Title: Cognitive task information is transferred between brain regions via resting-state network topology

Authors: Takuya Ito[1*], Kaustubh R. Kulkarni[1], Douglas H. Schultz[1], Ravi D. Mill[1], Richard H. Chen[1], Levi I. Solomyak[1], Michael W. Cole[1]

[1]Center for Molecular and Behavioral Neuroscience, Rutgers University

Corresponding Author:
Takuya Ito
Center for Molecular and Behavioral Neuroscience, Rutgers University
197 University Avenue, Newark, NJ 07102
taku.ito1@gmail.com

Abstract
The brain is thought to process cognitive task information in a widespread distributed manner, yet little is known about how such information is coordinated on a global scale. We built on recent advances in decoding cognitive task information and characterizing the flow of activity through large-scale brain networks to investigate region-to-region information transfer during complex cognitive tasks. We hypothesized that resting-state functional network topology – revealed via correlations among brain activity time series during task-free rest – describes the fine-grained mappings between brain regions that carry cognitive task information. Confirming this, we found that diverse task-rule information could be predicted in held-out brain regions based on estimated activity flow through resting-state network connections. Activity flow over fine-grained resting-state connections thus provides a large-scale network mechanism for cognitive task information transfer in the human brain.

Introduction

The human brain is thought to be a distributed information processing device, with its routes of information transfer as a core feature determining its computational architecture. Consistent with this, studies that investigate neural information content have decoded task-evoked activation patterns to show that task information representations are distributed throughout the brain (Haxby et al., 2006; Muhle-Karbe et al., 2016; Poldrack et al., 2009; Zhang et al., 2013). However, such studies have yet to reveal how these distributed representations are coordinated, and how information in any one brain region is used by other brain regions to produce cognitive computations (de-Wit et al., 2016). Studies investigating interdependence of brain regions have typically emphasized statistical dependencies between regional time series, rather than the transfer of task-relevant information content between those regions. For instance, resting-state functional connectivity (FC) with functional magnetic resonance imaging (fMRI) (Raichle, 2010) focuses on inter-region correlations in the absence of explicit task content. Other FC studies have included careful manipulations of task content (Cole et al., 2013a; Gratton et al., 2016; Sadaghiani et al., 2015), yet have not focused on how FC relates to the information content embedded in task activation patterns (Norman et al., 2006). Thus, it remains unclear whether or how the network topology described by FC is relevant to task information content.

Here, we provide evidence for a network mechanism underlying the transfer and coordination of distributed cognitive information during performance of a variety of novel tasks. Based on recent evidence that resting-state FC describes the routes of task-evoked activity flow (Cole et al., 2016) (Figure 1A) – the movement of task activations between brain regions – we hypothesized that resting-state network topology describes the mappings underlying task information transfer between brain regions. If true, this hypothesis implicates a network mechanism for an information-preserving mapping across brain regions involving communication "channels" (de-Wit et al., 2016; Shannon, 1948) described by resting-state network topology. Identifying such a mechanism would provide an important new window into the large-scale information processing architecture of the human brain.

We focus primarily on resting-state FC based on evidence that resting-state FC reflects the human brain's invariant global routing architecture (van den Heuvel et al., 2009; Marrelec et al., 2016). Supporting this, it has been recently demonstrated that most of the functional network topology variance present during task performance (80%) is already present during rest (Cole et al., 2014a; Krienen et al., 2014). Thus, resting-state FC primarily reflects an intrinsic functional network architecture that is present regardless of cognitive context, given that there are only moderate changes to functional network organization across tasks (Cole et al., 2014a; Krienen et al., 2014). We built upon these findings to test for a potential mechanistic role for intrinsic network topology in task-related information processing.

Our hypothesis required an approach to empirically derive the mapping between information representations of pairs of brain regions, similar to identifying the transformation weights between layers in a neural network model (Yamins et al., 2014). This approach contrasts with two previous approaches that describe the coordination of task-relevant information between brain regions. The first approach measures small

shifts in task-evoked FC according to task-relevant content (Cole et al., 2013a; Sadaghiani et al., 2015). The second approach measures the correlation of moment-to-moment fluctuations in information content between regions (Coutanche and Thompson-Schill, 2013). Critically, these approaches primarily describe time-dependent information coordination, rather than suggesting a large-scale mechanism by which task representations are mapped between brain regions. Thus, neither of these previous approaches were appropriate for our goal of characterizing a computational mechanism by which cognitive information is mapped between regions. Nonetheless, these previous approaches were important for demonstrating the basic phenomenon of large-scale task information coordination, which we sought to better understand via the recently developed activity flow mapping approach (Cole et al., 2016).

The hypothesis that fine-grained, resting-state FC describes the representational mappings between brain regions during tasks is compatible with several recent findings. First, resting-state FC topology was recently shown to be highly structured and reproducible, forming clusters of networks consistent with known functional systems (Gordon et al., 2014; Power et al., 2011; Yeo et al., 2011). Second, as already mentioned, these resting-state networks are likely task-relevant given recent demonstrations that the network architecture estimated by resting-state FC is highly similar to FC architectures present during a variety of tasks (Cole et al., 2014a; Krienen et al., 2014). Third, in addition to reflecting large-scale connectivity patterns, resting-state FC has been shown to reflect local topological mappings between retinotopic field maps in visual cortex, highlighting the specificity with which resting-state FC conserves functionally tuned connections (Haak et al., 2013; Heinzle et al., 2011). Finally, resting-state FC has been shown to systematically relate to task-evoked activations, allowing prediction of task-evoked activations in held-out individuals across a variety of tasks (Cole et al., 2016; Tavor et al., 2016). This suggests a strong role for resting-state FC in shaping task activations – a core feature of our hypothesis that resting-state FC carries the fine-grained activation patterns that represent task-relevant information.

Testing our hypothesis required the creation of a new approach, which quantifies the amount of information transferred between pairs of brain regions over resting-state FC (Figures 1B and 1C; see Methods). Briefly, this approach builds upon our recently-developed “activity flow mapping” framework (Cole et al., 2016) (Figure 1A), and allows us to predict the activation pattern in a held-out region based on a “source” region’s activation pattern (Figure 1B). This predicted activation pattern is then compared to that region’s actual activation pattern during the current task condition (Figure 1C). This “matched condition” predicted-to-actual similarity is then compared to the “mismatched condition” predicted-to-actual similarity, with the difference in similarity quantifying the amount of task-specific information present in the prediction. Since the prediction was based on activity flow over resting-state FC patterns, this quantifies the amount of task-relevant information transferred via resting-state FC. Note that it was important to compare the predicted with the actual activation pattern (rather than simply decoding the predicted activation pattern), given the possibility that even an inaccurate predicted pattern could contain task-related information from the source region. Thus, the predicted-to-actual similarity comparison is a more valid test of information transfer than simply decoding predicted activation patterns. Following computational validation of this

approach we apply it to test our primary hypothesis, in addition to several more specific hypotheses described below.

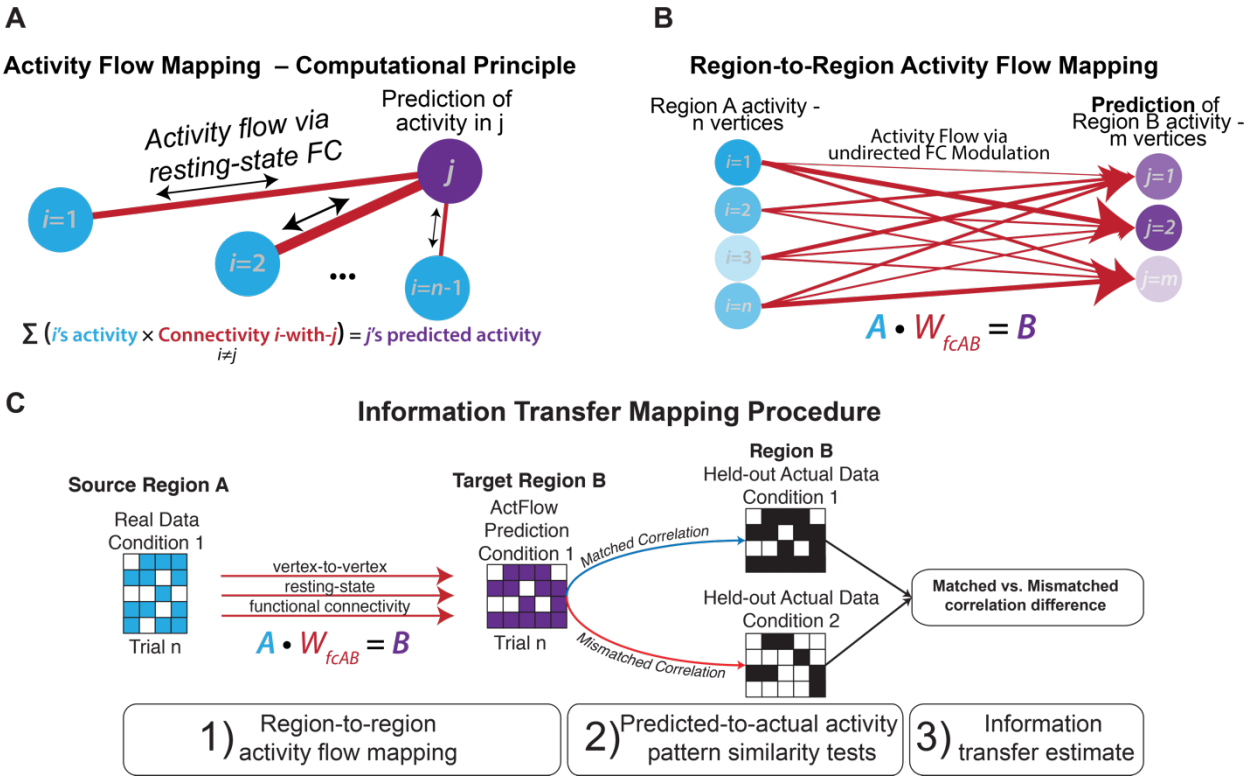


Figure 1. Activity Flow Mapping. A) Computational principle of activity flow mapping, as utilized by Cole et al. (2016). Activity in a held-out region is predicted by computing the linear weighted sum of all other regions' activity weighted by those regions' resting-state FC estimates to the held-out region. **B)** New approach involving activity flow mapping between vertices (or voxels) of isolated regions ("many-to-many" rather than "all-to-one" mapping of regions). We extend the underlying mechanism of activity flow mapping to compute the mapping of activity between regions. Mathematically, we predict the activation pattern in Region B by computing the dot product of Region A's activation pattern vector with the vertex-to-vertex resting-state FC matrix between Region A and B. This is conceptually similar to the formalism in (A), but extended to two distinct sets of regions. **C)** Mapping information transfer via region-to-region activity flow mapping and representational similarity analysis on held-out data. To test the transfer of task information for a given task block from Region A to Region B, we compare the predicted activation patterns from Region A to Region B to the actual, held-out prototypical task activation patterns of Region B for all task conditions using a spatial Spearman's rank correlation. For every block, spatial correlations to the task prototypes are computed and the information transfer estimate is measured by taking the difference of the correctly matched spatial correlation to the average of the incorrectly matched (mismatched) spatial correlations. Here we depict the approach for only two task conditions.

Information transfer mapping tests the ability of resting-state FC topology to compute mathematical transformations between brain regions. It further tests whether the involved matrix transformations preserve task-specific information. In other words, the mapping must preserve the predicted region's representational space such that task information is decodable after the matrix transformation on the source region's

activation pattern. This approach is related to recent efforts to map a given subject's representational space to a common space shared by all subjects (Haxby et al., 2011) using a matrix transformation that preserves a one-to-one correspondence across the two spaces. Here, we map representational spaces between regions for each subject, rather than mapping between subjects.

