (FD; Power et al., 2012) as the sum of temporal derivatives of translational and rotational displacement along the three (x,y,z) axes in mm, with the estimated motion parameters provided by HCP. Frames with FD>0.5mm were considered “misaligned” and were discarded (“scrubbed”) while estimating GC values. Because dGC is estimated based on lagged correlations, we also discarded one frame before and after every misaligned frame (AR model order was typically 1 for these data). We then repeated the SVM-based two-way classification of resting state from the seven different task states, with GC features estimated on the “motion scrubbed” data; we also repeated n-way classification among the 7 tasks. Comparison of classification (cross-validated) accuracies with and without motion scrubbing, across all 1000 subjects, is shown in ED Figure 1-6C.

Classification based on BOLD series

We tested how well the BOLD signal itself would classify among tasks, based on the mean and standard deviation of fMRI time series of each region, based on the Shirer et al parcellation (2012). Regional time series were truncated to common minimum number of timepoints for pair of task and resting state scans. LR and RL phase encoded data time series were concatenated, and mean and standard deviation were computed, for each of the 14 ROIs, providing 28 features for classification. Similarly, for n-way classification, time series of all tasks were truncated to the common minimum available number of timepoints across tasks, before computing the mean and standard deviation. Based on these 28 features, we sought to classify, as before, resting state from each task (two-way classification), and also among tasks (n-way classification).

Functional connectivity estimation and classification with partial correlations.

We compared the performance of classification based on GC measures with that based on partial correlations (PC). Partial correlations were computed based on the inverse of the covariance matrix as outlined previously (Marrelec et al., 2006; Ryali et al., 2012). Like iGC, the PC connectivity matrix is undirected and symmetric. Therefore, only the upper triangular portion of the matrix, including $(n*(n-1)/2)$ PC weights, was used as features in the classification analyses. Classification and cross-validation analyses followed the procedures described in the Methods section on “*Classification with linear support vector machines based on GC connectivity*”.

PC connectivity performed consistently better than GC connectivity for classifying task from resting state (Fig. 2A). We propose the following analytical explanation for this observation: PC, an estimator based on instantaneous covariance, is less susceptible to noise than GC, which is based on lagged covariance. This is due to the fact that the estimation of lagged-covariance is susceptible to errors from noise at multiple time-points. For illustration, consider a timeseries generated by a VAR(1) model: $\mathbf{x}(t) = A\,\mathbf{x}(t-1) + \mathbf{e}(t)$. The lagged (lag-1) covariance matrix (Σ_1) is estimated from the data as:

$$E\,[\mathbf{x}(t)\,\mathbf{x}(t-1)^T] = E\,[(A\mathbf{x}(t-1) + \mathbf{e}(t))\,\mathbf{x}(t-1)^T] = A\,E[\mathbf{x}(t-1)\,\mathbf{x}(t-1)^T] + E\,[\mathbf{e}(t)\,\mathbf{x}(t-1)^T]$$

Thus, when estimating the lagged covariance, the variance of the interaction term $E\,[\mathbf{e}(t)\,\mathbf{x}(t-1)^T]$ (second term in the right hand side) contributes to the variance of Σ_1 in addition to the variance in computing the instantaneous covariance $E[\mathbf{x}(t-1)\,\mathbf{x}(t-1)^T]$ (first term on the right hand side).

Classification based on GC connectivity in zero-lag correlation purged data.

To test the complementarity of information conveyed by GC functional connectivity versus functional connectivity based on instantaneous correlations we decorrelated the regional time series data to purge them of instantaneous correlations. We adopted two approaches for this purpose: i) zero-phase component analysis (ZCA) and ii) generalized eigenvalue decomposition (GEV).

i) Zero-phase component analysis (ZCA). Consider demeaned $t \times r$ data matrix $\mathbf{X}$ of regional timeseries with t timepoints and r regions, with covariance matrix $\mathbf{C}$. Decorrelating the data, to remove correlations among the columns of $\mathbf{X}$, is achieved with a whitening transformation. A common whitening transformation is based on principal components analysis (PCA): $\mathbf{Y} = \mathbf{W}_{\mathbf{PCA}}\mathbf{X}$, with $\mathbf{W}_{\mathbf{PCA}} = \mathbf{D}^{-1/2}\mathbf{E}^{T}$where $\mathbf{D}$ is a diagonal matrix, with the eigenvalues of $\mathbf{C}$ on its diagonals, and the columns of $\mathbf{E}$ contain the eigenvectors of $\mathbf{C}$. While the PCA transformation effectively decorrelates regional timeseries, there is no way to ensure one-to-one correspondence of the whitened dimensions across subjects, rendering subsequent classification analysis challenging. Consequently, here we chose a different whitening transformation based on zero-phase component analysis (ZCA), also known as the Mahalanobis transformation. Based on this transformation, whitening is achieved as: $\mathbf{Y} = \mathbf{W}_{\mathbf{ZCA}}\mathbf{X}$, with $\mathbf{W}_{\mathbf{ZCA}} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^{T} = \mathbf{C}^{-1/2}$. A particular advantage of the ZCA transformation is that it yields whitened data that is as close as possible to the original data, in a least-squares sense (Kessy et al., 2018). Therefore, each subject's data is projected on to a set of dimensions are most closely aligned with the underlying regional timeseries dimensions. Because the regions exhibit spatial correspondence across subjects (due to fMRI spatial normalization), the ZCA dimensions possess a natural, one-to-one correspondence across subjects, permitting subsequent classification. Before classification analysis ZCA dimensions were identified for each subject, separately for task and resting datasets. Regional time series for task and resting data were independently decorrelated by projecting onto their respective ZCA dimensions. GC (and PC) functional connectivity was estimated based on the these decorrelated timeseries, followed by classification analysis, as described previously (Methods section on "*Classification with linear support vector machines based on GC connectivity*"). As proof that the ZCA transformation was working effectively, classification accuracy based on PC (an instantaneous correlation measure) computed from ZCA components was at chance across all tasks (Fig. 2C top).

ii) Generalized Eigenvalue Decomposition (GEV). Although ZCA effectively purged correlations from the data, for the subsequent classification analyses task and resting state data were projected onto different, respective ZCA dimensions. Thus, the above-chance task versus resting state classification accuracy with GC features derived from ZCA components (Fig. 2C top) could perhaps be explained by, for example, systematic differences with how reliably ZCA dimensions were estimated across task and resting-state scans. We therefore sought an approach that could project both task and resting data into the same dimension while simultaneously decorrelating both. Such joint decorrelation may be achieved by projecting the data on to the generalized eigenvectors of the covariance matrices of the two datasets (Karampatziakis and Mineiro, 2014). Let $\mathbf{C_T}$ and $\mathbf{C_R}$ denote the covariance matrices of the task and resting datasets respectively. The generalized eigenvectors of these two symmetric matrices are given by the columns of $\mathbf{G} = \mathbf{E_T}\,\mathbf{D_T}^{-1/2}\,\mathbf{E_R}$, where, as before $\mathbf{D_T}$ is a diagonal matrix, with the eigenvalues of $\mathbf{C_T}$ on its diagonals, and the columns of $\mathbf{E_R}$ and $\mathbf{E_T}$ contain the eigenvectors of $\mathbf{C_R}$ and $\mathbf{C_T}$ respectively. It can be readily verified that $\mathbf{G}^T\mathbf{C_T}\mathbf{G}$ and $\mathbf{G}^T\,\mathbf{C_R}\,\mathbf{G}$ are both diagonal matrices. Therefore, $\mathbf{G}$ is a matrix that jointly diagonalizes both $\mathbf{C_T}$ and $\mathbf{C_R}$ and projecting either task or resting state data into the columns of G decorrelates the respective timeseries. So, for these analyses, the regional time series for the task and resting state conditions for each subject were jointly decorrelated by projecting them onto a single space spanned the generalized eigenvectors. This was followed by classification analysis with GC features obtained from the decorrelated time series. As before, we confirmed the effectiveness of the decorrelation by computing classification accuracy based on PC from GEV components, which was at chance across all tasks (Fig. 2C bottom).

Classification based on unweighted digraph representations of GC connectivity.

An unweighted directed graph (digraph) network representation shows the dominant direction (but not magnitude) of functional connectivity among brain regions. Obtaining significant directed connections with dGC is challenging due to number of multiple comparisons required for testing n^2-n connections. To identify significant directed connections, overcoming the multiple comparisons problem, we first subtracted the dGC connectivity matrix from its transpose and then applied the following two-stage procedure. In the first stage, the 1000 subjects were divided into five folds. For each two-way task versus resting state classification, recursive feature elimination (RFE, described in a later section titled "*GC feature selection based on Recursive Feature Elimination*") was performed based on dGC features of subjects from one fold (i.e. with 200 subjects). A minimal set of connection features identified by RFE, and their corresponding symmetric counterparts were then employed in the subsequent analyses; we term these connections K; the cardinality of K (the number of significant connections) was typically in the range of 2 - 86 (2.5th - 97.5th percentile). In the second stage, we identified statistically significant connections among these K features alone. For each of the subjects in the four remaining folds (i.e. 800 subjects), a null distribution for the dGC values of the features in K was obtained by estimating dGC following phase-scrambling the time series (Ryali et al., 2011). Next, we identified significant connections based on dGC values that occurred at the tail of the null distribution; the threshold for significant connections was determined based on a p-value of 0.05 with a Bonferroni correction for multiple comparisons. Classification performance based on digraph features was assessed with leave-one-out cross-validation.

GC connectivity in simulated fMRI time series.

To test the ability of GC measures to reliably recover functional interactions at different timescales, we simulated fMRI time series for model networks. Simulated fMRI time series were generated using a two-stage model. The first stage involved modeling latent neural dynamics with a stochastic, linear vector differential equation given by:

$$\tau d\mathbf{r}/dt = -\mathbf{r} + W\mathbf{r} + \boldsymbol{\varepsilon}$$

where $\mathbf{r}$ is the multivariate neural state variable representing the state of each neuron (or node) in the network (an N×1 vector, with N being the number of neurons), $d\mathbf{r}/dt$ is its temporal derivative, W is the neural ("ground truth") connectivity matrix (dimension N×N), τ is the time constant of each neuron (or node) and $\boldsymbol{\varepsilon}$ is i.i.d Gaussian noise ($N(0, \Sigma)$), with $\Sigma = I_N$ (N×N identity matrix). Although this model does not explicitly incorporate signal propagation delays, such vector Ornstein-Uhlenbeck models rank, arguably, among the most common models employed for simulating neural and fMRI time series, in many previous studies (Smith et al., 2011; Seth et al., 2013; Barnett and Seth, 2017).The multivariate time series $\mathbf{r}(t)$, sampled at discrete time points $r(k\Delta)$ with a sampling rate of Δ, were generated based on the discrete time (1-lag) connectivity matrix $A(\Delta)$ and a residual noise intensity $\Sigma(\Delta)$. Here:

$$A(\Delta) = e^{\Delta A}\ ; \qquad \Sigma(\Delta) = (1/\Delta)\ (\Gamma(0) - e^{\Delta A}\Gamma(0)\ e^{\Delta A'})$$

where $A = (1/\tau)\ (W - I_N)$, e^{A} denotes the matrix exponential, A' is the transpose of A, and $\Gamma(0)$ is the zero lag autocovariance which satisfies the continuous time Lyapunov equation $A\Gamma(0)+\Gamma(0)A'+\Sigma=0$ (Seth et al., 2013). In the second stage, the latent neural dynamics were convolved with the hemodynamic response function (HRF) to obtain the simulated fMRI time series: $\mathbf{y} = H \otimes \mathbf{x}$, where H is the canonical hemodynamic response function (hrf; simulated with *spm_hrf* in SPM8), $\otimes$ is the convolution operation and $\mathbf{y}$ is the simulated fMRI time series. Finally, following convolution with the hrf, the data were downsampled to 750 ms, to mimic the repeat time (TR) of the HCP fMRI scans used in this study. The same model was used for the different simulations used in the manuscript (third section of the Results). The parameters for the 2-node simulations, and for the 9-node (100 neurons per node) simulations are described in ED Figure 3-1.

For the 2-node simulations (Fig. 3A), iGC and dGC values were estimated by simulating the network for 200 timepoints, averaged across 25 repetitions. The 9-node simulations (Fig. 3B-C) were performed with a 900 neuron network, with 100 neurons per node. Each node had sparse, random excitatory/inhibitory connectivity among its neurons (parameters in ED Figure 3-1), whereas only 5% of neurons in each node were involved in inter-node connections, to mimic sparse, long-range connectivity in the neocortex (Knösche and Tittgemeyer, 2011). The network was simulated for 200 timepoints, and timeseries from all (100) neurons in each node were averaged to generate 9 node timeseries. iGC and dGC values were estimated from the node timeseries and averaged across 10 independent repetitions. Significance was assessed with a bootstrap approach that involved generating 1000 surrogates by phase scrambling the node timeseries to yield a null distribution of GC values (Ryali et al., 2011), followed by a Benjamini-Hochberg correction for multiple comparisons.

Simulations comparing PC and iGC connectivity (ED Figure 3-2 B-C) were performed as follows: We simulated a 7-node network with a 1-lag VAR model of the form: $\mathbf{X}_k = A\,\mathbf{X}_{k-1} + \boldsymbol{\varepsilon}_{k.}$ where $\mathbf{X}_k$ is the state of the discrete time process at discrete timestep ‘k’, *A* is the connectivity matrix, and $\boldsymbol{\varepsilon}$ is Gaussian noise with covariance matrix Σ_d. *A* was chosen to be a random matrix with spectral radius less than 1 to ensure stability. Σ was chosen such that the covariance between every pair of residuals was zero (independent residuals) except for the first two residuals. The correlation between these residuals, ε^1 and ε^2, was parametrically varied between -1.0 and 1.0 to systematically vary the strength of iGC connectivity. Note that, under this model, iGC between X^1 and X^2 vanishes only if and only if ε^1 and ε^2 are uncorrelated (Geweke, 1984).

GC feature selection based on Recursive Feature Elimination (RFE).

We performed features selection for analyses reported in Fig. 2D, 4B,C and ED Figure 4-2B, ED Figure 1-4E, based on Recursive Feature Elimination (RFE). RFE identifies a minimal set of features, which provide maximal cross-validation accuracy (Guyon and Elisseeff, 2003). Here, we implemented a two-level algorithm, described previously (Gel’fand and Yaglom, 1959; Chang et al., 2008). First, the data were divided into N_1 (here, 10) folds. Of these, N_1-1 folds were used as “training” data, and one fold was reserved as “test” data for quantifying the generalization performance of the classifier. Training data were pooled and further divided into N_2 (here, 5) folds. The SVM classifier was then trained on N_2-1 folds (leaving out one fold) and discriminative weights were obtained. The above procedure was repeated N_2 times by leaving out each fold, in turn. Average weights were then computed by averaging the absolute values of the discriminative weights across the N_2 runs. Next, 10% of the features (connections) contributing the lowest average weights were discarded, and the classifier was trained again with only the retained set of features. This procedure of feature selection and training was repeated until no more features remained. At this stage, the generalization performance for every set of retained features (each "RFE level") was assessed using the left out "test" data. The entire procedure was repeated N_1 times by leaving out each fold of the original data, in turn, as test data. Final generalization performances and discriminative weights of each RFE level were obtained as the average over N_1 folds. We selected the set of connections at the RFE level at which the generalization performance reached an "elbow": a minimal set of connections at which generalization performance dipped dramatically below its maximal level. To identify this elbow (e), we used a custom elbow fitting procedure, requiring a piecewise linear fit to the RFE curve, based on two lines, one for "x>e" and another for "x<=e", with the first line required to have a higher slope than the second. The first point in each RFE curve was excluded from the higher slope line fit (Fig. 4C, 4E, ED Figure 4-2B). RFE was typically repeated 5 times before determining peak accuracy and corresponding features.

