bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

How tasks change whole-brain functional organization to reveal brain-phenotype relationships

Abigail S. Greene[1], Siyuan Gao[2], Stephanie Noble[3], Dustin Scheinost[3,4,5], R. Todd Constable[1,3,6,7,*]

[1]Interdepartmental Neuroscience Program, Yale University
[2]Department of Biomedical Engineering, Yale School of Engineering & Applied Science
[3]Department of Radiology and Biomedical Imaging, Yale School of Medicine
[4]Department of Statistics and Data Science, Yale University
[5]The Child Study Center, Yale School of Medicine
[6]Department of Neurosurgery, Yale School of Medicine
[7]Lead contact
*Correspondence: todd.constable@yale.edu

Corresponding Author: R. Todd Constable
Yale University
The Anlyan Center
300 Cedar Street
New Haven, CT 06520

E-mail: todd.constable@yale.edu

Figure and table count: 5 figures
19 figure supplements

Keywords: Functional connectivity, brain-phenotype relationships, predictive modeling, fMRI, tasks

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Functional connectivity (FC) calculated from task fMRI data better reveals brain-phenotype relationships than rest-based FC, but why is unknown. In over 700 individuals performing 7 tasks, we use psychophysiological interaction (PPI) and predictive modeling analyses to demonstrate that FC and overall degree of task-induced signal change, but not task-evoked activation alone, drive phenotypic prediction, and their combination further improves prediction. Inter-subject PPI demonstrates that predictive utility is highest in distributed FC patterns that are dissimilar across individuals, except in regions of group-level task activation, suggesting that task FC better predicts phenotype than rest FC for two, regionally specific reasons: (1) tasks synchronize activated regions and amplify signal components that meaningfully vary across individuals; and (2) elsewhere, prediction is driven by nodal interactions that set individuals apart. These findings offer a framework to leverage both task activation and FC to reveal the neural bases of complex human traits, symptoms, and behaviors.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Functional connectivity analyses have offered sweeping insights into the macroscale neural circuits underlying complex cognitive processes, finding these circuits to be broadly distributed across the human brain (e.g., Dubois et al., 2018; Finn et al., 2015; Hsu et al., 2018; Rosenberg et al., 2015; Wager et al., 2013). Such analyses are typically performed using resting-state data (Biswal, Yetkin, Haughton, & Hyde, 1995; Power, Schlaggar, & Petersen, 2014), revealing "intrinsic connectivity networks" that recapitulate networks invoked during task execution (Stephen M Smith et al., 2009). This correspondence—along with demonstrations of the stability of functional connectivity (FC) patterns between resting and task states (Cole, Bassett, Power, Braver, & Petersen, 2014; Gratton et al., 2018; Krienen, Yeo, & Buckner, 2014)—suggests that the functional network architecture of the human brain is relatively state-invariant. Nevertheless, there is a growing consensus that FC contains useful dynamic, rather than just static, information (J. R. Cohen, 2018). Task-induced changes in patterns of FC have been shown to be widely distributed across the brain (Bolt, Nomi, Rubinov, & Uddin, 2017), to subserve the task at hand (Medaglia, Lynall, & Bassett, 2015), to make individuals more identifiable (Finn et al., 2017), and to improve FC-based prediction of both task performance (Rosenberg et al., 2015) and stable traits, such as intelligence measures (Greene, Gao, Scheinost, & Constable, 2018). Together, these findings suggest that task-induced changes in FC, while perhaps low-amplitude and/or local perturbations of a core functional architecture, are functionally significant and may amplify individual differences in brain functional organization.

Thoughtfully leveraging such changes therefore holds the promise of advancing individual differences research, but this will first require a more complete characterization of how tasks change patterns of FC. In particular, the question of whether these changes reflect task-evoked activation, changes in neural interaction, or some combination of the two has received substantial attention. While some have raised concerns that inadequate removal of task-evoked activation from node time courses may yield spurious patterns of FC (Cole et al., 2019), others have demonstrated that task-evoked activation and task-induced changes in FC can be cleanly dissociated (Di & Biswal, 2018), even when task-evoked activation is not removed from the BOLD signal (Kieliba et al., 2018), and that task-evoked FC (that is, task-induced changes in FC attributable to task-evoked activity) explains relatively little of the total task-induced change in FC (Lynch et al., 2018).

Here, we replicate and extend this growing literature, and in particular the work of Di and Biswal (Di & Biswal, 2018), by demonstrating that activation and FC calculated from in-scanner

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

task data (hereafter, "task FC") are spatially distinct and, critically, offer complementary insights into the neural bases of a given phenotype. That is, while we have previously demonstrated that task FC-based models better predict phenotypic measures than rest FC-based models (Greene et al., 2018), whether this improvement is attributable to sharpening of connectivity patterns irrespective of task condition (hereafter task "context"), task-evoked activation, and/or context-dependent changes in nodal synchrony (Figure 1a) remains an open question. To explore this question, we use the connectome-based predictive modeling (CPM) framework (Finn et al., 2015; Gao, Greene, Constable, & Scheinost, 2019; Xilin Shen et al., 2017), psychophysiological interaction (PPI) analyses (Cole et al., 2013; McLaren, Ries, Xu, & Johnson, 2012), and a novel extension of PPI and inter-subject FC (Simony et al., 2016) analyses to demonstrate that, in a range of tasks, the most informative single signal component for fluid intelligence (gF) prediction varies by task, but task activation fails to predict gF in all tasks, and model performance is improved by the inclusion of multiple signal components. This is true even when no single component successfully predicts gF, suggesting that each component carries independent, trait-relevant information that together is greater than the sum of its parts. Finally, given the recent finding that tasks increase both the similarity (i.e., correlation) of individuals' patterns of FC and the identifiability of individuals on the basis of these patterns (Finn et al., 2017), we investigated the relationship between consistency of BOLD signals across individuals and their predictive utility. Across the brain, inter-subject consistency of moment-to-moment BOLD fluctuations decreases the predictive utility of incident edges, with one exception: for edges that connect focal, activated regions, inter-subject consistency boosts predictive utility. This finding highlights that while changes in FC, not task activation, reveal brain-phenotype relationships, informative task-induced changes in FC take two forms: distributed and individual-specific, and focal and stereotyped.

Results

During a task, FC and overall degree of task-induced signal change—but not task-evoked activation—predict phenotype

To explore why task FC-based models outperform resting-state FC-based models, we used data from the Human Connectome Project (HCP; Van Essen et al., 2013) S1200 release (n = 703; see Methods for inclusion criteria). Each subject completed seven in-scanner tasks, providing an internal validation of results' generalizability; for each task run, fMRI data were parcellated into 268 nodes (Finn et al., 2015; X. Shen, Tokoglu, Papademetris, & Constable,

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2013) and a mean time course was calculated for each node. Each node's time course was decomposed via multilinear regression, using a validated psychophysiological interaction (PPI) framework (Cole et al., 2013; McLaren et al., 2012), into terms that reflect its context-independent FC with the predictor node (ciFC), its context-dependent FC with the predictor node (cdFC), its task activation, and its overall degree of task-induced signal change (i.e., intercept, which effectively combines the task activation and interaction terms, as well as potentially relevant, unmodeled information; for further discussion, see Figure 2—figure supplement 1). Note that a non-zero, informative intercept term was made possible by the choice to zero-center (i.e., set condition-on to 0.5, condition-off to -0.5), not mean-center (i.e., *z*-score), task-timing regressors. The analysis was repeated with mean-centered task regressors, with comparable results (Figure 2—figure supplement 1). For an explanation of why this choice does not affect results of this analysis, as well as more details on the relevant methods, see *Methods, Investigating potential confounds* and Figure 2—figure supplement 1. This regression was performed for every node pair for each task and subject, and linear contrasts were calculated for task activation and cdFC terms, yielding four PPI beta matrices per subject per task (Figure 1b).

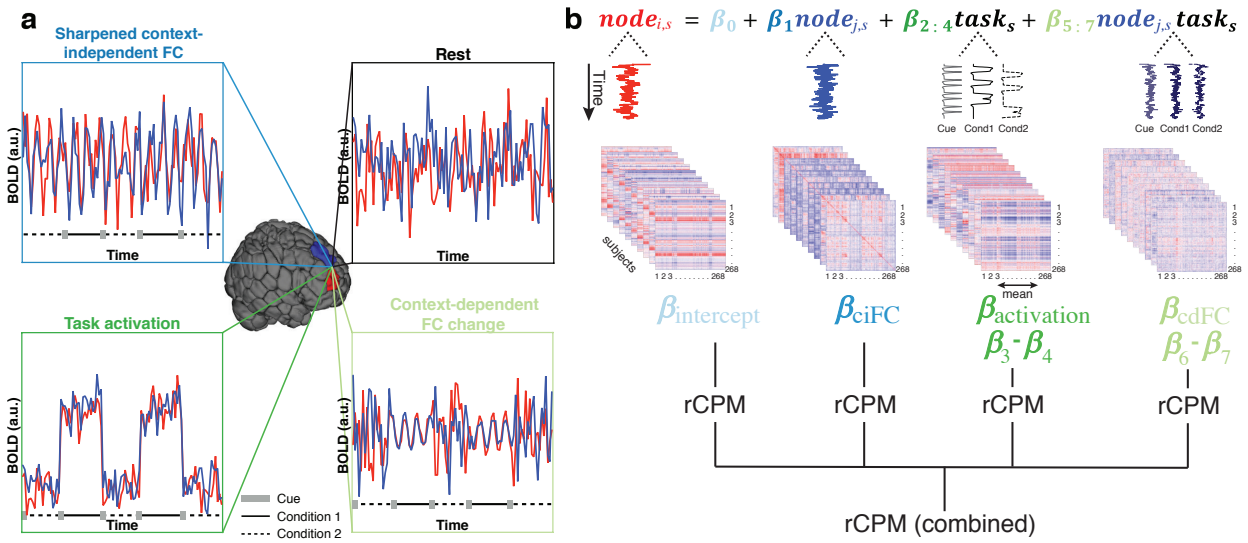

Figure 1 with 2 supplements. Pairing psychophysiological interaction (PPI) analysis with prediction permits characterization of how tasks change patterns of brain activity to reveal brain-phenotype relationships. (a) Relative to rest, in-scanner tasks may sharpen context-independent FC ("ciFC"), elicit baseline shifts in activity ("task activation"), and/or induce context-dependent changes in nodal synchrony ("cdFC"). (b) These components of each node's time course (here, examples taken from nodes in the WM task; see Methods) can be modeled in an adapted gPPI framework, yielding one node-by-node matrix of PPI betas for each term in each subject *s*. After calculating condition contrasts for the task activation and cdFC terms (indicated schematically by subtracted betas) and collapsing task activation via averaging into a single value per node per subject, these matrices (and vector, in the case of activation) were then

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

submitted, individually and in combination, to the ridge CPM pipeline (rCPM; Gao et al., 2019) to yield predictions of fluid intelligence (gF). See Methods for details. Cond, condition; *i, j*, two nodes in the Shen parcellation (Finn et al., 2015; X. Shen et al., 2013).

These matrices were then submitted, individually ("individual-term model") and in combination ("combined model"), to a ridge regression-based version of the connectome-based predictive modeling pipeline (rCPM; Gao et al., 2019; Shen et al., 2017) to predict phenotype (here, gF) scores (Figure 1b). In brief, this cross-validated machine learning approach selects features on the basis of their correlation with the predicted measure, regresses (with ridge regularization) phenotype scores on selected features' values (here, PPI beta estimates), and uses resulting ridge regression coefficient estimates to construct a linear model relating brain data to phenotype measures. This model is then applied to the left-out fold, and the process is repeated iteratively until all folds have been used as the test group (see Methods for details). Model performance was quantified as 1 – normalized mean squared error (Methods) of each model (q^2; higher values indicate better performance). This analysis was repeated 100 times with different assignments of participants to folds, and performance is presented for every iteration (Figure 2a) and as the mean and standard deviation across iterations (Figure 2b).

Model performance was assessed for significance using non-parametric permutation tests. That is, the analysis was repeated 100 times with gF permuted across subjects each time; given the existence of many sibships in the dataset, allowed permutations respected family-related limits on exchangeability (Winkler, Ridgway, Webster, Smith, & Nichols, 2014; Winkler, Webster, Vidaurre, Nichols, & Smith, 2015). *P* values were calculated as the fraction of iterations on which the unpermuted gF-based models performed worse than the best-performing corresponding null model. For each task, the best-performing (i.e., highest mean q^2) unpermuted gF-based model was also compared to the second best-performing model via Wilcoxon signed-rank test. All *P* values were corrected for multiple comparisons using the Bonferroni correction.

Consistent with previous work demonstrating that HCP task-based models successfully predict gF (Greene et al., 2018), for each task, at least one model significantly predicted gF (all $P < 0.001$, corrected). Predictions were most accurate with the combined model for the language task (mean $q^2 = 0.12$). Notably, which individual term best predicted gF was task-dependent. For some tasks, the intercept best predicted gF (gambling: intercept mean $q^2 = 0.05$, $P < 0.001$, corrected; WM: intercept mean $q^2 = 0.02$, $P < 0.001$, corrected; social: intercept mean $q^2 = 0.03$, $P < 0.001$, corrected). For other tasks, ciFC best predicted gF (motor: ciFC mean $q^2 = 0.03$, $P < 0.001$, corrected; emotion: ciFC mean $q^2 = 0.03$, $P < 0.001$, corrected). In the language task,

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

intercept and cdFC predicted gF about equally well (intercept mean $q^2 = 0.05$, cdFC mean $q^2 = 0.06$, both $P < 0.001$, corrected). Finally, for the relational task, none of the individual-term models successfully predicted gF (all $P > 0.05$, corrected). Notably, task activation did not predict gF for any of the tasks; this result was replicated using HCP-released, individual-level GLM-based node activation for prediction (see *Methods, Investigating potential confounds* and Figure 2—figure supplement 2). We note that model performance is lower than previously reported (Greene et al., 2018), a finding that holds when "standard FC" matrices (i.e., without task modeling) were passed through the same preprocessing and prediction pipelines (gray bars in Figure 2b, Figure 2—figure supplement 3); in fact, the best-performing PPI-based models often outperformed the standard FC-based models. However, the addition of global signal regression (GSR) to the FC matrix preprocessing pipeline substantially improved prediction performance, rendering it comparable to previous results (Figure 2—figure supplement 3). While GSR was not appropriate for main analyses due to potential task-related fluctuations of the global signal (e.g., due to fluctuating arousal or vigilance [Liu et al., 2017]), and while these analyses depend on relative, rather than absolute, prediction performance, this finding supports the utility of GSR for FC-based prediction (Greene et al., 2018; Li et al., 2019).

However, for all tasks except emotion, combining all terms (intercept, ciFC, cdFC, and task activation) yielded a model that significantly outperformed the best-performing individual-term model for that task (all $P < 0.001$, corrected; Figure 2a,b). That the combined model was not most predictive for the emotion task may be related to the fact that the best-performing model for that task (ciFC alone) attained relatively low accuracy. The relatively high performance of combined models, but critically not cdFC- and activation-based models, persisted even when task regressors were mismatched with brain data (Figure 2—figure supplement 4; Methods), an analysis in which the combined model, and to a lesser degree ciFC (Figure 2—figure supplement 2), would be expected to largely recapitulate standard FC.

Strikingly, these results suggest that the combined models are greater than the sum of their parts. That is, even for tasks where only one individual-term model significantly predicted gF, a successful combined model still significantly outperformed that individual-term model, and even significantly predicted gF when no individual-term models did (relational task: combined model mean $q^2 = 0.07$, $P < 0.001$, corrected). Further, examining the relative contributions of each term to the combined model for all tasks (Figure 2c; see Methods for derivation) demonstrates that these contributions often do not follow the performance of individual-term models for that task, highlighting the importance of information uniqueness. For example, for the

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

language task, the intercept- and cdFC-based models are the only individual-term models that significantly predicted gF, but in the combined model, ciFC contributed the most predictive information, followed by cdFC and then intercept terms. This finding also holds in reverse; that is, term contribution makes only a slight difference in the performance of the model when that term is dropped (Figure 2—figure supplement 5). Notably, the predictive contributions of task activation to the combined models are negligible for all tasks.