Going beyond our general hypothesis, we additionally sought to focus on the contribution of particular features of resting-state network topology in contributing to task-related information transfer. Recent studies have identified domain-general "flexible hub" networks with widespread resting-state FC and high activity during cognitive control tasks (Cole et al., 2010a, 2013a; Power et al., 2013). The high involvement of these cognitive control networks – the frontoparietal network (FPN), cingulo-opercular network (CON), and dorsal attention network (DAN) – in cognitively-demanding processes suggests a role for flexibly transferring task information across regions and networks.

We sought to isolate cognitive representations that would likely involve cognitive control networks by using a cognitive paradigm that involves multiple features thought to be central to cognitive control. First, we used novel tasks given the need for control to specify behavior in such under-practiced scenarios (Cole et al., 2013b; Rabbitt, 1997). Second, we used complex tasks given the need to deploy additional cognitive control resources when working memory is taxed (Miller and Buschman, 2015). Finally, we used a variety of abstract rules given that such rules are thought to be represented within cognitive control networks (Cole et al., 2011, 2015; Muhle-Karbe et al., 2016). Using many rules also allowed us to test our hypotheses across a variety of task conditions. These features converge in the Concrete Permuted Rule Operations paradigm (C-PRO; Figure 2). This paradigm was developed as part of this study, and is a modified version of the PRO paradigm (Cole et al., 2010a). We utilized this paradigm to observe how task-rule representations were transferred across brain regions. We predicted that cognitive control networks would flexibly represent C-PRO rule information and transfer that information to other regions and networks through their widespread intrinsic connections.

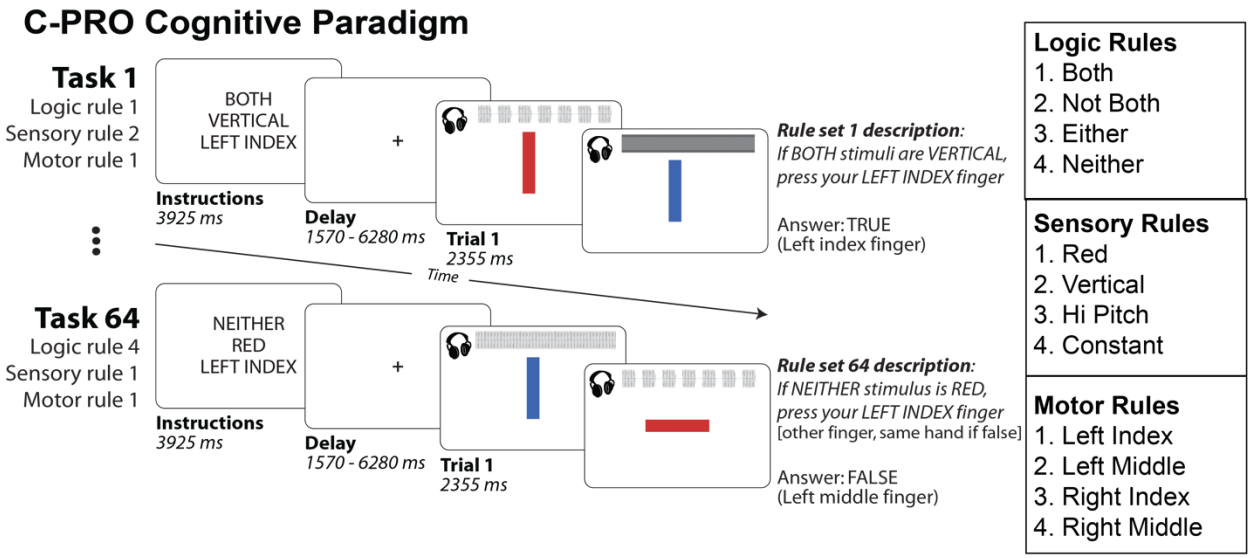


Figure 2. Concrete Permuted Rule Operations (C-PRO) experimental paradigm. For a given task, subjects were presented with an instruction set (i.e., a task-rule set), in which they were presented with three rules each from a different rule domain (logic, sensory, and motor rule domains). Subjects were then asked to apply the presented rule set to two consecutively presented stimulus screens and respond accordingly. Auditory and visual stimuli were presented simultaneously for each stimulus screen. A mini-block design was used, in which for a given set of instructions three trials were presented consecutively. The inter-trial interval was set to a constant 1570ms (2 TRs), with a jittered delay following the three trials prior to the subsequent task block (see Methods for more details). Task blocks lasted 28.260 seconds, or 36 TRs.

We began by replicating previously established properties of cognitive control networks, such as widespread resting-state FC (Cole et al., 2010a; Power et al., 2011). We then used this replication to motivate a computational model that validates the effectiveness of the information transfer mapping procedure for estimating the role of resting-state network topology in transferring task information. Finally, we apply this framework to empirical fMRI data, allowing us to test our hypotheses that (1) resting-state FC describes the channels of task information transfer and (2) cognitive control networks play an integral role in transferring task information to other regions based on their intrinsic functional network properties.

Results

Network organization of cognitive control networks

Cognitive control networks have been shown to exhibit connectivity profiles typical of hubs, indicating that they are optimally positioned to receive and distribute large amounts of information throughout the brain (Cole et al., 2010a; van den Heuvel and Sporns, 2013; Power et al., 2011, 2013). Additionally, regions within these networks adapt flexibly across tasks (Cole et al., 2013a; Stokes et al., 2013; Yeo et al., 2015). Thus, converging evidence indicates that cognitive control networks behave as flexible hub networks critical for coordinating cognition throughout the human brain. Given the recent replication crisis in neuroscience and other fields (Button et al., 2013; Szucs and Ioannidis, 2016), we sought to replicate the hub-like characteristic of cognitive control networks before moving forward with analyses that build on these previous findings.

Note that we used a recently developed set of functionally defined cortical regions (Glasser et al., 2016a). This region set is potentially more accurate than previous definitions, given that it is based on agreement across multiple neuroimaging modalities. This is important given that each modality has limitations that are often compensated for by other modalities, such that areal borders agreeing across modalities may be more accurate. Since cognitive control networks were not yet defined with this region set, we used community detection with resting-state FC to define these and other resting-state networks (Figure 3A; see Methods for details).

In our effort to replicate the hub-like properties characteristic of cognitive control networks (Cole et al., 2010a; Power et al., 2011), we tested whether cognitive control networks are hubs, based on having widespread intrinsic out-of-network FC estimated

at rest. We constrained our analyses to seven networks (Figure 3A), identified by being replicated across multiple previously published functional network atlases (Gordon et al., 2014; Power et al., 2011; Yeo et al., 2011). We focused on out-of-network connectivity to reduce the bias toward larger mean connectivity (i.e., weighted degree centrality, or global brain connectivity; (Cole et al., 2010a) for larger networks (Power et al., 2011, 2013). We found that the top three networks with highest out-of-network intrinsic FC were the three cognitive control networks (Figure 3D; all cognitive control networks greater than all non-cognitive control networks networks; FDR-corrected $p < 0.0001$). Given our hypothesis that resting-state connections describe the channels by which information transfers occur, the underlying resting-state network topology should shape the flow of information during task states. Specifically, we hypothesized that networks with high out-of-network resting-state FC (i.e., cognitive control networks) play a disproportionate role in shaping information transfer between regions belonging to distinct networks.

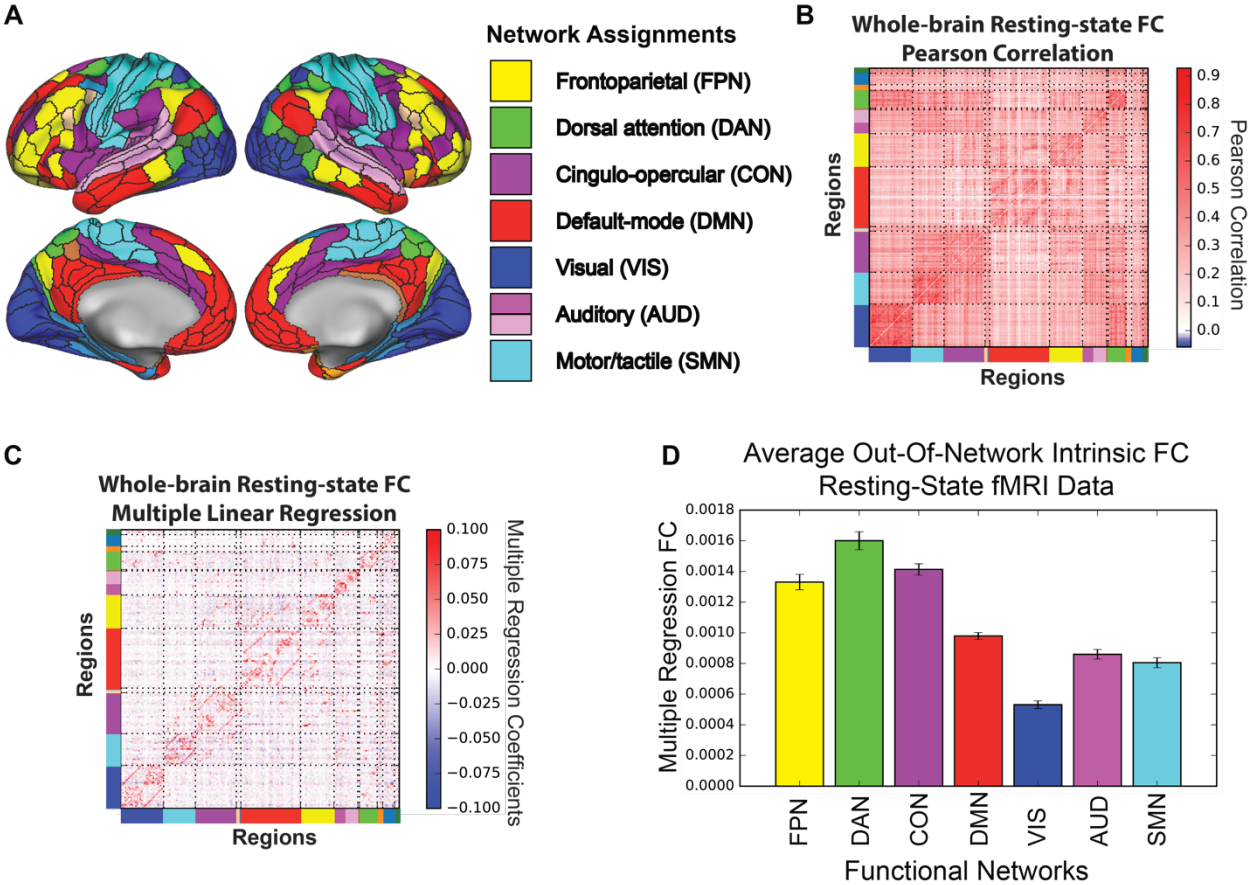


Figure 3. Large-scale network organization during rest. A) Surface visualization of network organization of the human cerebral cortex. Using a recently released, multi-modal parcellation of the human cerebral cortex (Glasser et al., 2016a), we assigned each region to a functional network using the Generalized Louvain method for community detection with resting-state fMRI data. We designated functional labels to seven networks that were replicated with other network assignments (Gordon et al., 2014; Power et al., 2011; Yeo et al., 2011). **B)** Whole-brain resting-state FC matrix computed using a Pearson correlation for every pair of regions in the Glasser et al. parcels. Color bars along the rows and columns denote network assignments. **C)** Whole-

brain resting-state FC matrix computed using multiple linear regression. For every parcel's time series, we fit a multiple linear regression model using the time series of all other parcels in the atlas as regressors of the target parcel. **D)** Averaged out-of-network resting-state FC for each defined functional network. Using multiple regression to estimate intrinsic FC during resting-state scans, we found that cognitive control networks had higher average out-of-network intrinsic FC estimates relative to non-cognitive control networks (i.e., DMN and sensorimotor networks). Error bars reflect across-subject standard error.

