

Dissociable changes in functional network topology underlie early category learning and development of automaticity



Fabian A. Soto ^{a,*}, Danielle S. Bassett ^{b,c}, F. Gregory Ashby ^d

^a Department of Psychology, Florida International University, Miami, FL 33199, USA

^b Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, USA

^c Department of Electrical & Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA

^d Department of Psychological & Brain Sciences, University of California, Santa Barbara, CA 93106, USA

ARTICLE INFO

Article history:

Received 27 October 2015

Accepted 14 July 2016

Available online 21 July 2016

Keywords:

Functional network

Network science

Category learning

Automaticity

Multiple memory systems

ABSTRACT

Recent work has shown that multimodal association areas—including frontal, temporal, and parietal cortex—are focal points of functional network reconfiguration during human learning and performance of cognitive tasks. On the other hand, neurocomputational theories of category learning suggest that the basal ganglia and related subcortical structures are focal points of functional network reconfiguration during early learning of some categorization tasks but become less so with the development of automatic categorization performance. Using a combination of network science and multilevel regression, we explore how changes in the connectivity of small brain regions can predict behavioral changes during training in a visual categorization task. We find that initial category learning, as indexed by changes in accuracy, is predicted by increasingly efficient integrative processing in subcortical areas, with higher functional specialization, more efficient integration across modules, but a lower cost in terms of redundancy of information processing. The development of automaticity, as indexed by changes in the speed of correct responses, was predicted by lower clustering (particularly in subcortical areas), higher strength (highest in cortical areas), and higher betweenness centrality. By combining neurocomputational theories and network scientific methods, these results synthesize the dissociative roles of multimodal association areas and subcortical structures in the development of automaticity during category learning.

© 2016 Elsevier Inc. All rights reserved.

Introduction

Network science provides a set of robust tools that are increasingly used to describe and understand neural systems (Bullmore and Sporns, 2009; Sporns, 2014). Neurons or brain regions are represented as network nodes, and structural or functional connections between regions are represented as network edges. Recent studies demonstrate that the topology of functional brain networks can reconfigure quickly as the result of learning (Bassett et al., 2013b; Bassett et al., 2011, Bassett et al., 2015) and task engagement (Bassett et al., 2006; Ekman et al., 2012; Fornito et al., 2012; Kitzbichler et al., 2011). In several cases, this reconfiguration leads to more integrated and less segregated processing (Cole et al., 2014; Ekman et al., 2012; Kitzbichler et al., 2011) and involves strong reconfiguration in some nodes, while global network properties can remain relatively stable (Bassett et al., 2006; Moussa et al., 2011; Rzucidlo et al., 2013; Braun et al., 2015). In particular, nodes in multimodal association areas—within frontal, temporal,

and parietal cortex—flexibly change their community affiliation during learning (Bassett et al., 2013b; Bassett et al., 2011, Bassett et al., 2015), their connectivity pattern during rule application and preparatory attention (Cole et al., 2013; Ekman et al., 2012), and the cost-efficiency of their connectivity during accurate performance of working memory tasks (Bassett et al., 2009; Braun et al., 2015).

Despite these results, it is unlikely that connectivity changes in cortical association areas underlie functional network reconfigurations across all tasks. For example, connectivity changes and integrative processing in the basal ganglia are likely to be of utmost importance during initial learning of some categorization tasks (Ashby and Ennis, 2006). A body of behavioral and neurobiological evidence suggests that the brain areas associated with categorization are organized in relatively separate category learning systems and that different categorization tasks engage the systems differently (Ashby and Maddox, 2005; Nomura and Reber, 2008; Poldrack and Foerde, 2008; for a formalization of this multiple-systems hypothesis in a neurocomputational model, COVIS, see: Ashby et al., 1998; Ashby, Paul, & Maddox, 2011). Rule-based tasks, in which the optimal strategy is easy to verbalize and can be learned through a logical reasoning process, recruit a declarative-learning system that is based on explicit reasoning and hypothesis

* Corresponding author at: Department of Psychology, Florida International University, Modesto A. Maidique Campus, 11200 SW 8th St, AHC4 460, Miami, FL 33199, USA.

E-mail address: fabian.soto@fiu.edu (F.A. Soto).

testing. Learning in this system is implemented in a network of areas including prefrontal cortex, basal ganglia, and hippocampus. Many of the previous studies reporting network reconfiguration during learning and task performance are similar to rule-based tasks in that they seem to rely heavily on executive function (e.g., Bassett et al., 2011; Braun et al., 2015; Cole et al., 2013; Ekman et al., 2012). This might explain why connectivity changes in cortical association areas accompanied network reconfigurations in such studies.

On the other hand, learning of information-integration categorization tasks does not require executive function. Information-integration tasks require the integration of information from two or more stimulus components at a pre-decisional stage, and they recruit a procedural-learning system implemented in the circuitry of the basal ganglia (caudate, putamen, pallidum, and related thalamic nuclei). Thus, it is likely that changes in connectivity in the basal ganglia and related subcortical structures underlie network reconfiguration during learning of information-integration categorization tasks.

Even so, Ashby et al. (2007) proposed that in contrast to early learning, automatic categorization is mediated entirely within cortex and that the development of automaticity is associated with a gradual transfer of control from the basal ganglia to cortical–cortical projections from the relevant sensory areas directly to the premotor areas that initiate the behavior (see also, Ashby et al., 2010; Helie et al., 2015). Some neuroimaging results support this view of how automaticity develops (DeGutis and D'Esposito, 2009; Waldschmidt and Ashby, 2011).

During the acquisition of virtually all skills, improvements in accuracy asymptote long before improvements in response time (e.g., Crossman, 1959; Helie et al., 2010). Numerical simulation studies show that the relatively fast changes in accuracy that occur during early skill acquisition are likely to reflect learning-related changes in the basal ganglia and related subcortical areas, whereas the slower changes in the speed of correct responding likely reflect the switch to cortically controlled automatic performance (Ashby et al., 2007). This dissociation in behavioral measures can be used to study whether and how changes in functional networks are related to different stages of category learning. We can expect changes in the connectivity of subcortical areas—instead of cortical association areas—to predict initial category learning best. Furthermore, this central role of the basal ganglia should be more apparent in the prediction of accuracy than in the prediction of response times.

Here we explore these predictions using a combination of network science and multilevel regression (Gelman and Hill, 2007). We study how changes in the connectivity of brain regions can predict behavioral changes during extensive training in a task known to foster procedural category learning (Ashby et al., 2003). Our analysis approach is illustrated in Fig. 1. The red broken-line boxes represent points where data entered the analysis. Structural images were used to define 742 small clusters of voxels that were used as units for subsequent analyses (i.e., network nodes). Only nodes localized in a number of regions of interest (ROI) were included in the analysis. These ROIs were chosen based on neurocomputational theory and previous research on the neural correlates of category learning (see Soto et al., 2013). Thus, the analyses focus specifically on the brain network thought to be involved in category learning.

Functional scans from each block of training were preprocessed and the average BOLD signal was computed from each cluster of voxels defining an individual node. Functional connectivity matrices were built by computing the wavelet correlation between average BOLD signals and then thresholding these correlations. The functional connectivity matrices were then used to compute a number of graph measures (for a summary description of each measure, see Table 1) for each node of the network, providing a characterization of the node's topological role in the functional network at a particular point during categorization training (that is, during each block).

Changes in network measures across training were used in regression analyses to predict corresponding changes in accuracy and

response times. Based on our hypotheses, we expected that measures computed from subcortical nodes, instead of nodes located in cortical association areas, would predict initial category learning best, and that the importance of subcortical areas would be more apparent in the prediction of accuracy than in the prediction of response times.

Finally, regression coefficients were analyzed further to explore the specific relation between each predictor measure and behavior.

Materials and methods

Experimental procedures

Participants

Ten healthy undergraduate students from the University of California, Santa Barbara (6 males, 4 females), voluntarily participated in this study in exchange for course credit or a monetary compensation. This is a small but sufficient sample size (Snijders and Bosker, 2012) that has been shown to provide unbiased estimates of regression coefficients in multilevel regression (Bell et al., 2014; Maas and Hox, 2005; see discussion in the supplementary material). All participants gave their written informed consent to participate in the study. The institutional review board of the University of California, Santa Barbara, approved all procedures in the study.

Standard univariate and multivariate analyses of the imaging data acquired on this sample have been previously reported (Soto et al., 2013; Waldschmidt and Ashby, 2011). We excluded one person from the full sample of eleven participants due to incomplete data.

Behavioral task

The stimuli were circular sine-wave gratings of constant contrast and size (see example in Fig. 3A) that varied in orientation from 20° to 110° and in frequency from 0.25 to 3.58 cycles per stimulus width. Fig. 3B shows the category structure used to train participants; each dot in the figure represents a different stimulus and the dotted line represents the boundary separating the two categories. Previous research suggests that this task is mastered through procedural learning (e.g., Ashby et al., 2003; Maddox et al., 2004). During each trial, participants were presented with one of these stimuli and had to identify the category to which the stimulus belonged by pressing a button; this was followed by feedback indicating the accuracy of the response. Stimuli were presented and responses were recorded using MATLAB augmented with the Psychophysics Toolbox (Brainard, 1997), running on a Macintosh computer. For a more detailed description of the stimuli and apparatus, see Helie et al. (2010).

The experiment consisted of 23 sessions of training in the categorization task, four of which were conducted in the MRI scanner. The training sessions were carried out over 23 consecutive workdays, one session per day. The scanning sessions were sessions 2, 4, 10, and 20, and each consisted of 6 blocks of 80 stimuli, for a total of 480 stimuli per session. Participants selected their responses through response boxes, where the button box in their left hand was correct for the category at the top-left of the bound in Fig. 2B, and the button box in their right hand was correct for the category at the bottom-right of the bound in Fig. 2B. Feedback was displayed for 2 s and consisted of a green check mark for correct responses or a red “X” mark for incorrect responses. If it took more than 2 s for the participant to respond, a black dot was displayed indicating that the response was too slow. Half of the trials included the presentation of a cross-hair before the stimulus presentation.

The 19 sessions of categorization training outside the scanner were similar to the scanner session but carried out on a Macintosh computer. For a more detailed description of these sessions, see Helie et al. (2010).