Simulating hemodynamic lag variations across nodes.

We simulated systematic differences in hemodynamic lags across nodes by varying the onset parameter of the *spm_hrf* function (SPM8; Penny et al., 2007). For network configurations A and B described in Figure 4A, we simulated 4 scenarios: a) same mean HRF onset (μ_L= 3s) across nodes; b) source node HRF onset lagging the destination node by 1s ($\mu_{L\text{-src}} > \mu_{L\text{-dst}}$); c) source node HRF onset leading destination node by 1s ($\mu_{L\text{-src}} > \mu_{L\text{-dst}}$); and d) mixed latencies of lead and lag across source and destination nodes (see next). GC was estimated for 100 simulated participants, by sampling onset latencies for each of the 6 nodes (A-F) from normal distributions (truncated to have only positive latency values), over a range of different standard deviations (σ_L=0-1s, in steps of 0.2s). Onset latencies were sampled independently across participants, but were sampled such that the relative latency between each pair of source and destination nodes, across corresponding network configurations, remained the same for each participant. For example, if the onset latency difference between nodes A and B was 0.7s ($\mu_{L\text{-B}}-\mu_{L\text{-A}}$=0.7s) for a particular subject, the same difference in onset latency was also maintained between nodes B and C ($\mu_{L\text{-C}}-\mu_{L\text{-B}}$=0.7s). For simulations with mixed latencies (case d), 50% of simulated participants had onset latencies drawn from distributions with the source node lagging the destination node (case b) and the remaining 50% with the source node leading the destination node (case c). GC values were averaged over 5 runs for each simulated participant. Finally, we performed RFE to identify key connections that distinguished the two network configurations (same procedure as in Fig. 4B). Connections weights of the most discriminative connections following RFE are shown in ED Figure 3-2E (for σ_L=0.4s). Difference of dGC connections strengths as well as iGC connection strengths, for various values of σ_L, are shown in ED Figure 3-2D.

Identifying "task-generic" and "task-discriminative" GC connections.

To identify a minimal set of connections that occurred consistently across tasks ("task-generic" connections), we adopted the following approach. We performed RFE analysis for task versus resting state classification for each of the six tasks (all tasks except motor); we expected each of these tasks to recruit common cognitive control mechanisms. We then performed a binomial test to identify connections that were consistently activated across tasks. Briefly, the presence or absence of a connection in the set of RFE features for a given task versus resting state classification was considered as a Bernoulli trial, with probability of success (its presence) p being the mean number of RFE features identified across all six classifications. The number of trials n was the number of tasks versus resting state classifications (here n=6). The probability of a randomly picked connection being present in more than k such RFE sets is given by the cumulative distribution function for the binomial distribution $F(k; n, p)$. Significant connections were identified as those that occurred in k or more tasks, with threshold at the p=0.05 level.

To identify a minimal set of connections that maximally differed across tasks ("task-discriminative" connections), we used RFE with an n-way classifier, to classify among all six tasks (again, except the motor task). The n-way classifier is based on training n (here, 6) one-vs-all binary learners. At the second level of the RFE procedure described above, average weights were computed for each of these n binary learners by averaging the absolute values of the discriminative weights across the N_2 runs. Next, a set of features obtained by taking union of 1% of the features (connections) contributing the lowest average weights in each learner was discarded, and the classifier was trained again with only the retained set of features.

While quantifying the overlap between task-generic and task-discriminating connections identified separately for dGC, iGC and PC, we converted the dGC matrix to a lower triangular matrix by reflecting all connections about the main diagonal. The degree of overlap between PC and GC connections was quantified as the number of overlapping connections as proportion of the total number of connections identified by PC. We then computed a null distribution of the degree of overlap by randomly permuting the connection identities within each matrix, while preserving the overall number of connections in each matrix, and generating 1000 surrogate samples. The significance of the overlap of task-generic or task-discriminating connections between each pair of metrics (PC-dGC or PC-iGC) was quantified as the fraction of overlapping connections in the data that exceeded this null distribution.

Predicting behavioral scores based on GC connectivity

We asked whether inter-individual differences in GC connectivity would be relevant for predicting inter-individual differences in behavioral scores. HCP provides a well-validated battery of behavioral scores assessed with a wide range of cognitive tasks. The task battery is based on the NIH Toolbox for Assessment of Neurological and Behavioral function (Gershon et al., 2013), developed to create a uniform set of measures for rapid data collection in large cohorts. The toolbox includes assessments of cognitive, emotional, motor and sensory processing scores in healthy individuals. We pre-selected, based on domain knowledge, a specific subset of 51 scores for these analyses, using age-adjusted scores, wherever available (listed in ED Figure 5-1). Next, we sought to predict subjects' behavioral scores based on GC connectivity with an established leave-one-out approach (Tavor et al., 2016). Briefly, we used linear regression to predict behavioral scores using, as features, GC estimates of functional connectivity, separately for iGC (91 features or connections) and dGC (182 features). The leave-one-out analysis was performed such that the support vector regressor was fit on all but one subject and the learned beta weights were used to obtain predictions of the left-out subject's behavioral score, using that subject's own GC connectivity weights. Predicted scores were correlated with the actual scores using robust correlations ("percentage-bend" correlations; Wilcox, 1994).

Next, we asked if GC connectivity could identify an individual based on a composite marker of her/his behavioral scores. Because 40 subjects did not have a full complement of behavioral scores, data from the remaining 960 subjects was included in this analysis. The 51 behavioral scores were, each, z-scored across subjects and formatted into a "composite behavioral score" vector. This vector served as an individual specific composite marker of behavioral scores, as revealed the weak off-diagonal values in the covariance matrix of this vector across subjects (Fig. 5D top). dGC and iGC features of individual tasks, as well as combination of tasks (Relational and Working memory), were used to then predict the composite score marker for individual subjects, using the same leave-one-out procedure as described above. The observed and predicted set of composite scores was correlated across subjects. The distribution of observed versus predicted correlation values for each subject (values on main diagonal; Fig. 5D bottom yellow) were compared against between-subject correlation values (off-diagonal values; Fig. 5D bottom grey) using a Kolmogorov-Smirnov test.

Results

GC estimated from slowly sampled fMRI data suffices to distinguish task and resting states

We asked if instantaneous GC (iGC) and directed GC (dGC) (ED Mathematical Note Section S1) connectivity would flexibly reconfigure with task demand, by testing if GC connectivity sufficed to accurately classify among seven different task states or the resting state (ED Figure 1-1; Methods; (Geweke, 1982, 1984)). Data were obtained from 1000 participants from the Human Connectome Project (HCP) database (Van Essen et al., 2012). We used connection weights among brain regions in each network (iGC or dGC) as feature vectors in a linear classifier based on Support Vector Machines (SVM) for high dimensional predictor data. Accuracies for classifying resting state from a working memory task (WM task) are described first; accuracies for other tasks are presented subsequently.

Both iGC and dGC connectivity were able to distinguish the working memory task from resting state significantly above chance (Fig. 1B; $p<0.001$, permutation test). Maximum accuracy (median, [95% CI]) was 97.3% [96.3 - 98.0%] with iGC and 92.0% [90.5 - 93.2%] with dGC (ED Figure 1-5B Yeo Parcellation; Statistical Table[a,b]), iGC: precision= 97.2, recall= 97.4; dGC: precision= 90.9, recall= 93.2). k-fold ($k=10$) cross-validation accuracy was comparable (iGC: 97.1% [96.2 - 97.9%], dGC: 91.7% [90.3 - 93.0%]). These numbers correspond to maximum cross validation accuracy across all five parcellations tested (ED Figure 1-4; ED Figure 1-5A); accuracies with each parcellation are shown in the Extended Data (ED Figure 1-5B).Non-linear classifiers, such as SVMs based on radial basis function kernels produced similar results, with comparably above chance classification accuracy for both iGC and dGC connectivity (ED Figure 1-5C).

We repeated these analyses by classifying the six other tasks (ED Figure 1-1) versus resting state. iGC and dGC connectivity could accurately classify each task from resting state significantly above chance. For iGC, maximum classification accuracies ranged from 90.1%, for emotion task vs. resting state classification, to 97.1%, for language task vs. resting state classification. Similarly, for dGC, accuracies ranged from 78.1%, for emotion task vs. resting state classification, to 92.8%, for language task vs. resting state classification (Fig. 1B; Statistical Table[c]). In general, classification accuracy increased with more scan timepoints for each task versus resting state classification (Fig. 1C), consistent with GC being an information theoretic measure; we confirmed this result with simulations also (ED Figure 1-5D).

In these analyses, classification accuracies based on dGC were systematically lower than those based on iGC. We asked if dGC accuracies were poorer due to noise corrupting the fit of the autoregressive model, and if a more consistent estimate could be obtained by averaging dGC connectivity features, to remove uncorrelated noise, across subjects. We addressed this question by partitioning the data into two groups -- a training (T) group and a test (S) groups – with 500 subjects each. We trained the classifier on group T and tested the classifier prediction by averaging GC matrices across several folds of S, each fold containing a few (m=2, 4, 5, 10, 25 or 50) subjects; the procedure was repeated by exchanging training and test datasets (see Methods). For the vast majority of tasks (6/7), dGC's classification accuracy was more than 95% with as few as m=5 subjects within each fold of the test set (Fig. 1D). These results suggest that averaging dGC matrices across a few subjects, yielded reliable estimates of dGC connectivity.

We considered other factors that, in addition to intrinsic connectivity differences, could have produced these superior classification accuracies. First,GC-based accuracies for classifying task versus resting state scans might arise from differences in brain regions activated during each of these scans. In addition to task-relevant sensory input, overt motor responses always occurred during task scans but were absent during resting state scans (Barch et al., 2013; Glasser et al., 2013). Could GC features discriminate among more subtle connectivity variations across/within tasks? Second, scan data from the HCP database was sampled at a TR (repetition time) of 720 ms, considerably faster than the TR for conventional fMRI scans. Would GC accuracies degrade if the data were sampled at much slower sampling rate (~2000 ms), in line with conventional fMRI TR?

We addressed the first question in two stages. First, we asked if GC connectivity features would be able to classify which of the seven tasks each subject was performing in the scanner. First, we performed a pairwise classification of each task from the other. Maximum classification accuracies for iGC (dGC) ranged from 87% (67%) for the emotion vs. gambling task classification to 98% (91%) for the language vs. social task classification. Again, the number of timepoints for each task proved to be a strong indicator of classification accuracies (Fig. 1E): average inter-task classification accuracies were highest for the language task (iGC: 97%, dGC:88%, n=316 timepoints) and lowest for the emotion task (iGC: 91%, dGC: 77%, n=176 timepoints).Next, we performed an n-way classification analysis across all 7 tasks, again using linear SVM (Methods). Accuracies were significantly above chance (14.3% for 1-in-7 classification) for classifying among the seven tasks (Fig. 1F; maximum accuracy, iGC: 74.4% [73.3%-75.4%]; dGC: 47.6% [46.4%-48.7%]; p<0.001, permutation test; Statistical Table[d,e,f]). These results indicate that functional connectivity was consistently estimated with GC, and reliably different across tasks.

Second, each of the different tasks in the HCP database comprised of blocks of contiguous trials, each corresponding to one of (at least) two different sub-tasks (Barch et al., 2013; ED Figure 1-2). For example, the motor task comprised of blocks of movements of the right or left hand interleaved with blocks of trials involving movement of the right or left foot. Similarly, the working memory task comprised of interleaved blocks of 0-back and 2-back tasks. We asked, therefore, if GC connectivity could distinguish among subtler variations in brain states across sub-tasks within each task. We sought to classify across two sub-tasks for each of six tasks (ED Figure 1-2). In all cases, except one, both iGC and dGC connectivity discriminated between each pair of sub-tasks with higher than chance accuracies (Fig. 1G; maximum accuracy, iGC: 89.2% [87.6% - 90.7%]; dGC: 80.1% [78.9% - 82.9%]; p<0.05 permutation test; Statistical Table[g]). These results indicate that GC functional connectivity could accurately distinguish among sub-tasks within each task as well.

Next, we tested whether GC connectivity estimated from slowly sampled fMRI data could accurately classify task and resting states. We downsampled the data to either one half (2x TR=1440 ms) or one third (3x TR=2160 ms) of its original sampling rate, by decimation, while also concatenating the decimated data to the end of the sub-sampled timeseries to preserve the overall number of timepoints (Methods). We repeated both of the previous classification analyses – pairwise task versus resting state classification (Fig. 1H left), as well as n-way inter-task classification (Fig. 1H right). Following downsampling, we observed that classification accuracies were marginally higher than accuracies in the original data, in the case of dGC (2x: p=0.02; 3x: p=0.06; Wilcoxon one-tailed signed rank test; Statistical Table[h]) and were even higher than those in the original data, in the case of iGC (2x: p=0.01; 3x: p=0.01; Statistical Table[i]), across tasks. These results indicate that the superlative sampling rate of the HCP fMRI data was not the primary reason for these high classification accuracies for GC-based classification.

We performed additional control analyses to confirm that these results were not due to data non-stationarity, biases in GC estimation or head motion artifacts.

As a first control analysis, we repeated the classification analyses including only subjects for whom the data passed tests of stationarity (Methods, ED Figure 1-7); typically, data from >99% of subjects passed three out of four tests of stationarity (except for the consistency test) across all tasks. Mean GC matrices for each task and resting scan closely resembled those of the population for subjects whose data passed all four tests of stationarity across all tasks (n=141, ED Figure 1-6A). Statistical tests revealed that dGC connectivity was only marginally different for this subset of subjects (proportion of significantly different connections: 6.3%±0.9%, mean ± std. error, across tasks; Kolmogorov-Smirnov test with Benjamini-Hochberg correction for multiple comparisons) whereas iGC connectivity was substantially different (80.6%±8.0%, mean ± std. error). Nevertheless, accuracies for classifying task versus resting state, as well as for classifying among tasks, were very similar and, in fact, marginally higher for the subjects who passed tests of stationarity compared to the population (ED Figure 1-6B).