Validation of information transfer mapping in a neural network model

We previously established that activity flow over resting-state networks was a non-trivial phenomenon in that it depends on sufficiently large global coupling (i.e., high average inter-region connectivity strength) relative to local processing (i.e., high recurrent within-region connectivity strength) (Cole et al., 2016). However, it remains unclear whether networks containing task information differentially transfer that information to other networks via activity flow over resting-state FC. We tested this possibility by shifting from an "all-to-one" activity flow approach (Cole et al., 2016) to modeling activity flow between each pair of networks separately. Note that we focus on network-to-network information transfer here, but later extend the approach to region-to-region information transfer.

We sought to validate this approach using a simple abstract neural network model with one hub network and four non-hub networks (see Methods; Figure 4A). This network organization was the basis for simulating fMRI BOLD dynamics during rest and task states, allowing us to establish a "ground truth" upon which to test the efficacy of the new information transfer mapping approach. This validation-via-modeling method was highly similar to the simple neural network model we previously used to validate the original activity flow mapping approach (Cole et al., 2016). Using Wilson-Cowan type firing rate dynamics (Cowan et al., 2016; Stern et al., 2014), we simulated resting state and four distinct task states, simulated the transformation of the simulated neural signals to fMRI data (see Methods), and estimated resting-state FC (Figure 4B) and task-evoked fMRI runs for each of the four task conditions (Figure 4C).

We found that simulated resting-state FC accurately reflected high out-of-network intrinsic FC topology for the hub network (Figure 4D). Further, given the underlying synaptic connectivity structure (Figure 4A) and the estimated intrinsic topology via resting-state FC (Figures 4B and 4D), we hypothesized that information transfer to and from the hub network would reliably preserve task-specific information. Using the information transfer mapping approach (Figure 1C; see Methods), we quantified the amount of information transfer via activity flow between every pair of networks (Figure 4E). We found that information transfer to/from the flexible hub network and non-hub networks preserved task-specific representations, while transfers between pairs of non-hub networks did not preserve statistically significant representations after FDR-correction (Figure 4E; FDR-corrected $p < 0.05$). These results suggest that FC estimates obtained during simulated resting-state fMRI dynamics adequately reflect underlying synaptic organization, and that these estimates are sufficient to describe the mappings that govern activity flow between functional networks.

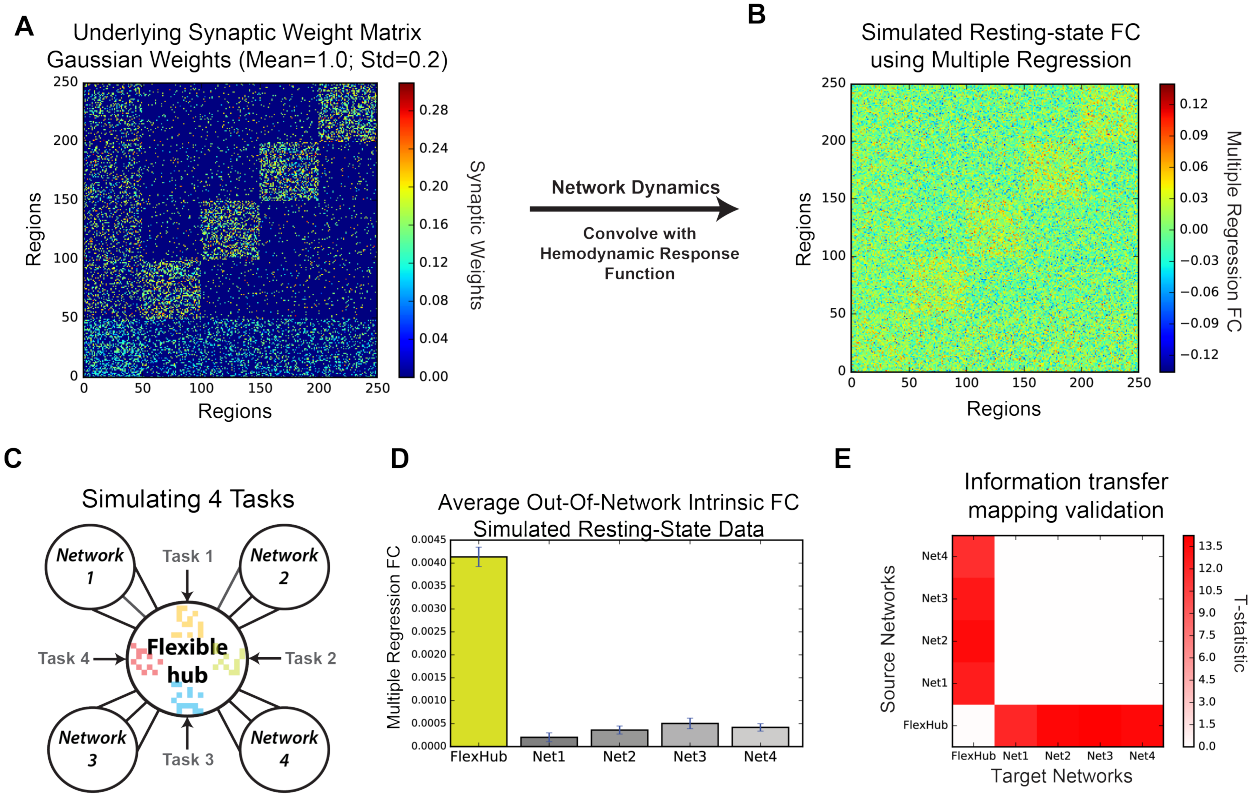


Figure 4. Computational validation of network-to-network information transfer mapping. A) Underlying synaptic weight matrix with four local networks and one hub network. We constructed an abstract neural network with a single hub network to see the relative effect of information transfer from the hub network to downstream local networks, similar to the hypothesized computational function of the cognitive control networks during task. **B)** Recovering large-scale synaptic organization via multiple regression FC estimates on a simulated resting-state time series. Using multiple regression FC, we recovered similar connectivity architecture to the underlying synaptic connectivity. **C)** We simulated four 'cognitive control tasks' by stimulating four distinct ensembles of regions within the flexible hub network. **D)** Increased out-of-network intrinsic FC reflects underlying synaptic organization. We validated that intrinsic FC estimates from rest preserve the underlying large-scale synaptic organization, where the flexible hub network has statistically greater out-of-network intrinsic FC than the other networks. Error bars represent across-subject standard error. **E)** Thresholded information transfer estimates between pairs of networks in a neural network model. Each row in the matrix corresponds to a source network from which we mapped activation patterns to other target networks using our information transfer mapping procedure (Figure 1C). Each column in the matrix corresponds to a target network to which we compared the predicted-to-actual activation patterns. FDR-corrected thresholded T-statistic map with $p < 0.05$. For every pair-wise network-to-network information transfer mapping, we performed an across-subject t-test of our information transfer estimates (Figure 1C) against zero.

These model simulations validate the plausibility of two hypotheses critical to the information transfer mechanism: (1) Resting-state FC is an adequate estimate of underlying intrinsic FC (reflecting aggregate synaptic connectivity and communication channel capacity); (2) Intrinsic FC describes the information-preserving mappings necessary to predict task-relevant activation patterns from one network to another.

These model results validate the analytical basis of estimating information transfer via activity flow, which is applied to network-to-network and region-to-region information transfer mapping with empirical fMRI data below.

Regions communicate task information via activity flow between fine-grained vertex representations

While our previous study demonstrated activity flow as a macroscopic mechanism for predicting whole-brain cognitive task activations (Cole et al., 2016), information is known to persist at lower levels of organization, such as in voxel/vertex patterns in fMRI data and neuronal ensembles in neurophysiological studies (Haxby et al., 2006; Stokes et al., 2013). Further, though propagation of task activity has been shown to be possible in animal models (Siegel et al., 2015; Timme et al., 2016), task-set information transfer between cortical regions and networks has not been shown in the human brain.

We were primarily interested in region-to-region information transfer given the finer spatial information involved, allowing for a more fine-grained mapping of information transfer throughout the brain. However, as in our computational model, we first focused on information transfer mapping between functional networks. Using the same information transfer mapping procedure as in the model, information was mapped between entire functional networks using region-level activation patterns and region-to-region FC measures (i.e., network-to-network information transfer; Figure S1; see Methods). As in our previous study (Cole et al., 2016), multiple regression FC was used for resting-state FC estimation. We first established that task-rule information is widely distributed across entire functional networks (Figure S1A). This allowed us to then evaluate whether decodable representations of information were transferred to other functional networks. We found that functional networks differentially transferred task-set information ($p<0.05$, FDR-corrected for multiple comparisons) depending on the task-rule domain and the networks involved (Figures S1B, S1C, and S1D). These network-to-network information transfers reinforce the computational principles underlying our computational model with empirical data.

We next sought to investigate whether fine-grained (i.e., vertex-to-vertex) resting-state connections between pairs of regions could approximate the information-preserving mapping of activation patterns between regions. To establish whether information transfer between individual pairs of regions could occur, we first needed to establish whether regions contained decodable task-rule representations. Thus, we first quantified the information content of each rule domain in the C-PRO paradigm (logic, sensory and motor rule domains) for each of the 360 parcels using activation patterns (at the vertex level) with a cross-validated representational similarity analysis (see Methods). We found that logic rules were relatively distributed, with highest-quality representations in frontal and parietal cortices (FDR-corrected $p < 0.05$; Figure 5A). Sensory rule information was also relatively distributed (FDR-corrected $p < 0.05$; Figure 5B), though the highest-quality representations were predominantly in visual areas. Lastly, we found that motor rule representations were significantly more localized, with the highest-quality representations in the somatomotor network (FDR-corrected $p < 0.05$; Figure 5C). The existence of distributed task-rule information in multiple cortical

regions allowed us to assess how task-rule-specific representations in one region might be transferred to other regions.

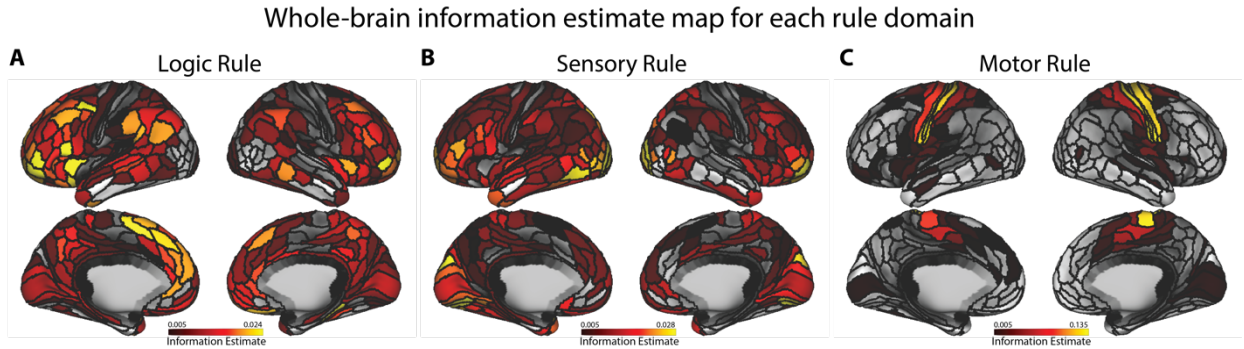


Figure 5. Representational content of each region for each task-rule domain, prior to information transfer mapping. A) Thresholded whole-brain logic rule information estimate map. A cross-validated representational similarity analysis (quantifying degree of information representation; see Methods) for the logic rule domain was computed using vertices within every parcel. For each parcel, an average information estimate (see Methods) was computed for each subject, and a one-sided t-test was computed against zero across subjects. We then corrected for multiple comparisons across all cortical regions using FDR-correction, and thresholded the statistical map with $p < 0.05$. **B)** Thresholded whole-brain sensory rule information estimate map. As in the logic rule analysis, rule representations were highly distributed across the entire cortex, though representations were especially prominent in visual areas. **C)** Thresholded whole-brain motor rule information estimate map. Unlike the logic and sensory rule representations, motor rule representations were significantly more localized to the motor cortex.