Neuroimaging

A rapid event-related fMRI procedure was used. Images were obtained using a 3T Siemens TIM Trio MRI scanner at the University of California, Santa Barbara Brain Imaging Center. The scanner was equipped

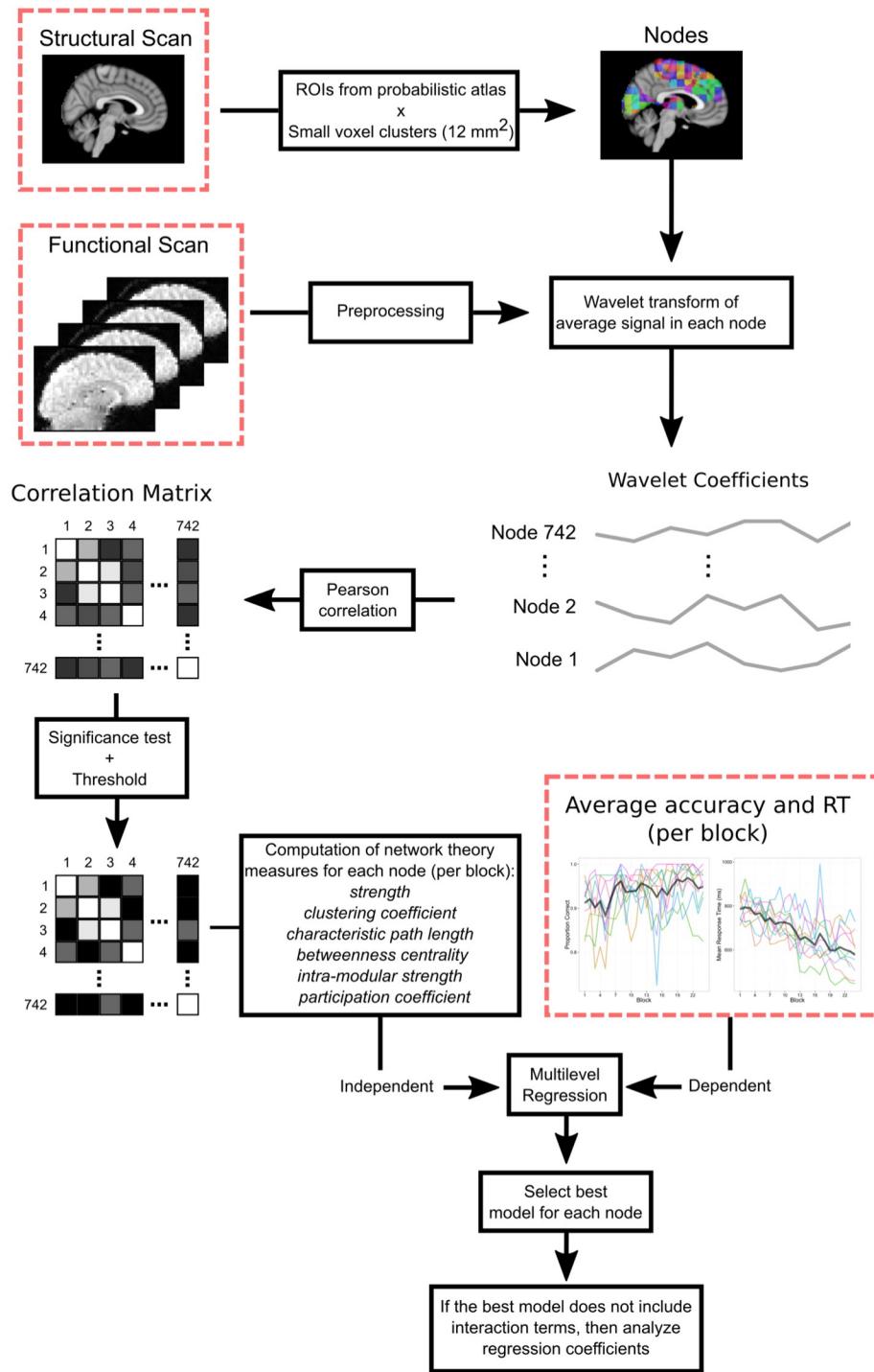


Fig. 1. A graphical summary of the data analysis procedures used in this study. Broken-line boxes represent points in the flowchart where data are entered into the analysis.

with an 8-channel phased array head coil. Cushions were placed around the head to minimize head motion. A localizer, a GRE field mapping (3 mm thick; FOV: 192 mm; voxel: $3 \times 3 \times 3$ mm; FA = 60°), and a T1-flash (TR = 15 ms; TE = 4.2 ms; FA = 20°; 192 sagittal slices 3-D acquisition; 0.89 mm thick; FOV: 220 mm; voxel: $0.9 \times 0.9 \times 0.9$ mm; 256×256 matrix) were obtained at the beginning of each scanning session, and an additional GRE field-mapping scan was acquired at the end of each scanning session. Functional runs used a T2*-weighted single shot gradient echo, echo-planar sequence sensitive to BOLD contrast (TR: 2000 ms; TE: 30 ms; FA: 90°; FOV: 192 mm; voxel: $3 \times 3 \times 3$ mm)

with generalized auto calibrating partially parallel acquisitions (GRAPPA). Each scanning session lasted approximately 90 min.

Data analysis

A graphical summary of the data analysis procedures is presented in Fig. 1. The starting points in this flowchart, where data are entered into the analysis, are highlighted with broken-line boxes. The procedure starts with the definition of 742 small clusters of voxels that were used as units of subsequent analyses (see first row in Fig. 1), which

Table 1

Summary of the network measures included in the present study and their observed relation to behavior. Symbols: i, j , and h index nodes; N is the total number of nodes; m indexes modules; M is the total number of modules; W_{ij} is the weighted connection between nodes i and j (measured as the absolute value of the correlation between nodes); k_i is the degree of node i (the number nodes in the network that are connected to node i); ρ_{hj} is the number of weighted shortest paths between nodes h and j ; $\rho_{hj}(i)$ is the number of those paths that pass through node i ; g_i is the module to which node i is assigned, $\delta(g_i, g_j)$ is the Kronecker delta (equal to one when g_i and g_j are equal, and zero otherwise); \bar{s}^m and σ^m are the mean and standard deviation (respectively) of the strength distribution within module m .

Network Measure	Equation	Typical interpretation	Average relation to accuracy	Average relation to speed	Interpretation of results
Strength	$S_i = \sum_{j=1}^N W_{ij}$	Magnitude of connectivity between two areas; nodes with a high strength are hubs of the system, thought to play an influential role in routing information.	Negatively related to accuracy across most ROIs. Positively related to accuracy in visual areas, particularly in extrastriate cortex.	Overall, slight bias toward a positive relation across the brain. Not localized in any particular ROI or area type.	Higher accuracy was accompanied by a reduction in connectivity across the brain, but an increase in the connectivity of visual areas. Higher speed was accompanied by a slight bias toward stronger connectivity across the brain, but this was not consistent in any given region.
Clustering coefficient	$C_i = \frac{\sum_{j=1}^N \sum_{h=1}^N \sqrt[3]{W_{ij} W_{ih} W_{jh}}}{k_i(k_i-1)}$	Measure of local information processing; nodes with a high clustering coefficient are thought to process information in local neighborhoods inside of the network.	Positively related to accuracy across all ROIs.	Positively related to speed across most ROIs. The relation was particularly strong in subcortical areas.	Higher accuracy and speed across training in the categorization task was accompanied by an increase in local information processing in the categorization network.
Characteristic path length	$L_i = \frac{\sum_{j=1}^N \sum_{u,v \in \gamma_{i,j}} \frac{w_{uv}}{w_{ij}}}{N-1}$	Measure of the efficiency of information transmission across the whole brain; nodes with a short characteristic path length are thought to be able to transmit information quickly to all areas of the brain.	Excluded from the analysis due to almost perfect negative correlation with strength. See results for strength.	Excluded from the analysis due to almost perfect negative correlation with strength. See results for strength.	Excluded from the analysis due to almost perfect negative correlation with strength. See results for strength.
Betweenness centrality	$b_i = \sum_{\substack{h,j \\ h \neq i, h \neq j}} \frac{\rho_{hj}(i)}{\rho_{hj}}$	Measure of bottlenecks or critical points in information transmission; nodes with high betweenness centrality are thought to be hubs of the system that are critical in routing information between otherwise disconnected sets of brain regions.	Inconsistent relation across nodes and area types. Some specific ROIs show stronger relations on average (see main text).	Negatively related to speed in most ROIs. A small number of ROIs had a small positive relation (see main text).	Higher speed across training was related to a decrease in routing of information transmission through nodes.
Intra-modular strength (standardized)	$\frac{z_i = \bar{s}_i^m - \bar{s}^m}{\sigma^m}$, where $\bar{s}_i^m = \sum_{j=1}^N W_{ij} \delta(g_i, g_j)$	Measure of the magnitude of connectivity that a node has to its own cognitive system, or functional module; nodes with high intra-modular strength are thought to process information predominantly related to the function of that module.	Positively related to accuracy in subcortical and motor areas. Negatively related to accuracy in visual areas, frontal areas and hippocampus.	Inconsistent relation across nodes and area types. Some specific ROIs show stronger relations on average (see main text).	Higher accuracy across training was related to stronger connectivity of subcortical and motor areas within their functional modules, but weaker connectivity of visual areas, frontal areas, and hippocampus within their functional modules.
Participation coefficient	$P_i = 1 - \sum_{m=1}^M \left(\frac{\bar{s}_i^m}{\bar{s}^m} \right)^2$	Measure of the degree to which a node communicates with other cognitive systems or functional modules; a node with a high participation coefficient is thought of as a connector node, that can potentially transmit or gate information between modules.	Positively related to accuracy across most ROIs.	Inconsistent relation across nodes and area types. Some specific ROIs show stronger relations on average (see main text).	Higher accuracy was accompanied by an increase in connectivity across functional modules in the categorization network.

was performed on a standard T1-weighted structural image (MNI152-T1-2 mm). Our analysis focused only on a number of ROIs known to be related to category learning from previous research, which are subdivided into a number of smaller (12 mm^2) voxel clusters, which served as nodes in the network analysis.

Functional scans from each block of training were separately preprocessed and transformed to standard MNI152-T1-2 mm space. Then the average BOLD signal from each node was computed and a wavelet transform was applied to this average, resulting in a vector of wavelet coefficients for each node (see second row in Fig. 1). Wavelet coefficients corresponding to a scale in the range 0.06–0.125 Hz were correlated across nodes, to obtain a 742×742 correlation matrix representing functional connectivity in the brain network associated with category learning (see third row in Fig. 1). Correlation matrices were then thresholded to obtain sparser connectivity matrices, from which a number of graph statistics were computed (see left part of fourth row in Fig. 1). Graph statistics were computed for each node of the network, providing a characterization of its topological role in the

network at a particular point in training (block). Note that our focus on local node measures (and exclusion of global measures) is related to the goals of this study: local statistics are much more sensitive to local changes in functional connectivity, which we expect to be more relevant than global changes, based on the processes known to drive category learning in humans.

From the raw behavioral data of each participant, we obtained mean response times and mean proportion of correct responses in each block of training (rightmost box in fourth row of Fig. 1). These measures of performance were separately used as a dependent variable in a multi-level regression, which had the previously computed graph statistics as independent variables. Thus, the regression analysis was repeated at each node, and it represented an attempt to determine to what extent changes in the topological role of a node, as quantified through network science tools, could explain changes in a specific measure of performance. Recall that changes in the two measures of performance, response times, and accuracy, are thought to reflect different learning processes (see “Introduction”).