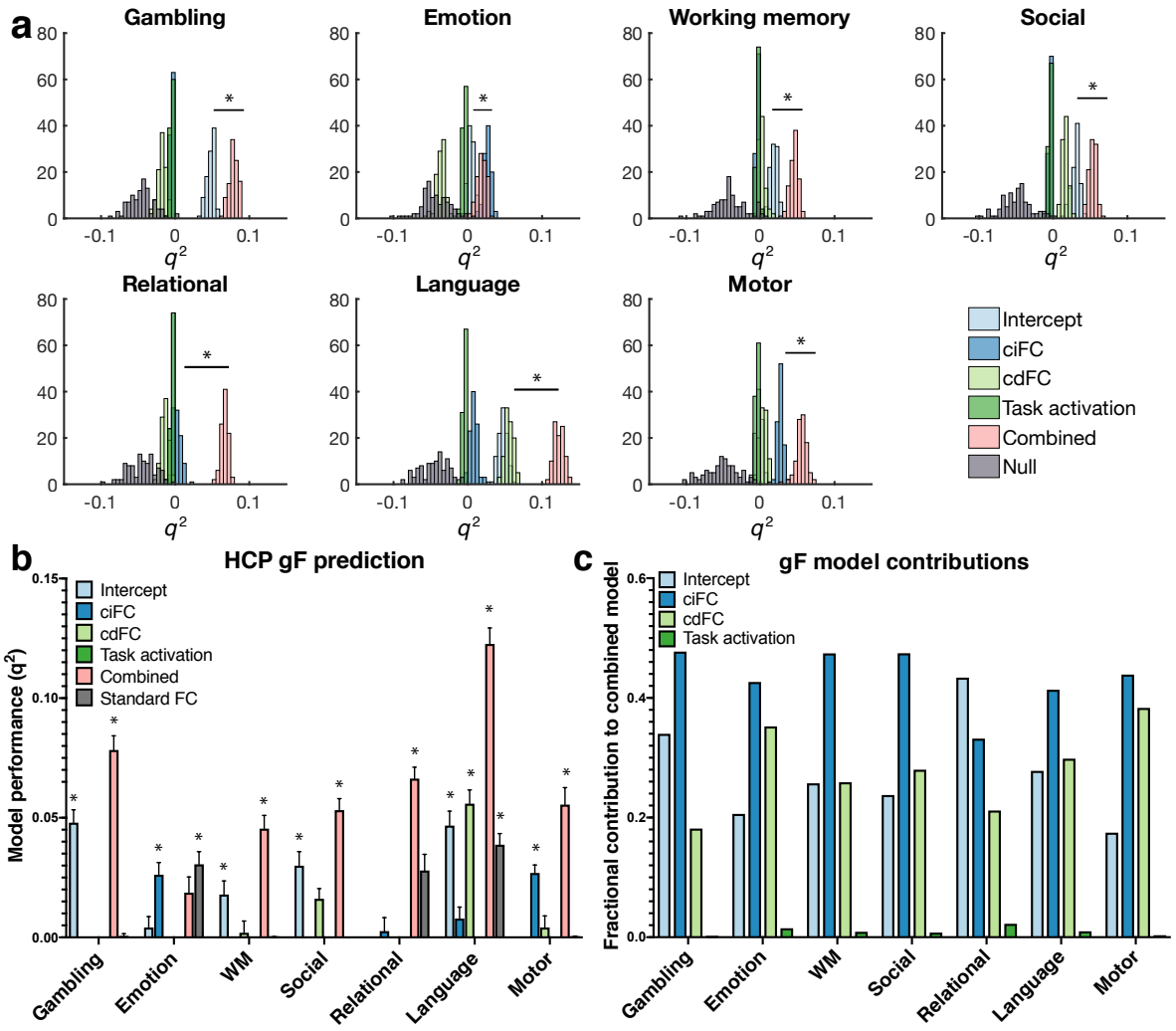

Figure 2 with 10 supplements. Context-independent FC, context-dependent FC, and overall task effect contain complementary information about phenotypic measures. (a) Histograms of model performance (quantified as q^2 [see Methods]) across 100 iterations of 10-fold rCPM with a P threshold of 0.1 for feature selection. In every task, the best-performing model significantly outperformed the second-best-performing model (*P < 0.001, corrected, via Wilcoxon signed rank test). "Null": prediction performance using permuted gF scores. (b) Summary (mean and s.d., indicated by error bars) of model performance for each task and term (*P < 0.001, corrected, via permutation test). Models with negative mean q^2 were set to 0. "Standard FC": prediction performance using FC calculated as Fisher-transformed Pearson

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

correlations, without task modeling (see Methods). (c) Percent contribution of each term to that task's combined model (see Methods for derivation). WM, working memory.

Further, given that mean motion was significantly correlated with observed gF for 3/7 tasks (Figure 2—figure supplement 6) and with mean predicted gF for 6/35 tasks and terms (Figure 2—figure supplement 7), several additional analyses were performed to ensure that in-scanner motion did not confound results. First, rCPM was repeated using partial correlation with individuals' mean motion per condition—rather than simple correlation—for feature selection (i.e., selecting features that are correlated with gF after controlling for their correlation with motion). Second, even more conservatively, rCPM was repeated after regressing individuals' mean motion for the given task from both gF and each edge (within the cross-validation loop); models were built from resulting residuals (see *Methods, Investigating potential confounds*). Results are comparable, both in terms of mean model performance (r(original, partial correlation) = 0.99, r(original, residualized) = 0.99, all $P < 0.001$; Figure 2—figure supplement 8) and feature weights (Figure 2—figure supplement 9), suggesting that modeling results are not confounded by in-scanner motion. Frame-to-frame motion was also found to be uncorrelated with task timing (Figure 2—figure supplement 10).

Model contributions of FC and activation terms are spatially distributed and distinct

We next sought to characterize the spatial distribution of features with high predictive utility (i.e., model contribution) for each term in each task's combined model. Predictive utility was quantified as the mean ridge coefficient across 100 iterations for features selected in 75% of analyses (10 folds * 100 iterations), scaled by the standard deviation of the PPI betas for that feature across all subjects. Except where otherwise noted, signed contributions were used to dissociate features that are positively and negatively related to gF. The seven tasks demonstrated substantial consistency in overall spatial patterns of predictive utility; for concision and clarity, the motor task is used to exemplify these patterns, with corresponding results for all tasks displayed in Figure 3—figure supplement 1.

We first summarized the distribution of predictive features at the node level, calculating the percent contribution of each node as the summed, absolute predictive utility of all edges incident to it (or, in the case of activation, as the absolute predictive utility of the node), normalized by the sum of all nodes' absolute predictive utility for the given task and term. Results (Figure 3a, d, g) demonstrate that predictive ciFC and cdFC features are broadly distributed

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

across the brain, while activation information content is sparser and concentrated more focally (for the motor task, primarily in visual and motor regions).

Next, we explored the distribution of predictive features at the network level, using ten canonical networks defined in an independent sample (Finn et al., 2015; Noble et al., 2017; for details, see Methods: *Functional parcellation and network definition* and Figure 3—figure supplement 2). For each network pair, ciFC and cdFC features' contributions were summed and normalized by the total number of features between that network pair. These values were normalized by the total, absolute sum of contributions for the given task and term, to yield a comparable scale across tasks. For task activation, features' (i.e., nodes') contributions were summed for each network, normalized by the size of (i.e., number of nodes in) the given network, and then all values were normalized by the absolute sum of these contributions for the given task. Resulting matrices and vectors are visualized in Figure 3b (ciFC), 3e (cdFC), and 3h (activation) with the difference between absolute ciFC (Figure 3b) and absolute cdFC (Figure 3e) network contributions visualized in Figure 3i. Results across tasks demonstrate the substantial and distributed contributions of ciFC and cdFC features to the combined models (c.f., Figure 3—figure supplement 1a, c, f), which contrasts with the more focal distribution of nodes with predictive activation (Figure 3g, h, and Figure 3—figure supplement 1e). Critically, predictive features are relatively non-overlapping for ciFC and cdFC (paired, two-sided Wilcoxon signed rank test of ciFC versus cdFC concatenated, vectorized, network matrices: $P < 0.001$; mean rank correlation between vectorized, absolute model contributions for ciFC versus cdFC across all tasks: overall $\bar{r}_s = 0.01$, intersection of predictive ciFC and cdFC features only $\bar{r}_s = 0.10$). In particular, visual inspection of these matrices reveals that, across tasks, predictive ciFC edges tend to be concentrated in medial frontal and frontoparietal networks, while predictive cdFC edges tend to be concentrated in motor and visual networks.

It is important to note that the predictive utility and distribution of useful context-dependent features will necessarily depend on the condition contrast that is applied, and these results demonstrate how findings can be interpreted with some cognitive specificity, given this choice. For example, while motor task ciFC and cdFC demonstrate similar overall patterns of predictive utility as other tasks (i.e., relative overrepresentation of edges in medial frontal and frontoparietal networks for ciFC, and in visual networks for cdFC), the motor task contrast—all motions (tongue, hands, and feet) versus fixation—reveals consistently negative cdFC ridge coefficients. That is, across the brain, the lower the FC during these body motions, the higher an individual's gF (Figure 3e). This pattern is in stark contrast to other tasks' patterns of cdFC

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

predictive contributions (Figure 3—figure supplement 1c), a finding with potentially interesting cognitive implications (see Discussion).

Before calculating the predictive contributions of network pairs (or networks, in the case of activation), ciFC and cdFC features were divided into edges with mean positive and mean negative ciFC (and activation features into nodes with mean positive and negative activation). This avoids a potential ambiguity: one could imagine a positive edge that is more positive (i.e., stronger) in individuals with higher gF, or a negative edge that is less negative (i.e., weaker) in individuals with higher gF. Both edges would receive a positive ridge coefficient but may represent physiologically distinct processes. Interestingly, we found that the ciFC and cdFC of edges with mean negative ciFC made essentially no contributions to combined models for any of the tasks (Figure 3—figure supplement 3a-b), and main results reflect only edges with mean positive ciFC (Figure 3b, e, i). Network-level contributions of cdFC patterns were further divided by mean cdFC (i.e., edges with positive mean cdFC and edges with negative mean cdFC); results are visualized in Figure 3—figure supplement 4. Conversely, both activated and deactivated nodes were consistently included in predictive models; given this, task activation network-level predictive utility is presented for both activated and deactivated nodes (Figure 3h; Figure 3—figure supplement 1e).

Visualizing predictive contributions at the network level, while useful, offers a necessarily coarse representation of where in the brain highly predictive features are located. To visualize the predictive utility of individual features, we present circle plots, with each network a different color on the outer track, and each line a ciFC (Figure 3c, Figure 3—figure supplement 1b) or cdFC (Figure 3f, Figure 3—figure supplement 1d) edge. This visualization again highlights that model contributions are distributed and relatively non-overlapping across terms. Specifically, it reveals that highly predictive cdFC features are not incident to the most predictive or most activated nodes (Figure 3f, Figure 3—figure supplement 1d): mean rank correlation of cdFC absolute contribution-based node degree (Methods) with absolute model contribution of each node's activation: $\overline{r_s} = 0.004$; mean rank correlation of cdFC node degree with absolute task effect size: $\overline{r_s} = 0.06$. Similarly, the contribution of a node's activation is only weakly related to its absolute task effect size: $\overline{r_s} = 0.11$. In sum, these results demonstrate that predictive utility—of ciFC, cdFC, or even task activation, itself—is not simply driven by task activation, and that predictive features from each of these terms are spatially highly distributed and relatively non-overlapping across terms.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Finally, it is worth noting that different insights can be gained from interpreting predictive contributions at the node, network, and edge levels (Horien, Greene, Constable, & Scheinost, 2019), as shown in Figure 3. The maximum correlation between vectorized, absolute model contributions of ciFC and cdFC is for the motor task (overall r_s = 0.07, intersection of predictive ciFC and cdFC features only r_s = 0.27). Interestingly, the network visualizations (Figure 3b and 3e) obscure this similarity, which is more evident when feature contributions are visualized at the edge level (Figure 3c and 3f), demonstrating the differential utility of fine- and coarse-scale spatial localization analyses.

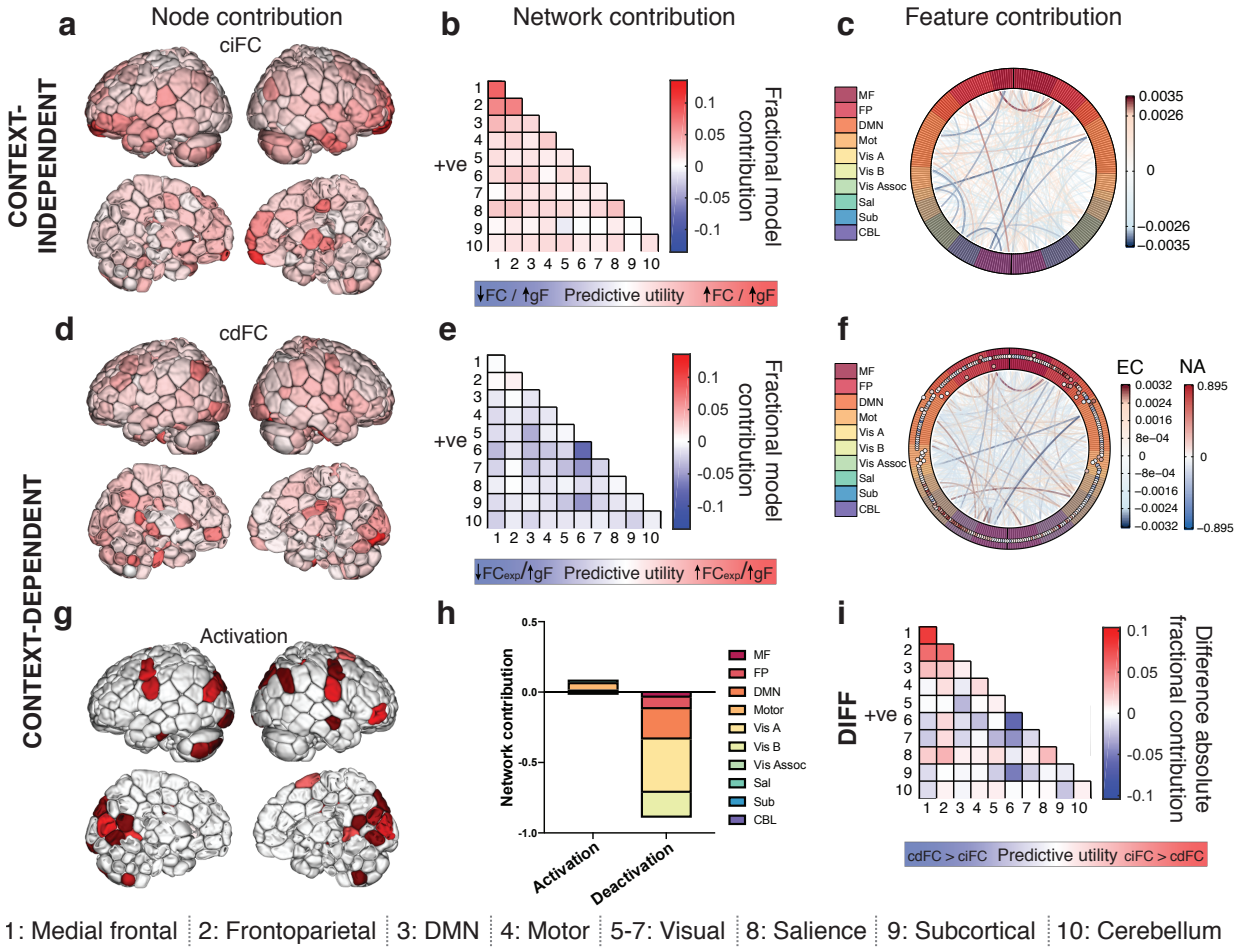

Figure 3 with 4 supplements. Context-independent FC, context-dependent FC, and activation predictive features are distributed and distinct. (a) Visualization of predictive ciFC features by node (i.e., absolute weighted node degree, normalized within task/term [see Methods], such that darker red indicates greater predictive utility). In this and all subsequent panels, results are depicted for the motor task; comparable results in all tasks can be found in Figure 3—figure supplement 1. (b) Visualization of predictive ciFC features by network for each task. Red = positive ridge coefficients, blue = negative ridge coefficients, shade = relative model contribution. In this and all subsequent figures: "+ve" indicates that results reflect only

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

contributions of edges with mean positive ciFC. 1-10 = network assignment. (c) Visualization of individual predictive ciFC features, with each consistently selected edge represented as a line; line color and thickness scale with predictive model contribution. In this and all subsequent figures, MF = medial frontal, FP = frontoparietal, DMN = default mode network, Mot = motor, Vis A = visual A, Vis B = visual B, Vis Assoc = visual association, Sal = salience, Sub = subcortical, CBL = cerebellum. (d) Visualization of predictive cdFC features by node (as in [a]). (e) Visualization of predictive cdFC features by network. FC_{exp}, FC during the experimental condition (i.e., condition of interest). Red = positive ridge coefficients, blue = negative ridge coefficients, shade = relative model contribution. 1-10 = network assignment. (f) Visualization of individual predictive cdFC (lines) and activation (outer track circles) features; line color and thickness scale with cdFC feature predictive utility, and circles represent the corresponding nodes, with their color indicating mean activation (red = positive, blue = negative) and distance from the x axis indicating their model contribution. EC, edge contribution; NA, node activation. (g) Visualization of predictive activation features (i.e., absolute node predictive utility, normalized within task/term, as in [a,d]). (h) Visualization of predictive activation features' (i.e., nodes') network assignments for nodes with mean positive PPI activation betas ("activation") and for nodes with mean negative PPI activation betas ("deactivation"). (i) Visualization of the difference between absolute, network-level ciFC model contributions (i.e., absolute value of matrix in [b]) and absolute, network-level cdFC model contributions (i.e., absolute value of matrix in [e]).

Context-independent FC is more consistent across individuals than context-dependent FC or activation

To explore the consistency of task effects on FC and activation across individuals, and the potential relationship between this consistency and predictive utility, we performed an inter-subject PPI analysis (Figure 4a) on the five tasks with consistent task timing across individuals (emotion, gambling, social, relational, and WM). This analysis revealed substantial consistency in activity patterns across individuals.