To test whether resting-state FC topology is capable of carrying task-rule representations between pairs of regions, we performed region-to-region information transfer mapping (Figure 4) using within-region vertex-level representations and vertex-to-vertex resting-state FC between regions (as in Figure 1C; also see Methods). We performed this procedure for every pair of 360 regions, and visualized our results as a 360x360 matrix for each rule domain (Figures 6A, 6C, and 6E). However, given the difficulty in visually interpreting information transfers between every pair of regions (due to sparseness), we collapsed the region-to-region matrix by network to better visualize significant region-to-region information transfers at the network level (FDR-corrected $p < 0.05$; Figures 6B, 6D, and 6F). In addition, to see the relative anatomical position of regions that transferred information, we computed the percent of significant transfers from each cortical region for each rule domain, and plotted these percentages on the cortical surface (Figures 7A, 7B, and 7C).

For logic rule mappings, while information transfers were highly distributed, most successful region-to-region information transfers predominantly involved the FPN and other frontoparietal regions (Figures 6B and 7A). In particular, regions within the FPN transferred information to other regions in the FPN, as well as regions in other cognitive control networks (i.e., CON and DAN) and DMN. These findings suggest that the FPN distributes logic rule representations both to other regions within the FPN as well as

other association cortex networks for task-set implementation and maintenance (Figure 6B).

For sensory rule mappings, we found high specificity and sparseness of region-to-region task information transfers (Figure 7B). Most notably, we found that most sensory rule representations are predominantly transferred among regions within the DAN and visual (VIS) networks, as well as between those networks (Figures 6C and 6D). While our results suggest the involvement of regions in other networks such as the CON, DMN, and FPN, previous studies have implicated a prominent role of the DAN and VIS in attentional processing of sensory information, consistent with the observed information transfers between these networks (Corbetta and Shulman, 2002). These findings potentially suggest sensory rule information is transferred between cognitive control networks, and transfers between regions in the DAN and VIS might implement these top-down information transfers.

Lastly, we found the most information transfer specificity for motor rule representations (Figures 6E, 6F, and 7C), consistent with the relatively localized representations of motor rule information (Figure 5C). In particular, transfer of motor rule representations largely involved regions in the SMN (Figure 6F and 7C), while between-network information transfer with the SMN primarily involved transfers between regions in the SMN and a cognitive control network (CON or DAN). Consistent with the previous rule domains, there was transfer from the FPN to other association cortex areas such as regions in the FPN, CON, and DMN.

To characterize the generality by which rule information transfer occurred between specific networks, we performed an analysis to quantify transfers between networks across all rule domains. We found that regions within the FPN consistently transferred task rule information to the CON across all three rule domains (Figure 6G). This finding is consistent with prior theories suggesting that the FPN implements task sets, while the CON is involved in task set maintenance (Power and Petersen, 2013). In addition, we wanted to see which networks consistently transferred information across all rule domains, regardless of the target region's network affiliation. We found that regions in the FPN and DAN were consistently involved in transferring information to other regions in all rule domains (Figure 6H). To assess whether these transfers were from the same regions within each network, we performed an analysis to see if any individual region consistently transferred information across all rule domains. We found no individual region that consistently transferred task rule information across the three rule domains, suggesting that transfers from the FPN and DAN across the rule domains were from distinct regions within each network. This suggests that the regions within the FPN and DAN collectively act as a hub network to communicate rules in different cognitive domains. Thus, despite large differences in information transfer across the three rule domains, cognitive control networks play a critical role in task rule transfers, regardless of rule domain.

These results uncover two key findings: (1) resting-state network topology describes the mappings underlying information transfer across distributed regions and functional networks, and (2) cognitive control networks play a critical role in transferring task-rule information to content-specific networks (e.g., motor rule information to the SMN) during complex cognitive tasks. Our results show that resting-state FC can be

used to describe the underlying information-preserving mappings between different cortical areas at multiple levels of spatial organization (i.e., regions and networks).

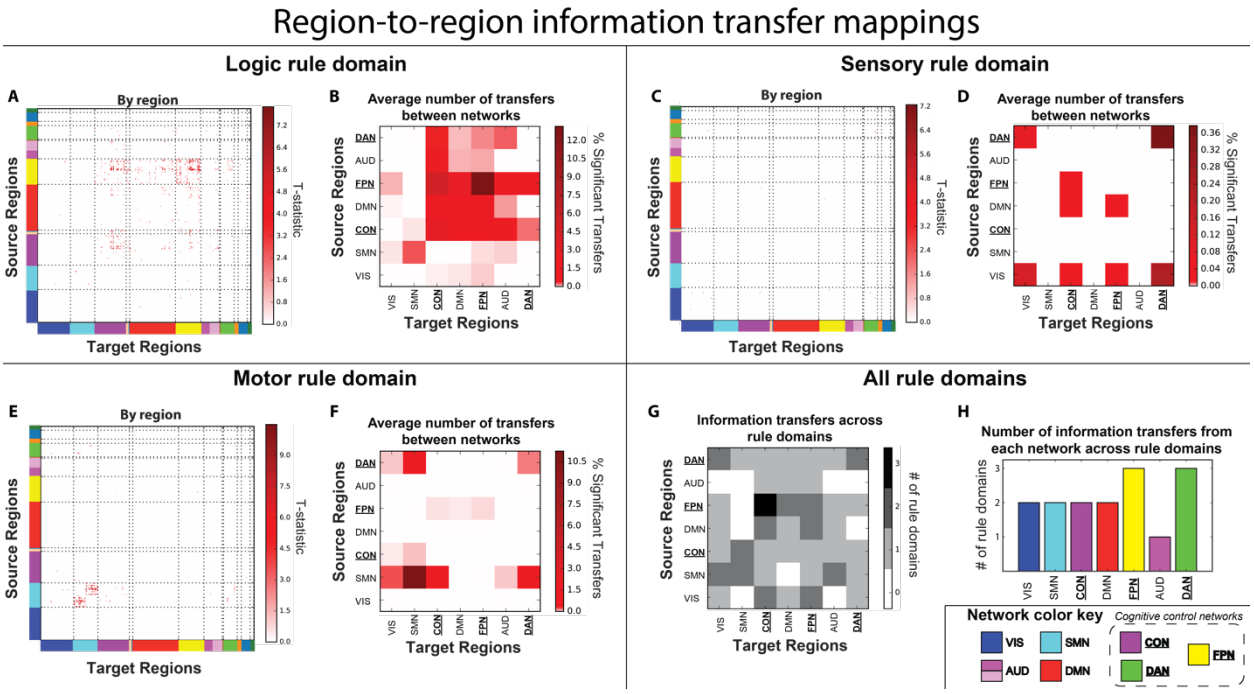


Figure 6. Information transfer mappings between all pairs of regions. **A)** Logic rule region-to-region information transfer mapping. We visualized region-to-region information transfer mapping results as a 360x360 region matrix, with source regions along the rows and target regions along the columns. **B)** Average number of region-to-region transfers by network affiliations. To better visualize and assess how region-to-region transfer mappings may have been influenced by underlying network topology, we computed the percent of significant rule transfers for every network-to-network configuration (i.e., the percentage of region-to-region transfers from a network A to a network B). Information transfer of logic rule information is highly distributed, with information transferred both within and between-networks (particularly with the FPN). **C,D)** Sensory rule region-to-region information transfer mapping. Region-to-region information transfer are significantly sparser for sensory rule mappings, likely contrasting the distributed nature of higher-order task-rule transfers observed with logic rule mappings. **E,F)** Motor rule region-to-region information transfer mapping. Motor rule mappings are noticeably more localized within the motor network, suggesting more localized processing of motor rule information within motor regions. **G)** Information transfers between regions by network affiliation across rule domainsNumber of rule domains each network configuration has at least one successful information transfer. For each network visualization panel (i.e., panels B, D, and F) we binarized each matrix to assess whether or not there were region-to-region transfers for a particular network configuration for each rule domain. We then summed these matrices to identify consistent region-to-region transfers between networks across the three rule domains. We found that transfers from the FPN to the CON was the only network configuration that consistently transferred information across the three rules. **H)** We performed a similar analysis as in panel G, but assessed whether a network was consistently involved in sending task rule information (as a source region) across the three rule domains. We found that regions in the FPN and DAN consistently transferred information across all rule domains.

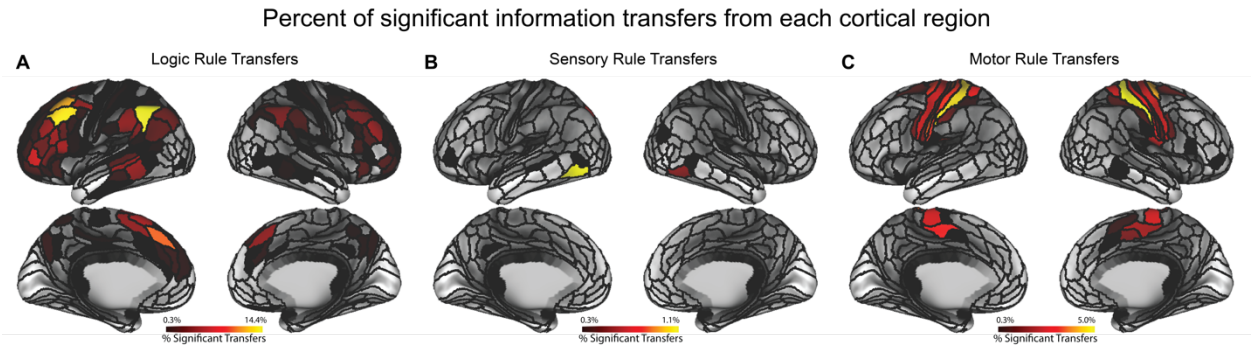


Figure 7. Percent of significant information transfers from each cortical region. **A)** Percent of significant transfers from each region for the logic rule domain. Percentages are computed by taking the number of significant transfers from each region, and dividing it by the total number of possible transfers from that region (359 other regions). Transfers are widely distributed, yet are predominantly transferred from frontoparietal regions and cognitive control regions. **B)** Percent of significant transfers from each region for the sensory rule domain. Transfers are much sparser than in the logic rule domain. Most transfers are from higher order visual areas and the DAN. **C)** Percent of significant transfers from each region for the motor rule domain. Transfers mostly come from within the motor network, with a few frontal regions and a right temporal region involved in transferring information.

Discussion

Studies from neurophysiology, fMRI, and computational modeling emphasize the distributed nature of information processing in the brain (Eliasmith et al., 2012; Huth et al., 2012; Siegel et al., 2015). However, fMRI studies often decode cognitive information from brain regions (Haxby et al., 2006) without considering how other brain regions might utilize that information (de-Wit et al., 2016). In other words, current neuroscientific findings emphasize an “experimenter-as-receiver” framework (i.e., the experimenter decoding information in a brain region) rather than a “cortex-as-receiver” framework (i.e., brain regions decoding information transferred from other brain regions; de-Wit et al., 2016), which clashes with the traditional understanding of information communication described by Shannon’s Information Theory (Shannon, 1948). Thus, understanding how cortical regions receive information from other regions bridges a crucial gap in understanding the nature of information processing in the brain. In light of recent findings relating resting-state fMRI to task-evoked cognitive activations (Cole et al., 2016; Tavor et al., 2016), we posited that resting-state FC describe the “channels” over which information can be communicated between cortical regions. More broadly, our findings suggest that resting-state network topology describes the large-scale architecture of information communication in the human brain.