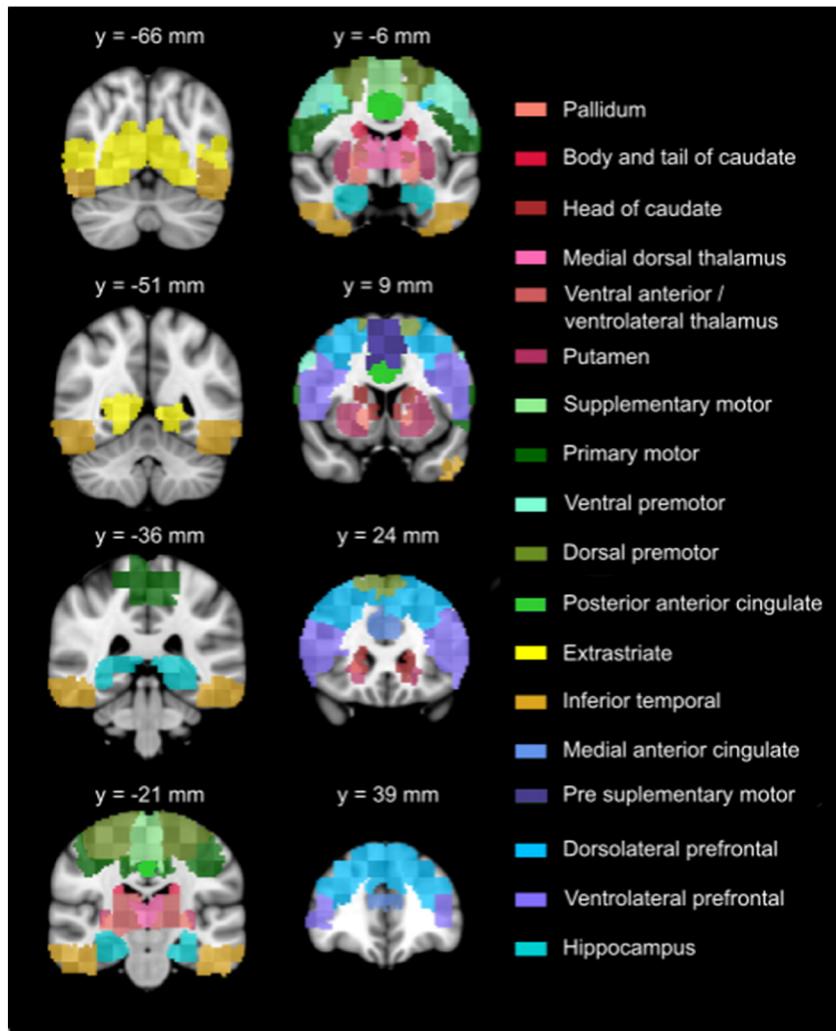


Fig. 2. Location of the ROIs and some of the nodes used in this study. Different ROIs are represented by different hues. ROIs belonging to the same anatomical category have been painted using similar hues, with hues close to red representing subcortical areas, hues close to green representing motor areas, hues close to yellow representing visual areas, and hues close to blue representing high-level cortical areas. Different nodes inside each ROI are painted using different levels of brightness.

A number of regression models varying in complexity (i.e., the level of interactions among graph statistics included in the model) were fitted to the data from each node, and standard model selection procedures were used to (a) determine whether a particular node was predictive of changes in behavior and (b) select the model that explained the data best at a predictive node (see bottom of Fig. 1). This first analysis allowed us to determine to what extent changes in the topological role of small brain regions could explain changes in behavior across training, and it included models of any complexity (i.e., any level of interactions among graph statistics) to explore the possibility that complex, nonlinear relations exist between graph statistics and behavior. An additional analysis of regression coefficients was carried out to determine the specific relation between each individual graph statistic and behavior. This analysis included only nodes with a best-fitting model that lacked interaction terms, because only in this case are regression coefficients easily interpretable.

In the following sections, each of the steps in our data analysis pipeline is described in more detail.

Definition of regions of interest

We were interested in measuring the functional connectivity of local clusters of voxels within ROIs that are thought to be part of the brain network involved in category learning. With this goal in mind, we

started by defining 36 anatomical ROIs (18 anatomical areas, one in each hemisphere) previously reported to be involved in visual category learning and automaticity. For a more in-depth discussion of the hypothesized functional role of each ROI in categorization and the empirical evidence, see Soto et al. (2013)¹.

The anatomical boundaries of each ROI were created in MNI152-T1-2 mm standard space using atlases included in FSL (Harvard–Oxford structural atlases, Oxford thalamic connectivity probability atlas, and Juelich histological atlas). Fig. 2 includes eight coronal slices showing the location of these ROIs in standard space. Different areas are filled

¹ The neurocomputational models that inspired the present study do not clearly identify what visual sensory areas are involved in category learning, as different tasks are likely to recruit different areas. The present analyses include extrastriate cortex because this area was active during stimulus processing according to previous analyses of the data (Waldschmidt and Ashby, 2011). Inferotemporal cortex is included due to its importance among visual areas in the literature on the neurobiology of visual categorization. Although primary visual cortex was included in previous analyses (Soto et al., 2013), this region was not included in the present study. Primary visual cortex is a quite large area that would have added many nodes to our networks, substantially increasing the computational cost of our analyses. On the other hand, primary visual cortex is neither recruited for stimulus processing in the current task (Waldschmidt and Ashby, 2011) nor consistently associated with visual categorization in the previous literature. Furthermore, unlike other visual areas, primary visual cortex does not send projections to the striatum of the basal ganglia.

with different colors, with similar hues for visual (close to yellow), motor (close to green), subcortical (close to red), and high-level cortical areas (close to blue). Visual areas included the extrastriate cortex and inferotemporal cortex. Motor areas included primary motor cortex and premotor cortex. The premotor cortex was divided into the supplementary motor area (SMA), pre-SMA, ventral premotor area, and dorsal premotor area as defined by (Picard and Strick, 2001). Subcortical areas included the caudate, putamen, pallidum, and thalamic areas. The caudate was divided into a head region and a body and tail region according to Nolte (2008). Using the Oxford thalamic connectivity probability atlas, we defined the medial dorsal nucleus of the thalamus as the thalamic area connected to prefrontal cortex, and the ventral anterior and ventral lateral nuclei of the thalamus as the thalamic area connected to primary motor and premotor cortices (Martin, 2003). Other cortical areas included the anterior cingulate cortex (ACC), prefrontal cortex (PFC), and hippocampus. The medial and posterior parts of the ACC were extracted following (Vogt et al., 2004). Following the definition given by Petrides and Pandya (2004), the dorsolateral PFC was extracted by joining the superior and middle frontal gyri and subtracting all premotor areas. Ventrolateral PFC was extracted using an inclusive definition from (Petrides and Pandya, 2004), which includes Brodmann areas 44, 45, and 47. For each of the aforementioned areas, we defined a left and a right ROI.

Definition of network nodes

The study of functional networks requires a definition of nodes (regions of interest) and a definition of edges (measures of the functional connectivity between nodes). Functional connectivity measures are stored in an adjacency matrix \mathbf{W} , with cell W_{ij} representing the functional connectivity between nodes i and j . The matrix \mathbf{W} is the starting point for all network-based analyses (see below). Two common ways to define nodes in the study of functional networks are (i) the areas obtained from a brain atlas and (ii) spheres of a given radius centered at coordinates of interest (Varoquaux and Craddock, 2013). Here we take an intermediate strategy between those two approaches. We focused on the atlas-based ROIs defined in the previous section, which are thought to cover most of the brain network involved in visual category learning (see Soto et al., 2013). This approach allowed us to cover as much as possible of the functional network related to category learning that has been identified in the previous literature. Simply defining nodes using statistical contrasts on our neuroimaging data would have reduced coverage of such network and it would have forced us to focus on regions that are important at the group level (it was necessary to use a single definition of nodes for all participants, to perform analyses at each node, see below), discarding important nodes at the level of individual participants. However, averaging the signal from such large areas could make it difficult to determine whether small clusters of voxels within each area are specifically related to categorization performance. Therefore, we subdivided each ROI into small clusters of voxels. Starting with a brain mask in MNI152-T1-2 mm standard space, we defined a rectangular box that enclosed this mask and divided it into $12 \times 12 \times 12$ mm cubes ($6 \times 6 \times 6$ voxels). This size is comparable to the spheres of 5–10 mm radius commonly used in the literature (Power et al., 2011; Varoquaux and Craddock, 2013). If at least 75 voxels (~35%) of a cube fell inside an ROI, then the overlap between the cube and the ROI was defined as a node (for a similar approach, see Meunier et al., 2009). There was no spatial overlap between nodes (i.e., each voxel was included in only one node). This resulted in a total of 742 nodes, an order of magnitude larger than the common Automated Anatomical Labeling (AAL) atlas (Achard et al., 2006), finely covering the visual category learning network.

fMRI data preprocessing

The data series from each block was preprocessed using FEAT (fMRI Expert Analysis Tool) version 5.98 in FSL (www.fmrib.ox.ac.uk/fsl). Preprocessing included motion-correction to the middle volume in

the series using tri-linear interpolation with six degrees of freedom in MCFLIRT (Jenkinson et al., 2002), slice timing correction (via Fourier time series phase-shifting), BET brain extraction, and a high pass filter with a cutoff of 50 s. The data were not spatially smoothed during pre-processing to avoid artificially increasing the correlation between adjacent network nodes (Achard et al., 2006). Each functional scan was registered to the corresponding structural scan using FLIRT (Jenkinson et al., 2002; Jenkinson and Smith, 2001) linear registration with its default settings. Each structural scan was registered to the MNI152-T1-2 mm standard brain using FNIRT (Andersson et al., 2007) nonlinear registration with its default settings. The resulting linear and nonlinear transformations were jointly used to transform the data series from subject space to standard space using tri-linear interpolation.

Computation of functional connectivity matrices

We built a functional connectivity matrix from each of the 24 preprocessed data series (from 6 separate blocks in each of 4 scanning sessions) separately for every participant. We chose to focus on task-related functional connectivity, and therefore removed all volumes before the first stimulus presentation and 20 s after the last stimulus presentation. The number of volumes in each time series varied across individual scans, with a mean of 298.53 and standard deviation of 13.16. The time series of each voxel in a node were averaged and the maximum-overlap discrete Daubechies-4 wavelet transform (Percival and Walden, 2000) of this averaged signal was computed using the WMTSA toolbox for MATLAB version 0.2.6 (Cornish, 2006). The wavelet transform was computed at three scales and, because the sampling frequency (TR) was 2 s, scale one corresponded approximately to 0.125–0.25 Hz, scale two corresponded approximately to 0.06–0.125 Hz, and scale three corresponded approximately to 0.06–0.03 Hz. All analyses focused on scale two, which contains the most relevant information for the goals of the present study and has been used in previous studies relating changes in network measures to behavior (e.g., Bassett et al., 2011; Ekman et al., 2012). Although scale one includes frequencies below the Nyquist frequency of 0.25, which could potentially carry information about functional connectivity, the hemodynamic response function acts as a low-pass filter on the underlying neural activity and the signal in high frequencies close to the Nyquist limit can become uninformative (Sun et al., 2004). This is confirmed by experimental estimates showing that both task-related functional connectivity (Richiardi et al., 2011; Sun et al., 2004) and resting-state functional connectivity (Cordes et al., 2001; Richiardi et al., 2011) are found predominately in the low-frequency band of 0.00–0.15 Hz and not in higher frequencies.

To build each functional connectivity matrix, we computed the Pearson correlation coefficient between the wavelet coefficients corresponding to each pair of nodes. We chose a threshold value above which the correlation coefficients were retained and below which the correlation coefficients were set to zero. The threshold was chosen separately for each correlation matrix, in two steps that had the goals of controlling false positive statistical associations and obtaining sparse networks. First, a *t*-test was used to determine whether the correlations deviated significantly from zero and the obtained *p*-values were corrected for multiple comparisons using a false discovery rate of 5% (Benjamini and Hochberg, 1995). The minimal possible threshold was set to the critical value in this omnibus test, which controlled the expected rate of false discoveries in each correlation matrix. Second, the value of the threshold was gradually increased in steps of 0.01 and the maximum value yielding a fully connected network was retained (i.e., there was a path between every two nodes in the network). This further reduced the number of edges in the graph, being consistent with the sparsity of anatomical connections in the brain (see Bassett et al., 2006; Meunier et al., 2009). The end result is an adjacency matrix \mathbf{W} representing a weighted functional network, with elements W_{ij} equal to zero for pairs of nodes with correlations below the threshold and W_{ij} equal to the wavelet correlation for pairs of nodes with correlations

above the threshold. In contrast to the common binary network construction, we chose to employ these weighted networks which retain neurophysiologically relevant information about the strength of functional interactions between network nodes (Bassett et al., 2012; Lohse et al., 2013; Rubinov and Sporns, 2011).