The network-level spatial distribution of inter-subject consistency is visualized for positive (and self-connecting; see Methods) edges' context-independent (Figure 4b) and context-dependent (Figure 4c) signals. As was the case for the prediction analyses, effects were almost entirely limited to edges with positive mean ciFC (i.e., positive mean PPI betas for the ciFC term, as were used previously); inter-subject consistency of mean negative edges' ciFC are presented in Figure 4—figure supplement 1c, and of mean negative edges' cdFC in Figure 4—figure supplement 1d. These visualizations reveal two trends. First, during a task, positive edges' ciFC and nodes' time courses become more similar across individuals (relative to intrinsic components of the BOLD signal, e.g., at rest, which should be uncorrelated across subjects [Simony et al., 2016]) in almost all networks for all tasks (Figure 4b), while context-dependent effects vary more by task and network, as would be expected given the varying designs and demands of the tasks (e.g., some network pairs' edges are consistently stronger during the

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

experimental condition relative to the control condition [red], while others are consistently weaker during the experimental condition relative to the control condition [blue]; Figure 4c). Second, ciFC inter-subject consistency is overall greater than cdFC or task activation inter-subject consistency (Figure 4d; median ciFC consistency: mean across tasks = 0.396, range = 0.27-0.63; median cdFC consistency: mean across tasks = -0.01, range = -0.04-0.01; median activation consistency: mean across tasks = -0.004, range = -0.03-0.01), suggesting that moment-to-moment fluctuations are more similar across individuals than are block-level changes in FC and activation.

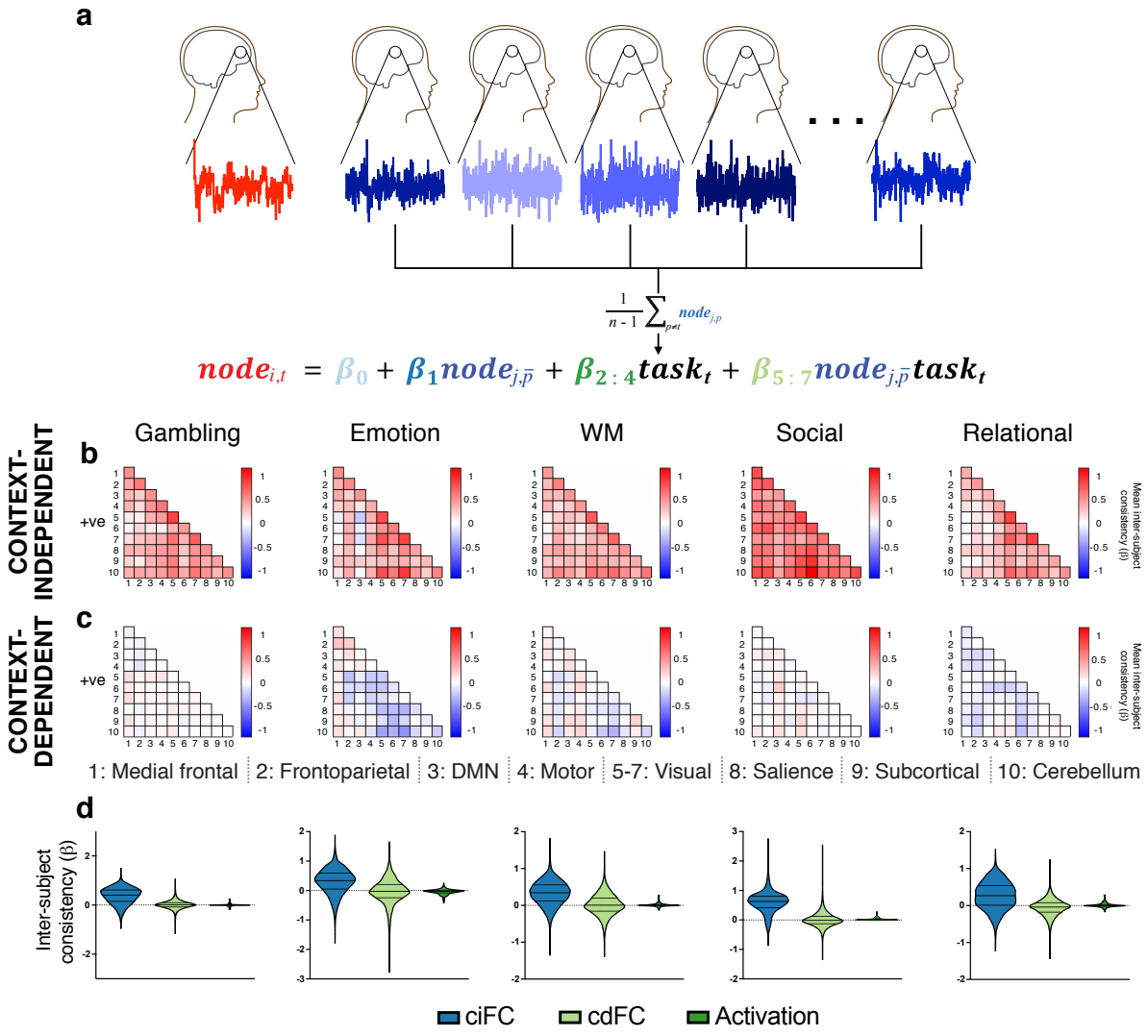

Figure 4 with 2 supplements. Inter-subject PPI analysis reveals consistent task-induced changes in context-independent and context-dependent activity across individuals. (a) Schematic depiction of the inter-subject PPI analysis pipeline, in which the target node *i* time course is taken from subject *t*, and the predictor node *j* time course is averaged across all remaining subjects *p*. (b,c) Network-level visualization of the substantial inter-subject

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

consistency for both context-independent and context-dependent signals. 1-10 = network assignment. (d) Violin plots (dashed line, median; dotted line, quartiles) of inter-subject consistency for all unique features (i.e., inter-subject PPI betas) reveal that inter-subject consistency of moment-to-moment fluctuations (ciFC) is greater than inter-subject consistency of block-level activation ("Activation") or FC (cdFC) changes.

Inter-subject similarities and differences in task response contribute differentially to predictive models

To explore the relationship between inter-subject consistency of task effects and predictive utility—and whether there is any spatial or functional structure to this relationship—we performed several analyses. First, we re-grouped reliably selected predictive features based on their absolute inter-subject consistency (see Methods) and visualized the network-level model contributions for these groups for ciFC (Figure 5a, top two rows) and cdFC (Figure 5a, bottom two rows). This analysis revealed that ciFC contributions are spatially distinct for high- and low-consistency edges, with predictive, high-consistency edges concentrated within medial frontal and visual networks, while predictive, low-consistency edges are more distributed across the brain. Differences in the spatial distribution of high- and low-consistency predictive cdFC edges are less pronounced, as demonstrated by higher rank correlations between vectorized high- and low-consistency predictive utility network matrices (Figure 5a) for cdFC than ciFC: ciFC r_s = -0.30-0.37; cdFC r_s = 0.26-0.73; mean within-task cdFC-ciFC difference = 0.3978, 95% CI = 0.15-0.65).

Next, we explicitly modeled the relationship between absolute inter-subject consistency and absolute predictive utility using a multilinear regression that included one additional term and the interaction between this term and inter-subject consistency (see Methods). To capture key anatomical and functional features for each edge, we used the following terms: mean absolute task effect size for the two nodes incident to the given edge (calculated as in Salehi et al., 2019; see Methods), edge membership between or within canonical networks, resting-state edge test-retest reliability (calculated as in Noble et al., 2017), edge length (i.e., Euclidean distance between incident nodes), and edge membership within or between hemispheres. Across these analyses, effects were more pronounced for ciFC than cdFC or task activation (where relevant; activation predictive utility was only modeled as a function of consistency and node activation, as the other edge-level metrics cannot meaningfully be applied to nodes). This is consistent with the finding that inter-subject consistency is greater for context-independent than context-dependent terms (Figure 4d), and thus would be expected to have a greater effect

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

on predictive utility for ciFC than for cdFC or activation; for completeness, we display results for all modeled terms (Figure 5b), but for clarity and concision, only ciFC effects are discussed. Overall, inter-subject consistency was negatively related to predictive utility, such that more consistent edges were less predictive. Interestingly, activation was also negatively related to predictive utility, suggesting that activated regions are connected by less predictive edges than non-activated regions. The interaction between inter-subject consistency and activation, however, was positive (Figure 5b, leftmost panel), suggesting that edges affected by task activity are more predictive when that effect is consistent across individuals. Network modeling results demonstrate that network membership does not itself affect predictive utility, but that predictive, consistent edges tend to be within-network, rather than between-network (Figure 5b, middle panel), consistent with the finding that high-consistency, predictive ciFC features tend to be concentrated within medial frontal and visual networks (Figure 5a, top row). Finally, reliability modeling results demonstrate that inter-subject consistency in this analysis does not simply recapitulate reliability, but that consistent, reliable edges are more predictive than consistent, unreliable edges (Figure 5b, rightmost panel). Because hemisphere and edge length were not clearly related to predictive utility, results for these analyses are displayed in Figure 5—figure supplement 1.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

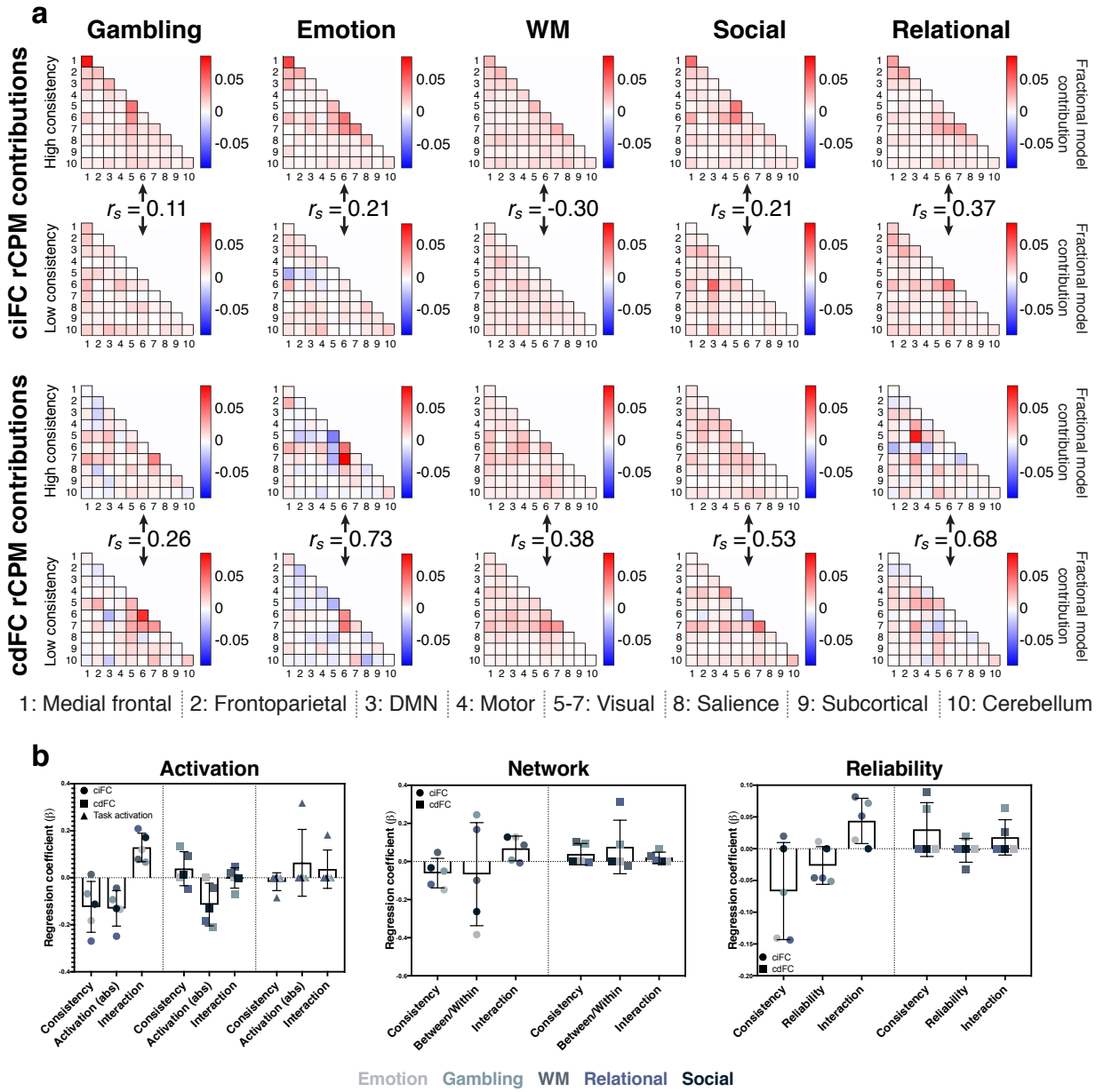

Figure 5 with 1 supplement. Task fMRI-based prediction is driven primarily by individualized, distributed FC patterns, except within networks activated by the task. (a) Visualization of predictive ciFC (top two rows) and cdFC (bottom two rows) features at the network level, with features divided by median inter-subject consistency. r_s, rank correlation between high- and low-consistency network matrices for the given task and term. 1-10 = network assignment. (b) Regression analyses relating the predictive utility of a feature's ciFC (all), cdFC (all), and activation ("Activation" only) to its inter-subject consistency, task activation, status between or within networks, resting-state test-retest reliability estimate, and relevant interactions. Results presented as regression coefficient for the given predictor in each of the five modeled tasks; bar height reflects mean coefficient; error bars indicate s.d. of coefficients.

Discussion

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In-scanner tasks have been found to amplify individual differences in patterns of FC and correspondingly improve FC-based prediction of phenotype (Greene et al., 2018), but the nature of this improvement—whether it is due to changes in ciFC, cdFC, and/or activation patterns—remains unexplored. In this work, we leverage intra- and inter-subject PPI and predictive modeling analyses to ask this question. Despite substantial differences in the nature and design of the analyzed tasks, we found a striking consensus: task FC better reveals brain-phenotype relationships than resting-state FC due not to task-evoked co-activation, but rather to task-induced changes in patterns of context-independent and -dependent FC, as well as to the overall degree of task-induced change in patterns of activity and FC. Further, tasks do not simply emphasize individual differences in FC patterns; while predictive utility is generally boosted by amplification of such individual differences, precisely the opposite is true for edges between regions activated by the task. That is, in regions activated by the task, it is the similarities between individuals that allow us to find the differences.

Anatomy of a successful predictive model

While the FC and intercept terms are the most useful for prediction, it is noteworthy that the components of a given node's signal contain complementary (i.e., unique), phenotype-relevant information, such that their combination further improves predictive model performance for a given task; in fact, in many cases, the combined model performs better than would be expected given the performance of corresponding individual-term models. This is consistent with reports of improved ridge regression-based CPM performance with the inclusion of more relevant features in the model (up to a *P* value of 0.5 [Gao et al., 2019]). That is, in evaluating and weighting each component of the signal separately, the combined model is able to capture more information than is contained in any of its component parts, or in comparably preprocessed standard FC (Figure 2—figure supplement 3), even when incorrect task regressors are used (Figure 2—figure supplement 4). Similar efforts to reveal brain-phenotype relationships may therefore benefit from the inclusion of more features—even if their relationship to the phenotype of interest is relatively weak—in models that use regularization (Gao et al., 2019), although it is always wise to exclude uninformative features to avoid overfitting (Scheinost et al., 2019).

Interrogation of these combined models revealed that they are broadly distributed and relatively non-overlapping across terms (i.e., ciFC, cdFC, and activation), further highlighting the distinct, phenotype-relevant ways in which tasks alter each signal component and replicating the finding that task-related activation and context-dependent FC are spatially distinct (Di &

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Biswal, 2018). Further, across all tasks, predictive FC patterns tend to be concentrated in medial frontal, frontoparietal, visual, and motor networks. These networks have all been previously implicated in successful FC-based predictive models of gF (Finn et al., 2015; Greene et al., 2018), but PPI-based prediction permits a more fine-grained evaluation of their involvement, revealing that medial frontal and frontoparietal networks are overrepresented in predictive *context-independent* FC edges, while visual and motor networks are overrepresented in predictive *context-dependent* FC edges (Figure 3i, Figure 3—figure supplement 1f). That is, while medial frontal and frontoparietal network FC is relevant to gF regardless of when in the task you look, block-level changes in visual and motor network FC predict gF. This is consistent with evidence that frontal and frontoparietal networks comprise domain-general, core components of task-set representations (Cole et al., 2013; Dosenbach et al., 2006), while visual and motor networks comprise domain-specific "data processing" systems (Posner & Petersen, 1990), which would be expected to adapt their operations to the nature and demands of the task at hand. This work is, to our knowledge, the first to parse and localize the differential predictive utility of context-independent FC and context-dependent FC. Results demonstrate the exciting potential of this analysis framework to understand the neural bases of successful predictive models, offering more nuanced insights into the neural representation of the predicted measure than would be accessible with standard FC-based models.

These findings are in line with the growing consensus that complex cognitive processes and constructs, such as fluid intelligence, are supported by distributed neural circuitry, rather than by circumscribed regions of interest (Turk-Browne, 2013), and demonstrate that rest-to-task FC changes, while perhaps small in magnitude (Cole et al., 2014), contain important information about phenotype independent of task-evoked activation. We of course do not suggest that co-activation cannot drive changes in FC (see, for example, Cole et al., 2019), but rather that *predictive* FC changes are not driven by co-activation.