As a second control, we repeated the same analyses by deriving GC estimates with a single full regression (one-stage GC), instead of with separate full and reduced regressions (two-stage GC; Methods); this analysis was necessary due to recent observations that the two-stage GC model can produce biased estimates, especially with incorrectly specified model orders (Stokes and Purdon, 2017; Barnett et al., 2018). Empirically, GC estimates for each of these methods were numerically different, but tightly correlated across subjects (ED Figure 1-6E) and tasks (ED Figure 1-6F): correlation values ranges 0.94-0.97 for dGC (p<0.001, ED Figure 1-6D). As before, we observed a very similar pattern of classification accuracies with the single full regression model (n-way classification accuracy among 7 tasks computed with the Shirer et al 14-network parcellation (2012): 48.3% based on dGC, 56.4% based on iGC), versus when GC was estimated with separate full and reduced regressions (47.6% based on dGC, 56.2% based on iGC; chance accuracy: 14.3% for 1-of-7 classification).

As a third control, we sought to remove the contribution of motion artifacts to these superlative classification accuracies. HCP's minimally pre-processed fMRI data are already motion corrected, based on FSL's MCFLIRT algorithm (Van Essen et al., 2012). We further controlled for motion artifacts using "motion scrubbing" (Power et al., 2012), by discarding frames with framewise displacement (FD) values greater than 0.5 mm (see Methods). Overall, across all task and resting state scans less than 2% of frames were discarded with this approach (ED Figure 1-6G). We recomputed GC values on the motion scrubbed data, for each of the 1000 subjects (Methods), and repeated the task-vs-rest and n-way task classification analyses. Classification accuracies following motion scrubbing were closely similar and marginally (albeit significantly) higher than accuracies obtained with the original data (ED Figure 1-6C; $p<0.01$ one-tailed signrank test).

As a fourth control, we sought to test how well the BOLD signal itself would classify among tasks, based on the mean and standard deviation of fMRI time series parcellated with the Shirer et al 14-network parcellation (2012) (see Methods). Accuracies for classifying a task state from rest were significantly lower [range: 62.7% - 67.7%; median: 65.9%] as compared to both dGC and iGC based classification accuracies ($p<0.01$ one-tailed signrank test;). In fact, n-way classification accuracy was 15.7%, only marginally above chance of 14.3%

These results demonstrate that both iGC and dGC yielded task-specific signatures of functional connectivity even with slowly sampled fMRI data (TR~2000 ms): these estimates were consistent across subjects and reliably different across tasks to permit successful classification. Furthermore, these superlative classification accuracies were obtained despite widely held caveats concerning the application of GC to fMRI data (Stokes and Purdon, 2017) suggesting that even if individual fMRI-GC network connections are unreliably estimated for a given task, the difference in fMRI-GC network connectivity across tasks was sufficiently reliable and informative to permit accurate classification among them.

Correlation-purged GC connectivity suffices for accurate task-state classification

Correlation-based (zero-lag) connectivity measures (e.g. partial correlations or PC) have been widely applied to estimate functional connectivity from fMRI data (Liang et al., 2012; Ryali et al., 2012). In fact, several previous studies(Smith et al., 2011; Seth et al., 2013) have argued that correlation-based measures are more reliable and should be preferred to lag-based measures like GC(Seth et al., 2015), for estimating functional connectivity with fMRI data. We tested this claim here with a three-fold analysis approach.

First, we asked how classification accuracies based on PC connectivity would compare with those reported above, based on GC connectivity. Maximum classification accuracies with PC connectivity ranged from 96-99% for task versus resting state classification, and were consistently higher than accuracies with GC connectivity (Fig. 2A). These results are along expected lines: estimators based on same-time covariance, such as PC, are less susceptible to noise than those based on lagged covariance, such as GC (derived analytically in the Methods, section on *Functional connectivity estimation and classification with partial correlations*). In addition, as mentioned previously, GC is an information theoretic measure: classification accuracy with iGC and dGC increased systematically with more scan time points, asymptotically matching PC accuracies (ED Figure 1-5D).

Second, we asked if lag-based connectivity could accurately classify task from resting state, once the data were purged of all instantaneous correlations. To accomplish this, we adopted two approaches: i) zero-phase component analysis (ZCA) and ii) generalized eigenvalue decomposition (GEV) (Methods). Briefly, ZCA (or the Mahalanobis transformation) produces whitened time series data that is closest, in a least squares sense, to the original regional time series data. As an alternative approach, we decorrelated both task and resting state time series jointly by projecting them onto a single set of generalized eigenvectors (GEV). These approaches provided empirical upper and lower bounds on GC's performance on correlation-purged data (Methods).

GC connectivity features sufficed to successfully classify all tasks from resting state, even in correlation-purged data. With ZCA, iGC accuracies ranged from 84% to 96% whereas dGC accuracies ranged from 82% to 96% across tasks. With GEV, iGC accuracies ranged from 60% to 71% whereas dGC accuracies ranged from 56% to 76% across tasks; in each case, classification accuracies were significantly above chance ($p<0.001$, permutation test; Statistical Table[j]). We confirmed that performance in each case was not an artifact of the decorrelation procedure (ZCA/GEV) by randomly interchanging task and resting state labels for each pair of datasets across subjects (Methods); shuffling labels reduced classification accuracy to chance. Note that in every case, classification performance based on PC connectivity was at chance (Fig. 2C), a direct consequence of removing instantaneous correlations from the data. Despite this, classification accuracies based on iGC connectivity were not at chance; in the next section, we discuss potential reasons for these differences between iGC and PC classification accuracies.

Third, we asked if an unweighted directed graph (digraph) network representation – whose edges indicated the dominant direction, but not magnitude, of connectivity (Fig. 2D) – would suffice to distinguish task from resting brain states (Methods). Again, dGC directed graphs successfully distinguished each task from resting state well above chance. Classification accuracies ranged from 56% for the motor task versus resting state classification to 68% for the relational task versus resting state; for each task, classification accuracies were significantly above chance ($p<0.001$; permutation test; Statistical Table[k]). Interestingly, we did not see a strong influence of the number of data points on classification accuracy in this case (Fig. 2D, purple dots). For instance the emotion task (n=176 timepoints) was classified with an accuracy of 62% from resting state, which was comparable to the classification accuracy of working memory (n=405 timepoints) from resting state (64%). Both iGC and PC, which are symmetric connectivity measures, could provide no directed connectivity information.

These results demonstrate that lag-based connectivity contained sufficient information to classify task from resting state even when instantaneous correlations were entirely purged from the data. Moreover, unweighted directed connectivity graphs alone, indicating the direction, but not scalar magnitude, of GC connectivity, sufficed to accurately classify task from resting brain states. These findings indicate that directed functional connectivity measures, like dGC, provide connectivity information that is distinct from, and complementary to, what can be obtained with undirected functional connectivity measures, like PC.

Instantaneous and directed GC identify complementary aspects of functional connectivity

What characteristics of functional connectivity are respectively identified by instantaneous and lag-based connectivity? And how can lag-based connectivity be reliably estimated with fMRI data, which is sampled at time scales orders of magnitude slower than neural timescales? We addressed both of these questions, first, with simulations (this section) and, then, with real data (next section).

First, we tested the ability of GC to reliably recover functional interactions in simple, two-node feedforward networks operating at different timescales (Fig. 3A). We simulated fMRI data using a two-stage model (Methods): i) a latent variable model that describes the dynamics of the nodes (vector Ornstein-Uhlenbeck process; Ganguli et al., 2008); ii) a convolution of these neural dynamics with a hemodynamic response function to obtain the simulated fMRI time series (Smith et al., 2011; Seth et al., 2013). Based on this model, we simulated activity in two 2-node networks. In the first network, individual node decay timescales were set to 50 ms, whereas in the second network, these were set to 1000 ms (parameters in ED Figure 3-1A). For convenience, we refer to these two network timescales as “fast” (50 ms) and “slow” (1000 ms). We then varied the sampling interval (T_s) of the simulated data from 50 ms to 1450 ms in steps of 100 ms. Connections at both “fast” and “slow” timescales were generally discovered by iGC regardless of sampling interval, although connections at slow timescales were less robustly detected than those at fast timescales (Fig. 3A). On the other hand, the connection in the “fast” timescale network was not discovered by dGC when the sampling interval was higher than 50 ms, in line with the results of Smith et al (2011). However, the connection in the “slow” timescale network was reliably discovered by dGC across a wide range of sampling intervals, upto, and exceeding 1000 ms (Statistical Table[l]). In each case, dGC failed to discover the underlying interaction when the sampling interval was much higher than the slowest timescale in each network, consistent with recent theoretical results (Barnett and Seth, 2017). These findings suggest that dGC can detect slow neural processes, which operate at a timescale slower than TR, in fMRI data.

How might such slow timescales, orders of magnitude slower than spike times and membrane time constants, arise in fMRI data? To answer this question, we availed of established results in random matrix theory. Connectivity in randomly connected E-I networks of neurons can produce slow timescales, without fine-tuning of network parameters (Rajan and Abbott, 2006; Ganguli et al., 2008; Friston et al., 2014). We modeled sparse, random, net excitatory connectivity in a small network of (N=100) neurons with connection parameters drawn from previous studies (ED Figure 3-1B; Markram, 2000; Holmgren et al., 2003; Ganguli et al., 2008). The eigen spectrum of the network revealed that each network exhibited one eigenvalue close to zero corresponding to a slow timescale (~1000 ms or greater, Fig. 3B bottom left);the latter constitutes an emergent timescale associated with the dominant eigenmode that is a property of network connectivity(Methods).

We modeled nine such networks, organized into three, non-interacting, clusters(Fig. 3B top right): a) a cluster with a purely feedforward connection across two networks, b) a cluster with recurrent excitatory (E-E) feedback connections among two networks and c) a cluster with recurrent excitatory-inhibitory (E-I) feedback connections among two networks. In each case, connectivity across networks was mediated by a small proportion (5%) of neurons in each network (parameters in ED Figure 3-1B). This configuration mimics "small-world" connectivity in brain networks (Bassett and Bullmore, 2006), with locally-connected brain regions interacting through sparse, long-range connections (Sporns et al., 2004). The eigenspectra revealed that dynamics in all clusters operated at timescales of around 6000 ms, comparable to or slower than the individual network timescales (Fig. 3B bottom right). To simulate fMRI data we averaged the activity across all 100 neurons in each network and convolved it with a canonical HRF. As before, these nine timeseries were then sampled at various sampling intervals, including a 750 ms interval mimicking the scan TR, and analyzed with GC to detect significant connections.

iGC and dGC identified complementary aspects of connectivity with these simulated data (Fig. 3C; Statistical Table^{m}). iGC robustly identified feedforward and excitatory (E-E) feedback connections. dGC also estimated these connections, albeit with the following differences: First, in the feedforward network dGC occasionally identified a spurious connection, albeit much weaker in magnitude, in the direction opposite to the true connection (Fig. 3C, left column, red dashed line). Second, when the E-E feedback connections were precisely balanced in strength (symmetric), dGC also failed to identify the connection reliably (ED Figure 3-2A). Yet, when these connections were of different strengths dGC reliably identified both connections, and their relative strengths (Fig. 3C, middle column, red). In contrast, when the connections were of different signs (E-I feedback) dGC robustly identified both connections, whereas iGC failed to reliably detect this connection (Fig. 3C, right column, blue). Yet, taken together, iGC and dGC identified all three connection types reliably.

Next, we compared the efficacy of connectivity estimation with partial correlations (PC). While PC robustly identified both feedforward and feedback E-E connections (Fig. 3C left and middle columns, black), it, surprisingly, failed to estimate feedback E-I connections, particularly when these were balanced in strength (Fig. 3C right column, black). When the E and I connection strengths were not balanced, but were strongly biased in favor of the E or the I connection, PC estimates varied with the sign of the more dominant connection (ED Figure 3-2B, right top). These results generalize beyond these particular simulations; in the ED Mathematical Note, Sections S2 and S3, we identify, analytically, network configurations for which PC estimates systematically deviates from ground-truth connectivity. We generated data with a seven node network, whose dynamics were described by a multivariate, autoregressive process. We systematically varied the covariance of the residuals of nodes 1 and 2 in the MVAR model (Υ), which is a key factor in determining iGC magnitude (ED Mathematical Note, Section S3, equations 11 and 21). Next, we computed the covariance between the residuals (K) in the regression of activities of nodes 1 and 2 against all other nodes (controlling variables), which is a key factor in determining PC magnitude. Although connectivity estimates based on iGC and PC were highly correlated, PC estimates systematically deviated from iGC estimates in value (ED Figure 3-2C left). In fact, for iGC covariance (Υ) values ranging from -0.3 to 0.0, indicating weak inhibitory functional connectivity, PC covariance (K) values were positive, ranging from 0 to 0.3, indicating excitatory functional connectivity (ED Figure 3-2C right, open squares). For these configurations, therefore, PC connectivity deviated systematically from ground-truth. The analytical relationship between PC connectivity and iGC connectivity explains this pattern of systematic deviations (ED Mathematical Note, Section S3, equation 23). Briefly, the relationship indicates that PC reflects a mixture of instantaneous and lagged connectivity rather than solely instantaneous interactions. Removing lagged interactions restores the identity between iGC and PC (ED Figure 3-2C right, open circles), as evidenced by setting the coefficients of the AR matrix to zero (ED Mathematical Note, Section S3, equation 11). These results highlight caveats with employing zero-lag correlation measures, like partial correlations, as compared to lag-based measures, like GC, for estimating connectivity with neural timeseries.

Taken together, these results indicate that instantaneous and lag-based connectivity measures can reveal complementary aspects of brain connectivity. In addition, the results challenge the notion that correlation-based measures, like PC, should be favored over lag-based measures, like dGC for measuring functional connectivity in the brain (Smith et al., 2011). Rather, the strengths and weaknesses of each measure (GC and PC) must be recognized when seeking to apply these to brain imaging data.

Identifying a cognitive core system based on GC connectivity

Our classification analyses and simulations suggested that iGC and dGC reliably recover task-specific brain networks, the latter when slow-timescale processes occur within the network. We asked whether iGC and dGC connectivity merely reflected reliable statistical patterns of brain activity, or whether it would be relevant for understanding the nature of information flow in the brain, and its relationship to behavior. To answer this question, we investigated whether each measure would identify brain networks with consistent outflow and inflow hubs across tasks.

Prior to analysis of real data, we validated RFE by applying it to estimate connectivity differences in two simulated networks (Fig. 4A,B). RFE accurately identified connections that differed in simulation ground-truth: specifically, differences in fast timescale connections were reliably identified by iGC, and in slow timescale connections by dGC (Fig. 4B bottom). We also tested whether dGC and iGC would be able to accurately identify differences in directed information flow among network configurations, even with systematic differences in hemodynamic lags among network nodes. For this we estimated GC for 100 simulated participants with the same two “ground-truth” network configurations (as shown in Fig. 4A), except with four different scenarios of hemodynamic lag differences (Methods): a) same mean HRF onset (μ_L= 3s) across all nodes; b) source node hemodynamic response function (HRF) onset lagging the destination node by 1s ($\mu_{L\text{-}src} > \mu_{L\text{-}dst}$); c) source node HRF onset leading destination node by 1s ($\mu_{L\text{-}src} > \mu_{L\text{-}dst}$); and d) mixed latencies of lead and lag such that 50% of simulated participants had the source node lagging the destination node and vice versa for the remaining 50% simulated participants. We performed these simulations by sampling the onset latency for each participant from a normal distribution, with standard deviations (σ_L) ranging from 0-1s (in steps of 0.2s) across simulations (ED Figure 3-2D-E). The relative magnitudes of these HRF onset latency differences, and their standard deviations, are comparable to their magnitudes observed in human data (Chang et al., 2008). RFE was then used to identify the most discriminative connections between the two networks.