Computation of network theory measures

We computed six network measures for each functional connectivity matrix using the Brain Connectivity Toolbox for MATLAB (Rubinov and Sporns, 2010): strength, clustering coefficient, characteristic path length, betweenness centrality, intra-modular strength z-score, and participation coefficient. Table 1 summarizes information about these measures, including the equations used to compute them, their typical interpretation in the neuroscientific literature, how they were related to performance measures in the present study (on average), and the interpretation of such results. Network measures were selected from two surveys on network diagnostics (Costa et al., 2007; Rubinov and Sporns, 2010) because collectively they parsimoniously capture the local structure of the network surrounding individual network nodes. To facilitate interpretation of results and avoid overly complex regression models, we chose to decrease the number of diagnostics employed by (i) focusing on measures that are commonly used in the neuroscientific literature to facilitate ease of comparison between studies, and (ii) not including multiple measures that provided significantly correlated information.

The *degree* of a node is the number of nodes to which it connects. Specifically, let a_{ij} be a binary variable (i.e., 1 or 0) representing whether or not a connection exists between nodes i and j , and let N represent the total number of nodes. Then the degree of node i is computed as

$$k_i = \sum_{j=1}^N a_{ij} \quad (1)$$

Node *strength* is an extension of the definition of degree to weighted networks (Barabási et al., 2004) and is defined as the sum of the weights of a node's connections to other nodes. Let W_{ij} represent the absolute value of the correlation between nodes i and j . The strength of node i is

$$s_i = \sum_{j=1}^N W_{ij} \quad (2)$$

The weighted *clustering coefficient* of a node is the fraction of its neighbors that are neighbors of each other. The weighted clustering coefficient of node i can be defined as (Onnela et al., 2005)

$$C_i = \frac{\sum_{j=1}^N \sum_{h=1}^N \sqrt[3]{W_{ij} W_{ih} W_{jh}}}{k_i(k_i-1)} \quad (3)$$

The weighted *characteristic path length* (Newman, 2010) of a node is the average of its weighted shortest path lengths, which are the sum of link lengths separating the node from all other nodes in the network. Let $\gamma_{i \rightarrow j}$ represent the set of weighted links along the weighted shortest path from node i to j . Then the weighted characteristic path length was computed as

$$L_i = \frac{\sum_{j=1, j \neq i} \sum_{u, v \in \gamma_{i \rightarrow j}} \frac{1}{W_{uv}}}{N-1} \quad (4)$$

The weighted *betweenness centrality* of node i is the number of shortest paths in the network that pass through node i . Let ρ_{hj} represent the number of weighted shortest paths between nodes h and j , and let

$\rho_{hj}(i)$ represent the number of those paths that pass through node i . Then the weighted betweenness centrality of node i is (Freeman, 1978):

$$b_i = \sum_{\substack{h, j \\ h \neq j, h \neq i, j \neq i}} \frac{\rho_{hj}(i)}{\rho_{hj}} \quad (5)$$

The calculation of additional network diagnostics first required the determination of a partition of network nodes into modules. As in previous research aimed at detecting the community structure of brain networks (e.g., Alexander-Bloch et al., 2012; Bassett et al., 2011; Chen et al., 2008; Meunier et al., 2009), the partition of nodes into modules was found by maximizing a modularity quality function. For a given partition, g_i and g_j represent the modules to which nodes i and j are assigned, respectively. The function $\delta(g_i, g_j)$ is the Kronecker delta, and therefore is equal to one when $g_i = g_j$ and zero otherwise. Then a modularity quality function can be defined as (Newman, 2004)

$$Q = \sum_{i,j} [W_{ij} - \frac{s_i s_j}{S}] \delta(g_i, g_j) \quad (6)$$

where

$$S = \sum_{i,j} W_{ij}$$

We used a Louvain-like locally greedy algorithm (Blondel et al., 2008) to find the partition of nodes into communities that maximized Q . Theoretical work (Good et al., 2010) has shown that maximization of Q is complicated by the fact that many different partitions yield near-optimal values of Q , together forming a high-modularity plateau in the optimization landscape. To deal with this near-degeneracy, we performed 100 optimizations of the modularity quality function using a Louvain-like locally greedy algorithm and we extracted a consensus partition from the resulting 100 partitions using the consensus partition method proposed by Bassett et al. (2013a). Additional results of the analysis of modularity can be found in the supplementary material. Importantly, these results showed that neither modularity nor the number of modules changed significantly as a function of training block, suggesting that global properties of the network partition into modules were relatively stable.

We were interested in determining how local network properties (i.e., the topological features of specific nodes) related to the partition into modules might be able to explain changes in behavior. With this goal in mind, we computed two additional network measures: intra-modular strength and participation coefficient.

Given a partition of the network into modules, the *intra-modular strength* is the sum of connection weights of a node with nodes from its own module m :

$$s_i^m = \sum_{j=1}^N W_{ij} \delta(g_i, g_j) \quad (7)$$

This value is standardized to obtain an intra-modular strength z-score (based on the binary version of this diagnostic first introduced in Guimera and Amaral, 2005):

$$z_i = \frac{s_i^m - \bar{s}^m}{\sigma^m} \quad (8)$$

where \bar{s}^m and σ^m represent the mean and standard deviation (respectively) of the strength distribution for module m .

Given a partition of the network into modules, the *participation coefficient* (based on the binary version of this diagnostic first introduced in Guimera and Amaral, 2005) measures the uniformity of the distribution of connections of a node to nodes from all partitions. Values close to one reflect a uniform distribution of connections, and values close to zero reflect a high concentration of connections to only one or a few modules. The measure is defined as

$$P_i = 1 - \sum_{m=1}^M \left(\frac{s_i^m}{s_i} \right)^2 \quad (9)$$

Multilevel regression

We used multilevel regression (Gelman and Hill, 2007) to explore the relationship between local network dynamics and behavioral changes across blocks of training in the categorization task. This is an extension of simple linear regression that allows us to define participant-specific terms in the model and better capture the nested structure of the data. As in simple linear regression, the outputs are regression weights that can be interpreted as quantifying the influence of the predictors (i.e., network measures) on the outcome variable (i.e., behavioral measures). These analyses were performed using the package *lme4* (Bates et al., 2014) within the statistical software R v. 3.0.2 (R Development Core Team, 2014). Accuracy and mean correct response time at each training block were used as outcome variables in separate analyses, with node measures as predictor variables.

Because collinearity and multicollinearity of the predictors can have adverse effects on the estimates of regression coefficients (Dormann et al., 2013; Rawlings et al., 1998), we performed a collinearity analysis (see supplementary material) that revealed that characteristic path length and strength were almost perfectly correlated across all nodes (with a median of -0.97 and a range of -0.85 to -0.99). For this reason, we decided to exclude characteristic path length from the regression analysis. Note that the almost-perfect negative correlation between strength and characteristic path length means that both measures carried the same information about integrative processing in our networks. The remaining predictors were strength, clustering coefficient, betweenness centrality, intra-modular strength z-score, and participation coefficient. The collinearity analysis did not reveal any remaining issues in this set of predictors.

Initial scrutiny of the behavioral data revealed extremely poor performance (near-chance performance and response times larger than 1000 ms) by a single participant during the first three blocks of training. Because performance jumped to high levels after these three blocks, it is likely that the initial outliers were produced by the introduction of the task in the scanner (e.g., by misunderstanding of instructions). The three outlier data points were removed from all analyses to avoid the influence of a likely artifact on the results.

To determine possible violations of the assumptions underlying linear regression, particularly homoscedasticity and normality of residuals, we performed an analysis of the distribution of residuals after performing an initial regression analysis using the untransformed percent of correct choices and mean response times (see supplementary material). This analysis revealed violations of the assumptions of the linear regression model. These violations were corrected by applying an arcsine-square-root transformation to the accuracy data and a power transformation to the response time data (Rawlings et al., 1998; see supplementary material).

Because our main interest was to explore how local network dynamics are related to behavioral changes across training in the categorization task, block was the main unit of analysis in the model. Note that each model was fit to a total of 237 data points (each data point corresponded to a block of 80 trials; there were 10 participants and 24 of such observations per participant; three points were excluded from the analysis of a single participant) and the model was built mainly to

explain variation in performance across training blocks as a function of corresponding changes in graph statistics. Some individual differences are also captured by the model, but the main focus is on variability due to learning of the categorization task. We used a varying-intercepts model (Gelman and Hill, 2007), which includes a group intercept and individual intercepts for each participant, but only a group regression weight for each predictor. This implements the assumption that different participants showed different baseline levels of performance in the task, but that the relation between network measures and behavior (represented by a single regression coefficient) was the same across participants.

For each node and outcome variable, six models were initially fit to the data: a null model and five explanatory models. It is important to note that model selection does not follow the logic of frequentist null hypothesis significance testing. However, after model selection, we did confirm whether the selected model was better than the null through a likelihood ratio test. This test does follow frequentist logic and is therefore corrected for multiple comparisons.

The null model included only intercepts and no predictors, representing a baseline of performance that was constant across blocks but could vary across participants. This model was used to test whether variations in the predictors across blocks could explain additional variation in behavior. The simplest explanatory model included each of the predictors, but no interaction terms. Although this model had the advantage of being easy to interpret, theoretical considerations suggest that specific combinations of values for two or more predictors could describe the role of a node in the network better than all predictors considered separately. For example, it has been suggested that specific combinations of values for the participation coefficient and intra-modular strength z-score might determine a set of discrete roles for network nodes (Guimera and Amaral, 2005). Because we had no hypotheses about the specific interaction terms that might be important for the prediction of behavior, our approach was to fit a sequence of models of increasing complexity, with each model in the sequence incorporating all interactions one level above those included by the previous model. Thus, one model included all interactions between two predictors, the next model included all interactions between two or three predictors, and so on. The most complex model that we considered included five-predictor interactions.

Each model was fit to the data using maximum likelihood estimation. Model selection was performed using the Akaike information criterion (AIC Akaike, 1974). However, as the AIC is known to be biased for small samples (Burnham and Anderson, 2004), we used a version that corrects for this bias (Hurvich and Tsai, 1989):

$$AIC = -2 \log L + 2p \left(\frac{T}{T-p-1} \right) \quad (10)$$

where L is the likelihood of the data given the model, p is the number of free parameters in the model and T is total number of data points.

Given a set of candidate models indexed by $d = 1, 2, \dots, D$, it is possible to obtain an estimate of the probability that each of them is the best model in the set, computing AIC weights (Burnham and Anderson, 2004):

$$AICw_d = \frac{\exp(-\Delta_d/2)}{\sum_{r=1}^D \exp(-\Delta_r/2)} \quad (11)$$

where

$$\Delta_d = AIC_d - AIC_{min} \quad (12)$$

and AIC_{min} is the lowest AIC value for the set of D models. The best-fitting model for a particular node is the one that produces the smallest AIC and largest AICw.

AIC selects a model with enough complexity to explain as much variability in the behavioral data as possible without overfitting (Burnham and Anderson, 2004). We used this approach to determine (i) whether including node measures as predictors in the model provided a better description of behavior than the null model, and (ii) whether or not complex models (with second- or higher-order interactions among predictors) would provide a better description of behavior than a simple model assuming independent influences from each predictor.

After model selection, nodes for which the best-fitting model was the null model were removed from the following analyses. When the best-fitting model was not the null, we confirmed that the selected model provided a better fit than the null model by performing a likelihood ratio test. The p -values from such tests were corrected for multiple comparisons using a false discovery rate of 5% (Benjamini and Hochberg, 1995). Only nodes in which the best-fitting model provided a significantly better fit than the null model were included in the subsequent analyses. In this way, the outcome of model selection with AIC was supported by an additional criterion based on a traditional statistical test. This procedure resulted in a sample of nodes that were deemed predictive of behavior according to a rather conservative criterion.