A framework to explore task-specific effects on functional organization

In addition to task-general predictive changes in FC, spatial localization of predictive features demonstrates the utility of this approach for drilling down into task-specific FC changes and their relationships to phenotypic measures. For example, the motor task is the only task for which context-dependent FC model contributions at the network level are consistently negative, indicating that the weaker an individual's predictive context-dependent FC edges during motion, the higher that individual's gF. Further, the finding that both edges that strengthen (i.e., mean

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

positive cdFC) and those that weaken (mean negative cdFC) during motion relative to fixation have negative context-dependent FC ridge coefficients (Figure 3—figure supplement 4) suggests that the relationship between context-dependent FC and gF depends more on total FC strength (i.e., globally weaker FC during motion predicts higher gF) than on the nature of task-induced change in FC (i.e., increased or decreased edge strength). It is possible that this task, given its low cognitive demands, does not require the widespread neural interactions that support higher-demand tasks (Di, Gohel, Kim, & Biswal, 2013), permitting mind wandering and decreased neural integration, which is energetically costly (Bullmore & Sporns, 2012; Di & Biswal, 2018).

While these preliminary results would be expected to depend to some extent on task design and modeling choices (Newell, 1973), they suggest a potentially fruitful direction for future investigation into the nature and cognitive implications of task-specific changes in functional brain organization. It is likely that such task-specific changes, by offering complementary insights into brain-phenotype relationships, explain the finding that combining FC data across task conditions often outperforms prediction using FC data from a single condition (Elliott et al., 2019; Gao et al., 2019). Better understanding these changes will enable more selective inclusion of data by condition, particularly when using data that include potentially less informative or noisy conditions (e.g., rest; Elliott et al., 2019).

Inter-subject PPI reveals two classes of predictive functional connections

The inter-subject correlation analyses are a novel extension of prior work on inter-subject correlation (Hasson, Nir, Levy, Fuhrmann, & Malach, 2004) and FC (Simony et al., 2016) that provide a complementary approach to study how tasks change patterns of FC to better reveal meaningful individual differences in them. Given the finding that tasks both increase inter-subject FC similarity and improve individual identifiability on the basis of FC patterns (Finn et al., 2017), we sought to explicitly explore the relationship between inter-subject time course synchrony (i.e., consistency) and predictive utility: do tasks constrain the state space (Buckner, Krienen, & Yeo, 2013; Elton & Gao, 2015; Leonardi, Shirer, Greicius, & Van De Ville, 2014), simultaneously making a relevant network more similar across individuals and amplifying signal components within it that vary across individuals, and/or are increased inter-subject consistency and increased predictive utility spatially separable? In fact, patterns of context-independent FC with high predictive utility are quite different for high- and low-consistency edges. While this was not the case for cdFC, inter-subject consistency was overall lower for cdFC than for ciFC, as might be expected given that interaction effects are less reliable than main effects (Di & Biswal, 2017);

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

future investigations may seek to explore these relationships in longer tasks to increase cdFC reliability. Regression analyses confirmed that the effect of inter-subject consistency on predictive utility is greater for context-independent than for context-dependent FC and activation. Specifically, more consistent context-independent FC patterns are less useful for prediction, but among consistent edges, those that are activated, more reliable, and within network tend to be most predictive.

Taken together, these results parallel two, complementary lines of human neuroscience research on task-induced changes in brain function: task activation studies, which identify clearly demarcated regions that are consistently activated by a task across individuals, and functional connectivity studies, which identify distributed, subtler task effects on patterns of brain activity outside of these consensus regions of activation (Cole et al., 2014; Horien et al., 2019; Salehi et al., 2019). The latter proved to be, overall, more useful for phenotypic prediction—in fact, task activation here failed to predict phenotype in every analysis. However, activation does have predictive relevance: in activated regions, functional connections that experience more similar task-induced changes across individuals are more predictive. When these regions are called upon to subserve a given task, their signals may become time-locked to the task (increasing inter-subject consistency [Hasson et al., 2004]) and/or more constrained (Buckner et al., 2013), amplifying signal components that may vary meaningfully across individuals (improving prediction accuracy). In non-activated regions, such consistency of moment-to-moment fluctuations was found to decrease the predictive utility of incident edges. These findings suggest that task-based FC better predicts phenotype than rest-based FC for two, regionally specific reasons: in regions where activity changes with task context, prediction is driven by edges between regions that change *in the same way* across people—that is, by edges connecting nodes that become time-locked to the task *and* more constrained to phenotypically relevant patterns of activity. Everywhere else in the brain, however, prediction is driven by diversity, by nodes that are doing different things in everyone, revealing inter-individual differences in FC patterns that reflect phenotype.

Additional considerations and future directions

It is worth noting that our characterization of task effects is intentionally limited; given that blocked task designs are common and better powered to reveal effects of interest than event-related or mixed designs (Chee, Venkatraman, Westphal, & Siong, 2003; Friston, Zarahn, Josephs, Henson, & Dale, 1999), we chose to use simple, well-studied task condition contrasts

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(Barch et al., 2013) to reveal fundamental, generalizable effects of task execution on FC and activation predictive utility. Further, regression analyses are here limited by the relatively small number of measurements (i.e., coefficient estimates for the five tasks that were used for both inter- and intra-subject analyses), which precludes the performance of rigorous statistical tests. Future application of the analysis framework presented here to independent datasets will provide opportunities to replicate, broaden (i.e., demonstrate their generalizability to different tasks and modeling approaches), and narrow (i.e., reveal task-specific changes in FC with phenotypic relevance) presented results.

Similarly, a deeper investigation of task-specific activation patterns may reveal that task-evoked activation here demonstrates little predictive utility because of the relatively large size of each node, which may blur informative, fine-scale patterns of activity and/or "wash out" activated voxels by grouping them with less activated voxels (Kriegeskorte, Goebel, & Bandettini, 2006; Norman, Polyn, Detre, & Haxby, 2006; Turk-Browne, 2013). We note, however, that we do not seek to make claims about the predictive utility of task activation, but rather to demonstrate that task-induced changes in standard FC that improve phenotype prediction are not driven by task activation. As such, we chose to use a conventional parcellation to calculate FC (Finn et al., 2015; X. Shen et al., 2013), but future work may seek to compare the predictive utility of FC and task activation at a finer spatial scale.

Conclusion

As task-based FC gains popularity for individual differences research, a better understanding of how tasks change patterns of FC is critical. By demonstrating that the success of task FC-based predictive models is attributable to task-induced changes in context-independent FC, context-dependent FC, and overall task effect, but not to task-evoked activation alone, and by characterizing how tasks change patterns of context-independent FC to improve prediction, these findings demonstrate that reconfiguration of the functional connectome during in-scanner tasks is real, meaningful, and useful. This lays the foundation for intentional, precise use of in-scanner tasks to amplify individual differences in functional brain organization and more effectively map the neural representations of behaviors, traits, and clinical symptoms.

Methods

Dataset

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Data used in this work were released as part of the Human Connectome Project (HCP) S1200 release, described below.

HCP participants. We restricted our analyses to those subjects who completed all seven fMRI tasks (WM, gambling, language, social, relational, motor, and emotion), whose grand mean root mean square (RMS) relative motion across all task runs was less than 0.1 mm and whose maximum mean RMS relative motion was less than 0.16 mm, and for whom gF measures were available. One subject was found to be missing data due to a download failure, and was excluded from all analyses to ensure consistency with previous results. A similarly conservative threshold for motion-based exclusion has been previously demonstrated to mitigate the relationship between FC and gF measures (Greene et al., 2018). In total, data from 703 subjects were used (342 males, ages 22-37 years [mean = 28.5, s.d. = 3.8, median = 29]).

HCP imaging parameters and preprocessing. Details of imaging parameters (Stephen M Smith et al., 2013; Uğurbil et al., 2013; Van Essen et al., 2013) and preprocessing (Glasser et al., 2013; Stephen M Smith et al., 2013) have been published elsewhere. In brief, all fMRI data were acquired on a 3T Siemens Skyra using a slice-accelerated, multiband, gradient-echo, echo planar imaging (EPI) sequence (TR = 720 ms, TE = 33.1 ms, flip angle = 52 degrees, resolution = 2.0 mm^3, multiband factor = 8). Images acquired for each subject include a structural scan and eighteen fMRI scans (WM task, incentive processing [gambling] task, motor task, language processing task, social cognition task, relational processing task, emotion processing task, and two resting-state scans; two runs per condition [one left/right (LR) phase encoding run and one right/left (RL) phase encoding run]; Barch et al., 2013; Smith et al., 2013) split between two sessions. Data from the seven HCP tasks were used for this work, and each task was a different length (WM, 5:01; gambling, 3:12; language, 3:57; social, 3:27; relational, 2:56; motor, 3:34; emotion, 2:16). The scanning protocol (as well as procedures for obtaining informed consent from all participants) was approved by the Institutional Review Board at Washington University in St. Louis. Use of HCP data for these analyses was deemed exempt from IRB review by the Yale Human Investigation Committee. The HCP minimal preprocessing pipeline was used on these data (Glasser et al., 2013), which includes artifact removal, motion correction, and registration to standard space. All subsequent preprocessing was performed in BioImage Suite (Joshi et al., 2011) and included standard preprocessing procedures (Finn et al., 2015), including removal of motion-related components of the signal; regression of mean time courses in white matter and cerebrospinal fluid; removal of the linear trend; and temporal smoothing with a Gaussian filter, σ = 0.18 (a relatively high low-pass filter designed to preserve potential high-

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

frequency, task-related signal components). Mean RMS relative motion was averaged for the LR and RL runs, yielding seven motion values per subject; these were used for subject exclusion and motion analyses (e.g., partial correlation-based feature selection). All subsequent analyses and visualizations were performed in BioImage Suite (Joshi et al., 2011), Matlab (Mathworks), R version 3.6.0 for macOS (packages: RColorBrewer [Neuwirth, 2014], ComplexHeatmap [Gu, Eils, & Schlesner, 2016], and circlize [Gu, Gu, Eils, Schlesner, & Brors, 2014]), and GraphPad Prism version 8.0 for macOS.

Functional parcellation and network definition

The Shen 268-node atlas (Finn et al., 2015; X. Shen et al., 2013) was applied to the HCP data, as described previously (Greene et al., 2018). This parcellation is derived from the application of a group-wise spectral clustering algorithm to an independent data set (X. Shen et al., 2013). Time courses of voxels within each node were averaged. Subjects without whole-brain coverage (i.e., with missing nodes) were excluded from all further analyses.

The same spectral clustering algorithm was used to assign these 268 nodes to eight networks (Finn et al., 2015; X. Shen et al., 2013), and the subcortical-cerebellar network was split into networks 8-10 (Noble et al., 2017; Figure 3—figure supplement 2). These networks are named based on their approximate correspondence to previously defined resting-state networks, and are numbered as follows: 1. Medial frontal, 2. Frontoparietal, 3. Default mode, 4. Motor, 5. Visual A, 6. Visual B, 7. Visual association, 8. Salience, 9. Subcortical, 10. Cerebellum.

Psychophysiological interaction (PPI) analysis

After parcellation, node time courses were submitted to an adaptation of the PPI pipeline developed and described by Cole and colleagues (Cole et al., 2013) and modeled after the generalized PPI framework (McLaren et al., 2012). In brief, the mean time course for each node was decomposed via multilinear regression with three regressor types: the mean time course of a predictor node (yielding a beta weight that reflects context-independent FC [ciFC] between the predictor and target nodes); zero-centered, block-level task boxcar regressors convolved with the canonical HRF (yielding a beta weight that reflects the influence of task activation on the target node time course), and the interactions of these terms (yielding a beta weight that reflects context-dependent FC [cdFC] between the predictor and target nodes). The canonical HRF (generated using SPM8) was used given its demonstrated efficacy in identifying patterns of task activation in these data (Barch et al., 2013). Task conditions and cues were modeled separately,

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

as relevant (WM: 0-back task, 2-back task, and cue regressors; language: story task and cue regressors; emotion: face block and cue regressors; gambling: reward block and loss block regressors; relational: relational block, matching block, and cue regressors; social: mental video block and random video block regressors; motor: left hand, right hand, left foot, right foot, tongue, and cue regressors). Regressors were calculated separately for each subject using HCP EV.txt files and downsampled given the HCP sampling rate (i.e., TR = 720 msec). Thus, for a given task with *k* conditions (including fixation and cue, if present), a given node's time course can be described by the following equation:

$$v_i = \beta_0 + \beta_1 v_j + \beta_{2:k} T_{1:k-1} + \beta_{k+1:2k-1} f(T_{1:k-1}) v_j$$

$$f(T) = \begin{cases} 0.5 \text{ if } T > 0 \\ -0.5 \text{ if } T \leq 0 \end{cases}$$

where $v_{i,j}$ are the *z*-scored time courses of target node *i* and predictor node *j*, and *T* is the relevant, HRF-convolved, zero-centered task timing regressor(s). As in previously published work (Cole et al., 2013), the task regressors used in the interaction term were binarized. Here, after convolution with the HRF, values greater than zero were set to 0.5, and values less than or equal to zero were set to -0.5 (to zero-center the binarized task regressors, as for the non-binarized task regressors) prior to their multiplication by the predictor node time course to yield the interaction regressor. (For a discussion of the choice to zero-center task regressors, see *Methods: Investigating potential confounds*). These steps are each depicted schematically in Figure 1—figure supplement 1. For all tasks except the emotion and language tasks (for which there were no fixation blocks, rendering the contrast implicit by modeling only one of the two task conditions) and the motor task (for which all motion conditions were summed to yield a motion versus fixation contrast), interaction and task activation beta weights were each combined via subtraction to yield one activation contrast beta weight and one interaction contrast beta weight per feature in each subject (McLaren et al., 2012; contrasts: *n*-back: 2-back – 0-back; gambling: reward – punish; emotion: fear [faces] versus neutral [shapes]; language: story versus math; relational: relation – match; social: TOM – random), and beta weights were calculated separately for each task run (i.e., LR and RL phase encoding runs) and then averaged. This process was repeated for every node pair (i.e., for a given target node, all other nodes were used as predictor nodes) and every subject, yielding, for each subject, four asymmetric, node-by-node matrices of beta weights (one each for the intercept, ciFC, task activation, and cdFC terms). Each task activation matrix was collapsed via averaging into a 268-element vector. All other matrices were symmetrized by averaging them with their transpose. All matrices (and

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

vectors, in the case of activation) of a given type were then submitted—alone and in combination—to the predictive modeling pipeline described below (*Cognitive prediction*).

Finally, to explore the effects of modeling task timing on resulting beta weights, we repeated the PPI using mismatched brain data and task regressors in an exhaustive fashion (e.g., emotion task fMRI data with gambling, language, motor, relational, social, and WM task regressors). Resulting betas were also submitted to the cognitive prediction pipeline; results are presented in Figure 2—figure supplement 4.

Cognitive prediction

Fluid intelligence was quantified using a 24-item version of the Penn Progressive Matrices test; this test is an abbreviated form of Raven's standard progressive matrices (Bilker et al., 2012). Integer scores indicate number of correct responses (PMAT24_A_CR, range = 5-24, mean = 17.70, s.d. = 4.43, median = 19).

A modified version of connectome-based predictive modeling (CPM; Finn et al., 2015; Shen et al., 2017) was used to predict gF from brain measures (i.e., beta matrices [see *Psychophysiological interaction analysis*]) using ridge regression (Gao et al., 2019). This pipeline predicts gF in novel subjects, validating the model through iterative, *k*-fold cross-validation; in this work, $k = 10$ to balance model bias and variance given the large sample size (Scheinost et al., 2019). Consistent with this motivation, split-half (i.e., $k = 2$) analyses yielded comparable patterns of results (e.g., best performance from combined models), but overall lower prediction performance (Figure 2—figure supplement 8). First, the sample was divided into ten groups, respecting family structure such that family members were always assigned to the same group. Nine of these groups were used as training data; in this training set, features (edges and/or nodes) were selected on the basis of their Pearson correlation with gF scores. A correlation P value of 0.1 was selected as the edge selection threshold, given evidence to suggest that more permissive feature selection yields improved regularized regression-based prediction results, and that $P = 0.1$ offers an acceptable compromise between model performance and computational demands (Gao et al., 2019). These edges were then submitted as predictors (with gF score as response) to an L2-constrained linear least squares regression (elastic net mixing value = 1e-6), using another, inner 10-fold cross-validation to find the regression coefficients that correspond to the largest regularization strength (lambda) that yields a MSE within one standard deviation of the minimum MSE. These fitted coefficients were then applied to the corresponding

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

edges in the left-out test subjects to predict their phenotype scores, and these steps were performed iteratively with each group left out once.

Model performance was quantified as cross-validated r^2: $q^2 = 1 - \frac{MSE}{var(y)}$, where y = observed gF scores. This whole pipeline was repeated 100 times with different group partitions, and model performance is reported as the mean across these 100 iterations (similarly, in graphical representations of results, bar heights represent mean performance, and error bars represent performance standard deviation). Significance of model performance was assessed via 100 iterations of nonparametric permutation testing (Finn et al., 2015), accounting for limits on exchangeability due to family structure (Winkler et al., 2014, 2015), and *P* values were calculated as the fraction of non-permuted iterations on which prediction accuracy was less than or equal to the accuracy of the best-performing null model for the given task and term. Two related subjects in our sample were missing family structure information; these subjects were excluded from permutation tests (n = 701). Resulting *P* values were Bonferroni corrected for multiple comparisons. Last, for each task, a paired Wilcoxon signed rank test was used to compare performance across all 100 iterations of the two models with the highest and second-highest mean performance.