First, we observed that across the different onset latency scenarios, GC connection strength magnitude generally decreased with increasing σ_L values (ED Figure 3-2D); an interesting exception was iGC connection strengths when source HRF onset led the destination HRF (case c, above; ED Figure 3-2D; lower row, dark blue curves). For sub-network ABC, with fast (50 ms) timescales, dGC revealed the correct directionality of connectivity (positive ΔdGC; ED Figure 3-2D, upper row) consistently in only one of the four cases (case c), when the source node onset systematically led the destination node (ED Figure 3-2D, upper row, odd columns: dark blue curves). On the other hand, for sub-network DEF, with slow (1000 ms) timescales, dGC revealed the correct directionality of connectivity in three of the four cases (ED Figure 3-2D, upper row, even columns: dark blue, light blue and black curves); all except case b, when the source node onset systematically lagged the destination node (ED Figure 3-2D, upper row, red).

In line with these results, RFE with dGC features correctly identified directionality of the most discriminative connections in no case for the fast sub-network (ED Figure 3-2E, rows 1-2, ABC sub-network), but correctly identified the directionality of these connections in three out of four cases for the slow sub-network (ED Figure 3-2E, rows 1-2, DEF sub-network). RFE with iGC features, identified maximally discriminative connections (albeit not their directionality) in all cases (ED Figure 3-2E, rows 3-4). Thus, RFE based on dGC and iGC accurately identified the relevant connections, but not always their directionality, even when systematic variations in hemodynamic lag occurred across regions. Taken together, these results indicate that fMRI-GC can identify differences in connectivity at slow timescales despite systematic differences and heterogeneity in HRF onset latencies across brain regions.

Next, with the real fMRI (HCP) data, we sought to identify a common core of “task-generic” connections across cognitive tasks. For this, we applied a feature selection approach – recursive feature elimination (Methods) – a technique that identifies a minimal set of features that provide maximal cross validation accuracy (generalization performance; Guyon and Elisseeff, 2003). We applied RFE to classify tasks versus resting state; we chose these six tasks (all tasks except the motor task) as being the most likely to engage common cognitive control mechanisms (Fig. 4C). For these RFE analyses we employed a 14 network functional parcellation (Shirer et al., 2012), as it consistently gave good classification accuracies with both iGC and dGC connectivity (ED Figure 1-5B). Following RFE, we applied a binomial test across tasks (Methods) to identify a common core of task-generic connections, separately for iGC and dGC.

RFE identified distinct task-generic networks with iGC and dGC, which comprised of connections that distinguished a majority of tasks from resting state. The iGC task-generic network revealed a visuospatial network hub, which connected with the anterior salience, dorsal DMN, higher visual and posterior salience networks (Fig. 4D, right). The dGC task-generic network confirmed the hub-like connectivity of the visuospatial network but, in addition, revealed consistent directed information outflow from the visuospatial network to the other networks (Fig. 4D, left). In addition, dGC revealed consistent inflow into the higher-visual network across tasks, including from the visuospatial, right executive control, and auditory networks, consistent with the ability of top-down inputs from these networks to strongly modulate sensory encoding in higher visual cortex (Gilbert and Li, 2013). Finally, the higher-visual network projected consistently to the sensorimotor network, suggesting a final common pathway, across these tasks, for influencing behavior. Interestingly, the only network providing inflow into the visuospatial network hub was the anterior salience network, in line with a previous study that indicated a role for the salience network in controlling other task positive networks (Sridharan et al., 2008).

Similarly, we asked whether iGC and dGC could identify connections that were maximally discriminative across tasks (“task-discriminative” networks). Because some network connections may not be present for any task, task-discriminative connections are not simply the complement of the task-generic connections. As before we repeated the RFE analysis, but this time based on an n-way classification across the six tasks (all except the motor task, Methods), seeking to identify connections that discriminated each task, from each of the other five tasks (Fig. 4E).

This analysis identified iGC and dGC connections among the vast majority of networks as being important for discriminating among tasks. Specifically, with iGC, basal ganglia connectivity was the most task-discriminative whereas for dGC, visuospatial network inflow and language network outflow were among the most discriminative (Fig. 4F). Connections with the precunues were strongly discriminative across both iGC and dGC networks. Notable exceptions to these trends were the sensorimotor network and ventral default mode network (vDMN). The sensorimotor network exhibited very few task-discriminative connections based on iGC (1/13) and dGC (3/26), whereas the vDMN exhibited only (1/13) task discriminative connections based on iGC. We further observed that each task recruited a distributed pattern of connectivity across networks (Methods), which was sufficiently characteristic of each task to permit accurate classification (ED Figure 4-2A). We also correlated the beta weights of the 11 overlapping connections across iGC and dGC and found no significant correlations ($r=-0.18$, $p=0.59$). The results indicate that the task-discriminative information flow patterns, as measured by iGC or dGC connectivity, arise from distinct, distributed networks across the entire brain.

We also tested whether partial correlation (PC) would identify task-generic and task-discriminative connections that were more in line with those identified by iGC or dGC or both (ED Figure 4-2B-D). Both task-generic and task-discriminative connections identified with PC revealed significant overlap with both iGC (task-generic: 75% overlap, and task-discriminative: 65.2% overlap, $p<0.05$ randomization test) and dGC (task-generic: 100% overlap, and task-discriminative: 78.3% overlap, $p<0.05$). These findings are consistent with our theoretical result that PC connectivity reflects a mixture of iGC and dGC connectivity.

Predicting behavioral scores with GC connectivity

To address GC's relevance for understanding brain-behavior relationships we tested whether the strength of functional connections estimated with iGC and dGC could predict inter-individual variations in behavioral scores as measured by a standard cognitive battery (Methods; ED Figure 5-1).We employed a leave-one-out prediction analysis based on multilinear regression followed by robust correlations of predicted and observed scores (Fig. 5A; $p<0.05$ with Benjamini-Yekutieli correction; Methods).

Both iGC and dGC predicted key behavioral scores (Fig. 5C; Statistical Table[n]). Several scores were predicted uniformly well by iGC across tasks (Fig. 5B, right; Fig. 5C, bottom; ED Figure 5-2B). Scores of fluid intelligence (Penn progressive matrices), spatial orientation (Penn line orientation test), grip strength, endurance, and language (picture-vocabulary and reading; Fig. 5B right), were all well predicted by iGC .(ED Figure 5-2B, r:0.104 - 0.363; p<0.01). On the other hand, dGC-based predictions were more selective, in that several behavioral scores were best predicted by dGC based on specific tasks alone (Fig. 5B, left; Fig. 5C, top; ED Figure 5-2A). For instance, dGC in the gambling task alone predicted self-report scores of fear (r=0.139, p<0.001) and dGC in the motor task alone predicted median reaction time in the fluid intelligence test (r=0.123, p<0.001) and self-reported scores of perceived emotional support (r=0.113, p<0.001). In addition, dGC in the working memory task predicted a range of scores in the "cognition" category including list sorting (Fig. 5B, left,pink; r=0.119, p=0.000), fluid intelligence, picture discrimination speed (Fig. 5C top, ED Figure 5-2A).

Similarly, we employed PC functional connection strengths as features for predicting inter-individual differences in behavioral scores. We observed that 129 behavioral scores were successfully predicted based on PC connectivity (ED Figure 5-2C, following BY correction for multiple comparisons), as compared with 39 scores based on dGC connectivity (ED Figure 5-2A) and 92 scores based in iGC connectivity (ED Figure 5-2B). Around 54% of the scores predicted well by PC (70/129) were also predicted well by either dGC or iGC. On the other hand, behavioral scores were predicted well by PC, but not by GC included reaction times in the Penn word memory test and Penn emotion recognition test, as well as several scores of the language task (ED Figure 5-2C).

Next, we compared the connection features that led to successful predictions based on GC and PC. For this, we z-scored the connection strengths (individually) and repeated the prediction process (Methods) separately with dGC features, iGC features and PC features derived from each of the 7 tasks. 17 of these predictions were significant (following BY correction) across all three connectivity features (ED Figure 5-2A,B,C). We then correlated the beta weights for each entry of the iGC matrix, with those of the PC matrix, across these 17 predictions. For dGC, the upper and lower triangular portions of the matrix were correlated separately, with the corresponding PC connection weights. We then computed the mean correlation (r) values across all 91 features (iGC versus PC) and 182 features (dGC lower and upper matrix versus PC).

We observed an interesting dissociation between PC, iGC and dGC. Connection features that were relevant for behavioral predictions with PC overlapped highly with iGC features, but not with dGC features (PC vs. iGC: r=0.39±0.02, mean±std; PC vs. dGC: r=0.03±0.02, $p<0.001$, ranksum test). The results provide further empirical evidence for a clear distinction between connectivity computed with instantaneous (PC, iGC) and lag-based (dGC) measures.

Finally, we tested whether GC connectivity could predict a combined set of behavioral scores unique to each subject. For this, we created a vector of all independent behavioral scores (composite score; Methods), and confirmed that this composite behavioral score uniquely identified each subject in the database, as evidenced by the highest values along the main diagonal of the inter-subject correlation matrix (Fig. 5D top). Following this, we performed the leave-one-out prediction, as before, except that we used dGC and iGC connectivity features from two of the tasks alone (working memory and relational; also see ED Figure 5-2D). We then tested whether each subject's predicted composite score would correlate best with her/his own observed composite scores. Although we did not observe the highest correlation values consistently along the main diagonal, the distribution of correlation coefficients along the diagonal were significantly different (and higher) than the distribution of off-diagonal correlation coefficients (Fig. 5D bottom; $p<10^{-15}$, Kolmogorov-Smirnov test; Statistical Table^{o}). Inter-individual variation GC connectivity, therefore, contained sufficient information to accurately identify subject-specific behavioral scores in this cohort of subjects.

In summary, the ability to successfully predict subject-specific behavioral scores suggests that GC functional connectivity is relevant for understanding brain-behavior relationships. Moreover, connection features that were relevant for behavioral predictions with PC overlapped highly with iGC, but not with dGC, thereby validating our simulation results regarding the complementarity of iGC and dGC connectivity estimates.

Table 1. Statistical Table

	Figure	Comparison	Type of Test	Statistic	Confidence Interval or Power
a	1B	Rest vs working memory best iGC classification accuracy value	Binomial test	Clopper-Pearson confidence intervals	[96.3 - 98.0%]
b		Rest vs working memory best dGC classification accuracy value			[90.5 - 93.2%]
c		Rest vs Task maximum classification accuracy (each bar) vs. chance	Permutation test	p value	p <0.001
d	1F	n-way task classification maximum iGC accuracy value	Binomial test	Clopper-Pearson confidence intervals	[73.3%-75.4%]
e		n-way task classification maximum dGC accuracy value			[46.4%-48.7%]
f		n-way task classification maximum accuracy values (each bar) vs. chance	Permutation test	p value	p <0.001
g	1G	Subtask classification maximum accuracies (each bar) vs. chance	Permutation test	p value	p <0.05
h	1H (left)	Rest vs Task dGC classification accuracies with 2x, and 3x sampling rate (vs 1x)	Wilcoxon one-tailed signed rank	p value	2x: p=0.02; 3x: p=0.06
i		Rest vs Task iGC classification accuracies with 2x, and 3x sampling rate (vs 1x)			2x: p=0.01; 3x: p=0.01
j	2C	ZCA, GEV classification accuracy values with dGC and iGC vs chance	Permutation test	p value	p <0.001
k	2D	Rest vs Task unweighted dGC classification accuracy value (each bar) vs chance	Permutation test	p value	p <0.001
l	3A	Each dGC, iGC, PC matrix connectivity value bound with black square vs corresponding null distribution	Phase-scrambling	p value	p <0.05
m	3C	Each dGC, iGC, PC matrix connectivity value bound with black square vs corresponding null distribution of phase-scrambled surrogates	Benjamini-Hochberg correction	p value	p <0.05
n	5C	Each prediction corr value with filled circle in stem plot	Benjamini-Yekutieli correction	p value	p <0.05
o	5D	Correlation coefficients between observed and predicted composite scores, for the same subject vs across different subjects	Kolmogorov-Smirnov test	p value	p <0.001

Discussion

Neural processes in the brain range from the timescales of milliseconds, for extremely rapid processes (e.g. sound localization), to timescales of several seconds to minutes, for processes that require coordination across diverse brain networks (e.g. when having a conversation) and hours to days for processes that involve large-scale neuro-plastic changes (e.g. when learning a new language). Coordinated activity among brain regions that mediate each of these cognitive processes should manifest in the form of functional connectivity among these regions at the corresponding timescales. Our results indicate that applying Granger-Geweke Causality (GC) with fMRI data permits estimating behaviorally relevant functional connectivity at a timescale corresponding to the sampling rate of fMRI data (seconds).

The application of GC to neuroscience is a contentious topic, for a variety of reasons (Chang et al., 2008; Friston et al., 2013; Seth et al., 2013; Wen et al., 2013; Stokes and Purdon, 2017). One particular challenge stems from the use of the word "causality": the notion of causality in GC is different from the notion of interventional causality (Pearl, 2011). Our use of the term Granger causality, here, purely reflects its application as a marker of information flow among brain networks (Roebroeck et al., 2005; Seth et al., 2013), and is not meant to indicate causality in a physical, interventional sense.

With this understanding, our results contain three key insights. First, we show that, either iGC or dGC connectivity suffices to reliably classify task-specific cognitive states with superlative accuracies (Fig. 1B). Instantaneous and directed GC – both measures of conditional linear dependence and feedback (Geweke, 1984) – were able to robustly estimate task-specific functional interactions even with slowly sampled fMRI data. Our simulations suggest that GC connectivity is relevant for estimating slow, emergent interactions among brain networks (Chang et al., 2008; Smith et al., 2011; Friston et al., 2013; Seth et al., 2013; Wen et al., 2013).