We devised a novel procedure to quantify information transfer between brain regions (Figure 1C). The procedure requires an information-preserving mapping between a sender region and a receiver region. In the computational modeling literature, such mappings are typically estimated through machine learning techniques to approximate synaptic weight transformations between layers of a network (e.g., an artificial neural network model using backpropagation; Yamins et al., 2014). However, given that artificial neural networks are universal function estimators (Hartman et al., 1990), we opted to take a more biologically principled approach. We utilized evidence

that patterns of spontaneous activity can be used to successfully estimate the flow of task-related activity in both local and large-scale brain networks (Cole et al., 2016; Smith et al., 2006; Timme et al., 2016) to obtain biophysically plausible, data-driven mappings between brain regions using resting-state fMRI. We used a computational model to validate the plausibility of this account of large-scale information transfer (Figures 4A and 4B), finding that despite the slow dynamics of the blood-oxygen level dependent signal, simulated resting-state fMRI accurately reflects the large-scale channels of information transfer. We then used empirical fMRI data to show that resting-state FC describes information-preserving mappings in cortex at two levels of spatial organization: brain regions and functional networks (Figures 6 and S1). In other words, mappings between sender and receiver regions obtained via resting-state FC preserve the same decodable task information content (in the same spatial organization) as in the actual, held-out data. Note that the organization of activity patterns is necessarily distinct between brain regions (given their distinct shapes), such that accurately predicting activation patterns in a held-out region based on activity in another region reflects accurate transformation of information-carrying activity patterns between those brain regions. These findings suggest that resting-state FC estimates reflect the actual mappings that are implemented in the brain during task information transfer.

The evidence that fine-grained resting-state FC describes the information-preserving mappings between regions is important for advancing neuroscientific theory for a number of reasons. First, it provides an empirically-validated, theoretical account for how cognitive representations in different regions are likely mechanistically related to one another. Second, it confirms the base assumption that decodable representations in a brain region are utilized by other regions through a biologically plausible construct – information transfer via fine-grained patterns of activity flow. Third, it expands the functional relevance of decades of resting-state FC findings (Biswal et al., 1995; Raichle, 2010), given resting-state FC's ability to describe cognitively meaningful fine-grained relationships between brain regions.

Previous studies have focused on the role of task-evoked FC in shifting distributed task representations (Cole et al., 2013a; Gratton et al., 2016). We recently utilized such findings to develop a "flexible hub" account of distributed task-set reconfiguration via cognitive control networks (Cole et al., 2013a, 2014b). The present results complement and advance these findings, providing an explicit computational mechanism involving resting-state FC topology (and cognitive control network hubs) in transferring these task representations throughout cortex. Importantly, recent findings have demonstrated that task-evoked FC changes tend to be small relative to resting-state FC topology (Cole et al., 2014a; Krienen et al., 2014). This suggests that the resting-state FC topology investigated here maps the bulk of the task-relevant information transfers, with task-evoked FC alterations to this topology contributing only small (but likely important) changes to this process. It will be important for future research to investigate the relative contributions of task-evoked FC and intrinsic FC (as estimated using resting-state FC) in shaping distributed task information transfers. Regardless of the outcomes of such future research, however, converging results from different FC methods suggest an important role for flexible hub regions within cognitive control networks in coordinating task-relevant information throughout cortex.

The information transfer mapping approach involves estimating linear information transfer. Critically, however, neural information processing is thought to often depend on nonlinear transformations (Eliasmith, 2007), such as face-selective neurons in the ventral visual stream responding to whole faces but not facial components (e.g., eyes and ears; Kanwisher et al., 1997; Tsao et al., 2006). The present findings represent an important step toward understanding the network mechanisms underlying information transformations between brain regions, setting the stage for future research to identify the role of resting-state FC in nonlinear information transformations. This would go beyond the information transfer processes investigated here to identify a role of resting-state FC in neural computation (not just communication).

Recent advances have allowed unprecedented characterization of the human brain's intrinsic network topology, yet it has remained unclear whether or how this network architecture relates to cognitive information processing. Here, we demonstrate how fine-grained intrinsic connectivity patterns relate to cognitive information transfer. Further, by estimating information transfer throughout cortex we found evidence that previously identified cognitive control networks play critical roles in global transfer of cognitive task information. We expect that these findings will spur new investigations into understanding the nature of distributed information processing throughout the brain, and reveal the contribution of these fine-grained information channels estimated at rest to task-relevant information transfers.

Acknowledgements

We thank Stephen J. Hanson, Catherine Hanson, and Gregg Ferencz in helping collect and set up MRI protocols for data collection. We also thank Merav Stern and Hiromichi Tsukada on helpful discussions regarding the computational model. We acknowledge support by the US National Institutes of Health under awards K99-R00 MH096801 and R01 MH109520. The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies.

Author Contributions

T.I. and M.W.C. conceived of the study. T.I. performed data preprocessing, data analysis, and developed the computational model under the supervision of M.W.C. T.I., D.H.S., L.I.S., and M.W.C. designed the experimental paradigm, and M.W.C. programmed the experiment. K.R.K. and M.W.C. constructed network definitions. T.I., K.R.K., D.H.S., R.D.M., R.H.C., and L.I.S. performed the behavioral and fMRI experiments under the supervision of M.W.C. T.I. and M.W.C. wrote the manuscript.

References

Barral, J., and Reyes, A.D. (2016). Synaptic scaling rule preserves excitatory–inhibitory balance and salient neuronal network dynamics. Nat. Neurosci.

Biswal, B., Yetkin, F.Z., Haughton, V.M., and Hyde, J.S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn. Reson. Med. *34*, 537–541.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. *10008*, 6.

Button, K.S., Ioannidis, J.P. a, Mokrysz, C., Nosek, B. a, Flint, J., Robinson, E.S.J., and Munafò, M.R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. *14*, 365–376.

Cole, M., Etzel, J., Zacks, J., Schneider, W., and Braver, T. (2011). Rapid Transfer of Abstract Rules to Novel Contexts in Human Lateral Prefrontal Cortex. *5*.

Cole, M.W., Pathak, S., and Schneider, W. (2010a). Identifying the brain's most globally connected regions. Neuroimage *49*, 3132–3148.

Cole, M.W., Bagic, A., Kass, R., and Schneider, W. (2010b). Prefrontal Dynamics Underlying Rapid Instructed Task Learning Reverse with Practice. J. Neurosci. *30*, 14245–14254.

Cole, M.W., Reynolds, J.R., Power, J.D., Repovs, G., Anticevic, A., and Braver, T.S. (2013a). Multi-task connectivity reveals flexible hubs for adaptive task control. Nat. Neurosci. *16*, 1348–1355.

Cole, M.W., Laurent, P., and Stocco, A. (2013b). Rapid instructed task learning: a new window into the human brain's unique capacity for flexible cognitive control. Cogn. Affect. Behav. Neurosci. *13*, 1–22.

Cole, M.W., Bassett, D.S., Power, J.D., Braver, T.S., and Petersen, S.E. (2014a). Intrinsic and task-evoked network architectures of the human brain. Neuron *83*, 238–251.

Cole, M.W., Repov , G., and Anticevic, A. (2014b). The Frontoparietal Control System: A Central Role in Mental Health. Neurosci. *20*, 652–664.

Cole, M.W., Ito, T., and Braver, T.S. (2015). The Behavioral Relevance of Task Information in Human Prefrontal Cortex. Cereb. Cortex bhv072-.

Cole, M.W., Ito, T., Bassett, D.S., and Schultz, D.H. (2016). Activity flow over resting-state networks shapes cognitive task activations. Nat. Neurosci.

Corbetta, M., and Shulman, G.L. (2002). Control of Goal-Directed and Stimulus-Driven Attention in the Brain. Nat. Rev. Neurosci. *3*, 215–229.

Coutanche, M.N., and Thompson-Schill, S.L. (2013). Informational connectivity: identifying synchronized discriminability of multi-voxel patterns across the brain. Front. Hum. Neurosci. *7*, 15.

Cowan, J.D., Neuman, J., and van Drongelen, W. (2016). Wilson-Cowan Equations for Neocortical Dynamics. J. Math. Neurosci. *6*, 1.

de-Wit, L., Alexander, D., Ekroll, V., and Wagemans, J. (2016). Is neuroimaging

measuring information in the brain? Psychon. Bull. Rev. 1–14.

Eliasmith, C. (2007). How to build a brain: from function to implementation. Synthese *159*, 373–388.

Eliasmith, C., Stewart, T.C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., and Rasmussen, D. (2012). A Large-Scale Model of the Functioning Brain. Science (80-.). *338*, 1202–1205.

Genovese, C., and Wasserman, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. J. R. Stat. Soc. Ser. B (Statistical Methodol. *64*, 499–517.

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., et al. (2013). The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage *80*, 105–124.

Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., et al. (2016a). A multi-modal parcellation of human cerebral cortex. Nature 1–11.

Glasser, M.F., Smith, S.M., Marcus, D.S., Andersson, J.L.R., Auerbach, E.J., Behrens, T.E.J., Coalson, T.S., Harms, M.P., Jenkinson, M., Moeller, S., et al. (2016b). The Human Connectome Project's neuroimaging approach. Nat. Neurosci. *19*, 1175–1187.

Gordon, E.M., Laumann, T.O., Adeyemo, B., Huckins, J.F., Kelley, W.M., and Petersen, S.E. (2014). Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. Cereb. Cortex.

Gratton, C., Laumann, T.O., Gordon, E.M., Adeyemo, B., and Petersen, S.E. (2016). Evidence for Two Independent Factors that Modify Brain Networks to Meet Task Goals. Cell Rep. *17*, 1276–1288.

Haak, K. V., Winawer, J., Harvey, B.M., Renken, R., Dumoulin, S.O., Wandell, B.A., and Cornelissen, F.W. (2013). Connective field modeling. Neuroimage *66*, 376–384.

Hartman, E.J., Keeler, J.D., and Kowalski, J.M. (1990). Layered Neural Networks with Gaussian Hidden Units as Universal Approximations. Neural Comput. *2*, 210–215.

Haxby, J. V, Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2006). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *2425*, 2425–2431.

Haxby, J. V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., Hanke, M., and Ramadge, P.J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron *72*, 404–416.

Heinzle, J., Kahnt, T., and Haynes, J.D. (2011). Topographically specific functional connectivity between visual field maps in the human brain. Neuroimage *56*, 1426–1436.

van den Heuvel, M.P., and Sporns, O. (2013). Network hubs in the human brain. Trends Cogn. Sci. *17*, 683–696.

van den Heuvel, M.P., Mandl, R.C.W., Kahn, R.S., and Hulshoff Pol, H.E. (2009). Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. Hum. Brain Mapp. *30*, 3127–3141.

Huth, A.G., Nishimoto, S., Vu, A.T., and Gallant, J.L. (2012). A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories

across the Human Brain. Neuron *76*, 1210–1224.

Jutla, I.S., Jeub, L.G.S., and Mucha, P.J. (2011). A generalized Louvain method for community detection implemented in MATLAB. URL Http//netwiki. Amath. Unc. edu/GenLouvain.

Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. J. Neurosci. *17*, 4302 LP-4311.

Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. *2*, 4.

Krienen, F.M., Yeo, B.T.T., and Buckner, R.L. (2014). Reconfigurable task-dependent functional coupling modes cluster around a core functional architecture. Philos. Trans. R. Soc. B Biol. Sci. *369*, 20130526–20130526.

Marrelec, G., Messé, A., Giron, A., and Rudrauf, D. (2016). Functional Connectivity's Degenerate View of Brain Computation. PLOS Comput. Biol. *12*, e1005031.

Miller, E.K., and Buschman, T.J. (2015). Working Memory Capacity: Limits on the Bandwidth of Cognition. Daedalus *144*, 112–122.