Evaluating and visualizing contributions to a predictive model

Given the improved performance of the combined models relative to the individual-term models and the opportunity to interrogate relative term contributions to these combined models, we performed several analyses to evaluate the contributions of individual features, networks, and terms to the combined model for a given task. First, to ensure that only reliably predictive features were analyzed, a given feature was required to have been selected on 75% of all feature selections (10 folds * 100 iterations = 1000 feature selections, for main analyses). The contribution of each selected feature was calculated as its mean ridge regression coefficients (b) across all 1000 analyses multiplied by the standard deviation of its PPI beta across subjects. That is, the contribution $\boldsymbol{c}$ for feature i was defined as

$$\boldsymbol{c_i} = std\left(\beta_i^{PPI}\right) * \bar{b}_i^{ridge}$$

For all reliably selected features from a given term t (e.g., ciFC), the absolute values of these contributions were summed and converted to fractional contributions,

$$C_t = \sum_{i=1}^{N_t} \left| std\left(\beta_i^{PPI}\right) * \bar{b}_i^{ridge} \right| \Big/ \sum_{t=1}^{T} \sum_{i=1}^{N_t} \left| std\left(\beta_i^{PPI}\right) * \bar{b}_i^{ridge} \right|$$

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

where N_t represents the number of features from the given term t (with features that were not selected on 75% of feature selections set to 0) and T represents the number of terms. This procedure was repeated for each task to yield four contribution fractions per task (one per term). These contribution fractions were in turn used to drop terms in order of ascending and descending contributions, after which prediction was repeated to determine the impact of term contribution on model performance (Figure 2—figure supplement 5). Contributions across terms were compared using rank correlation of all features' absolute contributions (ciFC and cdFC); of all nodes' absolute contributions (cdFC contribution degree $d_i = \sum_1^j abs(c_{i,j})$, with d_i the degree for node i, and $c_{i,j}$ the cdFC model contribution of the edge connecting nodes i and j) with absolute task activation contributions; and of absolute task activation contributions with absolute task activation, itself.

Contributions were visualized at several levels of analysis; results are displayed for the motor task in Figure 3 and for all tasks in Figure 3—figure supplement 1. First, signed contributions of each feature were visualized in circle plots (Figure 3c, f, and Figure 3—figure supplement 1b, d). In these plots, nodes are grouped by canonical network (see *Functional parcellation and network definition*), and reliably selected edges are represented by lines between these nodes. The color family of each edge indicates the sign of its predictive contribution (red for positive, blue for negative), and the shade and thickness of the line represent the magnitude of its predictive contribution (darker and thicker indicate greater contribution). Figure 3c and Figure 3—figure supplement 1b circle plots illustrate the predictive contributions of ciFC features for each task. Figure 3f and Figure 3—figure supplement 1d circle plots illustrate the predictive contributions of context-dependent features with predictive utility. That is, lines again represent edges (here, cdFC), with the addition of node-level information. Nodes, displayed as circles on each circle plot track, are colored by their mean activation difference for the given contrast (e.g., 2-back – 0-back), and their distance from the x axis indicates their signed predictive contribution. To avoid any bias in task activation beta estimates from the inclusion of additional predictors in the PPI analysis (but see *Investigating potential confounds* and Figure 2—figure supplement 2 for evidence that PPI task activation betas closely follow independently estimated task effect sizes), activation was calculated for this analysis and for the inter-subject consistency/utility regression (see *Modeling the relationship between inter-subject consistency and predictive utility*) as follows. As in work by Salehi and colleagues (Salehi et al., 2019), we used individual-level, volume-based task contrast of parameter estimate (COPE) files generated and described previously (Glasser et al., 2013) using FSL FEAT's FLAME (FMRIB's Local Analysis

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of Mixed Effects [Smith et al., 2004]) to calculate effect size for each voxel. One-sample *t*-statistics were calculated at each voxel and converted to Cohen's *d* coefficients as $d_s = \frac{t}{\sqrt{N}}$, where d_s is the sample *d* coefficient, *t* is the *t*-test statistic, and *N* is the sample size (J. Cohen, 2013). We then applied the initial 268-parcel group-level parcellation (X. Shen et al., 2013) to these voxel-level maps to calculate a mean task effect size per parcel for each of the tasks.

Second, predictive contributions were visualized at the network level, using ten canonical networks (see *Functional parcellation and network definition*; Figures 3b, e, h, i, 4b, c, 5a, and Figure 3—figure supplement 1a, c, e, f, and Figure 3—figure supplements 3 and 4). Given the potentially divergent interpretations of the predictive contribution of a positive and a negative edge or node (e.g., a positive contribution for a negative edge suggests that it is less negative [weaker] in those with higher fluid intelligence, while the same contribution for a positive edge suggests that it is more positive [stronger] in those with higher fluid intelligence), edges were first divided by their mean ciFC sign (i.e., edges that, at baseline, are on average positive across all subjects, and edges that, at baseline, are on average negative across all subjects; Figure 3b, e, i, Figure 3—figure supplement 1a, c, f, and Figure 3—figure supplement 3), and nodes by their mean task activation sign (Figure 3h, Figure 3—figure supplement 1e). For each group, the signed contributions of selected edges from each network pair were summed, and this sum was normalized by the total number of edges between the given networks to account for differences in network size, yielding the mean contribution for an edge in the given network pair. Finally, to increase interpretability of the scale for these network-level contributions, each network pair's contribution value was normalized by the summed absolute contributions of all network pairs for that term in both positive and negative edge groups. This analysis was repeated, with minor modifications, for task activation for each network, rather than network pair. That is, activation predictive utility was summed for all nodes in each network, and this value was normalized by the number of nodes in that network; resulting network contributions were then normalized by the summed absolute contributions of all networks for that task in both positive and negative node groups. To explore any differences in the spatial distribution of high-consistency predictive edges and low-consistency predictive edges, this analysis was repeated after splitting the reliably selected edges not by mean ciFC sign, but rather by the median absolute inter-subject consistency (see *Inter-subject psychophysiological interaction analysis*) for these selected edges for the given term and task.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Finally, predictive contributions were visualized at the node level (Figure 3a, d, g). Node percent contribution was quantified as the summed, absolute contributions (i.e., predictive utility) of all edges incident to it (or, in the case of task activation, as the absolute predictive utility of the given node), normalized by the sum of all nodes' absolute predictive utility for the given task. Because the percent predictive utility of node activation calculated in this way tended to be sparsely concentrated (i.e., high percentages for a few nodes), the colormap range was set so as to balance capturing the full range of percentages across all tasks and terms and the subtle differences among nodes' ciFC and cdFC percent predictive utility. To do so, the maximum was set such that a small number of nodes' percent predictive utility of activation was saturated (0-10 nodes per task).

Inter-subject psychophysiological interaction analysis

To evaluate inter-subject consistency separately for ciFC, task activation, and cdFC, the intra-subject PPI analysis was repeated with one modification: the predictor node's time course was averaged across the subset of all subjects that did not include the target subject (and that experienced the same task stimulus order; range across tasks = 608-703 subjects), and this process was repeated iteratively with each subject serving once as the target subject, yielding, again, one asymmetric matrix per term for each subject. That is,

$$v_{i,t} = \beta_0 + \beta_1 v_{j\bar{p}} + \beta_{2:k} T_{1:k-1} + \beta_{k+1:2k-1} bin(T_{1:k-1}) v_{j\bar{p}}$$

$$v_{j\bar{p}} = \frac{1}{n-1} \sum_{p \neq t} \boldsymbol{node}_{j,p}$$

where t is the target subject, p are the non-target subjects, and $\boldsymbol{node}_{j,p}$ represents the time course of the jth node for subject p. These matrices were averaged across all subjects, and, as before, contrasts were calculated, LR/RL contrast matrices averaged, and resulting beta contrast matrices symmetrized, yielding one matrix per term. Inter-subject consistency of task activation was defined as the main diagonal of the inter-subject interaction matrix. This was repeated for each of the five tasks for which task timing was meaningfully synchronized across subjects: gambling, emotion, WM, social, and relational. All subsequently described analyses using inter-subject consistency results were limited to these five tasks.

To explore the spatial distribution of ciFC and cdFC inter-subject consistency, the mean consistency of an edge in each canonical network pair was visualized (Figure 4b, c), using the same approach as in Figures 3 and 5 (see *Evaluating and visualizing contributions to a predictive model*), again after dividing these edges into those with mean positive ciFC and those with mean

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

negative ciFC (Figure 4—figure supplement 1c-d). We note that self-connections (e.g., node 1 – node 1) are constant in intra-subject analyses and thus non-contributory to predictive models. In inter-subject analyses, however, these self-connections correspond to inter-subject correlation (ISC; Hasson et al., 2004), or the similarity in a given node's time course across individuals. These connections are thus neither positive nor negative in the intra-subject analyses, but are included in Figure 4b-c, and context-dependent ISC is interpreted as consistency of task activation across individuals in subsequent analyses.

Modeling the relationship between inter-subject consistency and predictive utility

To investigate a potential relationship between predictive utility and feature consistency across individuals, we designed a regression-based analysis in which the combined model predictive utility (i.e., absolute contribution) of each reliably selected feature for the given task and term was used as the outcome variable, and the inter-subject consistency of each reliably selected feature (i.e., absolute inter-subject PPI beta) for the corresponding task and term was used as a predictor. We suspected that this relationship may interact with functional and anatomical relationships; to this end, we built five separate models, each with a different functional or anatomical variable explicitly modeled as a predictor: edge location within or between canonical networks (Figure 5b), edge location within or across hemispheres (Figure 5—figure supplement 1b), edge length (as measured by Euclidean distance between the nodes incident to that edge; Figure 5—figure supplement 1a), edge reliability (calculated in HCP resting-state data by Noble and colleagues [Noble et al., 2017]; Figure 5b), and task activation (i.e., mean absolute task effect size [see *Evaluating and visualizing contributions to a predictive model*] of the nodes incident to the given edge; Figure 5b). All non-dummy predictors were mean-centered. Both main effects and interaction terms of models with significant full-model *P* values are presented; betas of non-significant models were set to 0. Due to the limited number of measurements (i.e., five betas per model term, one per task) output by this analysis, results are discussed qualitatively, with the caveat that future work to replicate these findings with more tasks will permit more rigorous statistical testing of them (see *Discussion*).

Investigating potential confounds

To evaluate the potential impact of collinearity on PPI beta estimates, intra- and inter-subject PPI analyses were repeated with two partial models for each task: one without the ciFC term and one without the cdFC term. Partial-model and full-model betas for each term were

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

highly correlated, suggesting a minimal impact of collinearity on beta estimates (Figure 1—figure supplement 2 and Figure 4—figure supplement 2).

To ensure that ciFC and task activation beta estimates are comparable to standard measures of FC and task activation, respectively, ciFC betas were correlated with FC calculated using Fisher-transformed Pearson correlations across the entire node time courses ("standard FC"), mean task activation betas were correlated with group-level task effect size measures calculated for each node (see *Evaluating and visualizing contributions to a predictive model*), and HCP-released, individual-level GLM results were correlated with PPI task activation betas and averaged across subjects, for all subjects in the main sample for whom individual-level GLM results were available (n = 322; all results in Figure 2—figure supplement 2). These individual-level GLM-based task activation vectors (i.e., parcellated t-statistic maps generated in the task effect size analysis, yielding a 268-element vector of node activation values for each task for each subject) were also used to predict gF; results were comparable to prediction using PPI activation betas, confirming that task activation alone does not predict gF (Figure 2—figure supplement 2). For completeness, cdFC beta estimates were also correlated with standard FC. As predicted, results (Figure 2—figure supplement 2) demonstrate that ciFC is strongly correlated with standard FC and task activation betas with task effect size (both at the group and individual levels), but that cdFC is not significantly correlated with standard FC, validating the interpretation of PPI betas.

To further ensure that methodological choices did not affect main results, we repeated main analyses with mean-centered PPI regressors (that is, after z-scoring each PPI predictor immediately prior to the regression step). Prediction results were largely unchanged (Figure 2—figure supplement 1), though the intercept term, by definition, failed to predict gF in the mean-centered case (numerical error will yield intercept values that are close, but not equal, to zero, but this fluctuation around zero should not—and did not—predict gF). While at first surprising that prediction results are comparable using these two approaches, it follows from the similarity of task timing across subjects and from our choice to z-score node timecourses. That is, mean centering PPI regressors causes a linear scaling of resulting betas that is comparable across subjects, since $\hat{\beta}_{standardized} = \hat{\beta}_{unstandardized} / s_{regressor}$, and the standard deviation of task timing and of the interaction will be nearly identical across subjects. This linear shift in PPI betas will change their interpretation, but this change is not germane to the present work, as PPI betas themselves are not interpreted. However, with the exception of the intercept term, PPI betas'

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

predictive utility will be unchanged, since inter-subject relationships of PPI betas are relatively unchanged. Further, predictive utility estimates will be essentially unchanged, as ridge coefficients scale with PPI beta variance. The intercept will, of course, approach zero in the mean-centered case, but in the zero-centered case will reflect these linear shifts, scaled by subject-specific PPI activation and cdFC betas (Figure 2—figure supplement 1). If these betas meaningfully vary across subjects, then the intercept terms may predict individuals' phenotypes, as was found to be true in these analyses (Figure 2). Given the predictive utility of this "overall task effect" term (i.e., intercept), the unchanged prediction results for other terms, and the decreased collinearity among predictors after zero centering relative to mean centering, we present zero-centered results in the main text, and mean-centered results in the supplementary materials.

Finally, while standard approaches were taken to mitigate the effects of motion on fMRI data, we sought to more thoroughly explore any relationship of motion to task timing by correlating, for each subject, frame-to-frame displacement (HCP Movement_RelativeRMS.txt) for each task and phase encoding direction with the corresponding task timing regressors. Results (Figure 2—figure supplement 10) demonstrate no consistent relationship. We also correlated mean RMS relative motion for each subject and task (averaged over phase encoding runs) with observed gF (Figure 2—figure supplement 6) and predicted gF (averaged over 100 iterations for each task/term; Figure 2—figure supplement 7). Given several modest correlations, we repeated the prediction analyses using partial correlation-based feature selection with mean RMS relative motion (calculated for each subject and task) as a covariate. As an even more conservative motion control analysis, we also repeated the main analysis after regressing mean RMS relative motion out of gF and FC within the cross-validation loop. Regression coefficients were estimated for the training subjects and applied to the test subjects in the 10-fold analysis (to avoid the use of potentially unstable coefficient estimates in the smaller test sample), but were estimated separately for training and test subjects in the split-half analysis. Model performance and feature weights were relatively unchanged; the former is presented in Figure 2—figure supplement 8, and the correlations of feature weights from main analyses and partial correlation-based analyses are presented in Figure 2—figure supplement 9. Correlations of feature weights from main analyses and residual-based analyses were comparable.

Data and code availability

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The HCP data that support the findings of this study are publicly available on the ConnectomeDB database (https://db.humanconnectome.org). MATLAB code to run the ridge regression-based CPM analysis can be found at https://github.com/YaleMRRC/CPM. MATLAB code to run additional core analyses (PPI analyses, basic visualization, synchrony vs. predictiveness analyses, and family-based cross-validation) can be found at https://github.com/abigailsgreene/taskFC. BioImage Suite tools used for analysis and visualization can be accessed at www.bisweb.yale.edu. MATLAB and R scripts written to perform additional post-hoc analyses and visualizations are available from the authors upon request.

Acknowledgments

Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. The authors would like to thank Nick Turk-Browne and Daeyeol Lee for helpful discussions and suggestions.

Author contributions

R.T.C., D.S., and A.S.G. designed the study. S.N. developed the scripts used to preprocess the HCP data, and A.S.G. preprocessed the data with guidance from S.N. S.G. developed and validated the ridge regression-based predictive modeling pipeline. A.S.G. developed and validated the PPI and visualization code and performed all analyses, with the exception of task effect size calculation (designed and overseen by S.N.). D.S. and R.T.C. supported result interpretation. A.S.G. wrote the manuscript, with comments from all authors.