Analysis of regression coefficients

Because the best-fitting model in the large majority of predictive nodes did not include interaction terms (see [Results](#) section), we were able to focus on such nodes to study the relationship between node measures and behavior as expressed through the regression coefficients from the model. Because better performance is associated with higher accuracy but lower response times, we reversed the sign of the regression coefficients from the response time analysis to make them more comparable to those from the accuracy analysis.

In the first part of this analysis, we determined whether the overall distribution of regression coefficients across nodes was biased toward positive or negative values for different node measures. We would expect such biases if a particular node measure was predictive of specific changes in behavior across a majority of nodes. For example, an increase in the strength of many connections (i.e., more integrative processing) could be predictive of increments in accuracy. Under the null hypothesis of centered distributions (i.e., with median equal to zero), the number of coefficients larger than zero follows a binomial distribution with parameter $p = 0.5$. We used this fact to statistically determine whether the bias present in each distribution was larger than expected by chance.

Next, we determined whether the distribution of coefficients for a given predictor varied depending on whether the outcome variable was accuracy or response time. As indicated earlier, changes in these two behavioral measures should correlate with different learning mechanisms: accuracy with early category learning and response time with later learning in which the categorization responses become automatic (Ashby et al., 2007). To compare the two distributions, we performed a permutation test on the absolute difference in median regression coefficients, by randomly re-ordering the outcome label ("accuracy" and "response time") of each regression coefficient 10,000 times and computing the absolute difference in medians after this random re-ordering. This resulted in a distribution of difference values under the null hypothesis that the distribution of regression coefficients was the same for accuracy and response times, which was used to determine whether the observed difference was significantly higher than expected by chance.

Finally, to analyze the distribution of mean AICw and regression coefficients across ROIs, we ranked such values and computed mean ranks for different ROI types (subcortical, visual, motor, and high-level areas). To determine whether these mean ranks were different from what would be expected by chance, we performed permutation tests by randomly re-ordering the ROI types of all nodes that were used in a particular analysis 10,000 times and computing the mean rank of each ROI type after this randomization. This resulted in a distribution of mean

ranks under the null hypothesis of no effect of ROI type, which was used to determine whether the observed rank for each ROI type was significantly higher than expected by chance. The same distributions were used to determine whether the observed difference between accuracy and response time in mean rank of each ROI type was significant.

Results

Behavioral results

[Fig. 3C–D](#) shows behavioral data as a function of training blocks for individual participants and the group mean, demonstrating that performance improved across blocks. Interestingly, mean accuracy reached asymptotic levels around block 7 (second scanner session) while mean response time continued to improve across training. [Fig. 2E](#) confirms that accuracy and correct response time were weakly correlated across blocks and participants (Pearson's $r = -0.2$, $p < 0.01$), consistent with the possibility that different neurophysiological mechanisms might underlie changes in each behavioral variable. Note also that accuracy and speed (the inverse of response time) have a small positive correlation, a pattern opposite of what we would expect if there was a speed-accuracy tradeoff in our data (for a review, see Heitz, 2014). It is likely that the magnitude of such an effect, if present in our data, was very small in relation to the improvements in performance (both speed and accuracy) with practice in the categorization task. Thus, we can safely assume that any observed dissociation between accuracy and response times is not due to a speed-accuracy tradeoff.

Model selection and distribution of predictive nodes

Out of the 742 nodes in each network, 148 passed our criteria to be considered predictive of changes in accuracy across training, whereas 129 passed our criteria to be considered predictive of changes in response times. The best-fitting models at these nodes had variable complexity, as indexed by the level of interaction terms included in the model. Histograms of the frequency distributions of selected levels of interaction are shown in [Fig. 4A](#). For the large majority of nodes, the selected model was one in which no interaction terms were included (level 1), followed by a small proportion of nodes in which the best model included only interactions between pairs of predictors (level 2). Best models with interaction levels larger than 2 were comparatively infrequent. These results suggest that in most cases complex combinations of node measures are not required for the prediction of variations in behavioral measures. Rather, individual node measures can be related to behavioral changes in a straightforward, easily interpretable manner. We therefore focused on those nodes for which the best-fitting model had no interaction terms and we determined the relationship between each node measure and behavior variables through an analysis of regression coefficients (see below).

[Fig. 4B](#) shows the distribution of AICw (Eq. 11) for predictive nodes across cortical regions. The bar plots depict mean AICw across ROIs, with ROIs ranked from highest to lowest AICw. Nodes in which the best model was the null model were assigned an AICw of zero. The resulting values summarize how predictive of behavior each ROI was relative to its overall size (i.e., number of nodes in the ROI). Bars in this figure, and in all subsequent bar plots, are colored following the scheme presented in [Fig. 2](#). Different ROIs are represented with different colors, with ROIs of the same type sharing similar hues. Yellowish bars represent visual areas, greenish bars represent motor areas, bluish bars represent areas related to high-level cognition (prefrontal areas and the hippocampus), and reddish bars represent subcortical areas.

[Fig. 4B](#) shows that subcortical ROIs were the most highly predictive of behavior in the analysis of accuracy. Areas related to high-level cognition were the least predictive of behavior, whereas motor and visual areas tended to be in the middle of the ranking. A permutation test

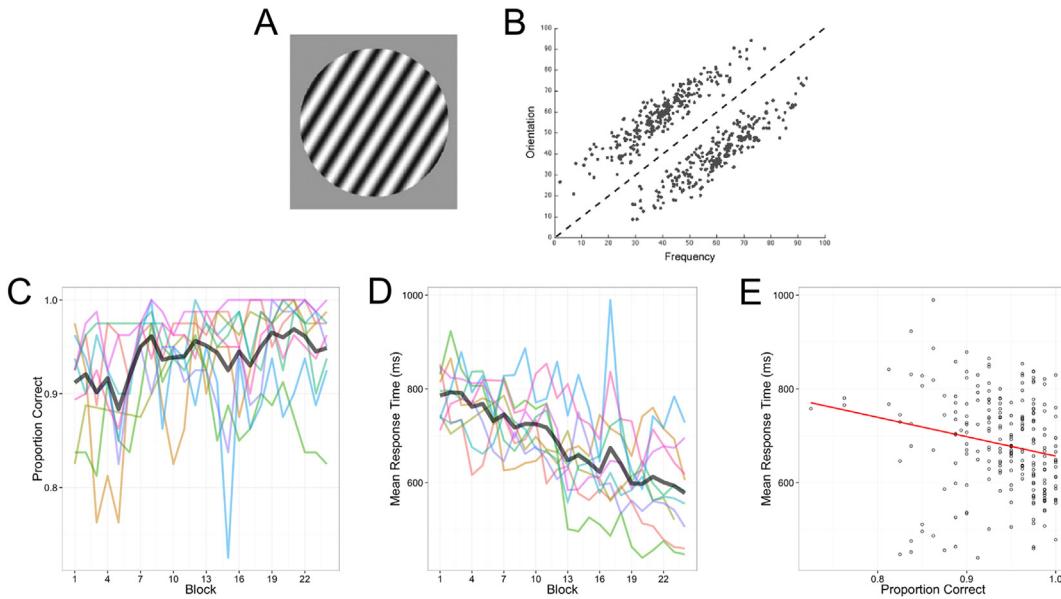


Fig. 3. Graphical summary of the behavioral task used in this study. (A) Example of a stimulus presented to participants in the categorization task. (B) Representation of the category structure learned by participants. (C) Proportion of correct responses as a function of scanning blocks; lines of different color represent different participants and the grey line represents the group mean. (D) Mean correct response time as a function of scanning block; lines of different color represent different participants and the grey line represents the group mean. (E) Scatterplot of the relationship between proportion of correct responses and correct mean response time across participants and scanning blocks.

revealed that the mean rank of subcortical areas was significantly higher than expected by chance, $\bar{x} = 9.83$, $p < 0.001$, the mean rank of areas related to high-level cognition was significantly lower than expected by chance, $\bar{x} = 26.5$, $p < 0.01$, and the mean rank of visual ($\bar{x} = 17$) and motor ($\bar{x} = 21.5$) areas were not significantly different from chance, $p > 0.1$.

Fig. 5 shows several coronal slices with the AICw obtained from selected individual nodes; the color of each node (red-yellow) represents the magnitude of its AICw, and the limits of ROIs have been drawn using lines colored according to the scheme presented in Fig. 2. The overall pattern of results previously described for the mean AICw can also be observed in individual nodes. For example, the highest-ranked ROI in Fig. 4B is the left head of the caudate. In the left side of Fig. 5, this structure is visible in two slices ($y = 9$ mm and $y = 24$ mm, at the top-right among subcortical structures), and in both cases the few nodes inside the ROI are classified as predictive and have relatively high AICw. A similar pattern repeats across subcortical ROIs (red-tone lines centrally located from $y = -21$ mm to $y = 24$ mm); given the relatively small size of most of these structures, the number of nodes within them that have high AICw is quite high. On the other hand, it can be seen in Fig. 4 that the frontal cortex and hippocampus (blue-tone lines) are relatively large areas, but not many of the nodes within them have a high AICw. This explains why these areas have a low average rank in Fig. 4A.

Fig. 4B reveals a similar ranking of areas in the analysis of response times, that is, subcortical areas were ranked relatively high, areas related to high-level cognition were ranked relatively low, and motor and visual areas fall between the two extremes. However, the separation of ranks by area type is less clean than that observed in accuracy. Permutation tests indicated that, for response times, the mean ranks of visual areas were significantly higher than expected by chance, $\bar{x} = 10.75$, $p < 0.05$, whereas the mean ranks of subcortical areas ($\bar{x} = 16.5$), motor areas ($\bar{x} = 19.2$), and areas related to high-level cognition ($\bar{x} = 23.3$), were not significantly different from what would be expected by chance (all $p > 0.05$). Furthermore, it was found that the mean rank of subcortical areas was significantly higher for accuracy than for response times, $p < 0.05$, but differences in other area types were not significant ($p > 0.05$).

Some of these results are also visible at the level of individual nodes, as shown in Fig. 5. For example, across all relevant slides in Fig. 5 (from $y = -21$ mm to $y = 24$ mm), the number of subcortical nodes that are good predictors of accuracy (left side) is consistently higher than the number of subcortical nodes that are good predictors of response times (right side).

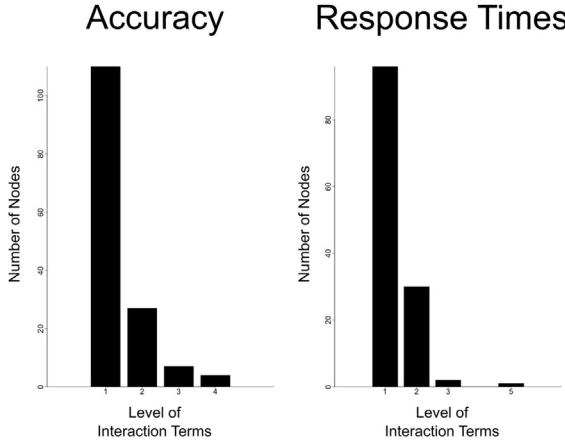
Relation between local network measures and behavior

Next, we were interested in characterizing the direction and magnitude of the relation between each node measure and behavioral performance, which is possible through exploration of the regression coefficients obtained from the best-fitting multilevel model at each node. The addition of interaction terms in the model makes the interpretation of each predictor difficult. Fortunately, as shown in Fig. 4A, for most nodes, the best-fitting model did not include any interaction terms. We therefore focused on these nodes and ignored the minority of nodes with best-fitting models that included interaction terms.