Muhle-Karbe, P.S., Duncan, J., De Baene, W., Mitchell, D.J., and Brass, M. (2016). Neural Coding for Instruction-Based Task Sets in Human Frontoparietal and Visual Cortex. Cereb. Cortex bhw032.

Norman, K.A., Polyn, S.M., Detre, G.J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. *10*, 424–430.

Oosterhof, N.N., Wiestler, T., Downing, P.E., and Diedrichsen, J. (2011). A comparison of volume-based and surface-based multi-voxel pattern analysis. Neuroimage *56*, 593–600.

Poldrack, R.A., Halchenko, Y.O., and Hanson, S.J. (2009). Decoding the large-scale structure of brain function by classifying mental states across individuals. Psychol. Sci. *20*, 1364–1372.

Power, J.D., and Petersen, S.E. (2013). Control-related systems in the human brain. Curr. Opin. Neurobiol. *23*, 223–228.

Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J. a., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., et al. (2011). Functional Network Organization of the Human Brain. Neuron *72*, 665–678.

Power, J.D., Schlaggar, B.L., Lessov-Schlaggar, C.N., and Petersen, S.E. (2013). Evidence for hubs in human functional brain networks. Neuron *79*, 798–813.

Rabbitt, P. (1997). Methodology of frontal and executive function.

Raichle, M.E. (2010). Two views of brain function. Trends Cogn. Sci. *14*, 180–190.

Sadaghiani, S., Poline, J.-B., Kleinschmidt, A., and D'Esposito, M. (2015). Ongoing dynamics in large-scale functional connectivity predict perception. Proc. Natl. Acad. Sci. U. S. A. *112*, 8463–8468.

Schneider, W., Eschman, A., and Zuccolotto, A. (2002). E-Prime: User's guide (Psychology Software Incorporated).

Shannon, C.E. (1948). A Mathematical Theory of Communication. Bell Syst. Tech. J. *27*, 379–423.

Siegel, M., Buschman, T.J., and Miller, E.K. (2015). Cortical information flow during flexible sensorimotor decisions. Science (80-.). *348*, 1352–1355.

Smith, V.A., Yu, J., Smulders, T. V., Hartemink, A.J., and Jarvis, E.D. (2006). Computational inference of neural information flow networks. PLoS Comput. Biol. *2*, 1436–1449.

Stern, M., Sompolinsky, H., and Abbott, L.F. (2014). Dynamics of random neural networks with bistable units. Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. *90*, 1–7.

Stokes, M.G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., and Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. Neuron *78*, 364–375.

Szucs, D., and Ioannidis, J.P.A. (2016). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. bioRxiv.

Tavor, I., Jones, O.P., Mars, R.B., Smith, S.M., Behrens, T.E., and Jbabdi, S. (2016). Task-free MRI predicts individual differences in brain activity during task performance. Science (80-.). *352*, 1773–1776.

Timme, N.M., Ito, S., Myroshnychenko, M., Nigam, S., Shimono, M., Yeh, F.-C., Hottowy, P., Litke, A.M., and Beggs, J.M. (2016). High-Degree Neurons Feed Cortical Computations. PLOS Comput. Biol. *12*, e1004858.

Traud, A.L., Kelsic, E.D., Mucha, P.J., and Porter, M.A. (2011). Comparing community structure to characteristics in online collegiate social networks. SIAM Rev. *53*, 526–543.

Tsao, D.Y., Freiwald, W.A., Tootell, R.B.H., and Livingstone, M.S. (2006). A Cortical Region Consisting Entirely of Face-Selective Cells. Science (80-.). *311*, 670 LP-674.

Yamins, D.L.K., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., and DiCarlo, J.J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc. Natl. Acad. Sci. *111*, 8619–8624.

Yeo, B.T.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zollei, L., Polimeni, J.R., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. *106*, 1125–1165.

Yeo, B.T.T., Krienen, F.M., Eickhoff, S.B., Yaakub, S.N., Fox, P.T., Buckner, R.L., Asplund, C.L., and Chee, M.W.L. (2015). Functional specialization and flexibility in human association cortex. Cereb. Cortex *25*, 3654–3672.

Zhang, J., Kriegeskorte, N., Carlin, J.D., and Rowe, J.B. (2013). Choosing the rules: distinct and overlapping frontoparietal representations of task rules for perceptual decisions. J. Neurosci. *33*, 11852–11862.

Methods

Region-to-region information transfer mapping

We extended the original activity flow mapping procedure as defined in Cole et al. (2016) (Figure 1A) to investigate transfer of task-related information between pairs of brain regions using vertex-wise activation patterns (i.e., region-to-region activity flow mapping; Figure 1B). The original activity flow mapping approach predicted the activity level of a single held-out region using the weighted sum of the task-evoked activity of all other regions. These activation estimates were obtained using a standard fMRI general linear model (GLM). The weights in the weighted sum were based on the resting-state FC from the source regions to the held-out region. The region-to-region activity flow mapping procedure developed here is computationally similar. However, instead of predicting the activity of a single held-out region based on all other regions, we predicted the activity of the vertices of a held-out target region based on the vertices of a "source" region. Mathematically, to predict Region B's activation pattern (a 1 x n vector) using Region A's activation pattern (a 1 x m vector), we take the dot product of Region A's activity vector and the vertex-to-vertex resting-state FC matrix between A and B (m x n matrix) to compute the prediction of Region B's activation pattern. This matrix transformation (i.e., weighted FC matrix) allowed us to map activation patterns in one region's spatial dimension to the spatial dimension of another region.

To test the extent that task representations are preserved in the region-to-region multivariate predictions, we quantified how much "information transfer" occurred between the two regions. Briefly, information transfer mapping comprises three steps, illustrated in Figure 1C: (1) Region-to-region (or network-to-network) activity flow mapping; (2) A cross-validated representational similarity analysis between predicted activation patterns and actual, held-out activation patterns; (3) Computing the difference between matched condition similarities and mismatched condition similarities.

More specifically, we quantified how much "information transfer" occurred between two regions by comparing the predicted activation pattern in the held-out region to the actual activation pattern in that region. For a given activation pattern in Region A for a single task block, we computed the prediction of Region B's activation pattern for that task block using activity flow mapping. Each block's predicted activation pattern was compared with the "prototype" of the predicted block's task-rule condition (the "matched" condition). Prototypes were computed as the average activation pattern across all other blocks of the same task-rule condition (excluding the to-be-predicted block). We used a leave-four-out cross-validation scheme to construct the prototypes within each cross-validation fold (randomly leaving out one block of each task-rule, without replacement) to keep the number of blocks included in each prototype equal across conditions. Thus, each task-rule prototype was constructed using 31 task blocks (32 blocks per rule in total). The goal of this was to hold out both the spatial (to-be-predicted region) and temporal (to-be-predicted task block). The predicted-to-actual comparisons were computed using Spearman's rank correlation, akin to a representational similarity analysis (Kriegeskorte et al., 2008). We compared the predicted block with the prototypes of the other conditions (the "mismatched" conditions). We then computed an "information transfer estimate" as the difference between the Fisher-Z transformed matched condition correlation and the average of the

mismatched condition correlations. Equivalently, in a 4x4 similarity matrix comparing the predicted-activation pattern with each of the four condition prototypes, the information transfer estimate is obtained by taking the difference between the average of the diagonal and average of the off-diagonal.

We also demonstrate that the predicted-to-actual similarity analysis in our information transfer mapping procedure can be substituted with a support vector machine (SVM) decoding scheme. Specifically, we show in our computational model that we could train a linear classifier on our predicted activation patterns that could decode the actual, held-out activation patterns in the target region (Figures S2B and S2C). We used the same leave-four-out cross-validation scheme as above to obtain these results, and we find that the information transfer mapping results with SVM decodings (Figure S2C) are identical to using representational similarity analysis (Figure 4E).

Note that information decoding was performed on the cortical surface, using vertices rather than voxels. This vertex-wise approach has been shown to provide better multivariate classifications than voxel-wise information decoding (Oosterhof et al., 2011), likely because surface analyses better reflect the underlying cortical anatomy.

Information transfer mapping was performed within subject between every pair of regions in the Glasser et al. (2016) atlas (360 regions in total). The success of this approach between all region pairs were then visualized via a 360-by-360 matrix (a total of 129,240 region-to-region mappings), where the regions along rows (source regions) indicated the activation patterns used to map onto another region's activation pattern (target region), which was indicated along the columns (Figures 6A, 6C, and 6E). Statistical tests were performed using a group one-sided t-test ($t > 0$) for every pair-wise mapping. A one-sided t-test was appropriate here given that our hypotheses were implicitly one-sided, since any significant deviation above 0 indicated a significantly higher matched correlation between predicted-to-actual activation patterns. Any information estimate that did not differ from 0 (or was negative) indicated that the predicted-to-actual similarity was at chance (akin to a chance decoding using classifiers). We tested for multiple comparisons using false discovery rate (FDR) for every region-to-region mapping, and significance was assessed using the FDR-corrected p-values of $p < 0.05$. Note that for region-to-region information transfer mapping, any vertices in a source region that fell into a 10mm radius of the to-be-predicted region (e.g., an adjacent region) would not contribute any activity flow to the to-be-predicted region (see FC estimation Methods section for details).

Given the visual sparsity of the region-to-region information transfer mapping visualization, we opted to down sample our matrix to provide a simpler visualization to assess how pairs of regions transfer information between and within functional networks (Figures 6B, 6D, and 6F). Thus, we computed the percent of statistically significant transfers for every pair of networks of the seven defined functional networks. This allowed us to better visually assess how region-to-region transfer mappings within specific network configurations may have been influenced by underlying network topology. To compute the percent of statistically significant transfers, we counted the number of significant transfers between every pair of networks and divided that by the total number of possible transfers within that network-to-network configuration. To characterize the generality with which information transfer mappings occurred between

specific network configurations, we computed the number of rule domains each network configuration contained at least one region-to-region transfer (Figure 6G). In other words, we took the matrices in Figures 6B, 6D, and 6F and binarized them with a 1 if a cell had a > 0 percentage of transfers, and a 0 otherwise. We then summed these matrices element-wise to obtain the number of rule domains each network configuration had a successful information transfer. To assess the number of rule domains each network contained at least one successful source region, we took the percent of significant transfers from each network to any other region in the brain (a 7-element array) and then binarized the array for each rule domain. We then summed across the three arrays (one for each rule domain) to obtain the number of rule domains each network had at least one successful source region used for information transfer (Figure 6H).

Lastly, to visualize the anatomical locations of the source regions for information transfer, we computed the percent of significant transfers from each cortical region for each rule domain (Figures 7A, 7B, and 7C). Percentages were obtained by taking the number of successful transfers from a region, and dividing it by total number of possible transfers (i.e., 359 other regions). We then plotted each of these percentages on the cortical surface using Connectome Workbench software (version 1.2.3) for each rule domain (Glasser et al., 2016b).

Network-to-network information transfer mapping

Network-to-network information transfer mapping in both the computational model (Figure 4E) and empirical data (Figures S1B, S1C, and S1D) was performed in the same computational framework as above, though instead of predicting region-level activation patterns using vertex-level activation patterns, network-level activation patterns were predicted using region-level activations (averaging across vertices within a given region). In other words, when predicting network B's region-level activation pattern, we computed the dot product between network A's region-level activity vector and the region-to-region resting-state FC matrix between regions in network A and B. We then submitted our 128 task block predictions for network B to our information transfer mapping procedure, as described above. This was repeated for every pair of the seven functional networks defined by our community-detection algorithm, resulting in 7-by-7 network-to-network mappings which were visualized as a 7x7 matrix (Figures S1B, S1C, and S1D). We tested for multiple comparisons using FDR for every network-to-network mapping within a rule domain, and significance was assessed using the FDR-corrected p-values of $p < 0.05$.