Declaration of interests

The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

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

Bilker, W. B., Hansen, J. A., Brensinger, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment*, *19*(3), 354–369. https://doi.org/10.1177/1073191112446655

Biswal, B., Yetkin, F. Z., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, *34*(4), 537–41. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8524021

Bolt, T., Nomi, J. S., Rubinov, M., & Uddin, L. Q. (2017). Correspondence between evoked and intrinsic functional brain network configurations. *Human Brain Mapping*, *38*(4), 1992–2007. https://doi.org/10.1002/hbm.23500

Buckner, R. L., Krienen, F. M., & Yeo, B. T. T. (2013). Opportunities and limitations of intrinsic functional connectivity MRI. *Nature Neuroscience*, *16*(7), 832–837. https://doi.org/10.1038/nn.3423

Bullmore, E., & Sporns, O. (2012). The economy of brain network organization. *Nature Reviews Neuroscience*, *13*(5), 336–349. https://doi.org/10.1038/nrn3214

Chee, M. W. L., Venkatraman, V., Westphal, C., & Siong, S. C. (2003). Comparison of block and event-related fMRI designs in evaluating the word-frequency effect. *Hum. Brain Mapping*, *18*, 186–193. https://doi.org/10.1002/hbm.10092

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Routledge. https://doi.org/10.4324/9780203771587

Cohen, J. R. (2018). The behavioral and cognitive relevance of time-varying, dynamic changes in functional connectivity. *NeuroImage*, *180*, 515–525. https://doi.org/10.1016/J.NEUROIMAGE.2017.09.036

Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S., & Petersen, S. E. (2014). Intrinsic and task-evoked network architectures of the human brain. *Neuron*, *83*(1), 238–251. https://doi.org/10.1016/j.neuron.2014.05.014

Cole, M. W., Ito, T., Schultz, D., Mill, R., Chen, R., & Cocuzza, C. (2019). Task activations produce spurious but systematic inflation of task functional connectivity estimates. *NeuroImage*, *189*, 1–18. https://doi.org/10.1016/J.NEUROIMAGE.2018.12.054

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience*, *16*(9), 1348–1355. https://doi.org/10.1038/nn.3470

Di, X., & Biswal, B. B. (2017). Psychophysiological interactions in a visual checkerboard task: reproducibility, reliability, and the effects of deconvolution. *Frontiers in Neuroscience*, *11*, 573. https://doi.org/10.3389/fnins.2017.00573

Di, X., & Biswal, B. B. (2018). Toward task connectomics: examining whole-brain task modulated connectivity in different task domains. *Cerebral Cortex*, 1–12. https://doi.org/10.1093/cercor/bhy055

Di, X., Gohel, S., Kim, E. H., & Biswal, B. B. (2013). Task vs. rest-different network configurations between the coactivation and the resting-state brain networks. *Frontiers in Human Neuroscience*, *7*, 493. https://doi.org/10.3389/fnhum.2013.00493

Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., … Petersen, S. E. (2006). A core system for the implementation of task sets. *Neuron*,

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

50(5), 799–812. https://doi.org/10.1016/j.neuron.2006.04.031

Dubois, J., Galdi, P., Han, Y., Paul, L. K., & Adolphs, R. (2018). Resting-state functional brain connectivity best predicts the personality dimension of openness to experience. *Personality Neuroscience*, *1*. https://doi.org/10.1017/pen.2018.8

Elliott, M. L., Knodt, A. R., Cooke, M., Kim, M. J., Melzer, T. R., Keenan, R., … Hariri, A. R. (2019). General functional connectivity: Shared features of resting-state and task fMRI drive reliable and heritable individual differences in functional brain networks. *NeuroImage*, *189*, 516–532. https://doi.org/10.1016/J.NEUROIMAGE.2019.01.068

Elton, A., & Gao, W. (2015). Task-related modulation of functional connectivity variability and its behavioral correlations. *Human Brain Mapping*, *36*(8), 3260–3272. https://doi.org/10.1002/hbm.22847

Finn, E. S., Scheinost, D., Finn, D. M., Shen, X., Papademetris, X., & Constable, R. T. (2017). Can brain state be manipulated to emphasize individual differences in functional connectivity? *NeuroImage*, *160*, 140–151. https://doi.org/10.1016/j.neuroimage.2017.03.064

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

Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N. A., & Dale, A. M. (1999). Stochastic designs in event-related fMRI. *NeuroImage*, *10*(5), 607–619. https://doi.org/10.1006/NIMG.1999.0498

Gao, S., Greene, A. S., Constable, R. T., & Scheinost, D. (2019). Combining multiple connectomes improves predictive modeling of phenotypic measures. *NeuroImage*, *201*, 116038. https://doi.org/10.1016/J.NEUROIMAGE.2019.116038

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

Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., … Petersen, S. E. (2018). Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron*, *98*(2), 439–452.e5. https://doi.org/10.1016/j.neuron.2018.03.035

Greene, A. S., Gao, S., Scheinost, D., & Constable, R. T. (2018). Task-induced brain state manipulation improves prediction of individual traits. *Nature Communications*, *9*(1), 2807. https://doi.org/10.1038/s41467-018-04920-3

Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, *32*(18), 2847–2849. https://doi.org/10.1093/bioinformatics/btw313

Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). circlize implements and enhances circular visualization in R. *Bioinformatics*, *30*(19), 2811–2812. https://doi.org/10.1093/bioinformatics/btu393

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*(5664), 1634–1640.

Horien, C., Greene, A. S., Constable, R. T., & Scheinost, D. (2019). Regions and connections: complementary approaches to characterize brain organization and function. *The Neuroscientist*, 1–17. https://doi.org/10.1177/1073858419860115

Hsu, W.-T., Rosenberg, M. D., Scheinost, D., Constable, R. T., & Chun, M. M. (2018). Resting-state functional connectivity predicts neuroticism and extraversion in novel individuals. *Social Cognitive and Affective Neuroscience*, *13*(2), 224–232.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

https://doi.org/10.1093/scan/nsy002

Joshi, A., Scheinost, D., Okuda, H., Belhachemi, D., Murphy, I., Staib, L. H., & Papademetris, X. (2011). Unified framework for development, deployment and robust testing of neuroimaging algorithms. *Neuroinformatics*, *9*(1), 69–84. https://doi.org/10.1007/s12021-010-9092-8

Kieliba, A. P., Madugula, S., Filippini, N., & Duff, E. P. (2019). Large-scale intrinsic connectivity is consistent across varying task demands. *PLoS ONE*, *14*(4), e0213861. https://doi.org/https://doi.org/10.1371/journal.pone.0213861

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863-3868. https://doi.org/10.1073/pnas.0600244103

Krienen, F. M., Yeo, B. T. T., & Buckner, R. L. (2014). Reconfigurable task-dependent functional coupling modes cluster around a core functional architecture. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1653). https://doi.org/10.1098/rstb.2013.0526

Leonardi, N., Shirer, W. R., Greicius, M. D., & Van De Ville, D. (2014). Disentangling dynamic networks: Separated and joint expressions of functional connectivity patterns in time. *Human Brain Mapping*, *35*(12), 5984–5995. https://doi.org/10.1002/hbm.22599

Li, J., Kong, R., Liégeois, R., Orban, C., Tan, Y., Sun, N., … Yeo, B. T. T. (2019). Global signal regression strengthens association between resting-state functional connectivity and behavior. *NeuroImage*, *196*, 126–141. https://doi.org/10.1016/J.NEUROIMAGE.2019.04.016

Liu, T. T., Nalci, A., & Falahpour, M. (2017). The global signal in fMRI: Nuisance or information? *NeuroImage*, *150*, 213–229. https://doi.org/10.1016/j.neuroimage.2017.02.036

Lynch, L. K., Lu, K.-H., Wen, H., Zhang, Y., Saykin, A. J., & Liu, Z. (2018). Task-evoked functional connectivity does not explain functional connectivity differences between rest and task conditions. *Human Brain Mapping*, *39*(12), 4939–4948. https://doi.org/10.1002/hbm.24335

McLaren, D. G., Ries, M. L., Xu, G., & Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. *NeuroImage*, *61*(4), 1277–86. https://doi.org/10.1016/j.neuroimage.2012.03.068

Medaglia, J. D., Lynall, M.-E., & Bassett, D. S. (2015). Cognitive network neuroscience. *Journal of Cognitive Neuroscience*, *27*(8), 1471–1491. https://doi.org/10.1162/jocn_a_00810

Neuwirth, E. (2014). RColorBrewer. *R Package Version1.1-2*. Retrieved from https://cran.r-project.org/package=RColorBrewer

Newell, A. (1973). *You can't play 20 questions with nature and win: projective comments on the papers of this symposium*. Retrieved from http://repository.cmu.edu/compsci

Noble, S., Spann, M. N., Tokoglu, F., Shen, X., Constable, R. T., & Scheinost, D. (2017). Influences on the test–retest reliability of functional connectivity MRI and its relationship with behavioral utility. *Cerebral Cortex*, *27*(11), 5415–5429. https://doi.org/10.1093/cercor/bhx230

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. https://doi.org/10.1016/j.tics.2006.07.005

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.

Power, J. D., Schlaggar, B. L., & Petersen, S. E. (2014). Studying brain organization via spontaneous fMRI signal. *Neuron*, *84*(4), 681–696.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

https://doi.org/10.1016/j.neuron.2014.09.007

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2015). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*, *19*(1), 165–171. https://doi.org/10.1038/nn.4179

Salehi, M., Greene, A. S., Karbasi, A., Shen, X., Scheinost, D., & Constable, R. T. (2019). There is no single functional atlas even for a single individual: Functional parcel definitions change with task. *NeuroImage*, 116366. https://doi.org/10.1016/J.NEUROIMAGE.2019.116366

Scheinost, D., Noble, S., Horien, C., Greene, A. S., Lake, E. M., Salehi, M., … Constable, R. T. (2019). Ten simple rules for predictive modeling of individual differences in neuroimaging. *NeuroImage*, *193*, 35–45. https://doi.org/10.1016/J.NEUROIMAGE.2019.02.057

Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., & Constable, R. T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, *12*(3), 506–518. https://doi.org/10.1038/nprot.2016.178

Shen, X., Tokoglu, F., Papademetris, X., & Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage*, *82*, 403–415. https://doi.org/10.1016/j.neuroimage.2013.05.081

Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., & Hasson, U. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature Communications*, *7*, 12141. https://doi.org/10.1038/ncomms12141

Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., … Glasser, M. F. (2013). Resting-state fMRI in the Human Connectome Project. *NeuroImage*, *80*(80), 144–168. https://doi.org/10.1016/j.neuroimage.2013.05.039

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., … Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(31), 13040–13045. https://doi.org/10.1073/pnas.0905267106

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., … Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, *23*(SUPPL. 1), S208–S219. https://doi.org/10.1016/j.neuroimage.2004.07.051

Turk-Browne, N. B. (2013). Functional interactions as big data in the human brain. *Science*, *342*(6158), 580–584. Retrieved from http://science.sciencemag.org/content/342/6158/580.full

Uğurbil, K., Xu, J., Auerbach, E. J., Moeller, S., Vu, A. T., Duarte-Carvajalino, J. M., … Yacoub, E. (2013). Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *NeuroImage*, *80*(80), 80–104. https://doi.org/10.1016/j.neuroimage.2013.05.012

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: an overview. *NeuroImage*, *80*(80), 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

Wager, T. D., Atlas, L. Y., Lindquist, M. A., Roy, M., Woo, C.-W., & Kross, E. (2013). An fMRI-based neurologic signature of physical pain. *New England Journal of Medicine*, *368*(15), 1388–1397. https://doi.org/10.1056/NEJMoa1204471

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, *92*, 381–397. https://doi.org/10.1016/j.neuroimage.2014.01.060

Winkler, A. M., Webster, M. A., Vidaurre, D., Nichols, T. E., & Smith, S. M. (2015). Multi-level

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

block permutation. *NeuroImage*, *123*, 253–268. https://doi.org/10.1016/j.neuroimage.2015.05.092

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

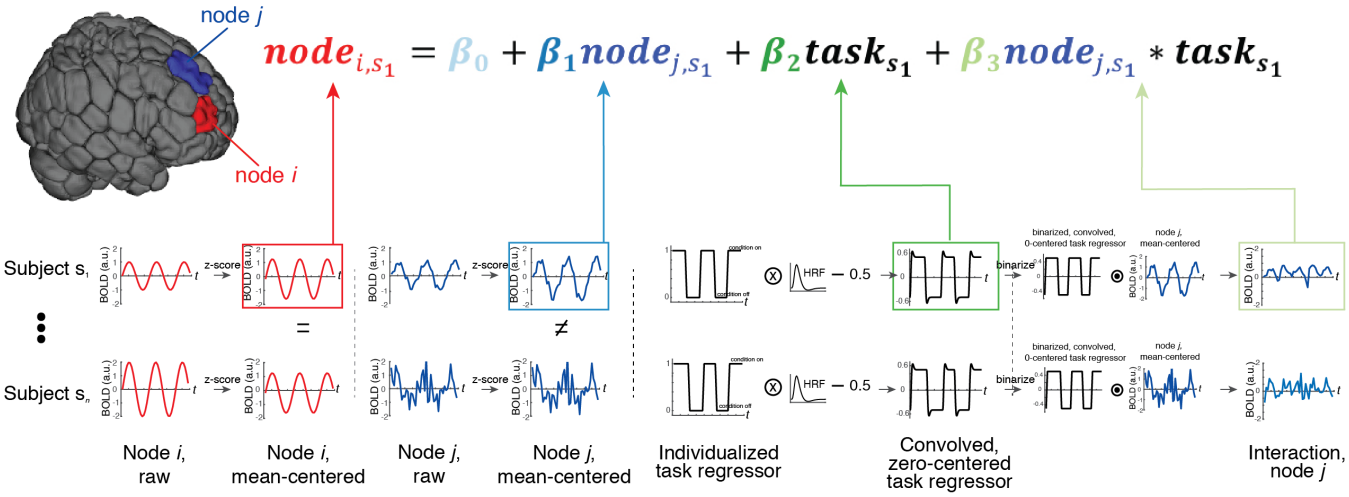

Figure 1—figure supplement 1. Schematic depiction of the PPI analysis pipeline.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

	Intercept, no cdFC	Intercept, no ciFC	ciFC, no cdFC	Activation, no cdFC	Activation, no ciFC	cdFC, no ciFC
Gambling	0.9910	0.9587	0.969	0.9941	0.9749	0.7439
Emotion	0.9892	0.9654	0.9918	0.9903	0.9569	0.9343
Language	0.9779	0.9337	0.9617	0.9877	0.9558	0.7679
Motor	0.9754	0.9961	0.5986	0.9751	0.9915	0.7329
Relational	0.9821	0.9438	0.9881	0.9878	0.9542	0.9290
Social	0.9868	0.9275	0.9661	0.9925	0.9570	0.7050
WM	0.9848	0.9228	0.9862	0.9891	0.9548	0.8920

Figure 1—figure supplement 2. Mean Pearson correlation coefficients between full-model PPI betas and partial-model PPI betas for a given term, computed within subject (separately for LR and RL runs) and averaged across subjects and then conditions (e.g. cue, 2-back, and 0-back terms for WM).

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

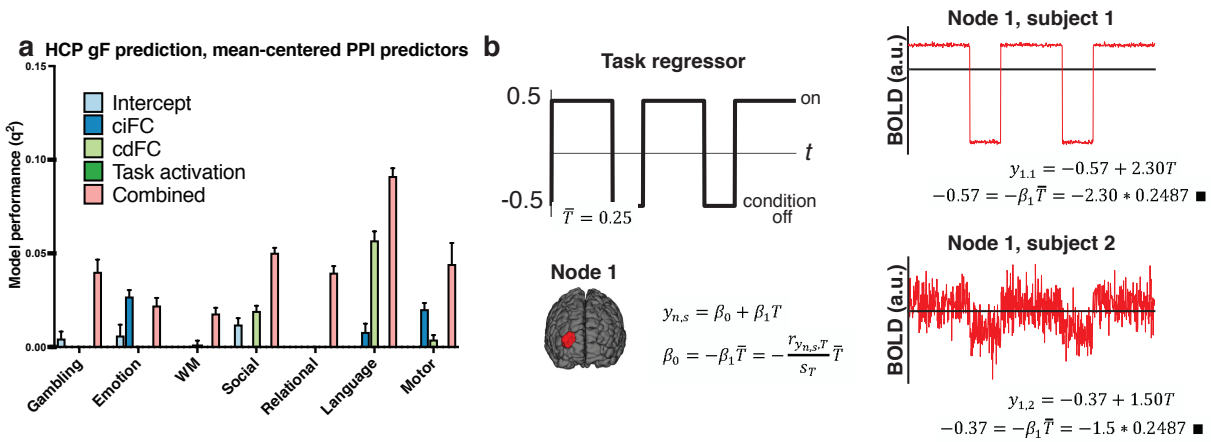

Figure 2—figure supplement 1. PPI intercept reflects overall degree of task effect on brain activity patterns, and has predictive utility. (a) Eliminating PPI model intercepts by mean centering all regressors (rather than zero centering task regressors and mean centering node time courses, as in main analyses) has little effect on prediction performance (mean and s.d., indicated by error bars) of ciFC, cdFC, activation, or combined models (c.f., Fig. 2b) because removal of a constant term (that is similar across subjects) from regressors scales beta estimates in a linear fashion, and thus does not affect their phenotypically relevant information content. However, the intercept reflects this constant term, weighted by activation and interaction betas, which may vary across subjects, reflecting inter-individual differences in overall degree of task effect on node activity and FC; should this inter-individual variance be related to gF, PPI intercept terms would be expected to successfully predict gF, as was found to be the case for 4/7 tasks (Fig. 2). (b) Schematic illustration of this interpretation of the intercept term. Here, task timing is represented as a zero-centered boxcar, yielding a mean task regressor value of 0.25 (because more time is spent in the "on" condition than in the "off" condition). Example node 1 time courses for two subjects are depicted in the rightmost panel; in subject 1, node 1 is strongly activated by the task, while in subject 2, node 1 is more weakly activated by the task, as reflected in the $ß_1$ estimates for these subjects. This difference, in turn, determines the intercept values for each subject. This simple case excludes the PPI interaction term, but the same logic would hold for this term, rendering the intercept a reflection of the overall degree to which the task affects activation and connectivity of the target node, a value which may vary meaningfully across individuals. T, task regressor; y, node time course of activity; n, node; s, subject; *r*, Pearson correlation; *s*, standard deviation.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

	ciFC/standard FC corr	cdFC/standard FC corr	Act. PPI/GLM grp corr	Act. PPI/GLM IS corr	GLM-based model performance
Emotion	0.973	-0.0975 (0.25)	0.9360	0.9196	-0.0093
Gambling	0.9364	-0.0049 (0.35)	0.7798	0.6725 (0.0001)	-0.0090
Language	0.9512	0.0206 (0.39)	0.9363	0.9369	-0.0100
Motor	0.6499	-0.0717(0.43)	0.8897	0.8259 (0.0001)	-0.0101
Relational	0.9546	-0.0123 (0.34)	0.9457	0.7760	-0.0092
Social	0.9278	-0.0064 (0.42)	0.92	0.8198	-0.0099
WM	0.9554	-0.0428 (0.35)	0.9472	0.8448 (0.0016)	-0.0087

Figure 2—figure supplement 2. Comparison of PPI results to "standard" FC and activation results. Columns 1-2: Pearson correlations of context-independent and context-dependent FC for each task with standard FC from that task, computed within subject and averaged across subjects to yield mean *r*(*P*, Bonferroni corrected). *P* value not reported indicates $P << 0.001$. Column 3: Pearson correlations of task activation PPI betas (averaged across subjects) with independently estimated, group-level HCP task effect sizes per node. All $P << 0.001$, Bonferroni corrected. Column 4: Pearson correlations of task activation PPI betas with independently estimated, individual-level HCP task effect sizes per node; intra-subject correlations averaged across subjects and presented as mean *r(P*, Bonferroni corrected). *P* value not reported indicates $P << 0.001$. Column 5: performance of gF predictive models trained and tested with GLM-based activation, rather than PPI-based activation. Results reported as mean q^2 across 100 iterations of 10-fold cross-validation. Act, activation; corr, correlation; grp, group; IS, intra-subject.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

	Prediction performance, no GSR	Prediction performance, with GSR
Emotion	0.0305 (0.0053)	0.0771 (0.0036)
Gambling	-0.0017 (0.0028)	0.1028 (0.0047)
Language	0.0387 (0.0046)	0.1046 (0.0035)
Motor	-0.0029 (0.0024)	0.0823 (0.0051)
Relational	0.0280 (0.0068)	0.0990 (0.0047)
Social	-0.0042 (0.002)	0.0872 (0.0044)
WM	-0.0029 (0.0022)	0.1092 (0.0037)

Figure 2—figure supplement 3. Prediction performance of rCPM (100 iterations, *P* threshold = 0.1) performed on "standard FC," with and without global signal regression (GSR). Results presented as mean q^2 (s.d. q^2).