Clustering coefficient

Fig. 6 shows the distributions of standardized regression coefficients for the clustering coefficient, a measure thought in some contexts to relate to the efficiency of local information processing (Bullmore and Sporns, 2009). Panel A shows violin plots with the overall distributions of regression coefficients across nodes. The contour of each violin plot represents an estimate of the density of nodes with particular regression coefficient values, the black bar at the center of each plot represents the interval containing the central 50% of the values in the distribution (bounded by the second and third quartiles), and the white circle inside this bar represents the median of the distribution. Distributions from both the accuracy and response time analysis are clearly unimodal and biased toward positive values (binomial $p < 0.001$ in both cases). This means that, in general, the presence of higher clustering coefficients across the network was related to better performance of the task. The strength of this bias was higher in the analysis of accuracy than in the analysis of response times (permutation test, $p < 0.001$), suggesting that the clustering coefficient was a stronger predictor of accuracy than of response time.

A. Level of interaction terms in best-fitting model



B. Measure of fit of best-fitting model

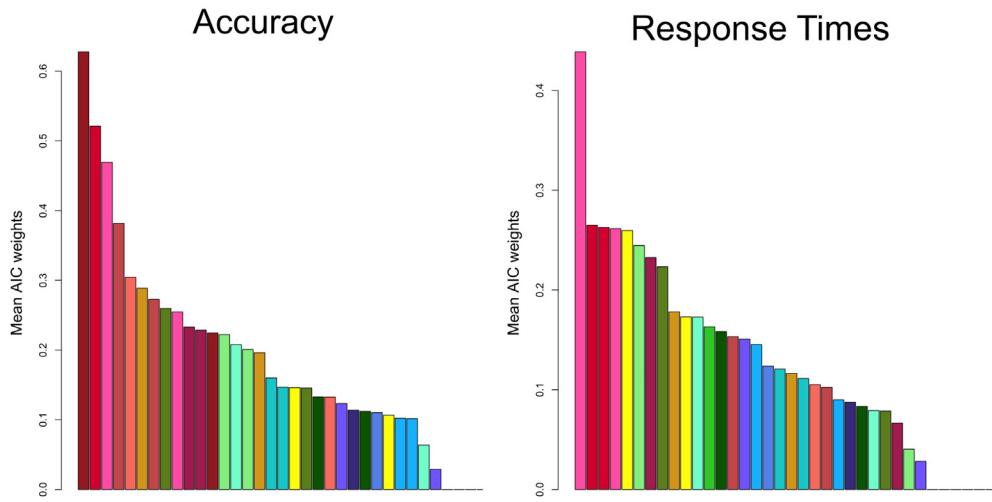


Fig. 4. Characterization of best-fitting models in predictive nodes in terms of their distribution of complexity and fit to the data. Predictive nodes are those for which variations in network measures could reliably predict variations in either accuracy or response times (according to our criteria, see “Materials and methods”). (A) Distribution of model complexity for predictive nodes. Model complexity is indexed here by the level of interaction terms in the best-fitting model for a given node. Models of level 1 are those without any interaction terms, models of level 2 are those including all interactions between pairs of independent variables, models of level 3 additionally include all interactions of three variables, and so on. (B) Distribution of AIC weights for the best-fitting model of predictive nodes across regions of interest (bar plots). AIC weight is a measure of the fit of the model to the data. Bars of different color represent different regions of interest, following the color scheme presented in Fig. 2.

Fig. 6B shows the mean value of regression coefficients for each ROI, which indicates where in the brain a clustering coefficient was a positive or negative predictor of task performance. Fig. 7 plots the values of regression coefficients obtained for selected individual nodes, different colors are used to plot positive (red-yellow) and negative (blue) weights. ROI limits have been drawn in Fig. 7 using the color scheme from Fig. 2. In the case of accuracy, Fig. 6B shows that almost all of the ROIs have a positive mean value (only left extrastriate cortex has a mean slightly below zero). All area types—each represented with a different bar color—tend to include ROIs with high and low ranks. Permutation tests indicated that mean rank was not significantly different from what would be expected by chance in any area type. The left side of Fig. 7 also clearly shows that regression coefficients of individual nodes were predominately positive across all ROIs.

In the case of response times, Fig. 6B shows that only two ROIs show negative mean values: right posterior ACC and right SMA. For all other ROIs, higher clustering coefficients were associated with faster response times on average. Subcortical areas (reddish bars) tended to have high ranks on average (permutation test: $\bar{x} = 8.11$, $p < 0.01$). All other

permutation tests were not significant. Furthermore, the mean rank of subcortical areas was significantly higher for response time than for accuracy, $p < 0.01$, whereas the mean rank of motor areas was significantly higher for accuracy than for response time, $p < 0.05$. No other differences in mean rank were significant. Many of these average results are also visible in the regression coefficients of individual nodes, some of which are shown in the right side of Fig. 7. Although a number of nodes have negative values, most nodes have positive values. In particular, subcortical nodes seem to be predominately positive (see slices at $y = -21$ mm, $y = -6$ mm, and $y = 9$ mm).

In sum, the relation between the clustering coefficient and task performance was quite straightforward: better performance in terms of both accuracy and speed was accompanied by higher clustering across all area types, with the relation between clustering and speed being particularly strong in subcortical areas.

Strength

Fig. 8 shows the distributions of standardized regression coefficients for strength, which measures the magnitude of functional connectivity

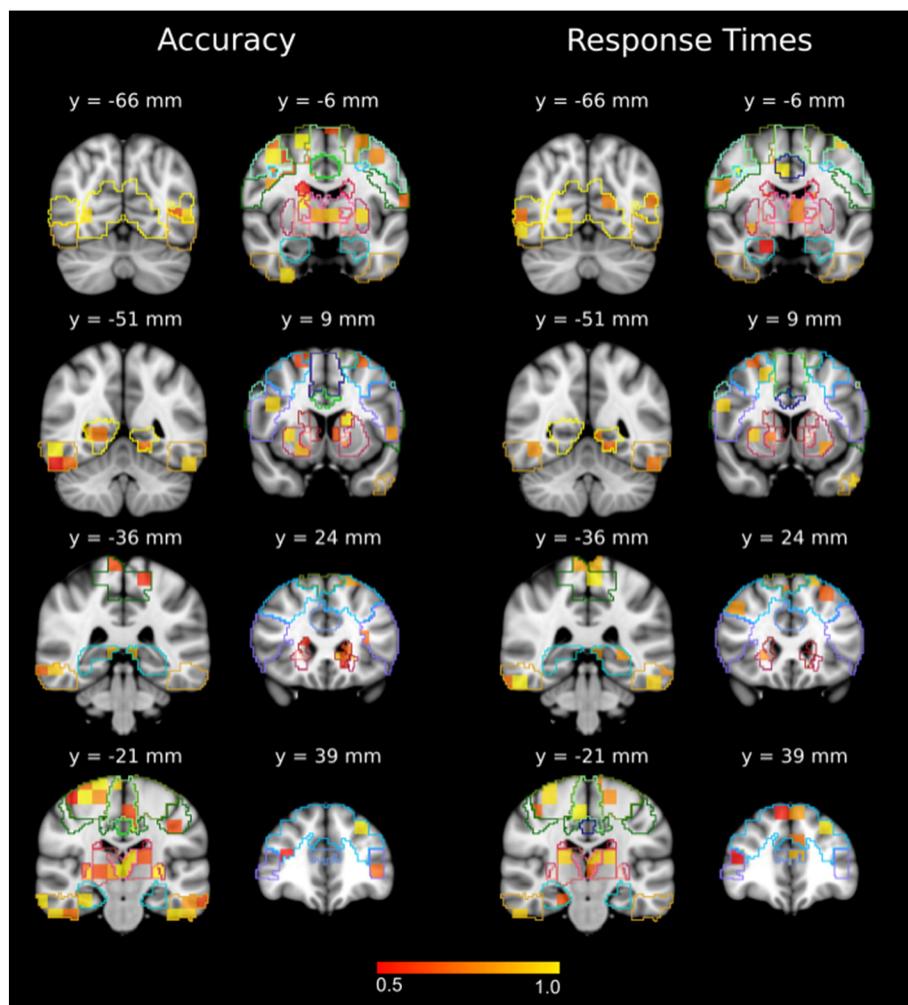


Fig. 5. AIC weights of individual selected nodes. The magnitude of a node's weight is represented by the red-blue color. Lines of different color represent ROI borders, following the color scheme presented in Fig. 2.

present between nodes in an ROI and the rest of the brain. Fig. 9 plots the corresponding values of strength regression coefficients obtained for selected individual nodes. The violin plots in Fig. 8A indicate that the distribution of strength regression coefficients was biased toward negative values when the behavioral outcome measure was accuracy (binomial $p < 0.001$) and toward positive values when the behavioral

outcome measure was response time (binomial $p < 0.01$). The difference between the two distributions was statistically significant as measured by a permutation test, $p < 0.01$.

As shown in Fig. 8B, an average positive relationship between accuracy and strength was found only in four ROIs (left extrastriate cortex, right head of the caudate, right SMA, and right dorsolateral PFC) and

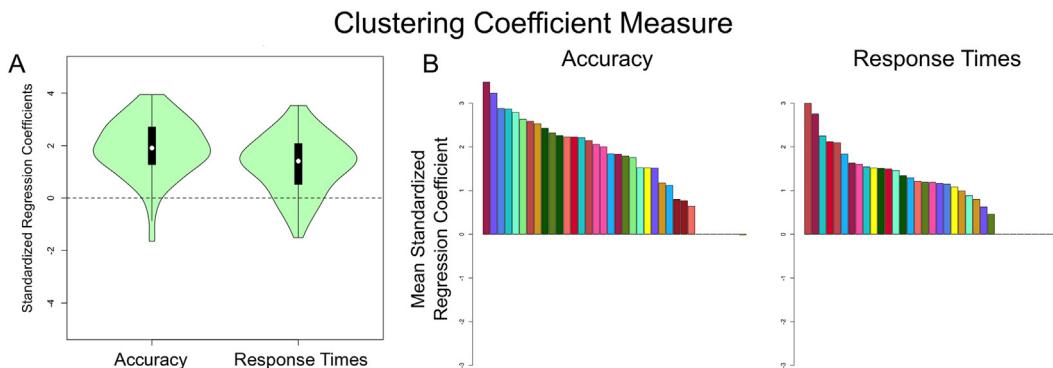


Fig. 6. Distributions of standardized regression coefficients for the clustering coefficient measure. (A) Violin plots with the overall distribution of regression coefficients across all predictive nodes. Predictive nodes are those for which variations in network measures could reliably predict variations in either accuracy or response times (according to our criteria, see "Materials and methods"). The contour of each violin plot represents an estimate of the density of nodes with particular regression coefficient values, the black bar at the center of each plot represents the interval containing the central 50% of the values in the distribution, and the white circle inside this bar represents the median. (B) Mean standardized regression coefficient for each region of interest, indicating whether clustering coefficient computed from nodes in that region was, on average, a positive (bars above dotted line) or negative (bars below dotted line) predictor of task performance. Bars of different color represent different regions of interest, following the color scheme presented in Fig. 2.