Behavioral paradigm

We used the Concrete Permuted Rule Operations (C-PRO) paradigm, which is a modified version of the original PRO paradigm introduced in Cole et al., (2010). Briefly, the C-PRO cognitive paradigm permutes specific task-rules from three different rule domains (logical decision, sensory semantic, and motor response) to generate dozens of novel and unique task sets. The primary modification of the C-PRO paradigm from the PRO paradigm was to use concrete, sensory (simultaneously presented visual and auditory) stimuli, as opposed to the abstract, linguistic stimuli in the original paradigm. Visual stimuli included either horizontal or vertical oriented bars with either blue or red

coloring. Simultaneously presented auditory stimuli included continuous (constant) or non-continuous (non-constant) beeps presented at high (3000Hz) or low (300Hz) frequencies. The paradigm was presented using E-Prime software version 2.0.10.353 (Schneider et al., 2002).

Each rule domain (logic, sensory, and motor) consisted of four specific rules, while each task set was a combination of one rule from each rule domain (Figure 2). A total of 64 unique task-sets (4x4x4) were possible, and each unique task-set was presented twice for a total of 128 task blocks. Identical task-sets were not presented in consecutive blocks. Each task block included three trials, each consisting of two sequentially presented instances of simultaneous audiovisual stimuli. A task block began with a 3925ms instruction screen (5 TRs), followed by a jittered delay ranging from 1570ms to 6280ms (2 – 8 TRs; randomly selected). Following the jittered delay, three trials were presented for 2355ms (3 TRs), each with an inter-trial interval of 1570ms (2 TRs). A second jittered delay followed the third trial, lasting 7850ms to 12560ms (10-16 TRs; randomly selected). A task block lasted a total of 28260ms (36 TRs). Subjects were trained on four of the 64 task-rule sets for 30 minutes prior to the fMRI session. The four practiced rule sets were selected such that all 12 rules were equally practiced. There were 16 such groups of four task-sets possible, and the task-sets chosen were counterbalanced across subjects. Subjects' mean performance across all trials performed in the scanner was 85% (median=86%) with a standard deviation of 8% (min=66%; max=96%). All subjects performed statistically above chance (25%).

fMRI Acquisition

Data were collected at the Rutgers Brain Imaging Center (RUBIC). 35 human participants (17 females) were recruited from the Rutgers University-Newark community and neighboring communities. We excluded three subjects, leaving a total of 32 subjects for our analyses; two subjects were excluded due to exiting the scanner early, and one subject was excluded due to excessive movement. All participants gave informed consent according to the protocol approved by the Rutgers University Institutional Review Board. The average age of the participants was 20, with an age range of 18 to 29. Whole-brain multiband echo-planar imaging (EPI) acquisitions were collected with a 32-channel head coil on a 3T Siemens Trio MRI scanner with TR = 785 ms, TE = 34.8 ms, flip angle = 55°, Bandwidth 1924/Hz/Px, in-plane FoV read = 208mm, 72 slices, 2.0 mm isotropic voxels, with a multiband acceleration factor of 8. Whole-brain high-resolution T1-weighted and T2-weighted anatomical scans were also collected with 0.8mm isotropic voxels. Spin echo field maps were collected in both the anterior to posterior direction and the posterior to anterior direction in accordance with the Human Connectome Project preprocessing pipelines (Glasser et al., 2013). A resting-state scan was collected for 14 minutes (1070 TRs), prior to the task scans. Eight task scans were subsequently collected, each spanning 7 minutes and 36 seconds (581 TRs). Each of the eight task runs were collected consecutively with short breaks in between (subjects did not leave the scanner).

fMRI Preprocessing

Imaging data was minimally preprocessed using the Human Connectome Project minimal preprocessing pipeline version 3.5.0, which included anatomical reconstruction and segmentation, EPI reconstruction, segmentation, spatial normalization to standard template, intensity normalization, and motion correction (Glasser et al., 2013). All subsequent preprocessing steps and analyses were conducted on CIFTI 64k grayordinate standard space for vertex-wise analyses and parcellated time series for region-wise analyses using the Glasser et al. (2016) atlas (i.e., one time series for each of the 360 cortical regions). We performed nuisance regression on the minimally preprocessed resting-state data using 12 motion parameters (6 motion parameter estimates plus their derivatives) and ventricle and white matter time series (extracted volumetrically), along with the first derivatives of those time series.

Task time series for task activation analyses were preprocessed in an identical manner to resting-state data. Task time series were additionally processed as follows. A standard fMRI GLM was fit to task-evoked activity convolved with the SPM canonical hemodynamic response function and the same 16 nuisance regressors as above. Block-by-block activity beta estimates were used for representational similarity analyses and activity flow mapping analyses. Task activity GLMs were performed at both the region-wise level and vertex-wise level for subsequent network-to-network and region-to-region information transfer mapping, respectively.

FC estimation

Given the success of FC estimation using multiple linear regression in our previous study (Cole et al., 2016), we employed multiple linear regression to estimate FC. To estimate FC to a given vertex, we used standard linear regression to fit the time series of all other vertices as predictors (i.e., regressors) of the target vertex. This provided an estimate of the contribution of each source vertex in explaining unique variance in the target region's time series. The same approach was used for region-to-region FC estimation, except the time series for each region were averaged across a given region's vertices prior to FC calculation. Multiple linear regression FC is conceptually similar to partial correlation, but is actually semipartial correlation, as the estimates retain information about scaling a source time series (i.e., regressor time series) into the units of the to-be-predicted time series (i.e., predicted variable).

For vertex-to-vertex FC estimation, due to computational intractability (i.e., more source regions/regressors than time points), we used principal components regression with 500 principal components. This involved reducing all source time series into 500 principal components and using the components as predictors/regressors to the target vertex. To avoid any possibility of spatial autocorrelation when estimating vertex-to-vertex FC, we excluded all vertices belonging to the same surface parcel (as determined by the Glasser atlas) as well as any vertices within 10mm of the border of that parcel in the principal components/regressors of the target vertex. (All vertices that fell within this criterion were given FC values of 0, preventing any vertices close to the target region from contaminating FC estimates.) Beta values obtained from the principal component regressors were then transformed back into the original 64k vertex space.

Network assignment of Glasser et al. (2016) parcels

Partitioning of the parcels (regions) into networks was based on the procedure used in Cole et al. (2014; see Supplemental Information). Specifically, we used the Louvain locally-greedy algorithm (Blondel et al., 2008; Jutla et al., 2011) for community detection. Data from the publically available Washington University-Minnesota Human Connectome Project "HCP100" dataset were used (N=100). Similar preprocessing procedures as used for the primary dataset were applied to the HCP100 dataset. Specifically, in addition to minimal preprocessing (Glasser et al., 2013), we ran a GLM nuisance regression using white matter, ventricles, and motion regressors (and their first derivatives). Global signal regression, motion scrubbing, and temporal filtering were not used. For each subject, all four resting state runs were concatenated and FC was estimated using standard Pearson correlations. The FC matrices were averaged across subjects to generate a group-mean resting-state FC matrix.

We searched over two free parameters to find a community partition for the group-mean resting-state FC matrix. The first parameter was the density threshold, whereby weak connections (based on the absolute value of FC strengths) were removed prior to running the community detection algorithm. The second parameter was the structural resolution parameter, which can be used to tune the number of communities identified in the FC matrix. The parameter search was conducted across combinations of these two parameters (density of 40% to 100% in increments of 5%, and resolution of 0.8 to 3 in increments of 0.05), with two criteria: 1) there should be a peak of partition similarity (z-score of the Rand coefficient; Traud et al., 2011) among adjacent locations in this two-dimensional parameter space, and 2) there should be distinct communities corresponding to visual, auditory, dorsal attention, default-mode, and motor/tactile systems (given decades of neuroscience research demonstrating their existence). Approximate locations of these systems were based on standard neuroscientific knowledge of these systems (given their strong establishment in the literature), in addition to their identification using resting-state FC in previous reports (Gordon et al., 2014; Power et al., 2011; Yeo et al., 2011). A five-community partition had the highest nearest-neighbor similarity in parameter space, but this did not separate out the auditory system. The next-highest nearest-neighbor similarity peak (density = 100%, resolution = 1.2) with distinct communities corresponding to auditory, visual, dorsal attention, default-mode, and motor/tactile systems was a 14-community partition. This partition was then visualized using Connectome Workbench software (Figure 3A). Labels were assigned to the seven most replicated networks identified using resting-state FC (Gordon et al., 2014; Power et al., 2011; Yeo et al., 2011). Colors were assigned to networks based on the colors used by Power et al. (2011).

Replication of network topological properties

We replicated a key property of resting-state network topology using our novel network assignments of the Glasser et al. (2016) parcels – high global connectivity of cognitive control networks. We included only functional networks which coincided with the seven most replicable functional networks found in three previously published network atlases (Gordon et al., 2014; Power et al., 2011; Yeo et al., 2011): the frontoparietal network (FPN), the dorsal attention network (DAN), the cingulo-opercular network (CON), the default mode network (DMN), the visual network (VIS), the auditory network (AUD), and the somatomotor network (SMN). We measured the average out-

of-network connectivity during resting-state FC, which was estimated using multiple linear regression (Figure 3D). To do this, we computed each individual parcel's average out-of-network intrinsic FC (i.e., out-of-network global brain connectivity), and then averaged across parcels within network to get an average statistic for each network. Out-of-network connections were defined as all connections from the source region to target regions outside the source region's network. To statistically test whether average out-of-network intrinsic FC was different for a pair of networks, we performed an across-subject paired t-test for every pair of networks. We corrected for multiple comparisons across pairs of networks using FDR (Genovese and Wasserman, 2002).

Neural network model

To validate our information transfer estimation approach we constructed a simple dynamical neural network model with similar network topological properties identified in our empirical fMRI data. We constructed a neural network with 250 regions, each of which were clustered into one of five network communities (50 regions per community). Regions within the same community had a 35% probability of connecting to another region (i.e., 35% connectivity density), and regions not assigned to the same community were assigned a connectivity probability of 5% (i.e., 5% out-of-network connectivity density). We selected one community to act as a "network hub", and increased the out-of-network connectivity density of those regions to 20% density. We then applied Gaussian weights on top of the underlying structural connectivity to simulate mean-field synaptic excitation between regions. These mean-field synaptic weights were set with a mean of $1.0/\sqrt{K}$ with a standard deviation of $0.2/\sqrt{K}$, where K is the number of synaptic inputs into a region such that synaptic input scales proportionally with the number of inputs. This approach was recently shown to be a plausible rule in real-world neural systems based on *in vitro* estimation of between-neuron synaptic-weight-setting rules (Barral and Reyes, 2016).

To simulate network-level firing rate dynamics, as similar to Stern et al. (2014), region x_i's dynamics for i=1,…,250 obeyed the equation

$$\frac{dx_i}{dt}\tau_i = -x_i(t) + s\ \phi\Big(x_i(t)\Big) + g\Big(\sum_{j\neq i}^{N} W_{ij}\ \phi\Big(x_j(t)\Big)\Big) + I_i(t)$$

.

We define φ as the transfer function tanh, x_j the dynamics of region j=1,…,250 for i≠j, $I_i(t)$ the input function (e.g., external spontaneous activity alone or both spontaneous activity and task stimulation) for i=1,…,250, W the underlying synaptic weight matrix, s the local coupling (i.e., recurrent) parameter, g the global coupling parameter, and τ_i the region's time constant. For simplicity, we set s, g, and τ_i equal to one, though we show in a previous study (Cole et al., 2016) that the activity flow mapping breaks down for parameter regimes s > g.