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

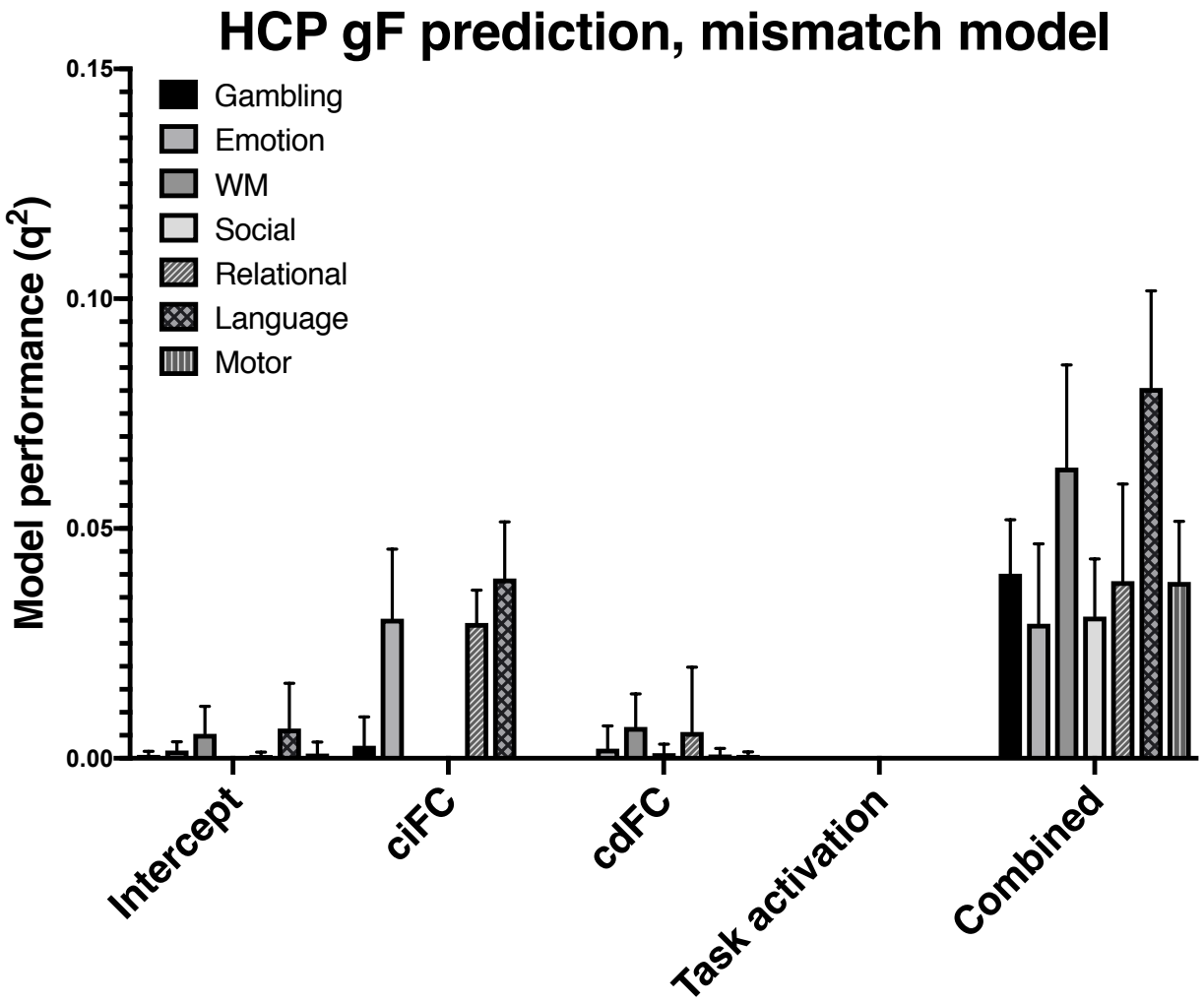

Figure 2—figure supplement 4. Combined and ciFC-based models, but not cdFC- or task activation-based models, successfully predict gF, even when incorrect task regressors are used. One iteration of 10-fold rCPM; 6 task/regressor combinations per task; models with $q^2<0$ were set to 0. Error bars indicated s.d. of prediction performance.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

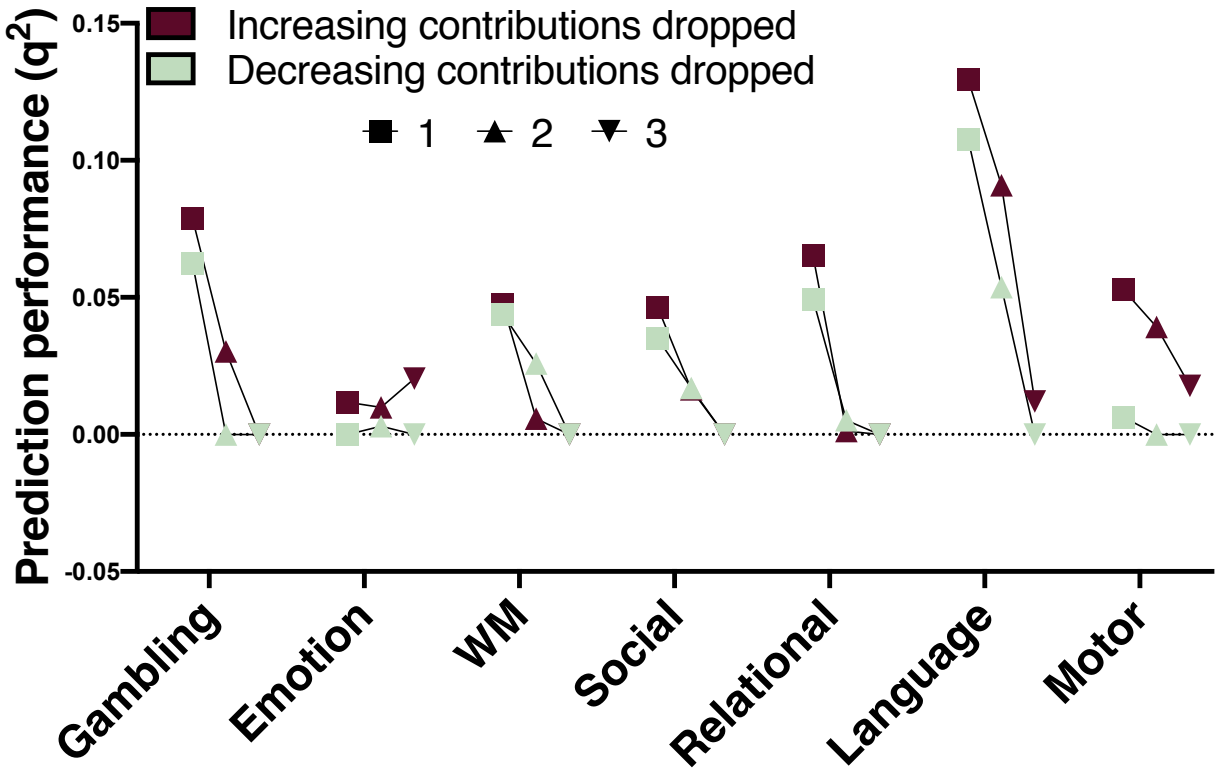

Figure 2—figure supplement 5. Combined model prediction performance with increasing numbers of individual terms dropped from the model in order of increasing (red) and decreasing (green) contributions. Numbers indicate number of dropped terms.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

	Motion/observed gF correlation
Emotion	-0.14*
Gambling	-0.06
Language	-0.09
Motor	-0.10
Relational	-0.15*
Social	-0.12*
WM	-0.08

Figure 2—figure supplement 6. Pearson correlation between mean RMS motion (averaged between LR and RL runs for each subject) and observed gF. *indicates significant at $P<0.05$, Bonferroni corrected.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

	Intercept	ciFC	cdFC	Activation	Combined
Emotion	0	-0.14*+	-0.02	0.11	-0.11
Gambling	0.04+	0.08	0	0.09	-0.03+
Language	-0.16*+	-0.09	-0.15*+	0.05	-0.17*+
Motor	-0.02	-0.20*+	0.04	0.11	-0.13*+
Relational	0.08	-0.05	0	0.08	-0.09+
Social	-0.02+	0.09	-0.01	0.09	-0.05+
WM	-0.06+	0	-0.03	-0.01	-0.09+

Figure 2—figure supplement 7. Pearson correlation between mean RMS motion (averaged between LR and RL runs for each subject) and predicted gF (averaged across 100 iterations; see main text) for all tasks and terms. *indicates significant at $P<0.05$, Bonferroni corrected; +indicates significant corresponding predictive model (see Fig. 2).

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

		Intercept	ciFC	cdFC	Activation	Combined
Emotion	*Parcorr (10/1)*	0.0076	0.0322	-0.0327	-0.0023	0.0255
	Residualized (10/1)	0.0061	0.0238	-0.0373	-1.02E-9	0.0211
	Original (10/100)	0.0042	0.0262	-0.0362	-0.005	0.0187
	Residualized (2/100)	0.0034	0.0176	-0.0301	1.78E-17	-0.0035
Gambling	*Parcorr (10/1)*	0.046	-0.0029	-0.015	-0.0029	0.0801
	Residualized (10/1)	0.0523	7.77E-4	-0.0104	-1.14E-9	0.0833
	Original (10/100)	0.0479	-0.0043	-0.0182	-0.0044	0.0783
	Residualized (2/100)	0.0302	0.0051	-0.0182	-3.44E-17	0.0499
Language	*Parcorr (10/1)*	0.0482	0.0088	0.0511	-0.0057	0.121
	Residualized (10/1)	0.0406	0.0079	0.0520	-2.39E-8	0.1198
	Original (10/100)	0.0467	0.0079	0.0559	-0.0044	0.1227
	Residualized (2/100)	0.0268	0.0231	0.0324	5.56E-07	0.0875
Motor	*Parcorr (10/1)*	-0.0027	0.0337	0.0087	-3.72E-04	0.051
	Residualized (10/1)	0.0024	0.0292	0.0060	-1.99E-7	0.0488
	Original (10/100)	-0.0025	0.0269	0.0041	-0.0046	0.0555
	Residualized (2/100)	7.94E-04	0.0217	5.59E-04	3.22E-17	0.0327
Relational	*Parcorr (10/1)*	-0.0039	-0.0039	-0.0142	-0.0039	0.0623
	Residualized (10/1)	-1.71E-8	0.0071	-0.0057	-1.71E-8	0.0622
	Original (10/100)	-0.0043	0.0026	-0.0138	-0.0042	0.0664
	Residualized (2/100)	0.0014	0.015	-0.0165	7.17E-06	0.0399
Social	*Parcorr (10/1)*	0.0326	-0.0018	0.0207	-0.0018	0.0621
	Residualized (10/1)	0.0309	-1.13E-5	0.0312	-8.12E-9	0.0596
	Original (10/100)	0.03	-0.0045	0.0162	-0.0046	0.0532
	Residualized (2/100)	0.0153	0.0035	0.0129	-5.55E-18	0.0449
WM	*Parcorr (10/1)*	0.0207	-0.003	0.0171	-0.0025	0.0526
	Residualized (10/1)	0.0216	4.38E-5	0.0011	-3.52E-7	0.0461
	Original (10/100)	0.0179	-0.004	0.0019	-0.0032	0.0454
	Residualized (2/100)	0.0186	0.0026	0.0179	6.77E-05	0.0536

Figure 2—figure supplement 8. Comparison of main results ("Original") to performance of predictive models built using partial correlation-based feature selection ("Parcorr," i.e., controlling for motion), and regression of motion values out of FC and gF ("Residualized"). In all cases, *P* threshold = 0.1. Numbers in parentheses indicate number of folds and iterations (e.g., "Parcorr (10/1)" indicates partial correlation-based feature selection with 10 folds and 1 iteration). For cases in which a single iteration of prediction was performed, results presented as q^2; for cases in which more than one iteration of prediction was performed, results presented as mean q^2.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

	Intercept	ciFC	cdFC	Activation	Combined
Emotion	NaN	0.9074	0.8881	NaN	0.9095
Gambling	0.9739	NaN	0.9300	NaN	0.9636
Language	0.9527	NaN	0.9368	NaN	0.9527
Motor	0.8852	0.9321	0.9265	NaN	0.9329
Relational	NaN	NaN	0.8916	NaN	0.9139
Social	0.9347	NaN	0.9267	NaN	0.9465
WM	0.8619	NaN	NaN	NaN	0.9779

Figure 2—figure supplement 9. Pearson correlation between consistently selected features' rCPM betas for the main models and the partial correlation-based models. NaN indicates that no features were selected on 75% or more feature selections for one or both model(s).

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

	Task timing/motion, LR correlation	Task timing/motion, RL correlation
Emotion	-0.0002	0.0030
Gambling	-0.0167	-0.0220
Language	-0.0132	-0.0124
Motor	0.0141	0.0065
Relational	0.0011	-0.0048
Social	-0.0292	-0.0344
WM	-0.0055	-0.0015

Figure 2—figure supplement 10. Rank correlation coefficients between frame-to-frame relative RMS motion and task timing, averaged across task conditions and subjects.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

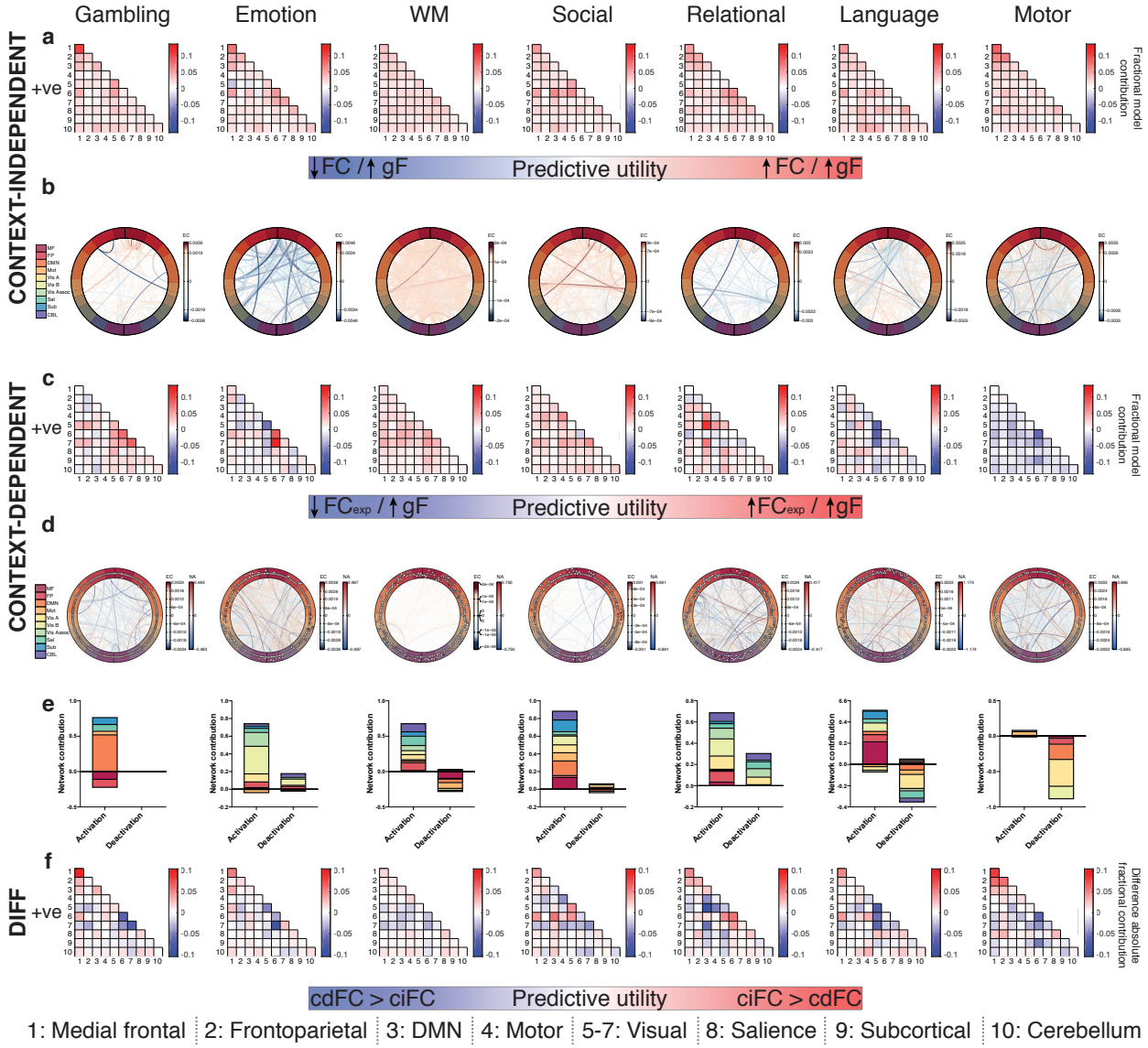

Figure 3—figure supplement 1. Context-independent FC, context-dependent FC, and activation predictive features are distributed and distinct. (a) Visualization of predictive ciFC features by network for each task. Red = positive ridge coefficients, blue = negative ridge coefficients, shade = relative model contribution. In this and all subsequent figures: "+ve" indicates that results reflect only contributions of edges with mean positive ciFC. 1-10 = network assignment. (b) Visualization of individual predictive ciFC features, with each consistently selected edge represented as a line; line color and thickness scale with predictive model contribution. In this and all subsequent figures, MF = medial frontal, FP = frontoparietal, DMN = default mode network, Mot = motor, Vis A = visual A, Vis B = visual B, Vis Assoc = visual association, Sal = salience, Sub = subcortical, CBL = cerebellum. EC, edge contribution. (c) Visualization of predictive cdFC features by network for each task. FC_{exp}, FC during the experimental condition (i.e., condition of interest). Red = positive ridge coefficients, blue = negative ridge coefficients, shade = relative model contribution. 1-10 = network assignment. (d) Visualization of individual predictive cdFC (lines) and activation (outer track circles) features; line color and thickness scale with cdFC feature predictive utility, and circles represent the corresponding nodes, with their color indicating mean activation (red = positive, blue = negative)