Second, we show that functional connections identified by iGC and dGC carry complementary information, both in simulated and in real fMRI recordings, and we demonstrate key caveats with employing correlation-based measures of functional connectivity like partial correlations, despite superior classification accuracies with these latter measures. First, PC fails to correctly infer reciprocal excitatory-inhibitory interactions, which can be accurately inferred with lag-based methods like dGC. Second, PC may yield incorrect estimates of functional connectivity that do not match ground truth (ED Figure 3-2C). In particular, when the data are well described by an autoregressive model framework our results suggest that instantaneous connectivity measures, like iGC, provide more accurate descriptions of functional connectivity than PC. Third, even with data completely purged of partial correlations, dGC connectivity was sufficient to classify task-specific cognitive states (Fig. 2C). In fact, unweighted directed connectivity alone sufficed to produce accurate classification at accuracies significantly above chance (Fig. 2D). These results indicate that information flow mapped by GC connectivity can be complementary to that of PC, and highlights the need for examining diverse measures, both instantaneous and lag-based, to obtain a complete picture of functional connectivity in the brain.

Third, differences in inter-individual iGC and dGC connectivity were able to explain inter-individual variation in behavioral scores on various cognitive tasks, and to identify an individual-specific composite marker of behavioral scores, with high accuracies. Perhaps because these behavioral scores were acquired in a separate testing session outside the scanning session (Barch et al., 2013), the effect sizes were small (albeit significant), and comparable to effect sizes in previous studies employing large, heterogeneous subject cohorts (Smith and Nichols, 2018). Nevertheless, the results suggest that GC connectivity was both individual-specific, and stable over timescales exceeding the scan session, to permit accurate prediction. Moreover, in our analysis, each subject's behavioral score was predicted based on her/his GC connectivity, whereas the regression beta weights – representing the relationship between GC connectivity and behavior – were computed from the population of all subjects excluding that subject (Fig. 5A). Successful predictions, therefore, indicate a consistent mapping between GC connectivity and behavioral scores across the population of subjects. These findings complement recent results showing that dynamic, resting-state functional connectivity, based on correlations, can explain significant variance in human behavioral data (Liégeois et al., 2019), and indicate the relevance of lag-based connectivity measures for understanding brain-behavior relationships.

Does GC's discriminatory power rely on directed functional connectivity in the underlying neural response or systematic distortions of this connectivity induced by subsampling (Seth et al., 2013) and hemodynamic filtering (Lin et al., 2009; Solo et al., 2018)? While our findings cannot completely rule out the latter hypothesis, we address, next, three key caveats raised by previous studies for estimating functional connectivity with fMRI-GC, and argue why our results support the former hypothesis.

First, several studies have shown that sub-sampling of neural time series, at the scale of fMRI TR, renders functional connections undetectable with GC (Lin et al., 2009; Smith et al., 2011; Seth et al., 2013, 2015). In these studies, GC was estimated with simulated fMRI time series, sampled at an interval (TR) of seconds, and failed to recover underlying neural interactions, which occur at millisecond timescales (e.g. Smith et al., 2011). However, these claims depended strongly on the nature and timescale of the connectivity in the networks employed in these simulations. For instance, a widely cited study (Smith et al., 2011) employed purely feedforward connectivity matrices with a 50 ms neural timescale in their simulations, and argued that functional connections are not reliably inferred with GC applied to simulated fMRI data. In addition to being neurally implausible, such purely feedforward network configurations yield eigenmodes whose slowest timescales are identical with the timescales of individual nodes (Sundaresan et al., 2017). Therefore, such a configuration rendered lag-based measures like GC, irrelevant for estimating neural interactions from slowly sampled fMRI data (Smith et al., 2011; Seth et al., 2013). Furthermore, such connectivity precludes the occurrence of slower, behaviorally relevant timescales of seconds, which readily emerge in the presence of feedback connections, both in simulations (Rajan and Abbott, 2006; Ganguli et al., 2008) and in the real brain (Friston et al., 2014; Runyan et al., 2017; Vidaurre et al., 2017). Our simulations show that slow timescale interactions emerge in networks with sparse, random, net excitatory connectivity, mimicking connectivity in the neocortex (Markram, 2000; Holmgren et al., 2003; Ganguli et al., 2008). While earlier studies have employed large-scale, biologically plausible models (Deco et al., 2009; Krishnan et al., 2018) to demonstrate the emergence of slow (<0.1 Hz) emergent functional interactions among brain networks, our results build upon these previous findings and show that such emergent, , functional interactions at slow timescales can be readily inferred from simulated fMRI data with GC. In fact, GC connectivity continued to robustly classify distinct task states even when data were sampled at 2x or 3x the original sampling interval of the fMRI data. Thus, while it is likely that GC applied to fMRI data is unable to detect connections at timescales faster than TR, our results show that sufficient distinguishing information occurs in slow-timescale connections to enable accurate inter-task classification with fMRI-GC. Sub-sampling alone may also produce spurious GC causality. The precise conditions under which spurious GC arises for continuous time vector autoregressive processes, possibly with time delay in between the nodes, is an area of active research, and must be addressed in future studies (Lin et al., 2009; Barnett and Seth, 2017).

Second, previous studies have shown that systematic differences in hemodynamic (HRF) lags (e.g. time to onset, or time to peak) among brain regions may produce spurious dGC estimates (Friston, 2009; Seth et al., 2013; Solo et al., 2018). With simulations we demonstrated that fMRI-GC could identify differences in slow-timescale network connectivity, despite systematic differences and heterogeneity in HRF onset latencies across nodes (ED Figure 3-2D-E). In all cases, applying recursive feature elimination with either dGC or iGC features identified the precise subset of connections that distinguished distinct network configurations. In a majority of cases, dGC also correctly identified the directionality of these connections. In our simulations, the only scenario in which dGC features failed to identify the directionality of connections correctly, was when the onset latency in the “destination” nodes were biased to be systematically earlier than those in the “source” nodes. Nevertheless, in the real data it is unlikely that systematic inter-regional HRF differences contributed to the observed superior classification accuracies. Variations in HRF delays would indeed confound dGC connectivity estimates – if they occurred consistently between brain regions across subjects and tasks (e.g. ED Figure 3-2D, red curves). Yet, such a scenario cannot account for the high classification accuracies among tasks and sub-tasks based on dGC connectivity alone. In other words, even if HRF latency differences did systematically bias dGC connectivity estimates, these estimates were sufficiently and reliably different across task cognitive states to permit accurate classification among them. To our knowledge, our study provides the first direct experimental validation of GC networks’ ability to distinguish cognitive states, as a marker of their potential utility for identifying these states. Moreover, network properties of key regions identified with fMRI-GC were consistent with their known functional properties of these regions. For instance, dGC identified the visuospatial network as an information outflow hub, across all six cognitive tasks (Fig. 4D left). The visuospatial network comprises frontal cortex regions, including the frontal eye field, as well as posterior parietal cortex, which are both widely implicated in visuospatial attention control (Corbetta et al., 1998; Behrmann et al., 2004; Schall, 2004; Thompson and Bichot, 2004). In addition, the only network that provided task-generic incoming connections to the visuospatial network was the anterior salience network comprising the anterior fronto-insular cortex and the anterior cingulate cortex (Dosenbach et al., 2008; Chen et al., 2013), regions implicated in feature-based attention and the suppression of distractors (Li et al., 2018). Information outflow from these key networks identified by dGC is consistent with their role in attention and executive control.

Third, simulations and theoretical results indicate that scanner noise can degrade, or even obliterate GC connectivity estimates (Seth et al., 2013).On the other hand, our classification accuracies suggest that GC estimates were sufficiently robust to scanner noise to permit accurate task and sub-task classification in these data. In fact, we show that averaging dGC connectivity across as few as 5 subjects' data improves classification accuracy to over 95% for nearly all tasks (Fig. 1D). Such superlative classification accuracies are unlikely to have occurred if scanner noise were to significantly degrade GC estimates.

In sum, these results suggest that lag-based methods like GC, applied to fMRI data, can be used infer slow functional interactions in the brain. While the directionality of interactions measured by GC may need to be interpreted with care (Seth et al., 2015; Solo et al., 2018), our results suggest that fMRI-GC may be useful for formulating hypothesis about the role of particular brain regions in providing "top-down" control signals, for modulating activity in other brain regions (Sridharan et al., 2008; Ryali et al., 2011), as well as for investigating the nature of information flow in cortical microcircuits with slow sampling rate techniques, such as calcium imaging (Fallani et al., 2015).The causal role of these brain regions in behavior can then be directly tested with interventional approaches such as transcranial magnetic stimulation, optogenetic inactivation or by examining patient populations with lesions in specific brain regions (Gaillard et al., 2006). Such a systematic analysis will pave the way for a mechanistic understanding of how flexible functional interactions among brain regions mediate complex cognitive behaviors.

References

Aiken, L. S., West, S. G. and Pitts, S. C. (2003) ‘Multiple Linear Regression’, *Handbook of Psychology*. doi: 10.1051/eas/1466005.

Arbabshirani, M. R. *et al.* (2017) ‘Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls’, *NeuroImage*, 145, pp. 137–165. doi: https://doi.org/10.1016/j.neuroimage.2016.02.079.

Barch, D. M. *et al.* (2013) ‘Function in the human connectome: task-fMRI and individual differences in behavior’, *NeuroImage*. Elsevier, 80, pp. 169–189.

Barnett, L., Barrett, A. B. and Seth, A. K. (2018) ‘Misunderstandings regarding the application of Granger causality in neuroscience’, *Proceedings of the National Academy of Sciences*. National Academy of Sciences. doi: 10.1073/pnas.1714497115.

Barnett, L. and Seth, A. K. (2014) ‘The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference’, *Journal of Neuroscience Methods*, 223, pp. 50–68. doi: https://doi.org/10.1016/j.jneumeth.2013.10.018.

Barnett, L. and Seth, A. K. (2017) ‘Detectability of Granger causality for subsampled continuous-time neurophysiological processes’, *Journal of Neuroscience Methods*, 275, pp. 93–121. doi: https://doi.org/10.1016/j.jneumeth.2016.10.016.

Bassett, D. S. and Bullmore, E. (2006) ‘Small-world brain networks’, *Neuroscientist*. doi: 10.1177/1073858406293182.

Bastos, A. M. *et al.* (2015) ‘Visual areas exert feedforward and feedback influences through distinct frequency channels’, *Neuron*. Elsevier, 85(2), pp. 390–401.

Behrmann, M., Geng, J. J. and Shomstein, S. (2004) ‘Parietal cortex and attention’, *Current Opinion in Neurobiology*. doi: 10.1016/j.conb.2004.03.012.

Buckner, R. L. *et al.* (2009) ‘Cortical Hubs Revealed by Intrinsic Functional Connectivity: Mapping, Assessment of Stability, and Relation to Alzheimer’s Disease’, *Journal of Neuroscience*. doi: 10.1523/JNEUROSCI.5062-08.2009.

Chang, C., Thomason, M. E. and Glover, G. H. (2008) ‘Mapping and correction of vascular hemodynamic latency in the BOLD signal’, *NeuroImage*, 43(1), pp. 90–102. doi: https://doi.org/10.1016/j.neuroimage.2008.06.030.

Chen, A. C. *et al.* (2013) ‘Causal interactions between fronto-parietal central executive and default-mode networks in humans’, *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1311772110.

Corbetta, M. *et al.* (1998) ‘A common network of functional areas for attention and eye movements’, *Neuron*. doi: 10.1016/S0896-6273(00)80593-0.

David, O. *et al.* (2008) ‘Identifying Neural Drivers with Functional MRI: An Electrophysiological Validation’, *PLOS Biology*. Public Library of Science, 6(12), pp. 1–15. doi: 10.1371/journal.pbio.0060315.

Deco, G. *et al.* (2009) ‘Key role of coupling, delay, and noise in resting brain fluctuations’, *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 106(25), pp. 10302–10307. doi: 10.1073/pnas.0901831106.

Dhamala, M., Rangarajan, G. and Ding, M. (2008) ‘Analyzing information flow in brain networks with nonparametric Granger causality’, *NeuroImage*, 41(2), pp. 354–362. doi: https://doi.org/10.1016/j.neuroimage.2008.02.020.

Ding, M. and Wang, C. (2014) ‘Analyzing MEG Data with Granger Causality: Promises and Pitfalls’, in Supek, S. and Aine, C. J. (eds) *Magnetoencephalography: From Signals to Dynamic Cortical Networks*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 309–318. doi: 10.1007/978-3-642-33045-2_15.

Dosenbach, N. U. F. *et al.* (2008) ‘A dual-networks architecture of top-down control’, *Trends in Cognitive Sciences*. doi: 10.1016/j.tics.2008.01.001.

Van Essen, D. C. *et al.* (2012) ‘The Human Connectome Project: A data acquisition perspective’, *NeuroImage*, pp. 2222–2231. doi: 10.1016/j.neuroimage.2012.02.018.

Fallani, F. D. V. *et al.* (2015) ‘Hierarchy of Neural Organization in the Embryonic Spinal Cord: Granger-Causality Graph Analysis of In Vivo Calcium Imaging Data’, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 23(3), pp. 333–341. doi: 10.1109/TNSRE.2014.2341632.

Fox, M. D. *et al.* (2005) ‘From The Cover: The human brain is intrinsically organized into dynamic, anticorrelated functional networks’, *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.0504136102.

Friston, K. (2009) ‘Causal Modelling and Brain Connectivity in Functional Magnetic Resonance Imaging’, *PLOS Biology*. Public Library of Science, 7(2), pp. 1–6. doi: 10.1371/journal.pbio.1000033.

Friston, K. J. *et al.* (2014) ‘On nodes and modes in resting state fMRI.’, *NeuroImage*, 99(100), pp. 533–547. doi: 10.1016/j.neuroimage.2014.05.056.

Friston, K., Moran, R. and Seth, A. K. (2013) ‘Analysing connectivity with Granger causality and dynamic causal modelling’, *Current Opinion in Neurobiology*, pp. 172–178. doi: 10.1016/j.conb.2012.11.010.

Gaillard, R. *et al.* (2006) ‘Direct Intracranial, fMRI, and Lesion Evidence for the Causal Role of Left Inferotemporal Cortex in Reading’, *Neuron*. doi: 10.1016/j.neuron.2006.03.031.

Ganguli, S. *et al.* (2008) ‘One-dimensional dynamics of attention and decision making in LIP’, *Neuron*. Elsevier, 58(1), pp. 15–25.

Gel’fand, I. M. and Yaglom, A. M. (1959) ‘Calculation of the amount of information about a random function contained in another such function’, *American Mathematical Society Translations*, 12(1), pp. 199–246.

Gershon, R. C. *et al.* (2013) ‘NIH Toolbox for Assessment of Neurological and Behavioral Function’, *Neurology*. doi: 10.1212/WNL.0b013e3182872e5f.

Geweke, J. (1982) ‘Measurement of linear dependence and feedback between multiple time series’, *Journal of the American Statistical Association*. Taylor & Francis, 77(378), pp. 304–313.

Geweke, J. F. (1984) ‘Measures of conditional linear dependence and feedback between time series’, *Journal of the American Statistical Association*. Taylor & Francis, 79(388), pp. 907–915.

Gilbert, C. D. and Li, W. (2013) ‘Top-down influences on visual processing’, *Nature Reviews Neuroscience*. doi: 10.1038/nrn3476.

Glasser, M. F. *et al.* (2013) ‘The minimal preprocessing pipelines for the Human Connectome Project’, *NeuroImage*. Elsevier, 80, pp. 105–124.