We first simulated spontaneous activity in our model by injecting Gaussian noise (mean of 0.0, standard deviation 1.0). Numerical simulations were computed using a Runge-Kutta second order method with a time step of dt=0.1s. We ran our simulation for 600 seconds (10 minutes). To simulate resting-state fMRI, we then convolved our time series with the SPM canonical hemodynamic response function and down sampled to a

1 second TR, resulting in 600 time points. We then computed resting-state FC using multiple linear regression. To replicate the empirical data, we computed the out-of-network resting-state FC (as in the empirical data) to validate that widespread out-of-network connectivity was preserved from synaptic to functional connectivity.

To model task-evoked activity, we simulated four distinct task conditions by injecting stimulation into four randomly selected but distinct sets of twelve regions in the hub network. Stimulation to the hub network was chosen to mimic four distinct top-down, cognitive control task-rules. Task stimulation coincided with spontaneous activity (e.g., for time points t during a task, I(t) = spontaneous activity at t + 0.5 constant task stimulation). We ran each task for 20 blocks, where each block lasted for 100 seconds. Each block contained five trials, each lasting for five seconds with an inter-trial interval of 15 seconds. In total, each task condition contained 100 task trials, with 500 seconds per task total. We then convolved these task time series with the SPM canonical hemodynamic response function and down sampled to 1-second TRs, as in the resting-state simulation.

We validated the usefulness of the model for characterizing hub-related dynamics by testing whether estimated resting-state FC preserved the hub network's higher out-of-network intrinsic FC (specified by its underlying synaptic connectivity). Out-of-network intrinsic FC was computed in the same way as in the empirical data (see above) for each of the network model's communities. For each of the five networks, we compared the out-of-network intrinsic FC between each network using an across-subject t-test. We corrected for multiple comparisons using FDR (Genovese and Wasserman, 2002) and significance was assessed with an FDR-corrected $p < 0.05$ threshold.

To perform network-to-network information transfer mapping in the model, we used the task-evoked activity (estimated by standard GLM beta estimates), and performed the information transfer mapping procedure between networks of regions using the resting-state FC matrix obtained via multiple linear regression. Network-to-network information transfer mapping is computationally identical to region-to-region information transfer mapping, and is described above. The information transfer mapping matrix (Figure 4E) was obtained using an FDR-corrected threshold of $p < 0.05$.

Baseline representational similarity analyses for regions and networks

To compute the baseline (i.e., unrelated to FC) multivariate representational content at the region level (Figure 5), we performed a within-subject cross-validated representational similarity analysis for every Glasser et al. (2016) parcel. We used vertex-level beta estimates for each task block, where each block was associated with a task-rule condition for a particular rule domain. Given the 128 task blocks, we computed a leave-four-out (one block of each task-rule held out) representational similarity analysis to identify whether a region's activity vector (consisting of vertex-level activity) preserved task-rule information. Prior to running the representational similarity analysis, all blocks were spatially demeaned to increase the likelihood that the representation we were identifying was a multivariate regional pattern (rather than a change in region-level mean activity).

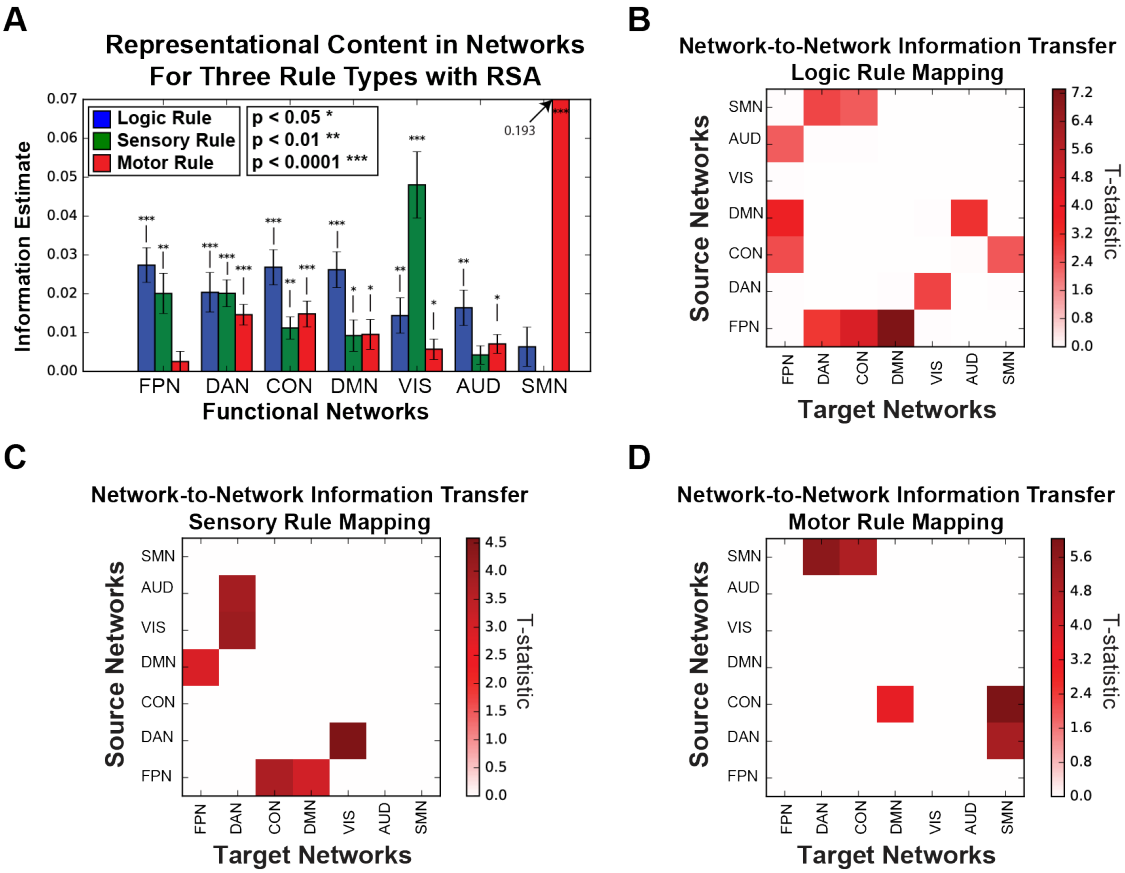
For each cross-validation fold, task blocks in the training set associated with the same task-rule (e.g., all 'both' rules) were averaged together to obtain four prototypical

region-level activity vectors. A spatial Spearman's rank correlation was computed between every pair of held-out (testing) data (a vector per task-rule condition) and the task-rule prototypes (a vector per condition). In our paradigm, since each rule domain contained four task-rules, we performed 4x4 comparisons resulting in a 4x4 similarity matrix. Correlations between rule-matched activity vectors were averaged together, while correlations between rule-mismatched activity were averaged together. (In the above similarity matrix, this would be computing the difference between the average of the diagonal and average of the off-diagonal.) The difference between the Fisher-Z transformed matched and mismatched correlation averages was taken for each fold to obtain our "information estimate". An average information estimate was then obtained from each subject by computing the average information estimate across all cross-validation folds (32 cross-validation folds for a four-way similarity analysis across 128 task blocks). Statistical significance was assessed by taking a group t-test against 0, since a greater than 0 difference of matches versus mismatches indicated significant preservation of task-rule representations. All p-values were corrected for multiple comparisons for all 360 parcels using FDR, and significance was assessed using a threshold of a corrected $p < 0\ .05$.

For network-level information estimates (Figure S1A), the same cross-validated representational similarity analysis procedure was conducted for the seven functional networks separately across the three rule domains, using region-level representations within each of the networks. Region-level beta estimates were obtained for every block by fitting the same GLM model as described above to every region separately. All p-values were FDR-corrected for multiple comparisons across seven networks, and significance was assessed using an FDR-corrected $p < 0.05$.

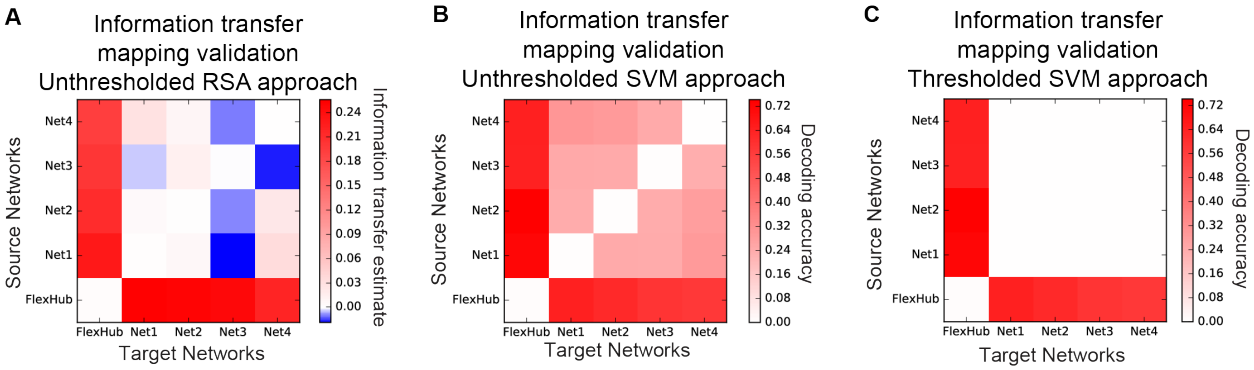
Region-to-region information transfer mapping, vertex-to-vertex FC estimation, baseline representational similarity analyses, and model simulations were performed on the Rutgers University supercomputer cluster (NM3).

Supplementary Figures



Supplementary Figure 1. Task information persist in functional networks and are transferred between networks via network-to-network activity flow mapping. **A)** Representational content of task-rule information across three rule domain prior to performing activity flow mapping. The seven networks contain statistically significant representations of at least two rule domain using a cross-validated representational similarity analysis method. In particular, the motor network contains the highest representational information for motor rule distinctions. In addition, all seven networks contain logic rule representations, suggesting that abstract rule representations are highly distributed across cortical networks. **B)** Network-to-network information transfer mapping of the logic rules. As in Figure 5, functional networks along the rows indicate the activation patterns that were projected to the networks indicated along the columns. All maps are thresholded using an FDR-corrected $p < 0.05$, and colors indicate the T-statistic. While most representational projections were bidirectional, the amount of representation projected from a networks A to B and B to A were not necessarily equal. For example, activity flow mappings *from* the FPN were on average greater than activity flow mappings *to* the FPN. Results from our computational model show that a better preserved representation in a Network A to B direction than B to A reflects the direction in which information flows. **C)** Network-to-network information transfer mapping of the sensory rules. Sensory rule representations flow between the FPN and other domain-general networks (DMN, CON), while sensory representations flow from sensory networks (AUD, VIS) to the DAN. **D)** Network-to-network activity flow mapping of motor rules. Information transfer mapping of motor

rule representations demonstrate the most specificity. Representational information only flows between the DAN <-> SMN, CON <-> SMN, and CON -> DMN.



Supplementary Figure 2. Computational validation of information transfer mapping with different decoding approaches. A) Unthresholded information transfer mapping validation using cross-validated representational similarity analysis. Here, we show the unthresholded map of the information transfer mapping as shown in Figure 4E. **B)** Unthresholded information transfer mapping validation using support vector machines (SVMs). In contrast to the representational similarity approach, we train a linear classifier using predicted activation patterns of a target network (given some source region's activation pattern), and then classify the actual, held-out activation patterns of the target network. This demonstrates that a classifier trained on predicted data can decode task information as in the real data. As in part (A), we use the same leave-four-out cross-validation scheme. **C)** Thresholded information transfer mapping validation using SVMs. For every network-to-network information transfer mapping, we perform an across-subject t-test against chance (25% for a four-way classification). To assess statistical significance, we corrected for multiple comparisons using FDR with a threshold of $p < 0.05$. We find qualitatively identical results as in our computational validation result (Figure 4E) using representational similarity analysis.