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and distance from the x axis indicating their model contribution. EC, edge contribution; NA, node activation. (e) Visualization of predictive activation features' (i.e., nodes') network assignments for nodes with mean positive PPI activation betas ("activation") and for nodes with mean negative PPI activation betas ("deactivation"). (f) Visualization of the difference between absolute, network-level ciFC model contributions (i.e., absolute value of matrices in [a]) and absolute, network-level cdFC model contributions (i.e., absolute value of matrices in [c]).

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

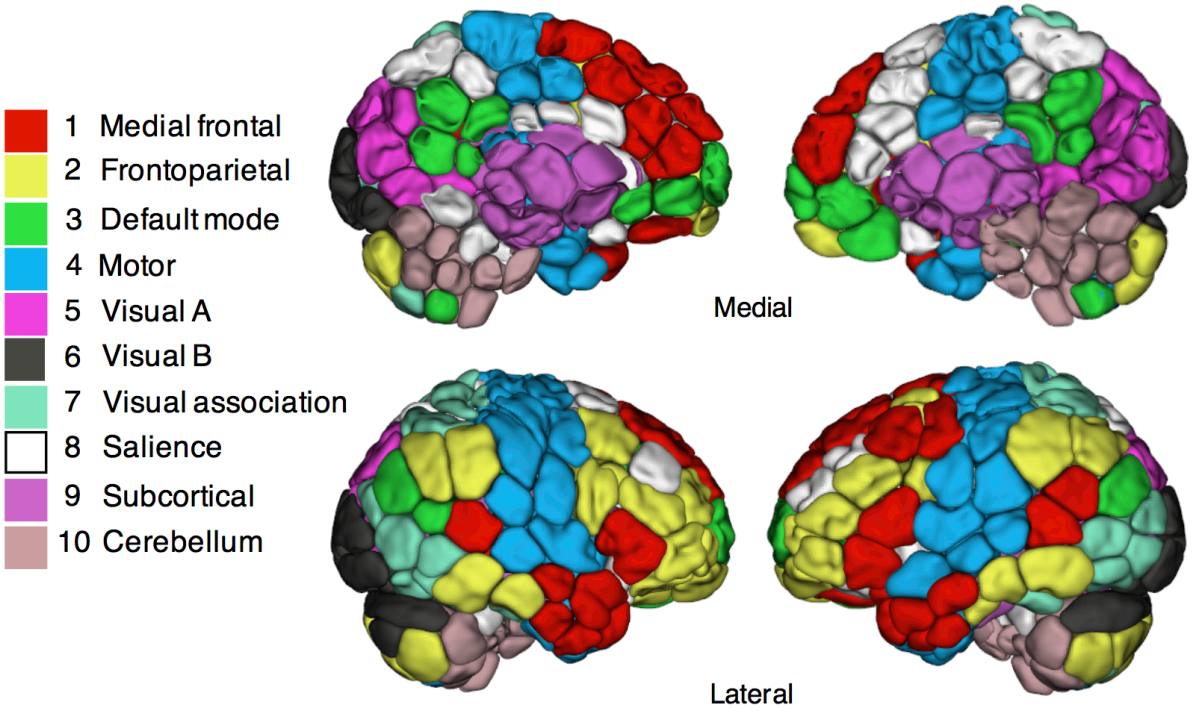

Figure 3—figure supplement 2. Ten canonical networks used for network analyses (see Methods for derivation). Figure adapted from Greene et al. (2018).

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

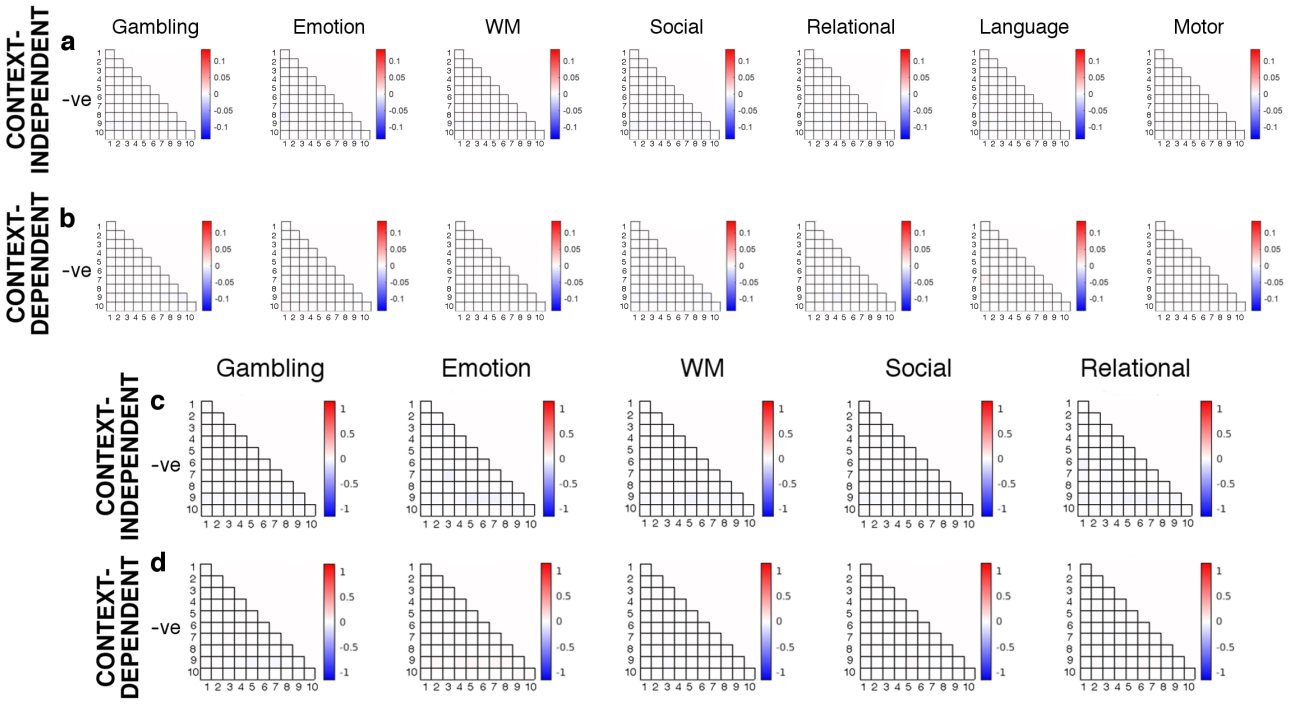

Figure 3—figure supplement 3. Predictive model contributions of features with mean negative ciFC (a,b) and inter-subject consistency of features with mean negative ciFC (c,d), summarized by canonical networks (Figure 3—figure supplement 2). Canonical network labels: 1 = medial frontal, 2 = frontoparietal, 3 = DMN, 4 = motor, 5-7 = visual, 8 = salience, 9 = subcortical, 10 = cerebellum.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

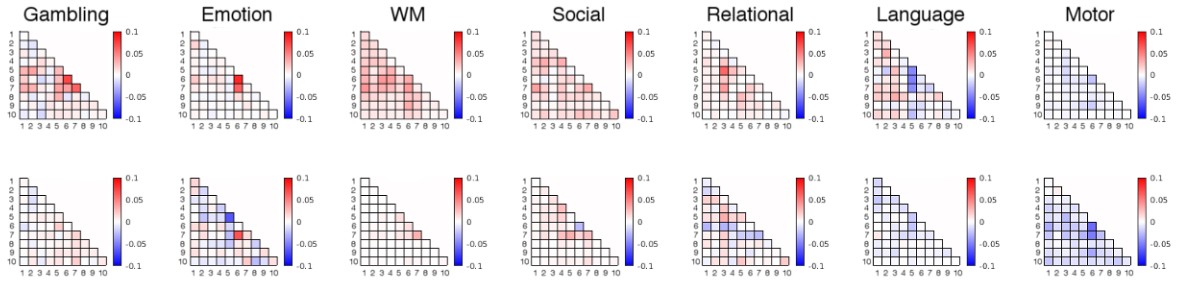

Figure 3—figure supplement 4. Further separating cdFC model contributions (for features with positive mean ciFC) by positive (top row) and negative (bottom row) cdFC offers additional insight into predictive task-induced changes in FC. Canonical network labels: 1 = medial frontal, 2 = frontoparietal, 3 = DMN, 4 = motor, 5-7 = visual, 8 = salience, 9 = subcortical, 10 = cerebellum.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

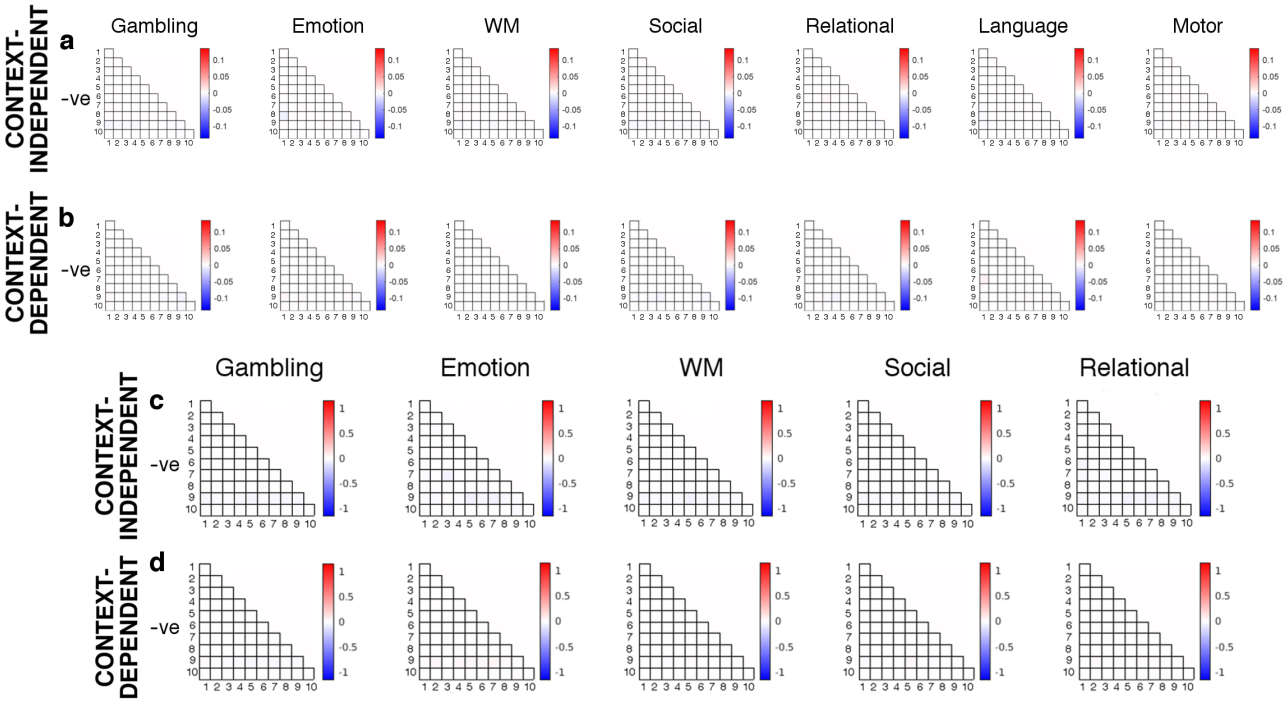

Figure 4—figure supplement 1. Predictive model contributions of features with mean negative ciFC (a,b) and inter-subject consistency of features with mean negative ciFC (c,d), summarized by canonical networks (Figure 3—figure supplement 2). Canonical network labels: 1 = medial frontal, 2 = frontoparietal, 3 = DMN, 4 = motor, 5-7 = visual, 8 = salience, 9 = subcortical, 10 = cerebellum. For clarity, duplicate of Figure 3—figure supplement 3.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

	Intercept, no cdFC	Intercept, no ciFC	ciFC, no cdFC	Activation, no cdFC	Activation, no ciFC	cdFC, no ciFC
Gambling	0.9830	0.7846	0.9368	0.9659	0.8441	0.9075
Emotion	0.9563	0.9045	0.9407	0.9477	0.8550	0.9370
Relational	0.9604	0.8388	0.9503	0.9428	0.8066	0.9655
Social	0.9749	0.7681	0.9308	0.9642	0.8190	0.8768
WM	0.9468	0.8164	0.9647	0.9515	0.8643	0.9684

Figure 4—figure supplement 2. Mean Pearson correlation coefficients between inter-subject PPI betas for full and partial models.

bioRxiv preprint doi: https://doi.org/10.1101/870287; this version posted December 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

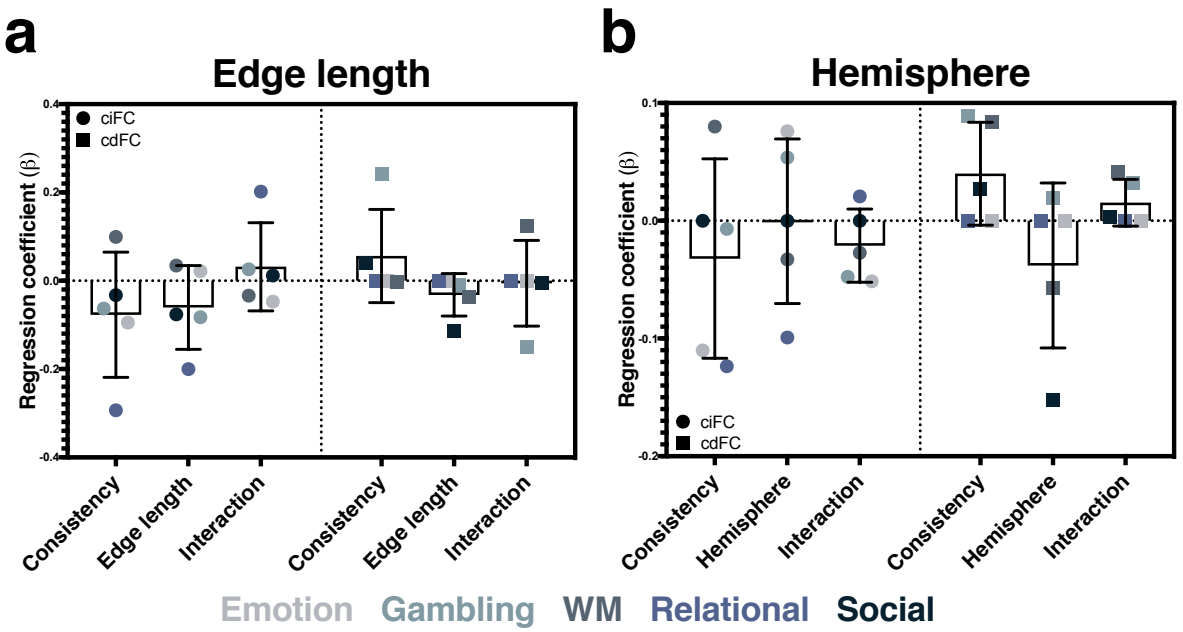

Figure 5—figure supplement 1. Relationships between inter-subject consistency, edge length (a), brain hemisphere (b), and predictive utility. Bar height reflects mean coefficient; error bars indicate s.d. of coefficents.