Guyon, I. and Elisseeff, A. (2003) ‘An Introduction to Variable and Feature Selection’, *Journal of Machine Learning Research (JMLR)*. doi: 10.1016/j.aca.2011.07.027.

Holmgren, C. *et al.* (2003) ‘Pyramidal cell communication within local networks in layer 2/3 of rat neocortex.’, *The Journal of physiology*. doi: 10.1113/jphysiol.2003.044784.

Karampatziakis, N. and Mineiro, P. (2014) ‘Discriminative Features via Generalized Eigenvectors’, *Proceedings of The 31st International Conference on Machine Learning*.

Kessy, A., Lewin, A. and Strimmer, K. (2018) ‘Optimal Whitening and Decorrelation’, *American Statistician*. doi: 10.1080/00031305.2016.1277159.

Knösche, T. and Tittgemeyer, M. (2011) ‘The Role of Long-Range Connectivity for the Characterization of the Functional–Anatomical Organization of the Cortex’, *Frontiers in Systems Neuroscience*, 5, p. 58. doi:

10.3389/fnsys.2011.00058.

Krishnan, G. P., González, O. C. and Bazhenov, M. (2018) ‘Origin of slow spontaneous resting-state neuronal fluctuations in brain networks’, *Proceedings of the National Academy of Sciences*. National Academy of Sciences. doi: 10.1073/pnas.1715841115.

Li, V. *et al.* (2018) ‘Gain control explains the effect of distraction in human perceptual, cognitive, and economic decision making’, *Proceedings of the National Academy of Sciences*. National Academy of Sciences. doi: 10.1073/pnas.1805224115.

Liang, X. *et al.* (2012) ‘Effects of Different Correlation Metrics and Preprocessing Factors on Small-World Brain Functional Networks: A Resting-State Functional MRI Study’, *PLOS ONE*. Public Library of Science, 7(3), pp. 1–16. doi: 10.1371/journal.pone.0032766.

Liégeois, R. *et al.* (2019) ‘Resting brain dynamics at different timescales capture distinct aspects of human behavior’, *Nature Communications*, 10(1), p. 2317. doi: 10.1038/s41467-019-10317-7.

Lin, F.-H. *et al.* (2009) ‘Dynamic Granger-Geweke causality modeling with application to interictal spike propagation.’, *Human brain mapping*, 30(6), pp. 1877–1886. doi: 10.1002/hbm.20772.

Markram, H. (2000) ‘Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex’, *Science*. doi: 10.1126/science.287.5451.273.

Marrelec, G. *et al.* (2006) ‘Partial correlation for functional brain interactivity investigation in functional MRI’, *NeuroImage*. doi: 10.1016/j.neuroimage.2005.12.057.

Nolte, G. *et al.* (2008) ‘Robustly Estimating the Flow Direction of Information in Complex Physical Systems’, *Phys. Rev. Lett.* American Physical Society, 100(23), p. 234101. doi: 10.1103/PhysRevLett.100.234101.

Ojala, M. and Garriga, G. C. (2010) ‘Permutation tests for studying classifier performance’, *Journal of Machine Learning Research*, 11, pp. 1833–1863.

Pearl, J. (2011) *Causality: Models, reasoning, and inference, second edition*, *Causality: Models, Reasoning, and Inference, Second Edition*. doi: 10.1017/CBO9780511803161.

Penny, W. *et al.* (2007) *Statistical Parametric Mapping: The Analysis of Functional Brain Images*.

Power, J. D. *et al.* (2012) ‘Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion’, *Neuroimage*. Elsevier, 59(3), pp. 2142–2154. doi: https://doi.org/10.1016/j.neuroimage.2011.10.018.

Rajan, K. and Abbott, L. F. (2006) ‘Eigenvalue spectra of random matrices for neural networks’, *Physical Review Letters*. APS, 97(18), p. 188104.

Roebroeck, A., Formisano, E. and Goebel, R. (2005) ‘Mapping directed influence over the brain using Granger causality and fMRI’, *NeuroImage*. Elsevier, 25(1), pp. 230–242.

Runyan, C. A. *et al.* (2017) ‘Distinct timescales of population coding across cortex’, *Nature*. Nature Research, 548(7665), pp. 92–96.

Ryali, S. *et al.* (2011) ‘Multivariate dynamical systems models for estimating causal interactions in fMRI’, *NeuroImage*. Elsevier, 54(2), pp. 807–823.

Ryali, S. *et al.* (2012) ‘Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty’, *NeuroImage*, 59(4), pp. 3852–3861. doi: https://doi.org/10.1016/j.neuroimage.2011.11.054.

Schall, J. D. (2004) ‘On the role of frontal eye field in guiding attention and saccades’, in *Vision Research*. doi: 10.1016/j.visres.2003.10.025.

Seth, A. K., Barrett, A. B. and Barnett, L. (2015) ‘Granger Causality Analysis in Neuroscience and

Neuroimaging', *Journal of Neuroscience*. doi: 10.1523/JNEUROSCI.4399-14.2015.

Seth, A. K., Chorley, P. and Barnett, L. C. (2013) 'Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling', *NeuroImage*. Elsevier, 65, pp. 540–555.

Shirer, W. R. *et al.* (2012) 'Decoding subject-driven cognitive states with whole-brain connectivity patterns', *Cerebral Cortex*. Oxford Univ Press, 22(1), pp. 158–165.

Smith, S. M. *et al.* (2011) 'Network modelling methods for fMRI', *NeuroImage*. Elsevier, 54(2), pp. 875–891.

Smith, S. M. *et al.* (2012) 'The danger of systematic bias in group-level FMRI-lag-based causality estimation', *NeuroImage*, 59(2), pp. 1228–1229. doi: https://doi.org/10.1016/j.neuroimage.2011.08.015.

Smith, S. M. and Nichols, T. E. (2018) 'Statistical Challenges in "Big Data" Human Neuroimaging', *Neuron*, 97(2), pp. 263–268. doi: https://doi.org/10.1016/j.neuron.2017.12.018.

Solo, V. (2016) 'State-Space Analysis of Granger-Geweke Causality Measures with Application to fMRI.', *Neural computation*, 28(5), pp. 914–949. doi: 10.1162/NECO_a_00828.

Solo, V. *et al.* (2018) 'Connectivity in fMRI: Blind Spots and Breakthroughs', *IEEE transactions on medical imaging*, 37(7), p. 1537—1550. doi: 10.1109/tmi.2018.2831261.

Sporns, O. *et al.* (2004) 'Organization, development and function of complex brain networks', *Trends in Cognitive Sciences*. doi: 10.1016/j.tics.2004.07.008.

Sridharan, D., Levitin, D. J. and Menon, V. (2008) 'A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks', *Proceedings of the National Academy of Sciences*. National Acad Sciences, 105(34), pp. 12569–12574.

Stokes, P. A. and Purdon, P. L. (2017) 'A study of problems encountered in Granger causality analysis from a neuroscience perspective', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1704663114.

Sundaresan, M., Nabeel, A. and Sridharan, D. (2017) 'Mapping distinct timescales of functional interactions among brain networks', *31st Conference on Neural Information Processing Systems*.

Tavor, I. *et al.* (2016) 'Task-free MRI predicts individual differences in brain activity during task performance.', *Science (New York, N.Y.)*. doi: 10.1126/science.aad8127.

Thomas Yeo, B. T. *et al.* (2011) 'The organization of the human cerebral cortex estimated by intrinsic functional connectivity', *Journal of Neurophysiology*. Bethesda, MD, pp. 1125–1165. doi: 10.1152/jn.00338.2011.

Thompson, K. G. and Bichot, N. P. (2004) 'A visual salience map in the primate frontal eye field', *Progress in Brain Research*. doi: 10.1016/S0079-6123(04)47019-8.

Vidaurre, D., Smith, S. M. and Woolrich, M. W. (2017) 'Brain network dynamics are hierarchically organized in time', *Proceedings of the National Academy of Sciences*. National Acad Sciences, 114(48), pp. 12827–12832.

Vincent, J. L. *et al.* (2008) 'Evidence for a Frontoparietal Control System Revealed by Intrinsic Functional Connectivity', *Journal of Neurophysiology*. doi: 10.1152/jn.90355.2008.

Wen, X., Rangarajan, G. and Ding, M. (2013) 'Is Granger Causality a Viable Technique for Analyzing fMRI Data?', *PLOS ONE*. Public Library of Science, 8(7), pp. 1–11. doi: 10.1371/journal.pone.0067428.

Wilcox, R. R. (1994) 'The percentage bend correlation coefficient', *Psychometrika*. doi: 10.1007/BF02294395.

Figure Legends

Figure 1. Discriminating between task and resting state with instantaneous and directed GC networks.

A. Schematic of task state classification based on instantaneous (iGC) and directed (dGC) Granger-Geweke Causality with fMRI data from 1000 subjects (see text for details; IDs in ED Figure 1-3).

B. Two-way classification accuracies (leave-one-out) for each of seven tasks versus resting state based on GC. Red unfilled bars and blue filled bars: accuracies based on dGC and iGC features, respectively (task key in ED Figure 1-1). Error-bars: Clopper-Pearson binomial confidence intervals. Chance accuracy: 0.5 (not shown).

C. Two-way task versus resting state classification accuracies based on dGC (red dots) and iGC (blue - dots), as a function of number of task scan time points (volumes). Dashed lines: linear fits.

D. Two-way task versus resting state classification accuracies based on dGC after averaging dGC matrices over different numbers of subjects (x-axis). Each task is represented with a different color. Colored dashed lines: biexponential fits. Black dashed horizontal and vertical lines: 95% accuracy and n=5 subjects' average, respectively.

E. Two-way classification accuracies across each pair of tasks. Cells: classification accuracies for each pair of tasks based on dGC (lower triangular matrix) or iGC (upper triangular matrix). Diagonal cells: number of task scan timepoints. Highlighted cells: lowest (dashed-line border) and highest (solid-line border) accuracies achieved with dGC (red) and iGC (blue).

F. N-way classification accuracies among all seven tasks. Dashed line: chance accuracy (14.3%). Other conventions are the same as in panel B.

G. Two-way sub-task classification accuracies (descriptions in ED Figure 1-2) based on GC. ns.: accuracy not significantly above chance. Other conventions are the same as in panel B.

H. (Left) Two-way task versus resting state classification accuracies obtained with regional time series sub-sampled at 2x (filled symbols) and 3x (open symbols) of the TR (720 ms) (y-axis) plotted against accuracies obtained with the original data (1x, x-axis) for each of 7 tasks. Red: dGC, Blue: iGC. Dashed diagonal line: Line of equality (x=y). (Right) N-way classification accuracies among all seven tasks with data sampled at 1x, 2x, 3x of the original TR. Other conventions are the same as in panel F.

For panels B,E,F: accuracies correspond to highest values, across all parcellations tested, and hyperparameter optimization was done for panel B. For panels C,G,H: accuracies correspond to Shirer et al (2012) 14-network parcellation. For panel D: accuracies correspond to Shirer et al (2012) 90-node parcellation. Further details and control analyses are presented in ED Figures 1-4 to 1-7.

Figure 2. Classification accuracies with GC purged of instantaneous correlations.

A. Two-way task versus resting state classification accuracies, based on partial correlations (PC; grey unfilled bars). Numbers reported correspond to highest leave-one-out classification accuracies across parcellations, obtained with hyperparameter optimization. Corresponding accuracies for dGC (red dots) and iGC (blue dots) are shown for comparison. Other conventions are as in Fig. 1B.

B. Schematic illustrating procedure for purging data of instantaneous correlations. fMRI regional timeseries were purged of instantaneous correlations by either whitening the data with zero-phase component analysis (ZCA), separately for each task and resting state scan, or by projecting data into a space spanned by the generalized eigenvectors (GEV), common to both task and resting state scans. GC and PC were then estimated with the ZCA or GEV projections of the timeseries data, followed by classification analysis based on GC or PC connection strength as features.

C. (Top) Two-way task versus resting state classification accuracies following ZCA-based decorrelation. Gray circles: Classification accuracies based on PC. Other conventions are as in Figure 1B. Dashed line: chance accuracy (50%).(Bottom) Same as in top panel, but for classification following GEV-based decorrelation.

D. (Top) Schematic showing unweighted directed graph obtained from dGC; this digraph representation encodes only the dominant direction of connectivity, and not its magnitude. (Bottom) Two-way task versus resting state classification accuracies based on dGC digraph representations. Secondary ordinate (y-axis on the right): number of scan timepoints for each task.(Panels C-D). GC features were estimated with the Shirer et al (2012) 14-network parcellation.

Figure 3. Robustness of GC estimates depend on network timescales in simulated hemodynamic data.

A. (Top) Two-node networks with fast (50 ms; left) or slow (1000 ms; right) decay timescales of individual nodes (parameters in ED Figure 3-1A). Each subpanel shows ground truth connectivity either as a schematic (left) or connectivity matrix (right). In the matrix, a non-zero entry at cell (i, j) corresponds to a connection from node j (source) to node i (destination).(Bottom) dGC (red), iGC (blue), and PC (black) connection strengths as a function of sampling intervals. Filled circles and solid lines: Strengths of true connections and curve (biexponential) fits, respectively. Open circles and dashed lines: Strengths of spurious connections and curve fits, respectively. Dashed vertical line: Sampling interval of 750 ms, mimicking the TR of the fMRI data. Matrices to the right of each plot show GC connection strengths estimated at sampling interval of 750 ms. Black squares surrounding matrix cells denote significant connections (Methods). For iGC and PC (symmetric connectivity), only the lower triangular matrix is shown, for clarity.

B. (Top left) Schematic showing a cluster of neurons, each with timescale 50ms, connected with sparse, random, net excitatory connectivity. Matrix: Connectivity among the 100 neurons in a representative cluster. Red: excitatory connections; blue: inhibitory connections. Each such cluster forms one of the nine nodes in the simulated network. (Top right) Connectivity among the nine nodes in the network (parameters in ED Figure 3-1B). (Bottom left) Eigenspectrum (upper panel) of a representative 100 neuron cluster, showing one slow emergent timescale corresponding to the real-part of one eigenvalue close to zero. Histogram (lower panel) showing timescales of all eigenmodes, with the slowest eigenmode at >2000ms. (Bottom right) Eigenspectrum (upper panel) of sub-network DEF exhibits multiple slow emergent timescales. Histogram (lower panel) showing timescales of all eigenmodes, with three slow eigenmodes at ~1000-6000 ms.

C. Same as in A, but for simulated 9-node networks (parameters in ED Figure 3-1B). (Left) Sub-network ABC, (middle) sub-network DEF (see also ED Figure 3-2), (right) sub-network GHI. Other conventions are as in panel A.

Figure 4. Recursive feature elimination (RFE) identifies task-generic and task-discriminative networks based on GC connectivity.

A. Schematic showing two simulated networks each with fast (50 ms; ABC) and slow (1000 ms; DEF) sub-networks, with distinct connectivity patterns. Network activity was simulated for 375 seconds with a sampling interval of 5 ms, convolved with the hemodynamic response function and sub-sampled at 750 ms to yield 500 simulated time points.