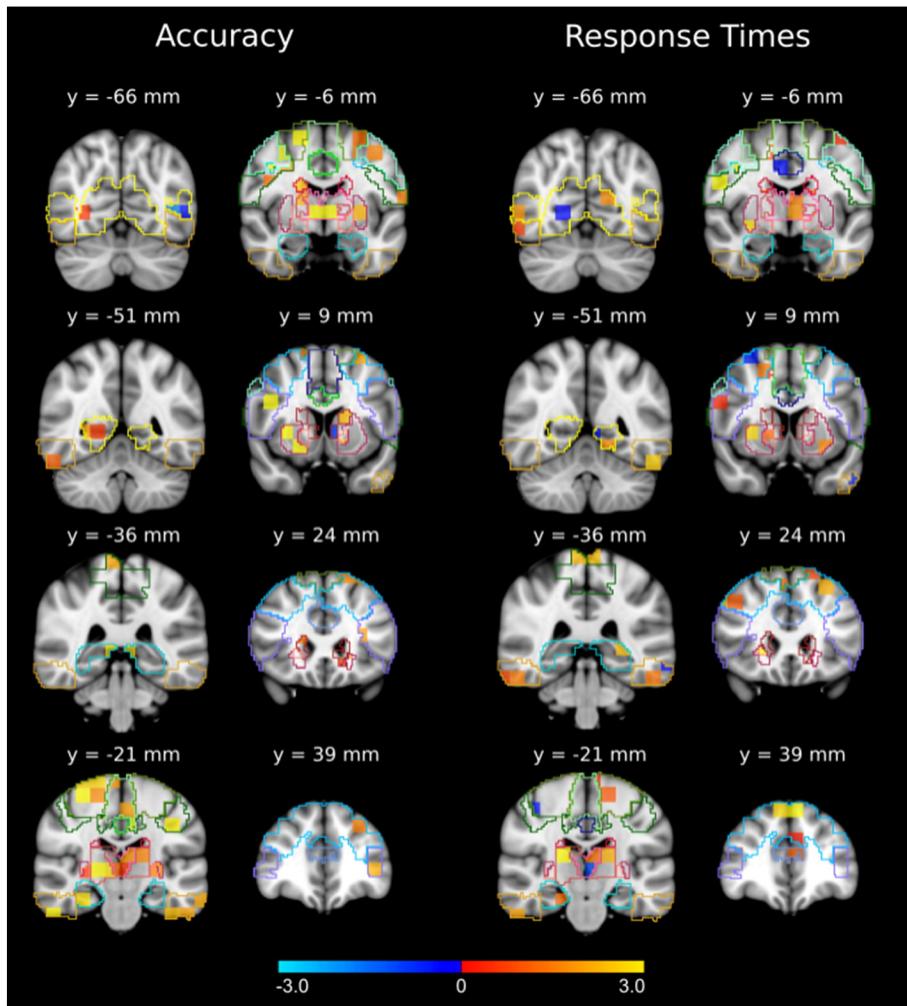


Fig. 7. Standardized regression coefficients for clustering coefficient, shown for individual selected nodes. Different colors are used to plot positive (red-yellow) and negative (blue) coefficients. Lines of different color represent ROI borders, following the color scheme presented in Fig. 2.

in all cases the mean regression coefficients had a relatively small absolute value. Visual areas (yellowish bars) showed significantly lower values than expected by chance (permutation test: $\bar{x} = 6.0$, $p < 0.01$). The mean ranks for all other area types were well within the distribution of values expected by chance (permutation test, $p > 0.05$). These

results are also easily observed in the left side of Fig. 9, where only a few nodes show a positive weight (red-yellow) and they tend to be localized posteriorly, in extrastriate and inferotemporal cortices (see $y = -66$ mm and $y = -51$ mm). Nodes in other areas have regression weights that are predominantly negative (blue).

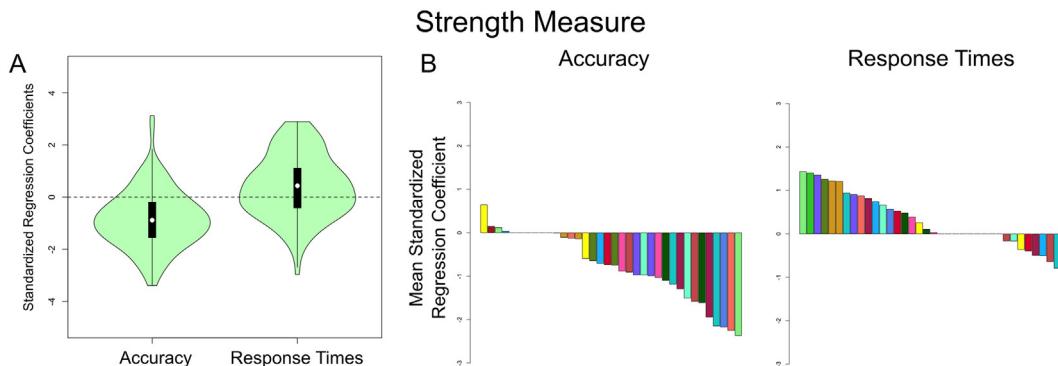


Fig. 8. Distributions of standardized regression coefficients for the strength measure. (A) Violin plots with the overall distribution of regression coefficients across all predictive nodes. Predictive nodes are those for which variations in network measures could reliably predict variations in either accuracy or response times (according to our criteria, see "Materials and methods"). The contour of each violin plot represents an estimate of the density of nodes with particular regression coefficient values, the black bar at the center of each plot represents the interval containing the central 50% of the values in the distribution, and the white circle inside this bar represents the median. (B) Mean standardized regression coefficient for each region of interest, indicating whether strength computed from nodes in that region was, on average, a positive (bars above dotted line) or negative (bars below dotted line) predictor of task performance. Bars of different color represent different regions of interest, following the color scheme presented in Fig. 2.

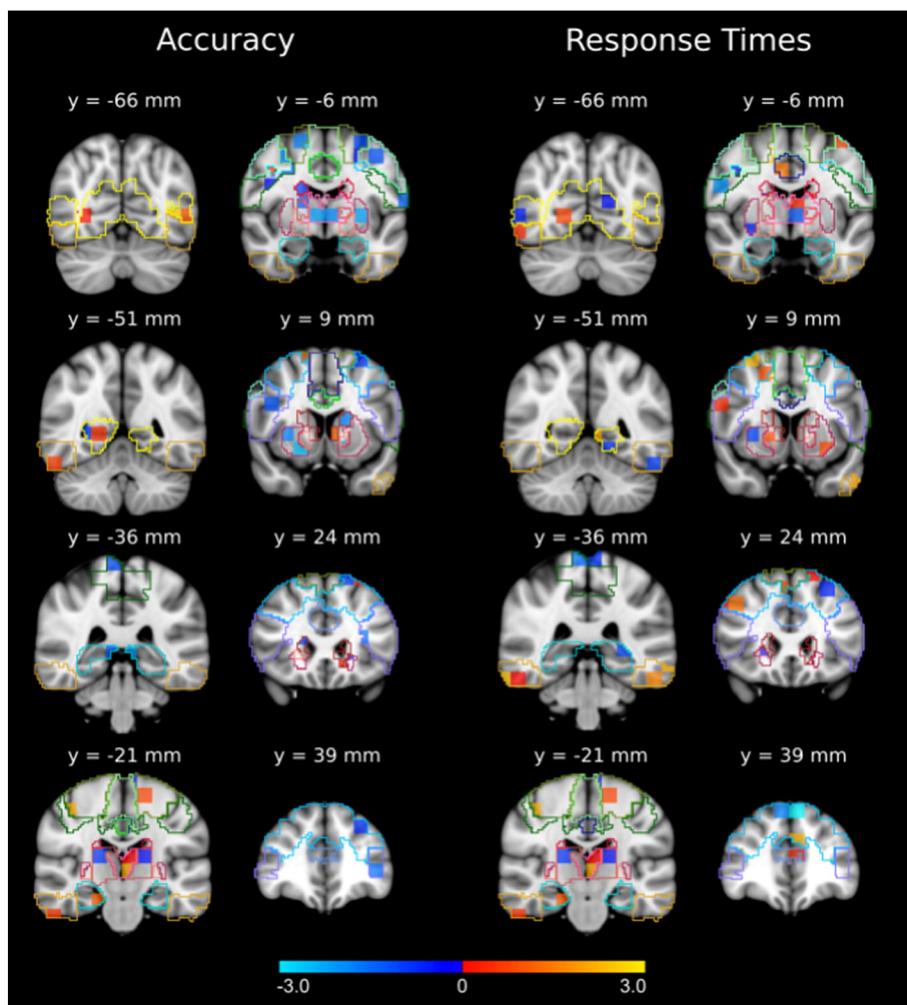


Fig. 9. Standardized regression coefficients for strength, shown for individual selected nodes. Different colors are used to plot positive (red-yellow) and negative (blue) coefficients. Lines of different color represent ROI borders, following the color scheme presented in Fig. 2.

In the case of response times, on average, an increase in speed was accompanied by an increase in strength in most ROIs, but there are several exceptions. The mean rank of each area type was not significantly different from what is expected by chance (permutation test, $p > 0.05$).

Results at the level of individual nodes for the analysis of response times are plotted in the right side of Fig. 9. At this level, it is difficult to see the slight bias toward positive regression weights observed in Fig. 8A, except perhaps in the slides corresponding to $y =$

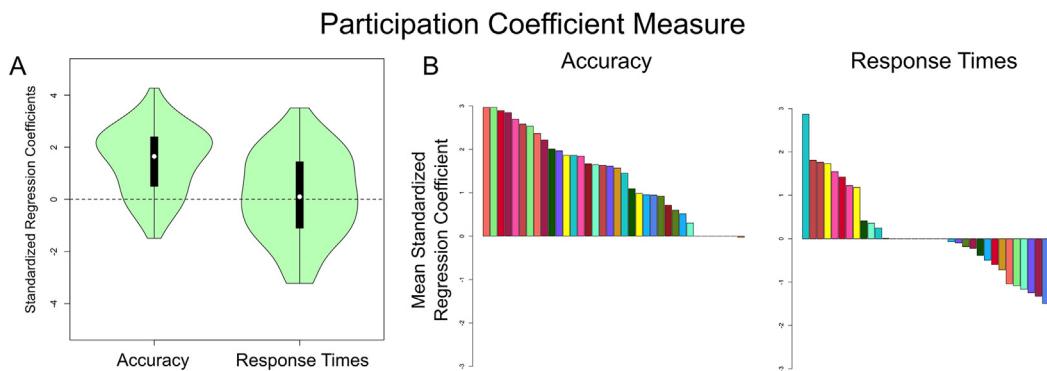


Fig. 10. Distributions of standardized regression coefficients for the participation coefficient measure. (A) Violin plots with the overall distribution of regression coefficients across all predictive nodes. Predictive nodes are those for which variations in network measures could reliably predict variations in either accuracy or response times (according to our criteria, see “Materials and methods”). The contour of each violin plot represents an estimate of the density of nodes with particular regression coefficient values, the black bar at the center of each plot represents the interval containing the central 50% of the values in the distribution, and the white circle inside this bar represents the median. (B) Mean standardized regression coefficient for each region of interest, indicating whether participation coefficient computed from nodes in that region was, on average, a positive (bars above dotted line) or negative (bars below dotted line) predictor of task performance. Bars of different color represent different regions of interest, following the color scheme presented in Fig. 2.

24 mm and $y = -21$ mm. The results are also mixed at the level of ROIs, with most ROIs including nodes with positive and negative weights. However, in agreement with Fig. 8B, the only ROIs showing a relatively consistent positive relation between strength and response time are visual (inferotemporal cortex, see bottom ROI from $y = -66$ mm to $y = -21$ mm) and motor (for example, dorsal premotor cortex at $y = -21$ mm).