B. (Top) RFE curves, with classification accuracy as a function of remaining features, for classification based on dGC (left) and iGC (right). (Bottom) Maximally discriminative features following RFE based on dGC (left) and iGC (right). Entries denote average beta weights across RFE iterations.

C. RFE curve for two-way classification of each of six tasks (all tasks except Motor) versus rest, based on dGC (top) and iGC (bottom). Color conventions are as in Figure 1D. Data points: RFE accuracies; solid lines: piecewise linear fits. Vertical dashed line: location of the elbow for each RFE curve.

D. Task-generic connections following task-versus-resting RFE, based on dGC (left) and iGC (right) features, using Shirer et al (2012) 14-network parcellation (details in ED Figure 4-1); each network is indicated with a different color and a label. Directed dGC connections are shown as tapered links, broad at the source node and narrow at the destination node. Undirected iGC connections are shown as bidirectional links between the respective pair of nodes. Colors of the connections represent the color of the destination node.

E. Same as in panel C, but for n-way classification across the six tasks. Color conventions are as in panel B.

F. Same as in panel D, but for task-discriminative connections (see also ED Figure 4-2), which maximally discriminated each task from the five others, following n-way RFE, based on dGC features (left) and iGC features (right). Other conventions are the same as in panel C.

Figure 5. GC connectivity explains inter-individual variations in behavioral scores.

A. (Left) Schematic of behavioral score prediction analysis. GC connectivity strengths for each task were used as independent factors to predict behavioral scores using linear regression with a leave-one-out approach. 51 different behavioral scores (descriptions in ED Figure 5-1) were predicted, compared against observed scores (upper right), and their correlation values plotted as a matrix (lower right).

B. Exemplar score predictions based on dGC (left panels) and iGC (right panels). In order (from left to right): List Sorting score predicted from Working memory task dGC connectivity Picture Vocabulary score from Language task dGC connectivity, Endurance score from Motor task iGC connectivity and Reading score from Language task iGC connectivity.

C. (Top) Prediction statistics for selected scores based on dGC connectivity (all scores shown in ED Figure 5-2). Correlation coefficients (r values) between the predicted and observed scores are plotted in the top half of each stem plot, and significance (p values) are plotted in the bottom half. Each score is denoted by a different color, and each sub-panel shows predictions based on GC connectivity for a different task; Stems with open symbols represent non-significant correlation coefficients, whose corresponding p-values are not shown. p values are floored at 10^{-4} for ease of visualization. (Bottom) Same as in top panel, but predictions based on iGC connectivity.

D. (Top) Inter-subject correlation matrix of composite behavioral scores. Row and column indices: subjects. (Bottom)Cumulative distributions (solid lines) and density function estimates (filled area) of correlation coefficients between observed and predicted composite scores, for the same subject (yellow) or across different subjects (grey). Predictions were based on GC estimates from the relational and working memory tasks. p-value: Kolmogorov-Smirnov test.

Table Legend

Table 1. Statistical Table

Extended Data Legends

ED Figure 1-1. Task descriptions.

Description of fMRI scans and tasks used in the analysis

ED Figure 1-2. Description of sub-tasks

ED Figure 1-3. Subject identifiers

HCP IDs of 1000 subjects whose data was employed in the analysis. Relational processing scans were not available for IDs marked in grey

ED Figure 1-4. Parcellations used in the analysis

ED Figure 1-5. GC classification accuracies for different parcellations and alternative classifiers

A. Surface renderings showing the 5 anatomical and functional parcellations employed in this study network (ED Figure 1-3).

B. (Top row) Same as in Figure 1B (main text), but showing two-way task versus resting state leave-one-out classification accuracies based on each of the five parcellations (panel A), each in one column. (Second and third rows) Same as top row, but showing precision (second row) and recall (third row). (Last row) Same as top row, but showing K-fold (10-fold) cross-validation accuracies. Other conventions are as in Figure 1B (main text).

C. Same as in Figure 1B (main text), but showing two-way task versus resting state classification accuracies obtained using an SVM with an RBF (radial basis function) kernel (y-axis) against a conventional SVM (x-axis). Classification accuracies were computed with the Shirer et al (2012) 14-network parcellation (2012). Red and blue data: accuracies based on dGC and iGC features, respectively. Dashed diagonal line: Line of equality (x=y).

D. (Top) Two-way classification accuracy for the Working memory task versus Resting state classification, as a function of number of scan time points used to estimate GC. Red dots: dGC. Blue dots: iGC. Curves: Sigmoid fits. Dashed horizontal line: chance accuracy (0.5). (Bottom) Same as in the top panel, but two-way classification accuracies for distinguishing between two simulated networks (shown in Fig. 4A, main text). Other conventions are the same as in the left panel.

ED Figure 1-6. Control analyses

A. Comparison of average GC connection strengths of all subjects (even rows), and subjects who passed all tests of stationarity (odd rows), shown for each task (each column). Each 14x14 matrix depicts connections between all pairs of the 14 networks in the Shirer et al parcellation (Shirer et al., 2012). Entry in cell (i,j) corresponds to dGC connection from node j (source) to node i (destination) or iGC connections between nodes i and j. Source of connection at column, and destination at row.

B.(Left) Comparison of two-way task versus resting state classification accuracies, for the cohort of all subjects (x-axis) vs. for subjects who passed all tests of stationarity (y-axis). Other conventions are the same as in ED Figure 1-5C. (Right) Comparison of n-way classification accuracies across all 7 tasks, for all subjects (right bars) and for subjects who passed all tests of stationarity (left bars). Other conventions are the same as in Figure 1F (main text).

C. (Left) Comparison of two-way task versus resting state classification accuracies, without motion scrubbing (x-axis) vs. after motion scrubbing (y-axis) for all subjects. (Right) Comparison of n-way classification accuracies across all 7 tasks, without motion scrubbing (right bars) and after motion scrubbing (left bars) for all subjects. Other conventions are the same as in panel B.

D. Comparison of dGC connection strengths calculated using 1-stage versus 2-stage methods, for all seven tasks and resting state. Each point corresponds to the average strength, across subjects, of each one of the 182 dGC connections (panel A). Diagonal line: line of equality (x=y).

E. Distribution, across all subjects, of correlation coefficients (r values) obtained by correlating 1-stage versus 2-stage dGC estimates across connections for each subject. Distribution for each of the seven tasks, and resting state are shown in different colors.

F. Distribution, across all connections, of correlation coefficients (r values) obtained by correlating 1-stage versus 2-stage dGC estimates across subjects for each connection. Distribution for each of the seven tasks, and resting state are shown in different colors.

G. Distribution of frame-wise displacement (FD) values (log-scale) across all tasks and resting scans of all 1000 subjects. Each color denotes one of the seven task (or resting) scans. Dotted line: threshold FD value of 0.5mm. For panels A-F, accuracies and connectivity estimates were computed with the Shirer et al 14-network parcellation (2012).

ED Figure 1-7. Number of subjects passing stationarity tests

ED Figure 3-1. Parameters of simulated networks

Parameters of 2-node and 9-node networks

ED Figure 3-2. Relationship between network connectivity, GC and partial correlations

A. Same as in Figure 3C (middle column panel; main text) except for a network with balanced recurrent excitatory (E-E) feedback (Middle). Matrix shows the connections estimated at a sampling interval of 750 ms (Bottom). Other conventions are the same as Figure 3C (main text).

B. (Left) Schematic of a two-node network simulated with a discrete time vector autoregressive model. c and d denote the strength of internode connections, and a and b denote strength of recurrent connections within each node. Here, for simplicity, we assume a=b. (Right, top) Variation of zero-lag covariance (σ_{12}), which is the basis of computing PC, with varying values of c+d for three different values of a. Note that σ_{12} is zero when c=-d, regardless of a. (Right, bottom) Example simulations of node dynamics for c=-0.2 and d=0.2.

C. Covariation of iGC (blue triangles), PC (filled black circles) and PC covariance (K; open squares) with iGC covariance (Y) for simulations with a first-order vector autoregressive (AR) model, with both instantaneous and lag-based connectivity (see ED Mathematical Note, Section S3, equation 11 and equation 23). Open circles: PC covariance (K) for a system with no lag-based connectivity (AR coefficients zero).

D. (Top) Difference in dGC estimates (ΔdGC) between connection in actual direction and connection in the reverse direction, plotted against different standard deviations in onset latencies (σL). Each color denotes one particular scenario of differences in onset HRF latencies (see text for details). Columns 1-2: Network A; columns 3-4: Network B; odd columns: fast timescale (50 ms) sub-network (ABC); even columns: slow timescale (1000 ms) sub-network (DEF) (refer Fig. 4A, main text). (Bottom) Same as in top panel, but for iGC estimates. Errobars: standard error of the mean.

E. (Leftmost column). Ground truth connectivity matrix for the two networks. (Other columns) Same as in Figure 4B (main text), but showing RFE curves (top sub-panel) and maximally discriminative features following RFE (bottom sub-panel). Rows 1-2: RFE based on dGC features; Rows 3-4: RFE based on iGC features. Filled circles: number of features at the “elbow” of each RFE curve. Each column corresponds to one of the four scenarios of onset latency differences (see panel D, and text for details). All panels: HRF onsets sampled from a distribution with σL=0.4s. Other conventions are the same as in Figure 4B (main text).

ED Figure 4-1. Network labels in the Shirer et al (2012) 14-network parcellation

ED Figure 4-2. Task generic and discriminative connections based on partial correlations (PC)

A. Task-discriminative connections based on dGC (top row), iGC (middle row) and PC (last row). Other conventions are as in ED Figure 1-6A.

B. (Top) Same as in Figure 4C (main text), but for RFE based on PC. (Bottom) Same as in Figure 4E (main text), but for RFE based on PC.

C. Same as in Figure 4D (main text), but for task-generic connections based on PC.

D. Same as in Figure 4F (main text), but for task-discriminative connections based on PC.

ED Figure 5-1. Behavioral scores and descriptions

ED Figure 5-2. Behavioral score predictions based on GC connectivity strengths

A. Correlation between predicted and observed behavior scores based on dGC connectivity strengths. Rows: Task scans from which GC estimates were derived; columns: behavior scores predicted (key: ED Figure 5-1). Red-blue colorscale indexes positive and negative correlations, respectively. Black highlighted squares: Significant p-values ($p<0.05$) following Benjamini-Yekutieli correction for multiple comparisons.

B. Same as in A, but predictions based on iGC connectivity strengths.

C. Same as in A, but predictions based on PC connectivity strengths.

D. Same as in Figure 5D bottom, but cumulative distributions of correlation coefficients, for composite score predictions based on GC estimates from each task. Other conventions are the same as in Figure 5D (main text).

ED Mathematical Note

Extended Data 1

The MATLAB codes to reproduce the results are available at https://figshare.com/s/9d9131a6780fc8197cf1 Separate folders correspond to each figure, and subfolders contain scripts for generating each panel in the respective figure. The filenames are alphabetically ordered to provide a sequence for running the scripts. The Multivariate Granger Causality toolbox (mvgc_v1.0; available at http://users.sussex.ac.uk/~lionelb/downloads/mvgc_v1.0.zip) is a pre-requisite. Data necessary to run the scripts (both input and output) are placed in a 'data' subfolder within each figure folder.

Figure 1

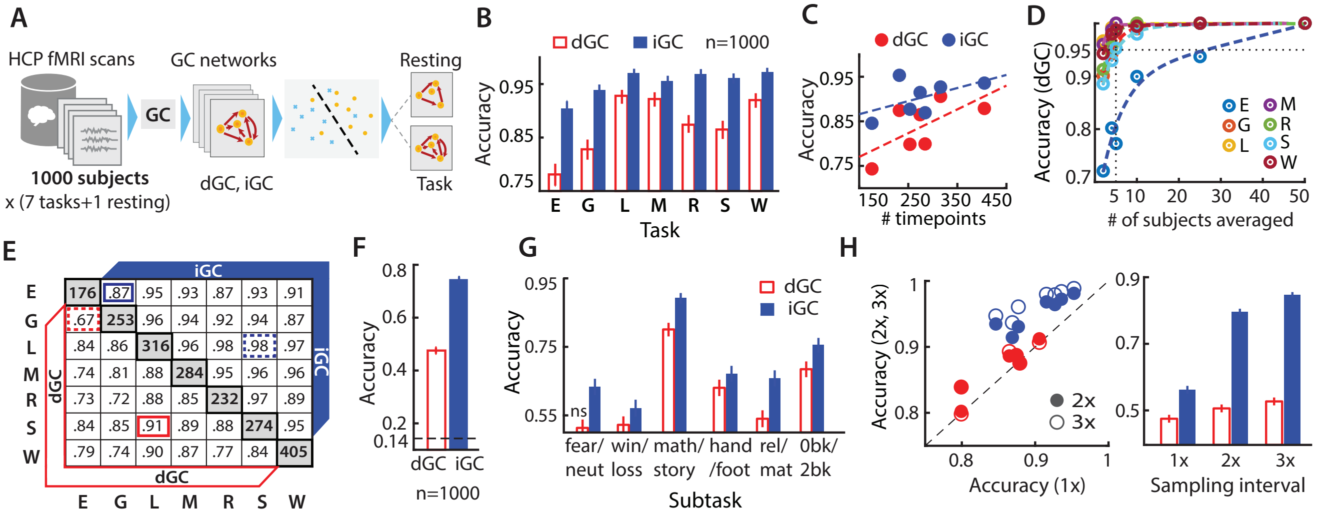

A
HCP fMRI scans
1000 subjects
x (7 tasks+1 resting)
GC
GC networks
dGC, iGC
Resting
Task
B
dGC
iGC
n=1000
Accuracy
0.95
0.85
0.75
E G L M R S W
Task
C
dGC
iGC
Accuracy
0.95
0.85
0.75
150 250 350 450
timepoints
D
Accuracy (dGC)
0.95
0.9
0.8
0.7
5 10 20 30 40 50
# of subjects averaged
E G L M R S W
E
iGC
dGC
E 176 .87 .95 .93 .87 .93 .91
G .67 253 .96 .94 .92 .94 .87
L .84 .86 316 .96 .98 .98 .97
M .74 .81 .88 284 .95 .96 .96
R .73 .72 .88 .85 232 .97 .89
S .84 .85 .91 .89 .88 274 .95
W .79 .74 .90 .87 .77 .84 405
E G L M R S W
F
Accuracy
0.8
0.6
0.4
0.2
0.14
dGC iGC
n=1000
G
dGC
iGC
Accuracy
0.95
0.75
0.55
ns
fear/ neut
win/ loss
math/ story
hand /foot
rel/ mat
0bk/ 2bk
Subtask
H
Accuracy (2x, 3x)
1
0.9
0.8
2x
3x
0.8 0.9 1
Accuracy (1x)
0.9
0.7
0.5
1x 2x 3x
Sampling interval

Figure 2

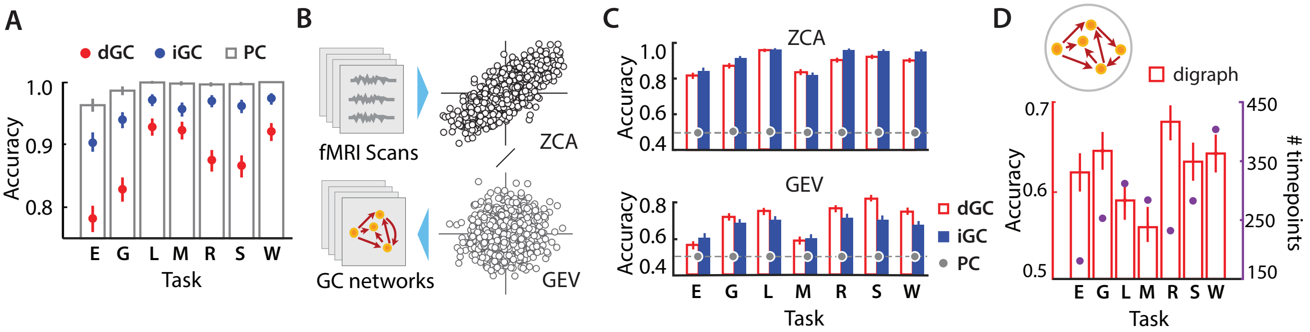

A
dGC
iGC
PC
Accuracy
1.0
0.9
0.8
E G L M R S W
Task
B
fMRI Scans
ZCA
GC networks
GEV
C
ZCA
Accuracy
1.0
0.8
0.6
0.4
GEV
Accuracy
0.8
0.6
0.4
dGC
iGC
PC
E G L M R S W
Task
D
digraph
Accuracy
0.7
0.6
0.5
450
350
250
150
timepoints
E G L M R S W
Task

Figure 3

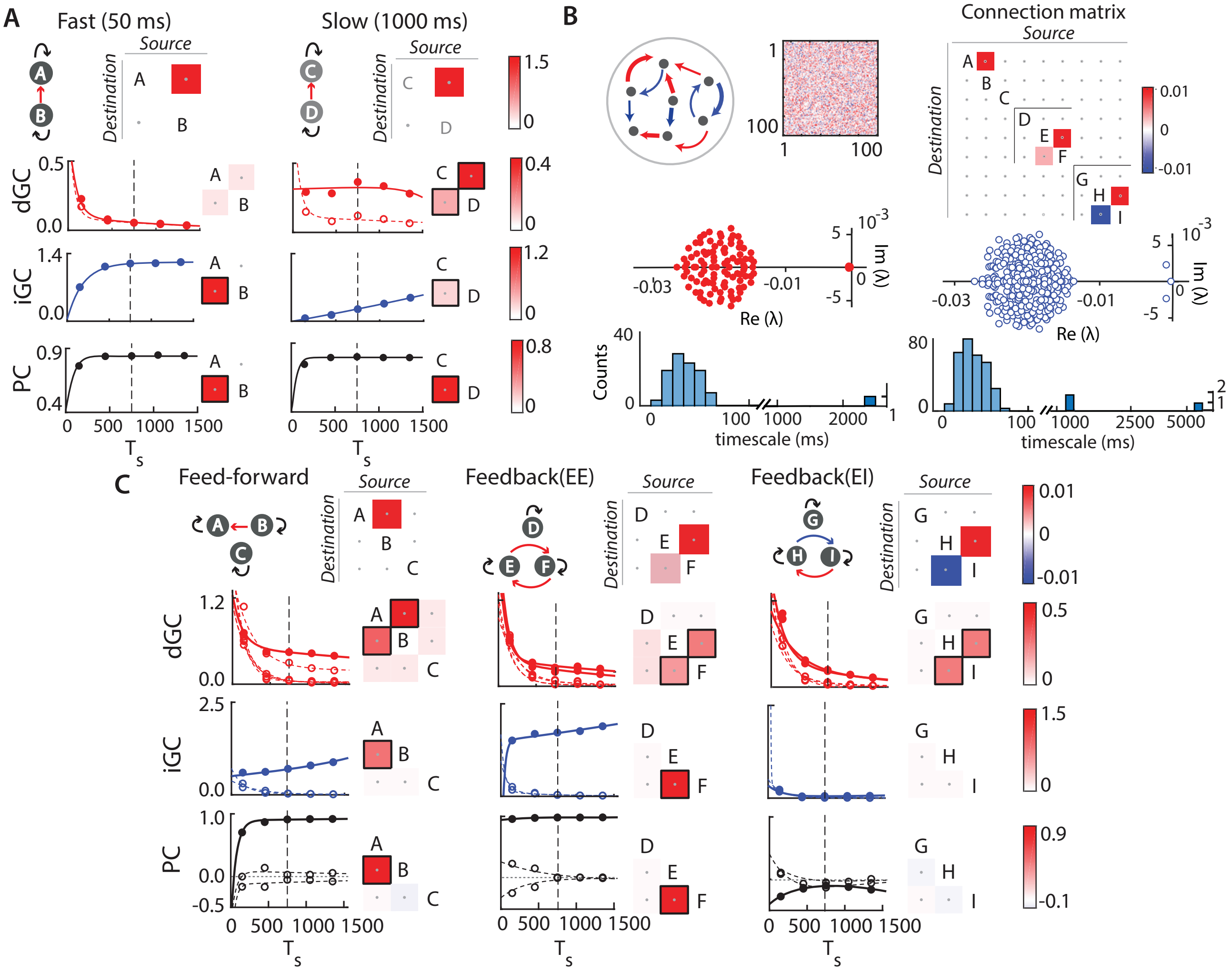

A
Fast (50 ms)
Slow (1000 ms)
Source
Destination
A
B
C
D
1.5
0
dGC
iGC
PC
0.5
0.0
1.4
0.0
0.9
0.4
0.4
0
1.2
0
0.8
0
0
500
1000
1500
T_s
B
Connection matrix
Source
Destination
1
100
1
100
A
B
C
D
E
F
G
H
I
0.01
0
-0.01
10^-3
5
0
-5
Im (λ)
Re (λ)
-0.03
-0.01
40
80
Counts
0
100
1000
2000
2500
5000
1
2
timescale (ms)
C
Feed-forward
Feedback(EE)
Feedback(EI)
Source
Destination
1.2
0.0
2.5
0.0
1.0
0.0
-0.5
0.5
0
1.5
0
0.9
-0.1
0
500
1000
1500
T_s

Figure 4

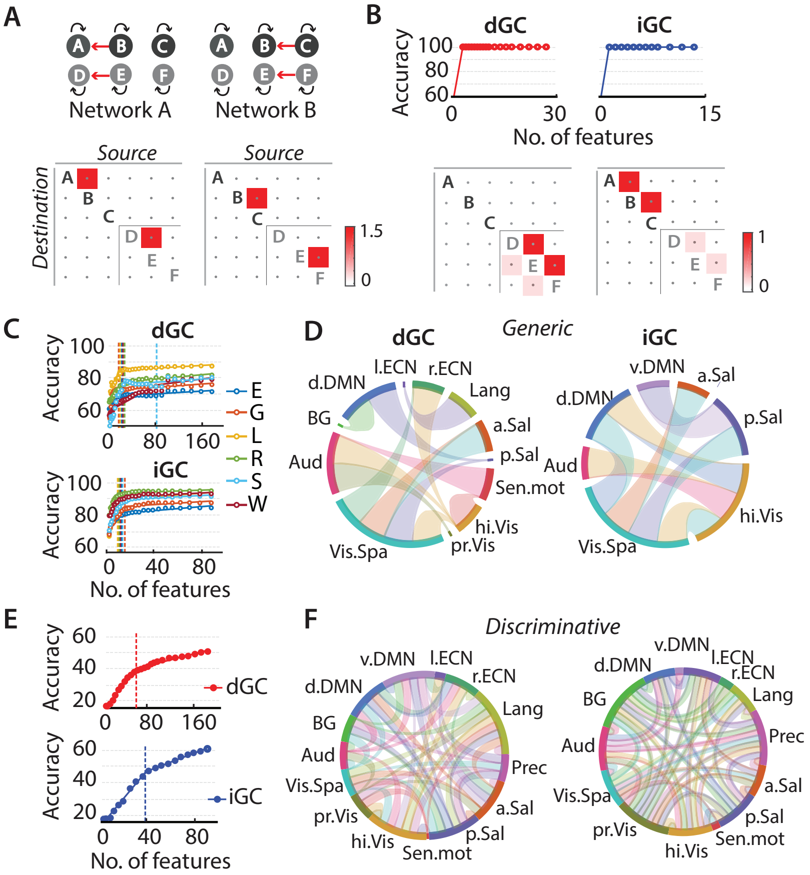
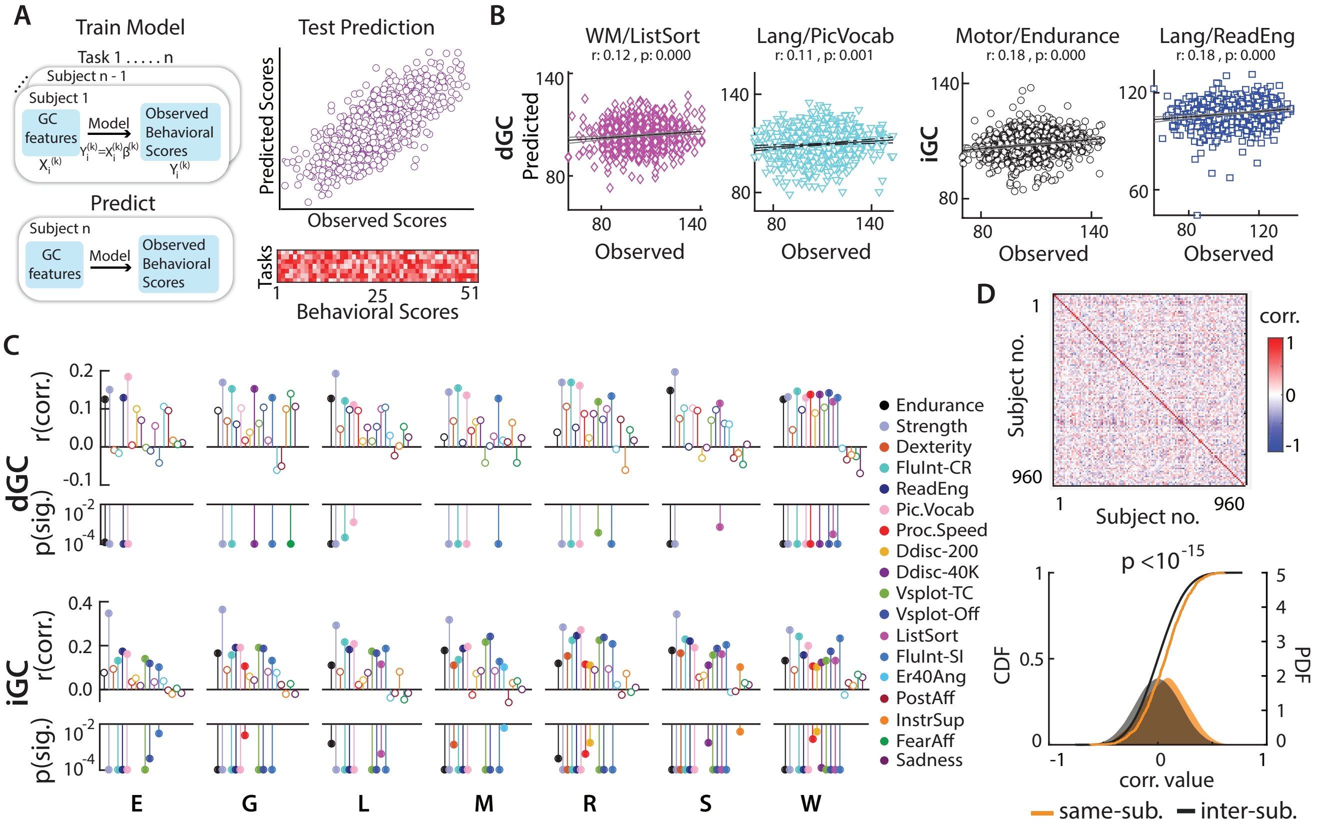

A
A
B
C
D
E
F
Network A
A
B
C
D
E
F
Network B
Source
Destination
A
B
C
D
E
F
Source
A
B
C
D
E
F
1.5
0
B
dGC
iGC
Accuracy
100
80
60
0
30
0
15
No. of features
A
B
C
D
E
F
A
B
C
D
E
F
1
0
C
dGC
Accuracy
100
80
60
0
80
160
E
G
L
R
S
W
iGC
Accuracy
100
80
60
0
40
80
No. of features
D
dGC
Generic
iGC
l.ECN
r.ECN
d.DMN
Lang
BG
a.Sal
p.Sal
Aud
Sen.mot
hi.Vis
pr.Vis
Vis.Spa
v.DMN
a.Sal
d.DMN
p.Sal
Aud
hi.Vis
Vis.Spa
E
Accuracy
60
40
20
0
80
160
dGC
Accuracy
60
40
20
0
40
80
iGC
No. of features
F
Discriminative
v.DMN
l.ECN
r.ECN
d.DMN
Lang
BG
Aud
Prec
Vis.Spa
a.Sal
pr.Vis
p.Sal
hi.Vis
Sen.mot
v.DMN
l.ECN
d.DMN
r.ECN
BG
Lang
Prec
Aud
a.Sal
Vis.Spa
p.Sal
pr.Vis
Sen.mot
hi.Vis

Figure 5


A
Train Model
Task 1 . . . . . n
Subject n - 1
Subject 1
GC features
$X_i^{(k)}$
Model
$Y_i^{(k)}=X_i^{(k)}\beta^{(k)}$
Observed Behavioral Scores
$Y_i^{(k)}$
Predict
Subject n
GC features
Model
Observed Behavioral Scores
Test Prediction
Predicted Scores
Observed Scores
Tasks
1
25
51
Behavioral Scores
B
WM/ListSort
r: 0.12 , p: 0.000
Lang/PicVocab
r: 0.11 , p: 0.001
Motor/Endurance
r: 0.18 , p: 0.000
Lang/ReadEng
r: 0.18 , p: 0.000
dGC
Predicted
iGC
140
80
120
60
Observed
C
dGC
r(corr.)
p(sig.)
iGC
0.2
0.1
0.0
-0.1
0.4
10^{-2}
10^{-4}
E
G
L
M
R
S
W
Endurance
Strength
Dexterity
FluInt-CR
ReadEng
Pic.Vocab
Proc.Speed
Ddisc-200
Ddisc-40K
Vsplot-TC
Vsplot-Off
ListSort
FluInt-SI
Er40Ang
PostAff
InstrSup
FearAff
Sadness
D
Subject no.
1
960
corr.
1
0
-1
$p < 10^{-15}$
CDF
PDF
0.5
corr. value
same-sub.
inter-sub.