The mean rank of visual areas was significantly lower in the accuracy analysis than in the response time analysis (permutation test, $p < 0.05$).

In sum, an increase in accuracy was predicted by a decrease in the connectivity of nodes across the whole network, except for visual areas. On the other hand, an increase in speed was accompanied by a slight increase in the connectivity of many nodes, but this bias was not consistently observed in any specific region. On average, although many ROIs showed a positive relation between speed and strength, there were eight ROIs that showed the reverse relation (see Fig. 8B), and they were distributed across all area types. These findings suggest that accuracy and speed are differentially driven by stronger versus weaker connectivity between nodes: accuracy requires a cost-efficient decrease in strength whereas speed is more inconsistently related to strength, showing a tendency to require a non-localized increase in strength.

Participation coefficient

A weaker form of dissociation between accuracy and response times was found for all other predictors, with the distribution of regression coefficients being biased for one behavioral variable, but not the other. As

seen in Fig. 10A, in the case of the participation coefficient, the distribution is clearly biased toward positive values when the outcome is accuracy, $p < 0.001$, but it shows no bias when the outcome is response time, $p > 0.50$. The difference between the two distributions was not significant, $p > 0.05$.

Fig. 10B confirms that, on average, an increase in accuracy was accompanied by an increase in participation coefficient across practically all ROIs, with this relation being particularly strong in subcortical areas and SMA. Subcortical areas (reddish bars) had a mean rank significantly higher than expected by chance (permutation test: $\bar{x} = 9.82$, $p < 0.01$). Permutation tests for other area types were not significant. These results can also be observed at the level of individual nodes, in the left side of Fig. 11. It can be seen there that most nodes showed a positive relation between participation coefficient and accuracy. This relation was particularly strong and consistent in subcortical nodes (most bright yellow nodes are located in subcortical areas, at $y = -21$ mm, -6 mm and 9 mm).

In contrast, Fig. 10B shows that the direction of the relationship between speed and participation coefficient depended on each specific ROI. Area type did not have an influence in the direction of this relationship, as revealed by non-significant permutation tests. The right side of Fig. 11 confirms that there is no consistent relation between participation coefficient and speed in any area type, although some specific ROIs seemed to be more consistent. For example, extrastriate cortex (at $y = -66$ mm and $y = -51$ mm) and some subcortical areas ($y = -21$ mm and $y = -6$ mm) show mostly positive regression

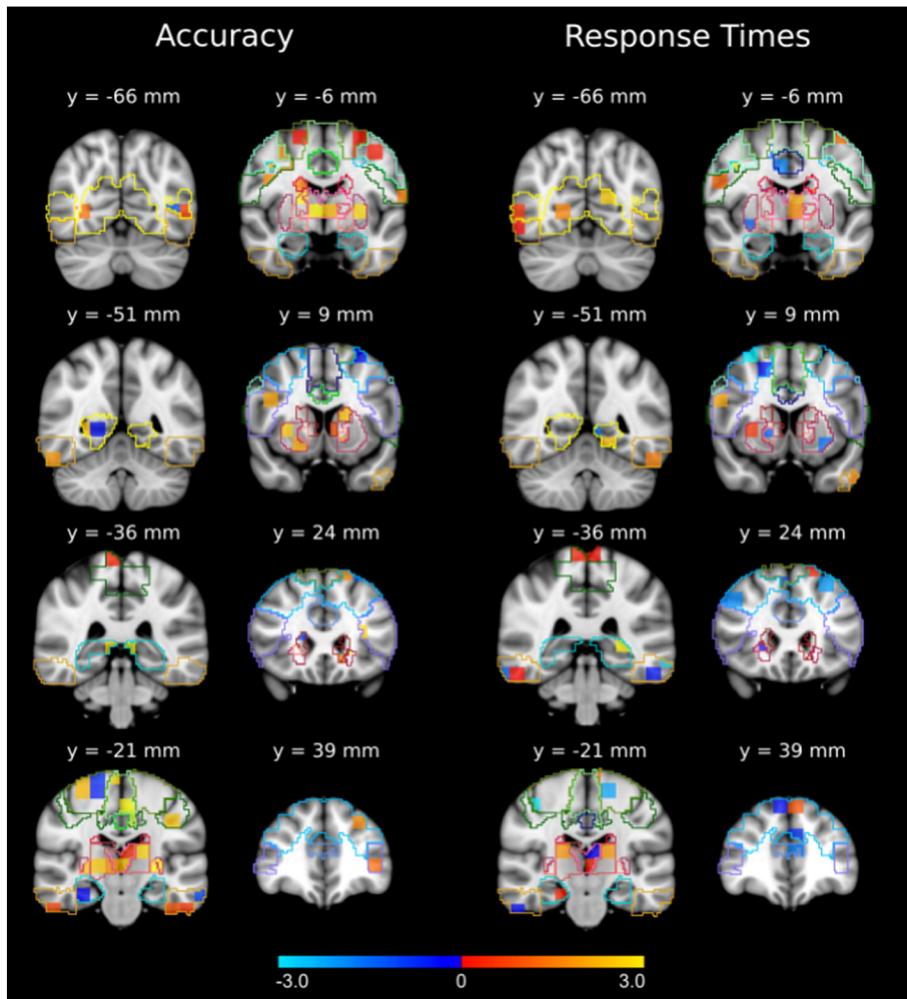


Fig. 11. Standardized regression coefficients for participation coefficient, shown for individual selected nodes. Different colors are used to plot positive (red-yellow) and negative (blue) coefficients. Lines of different color represent ROI borders, following the color scheme presented in Fig. 2.

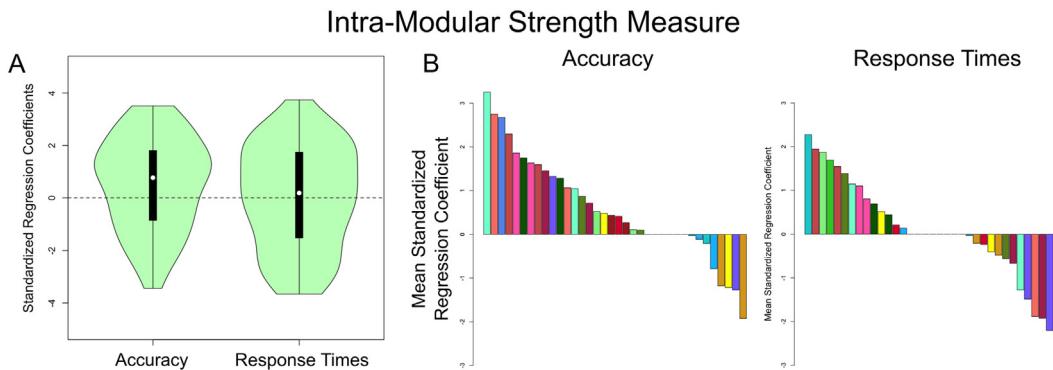


Fig. 12. Distributions of standardized regression coefficients for the intra-modular strength measure. (A) Violin plots with the overall distribution of regression coefficients across predictive nodes. Predictive nodes are those for which variations in network measures could reliably predict variations in either accuracy or response times (according to our criteria, see "Materials and methods"). The contour of each violin plot represents an estimate of the density of nodes with particular regression coefficient values, the black bar at the center of each plot represents the interval containing the central 50% of the values in the distribution, and the white circle inside this bar represents the median. (B) Mean standardized regression coefficient for each region of interest, indicating whether intra-modular strength computed from nodes in that region was, on average, a positive (bars above dotted line) or negative (bars below dotted line) predictor of task performance. Bars of different color represent different regions of interest, following the color scheme presented in Fig. 2.

weights, whereas some motor areas (at $y = -21$ mm) and prefrontal areas (from $y = 9$ mm to $y = 39$ mm) show mostly negative regression weights. The mean rank of visual areas was significantly lower for response time than for accuracy (permutation test, $p < 0.05$), but other differences were not significant.

Intra-modular strength z-score

The distribution of regression coefficients for intra-modular strength z-score in Fig. 12A shows a pattern of results similar to that found for the participation coefficient: the distribution is slightly biased toward positive values when the outcome is accuracy ($p < 0.05$) and shows no

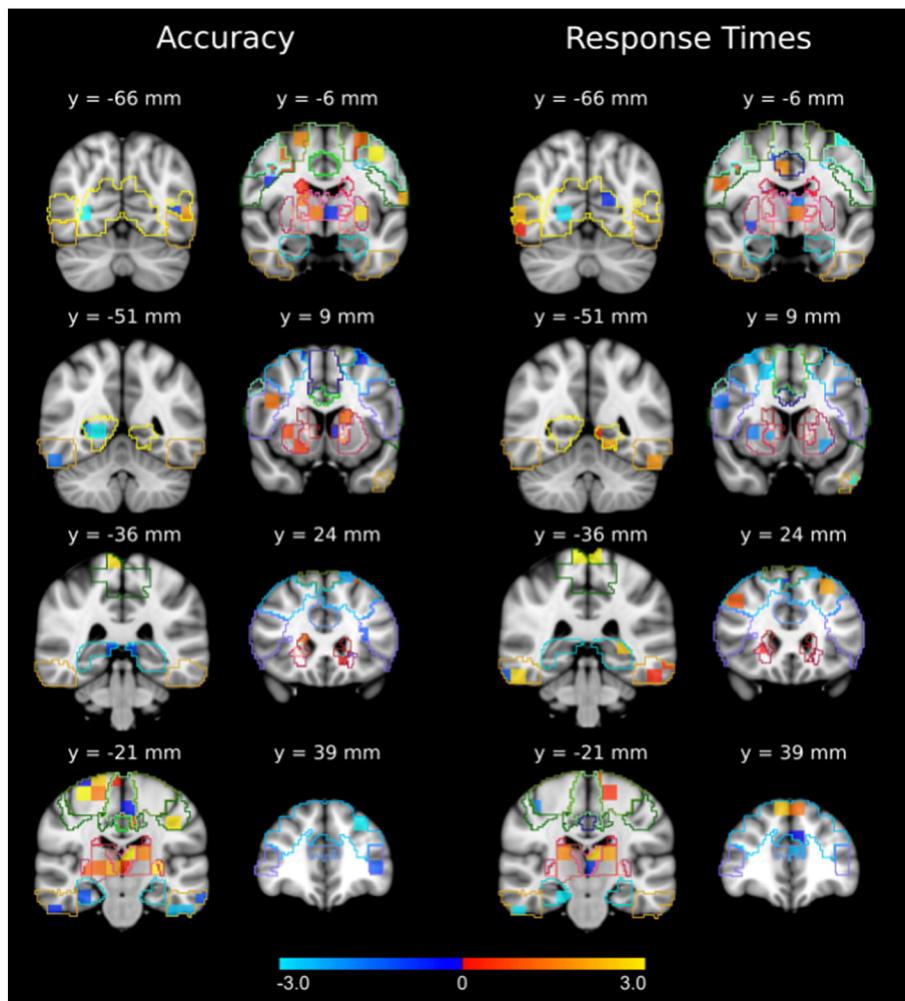


Fig. 13. Standardized regression coefficients for intra-modular strength, shown for individual selected nodes. Different colors are used to plot positive (red-yellow) and negative (blue) coefficients. Lines of different color represent ROI borders, following the color scheme presented in Fig. 2.

bias when the outcome is response time ($p > 0.10$) and the difference between both distributions was not significant ($p > 0.05$). Note that this qualifies the previously described negative relationship between strength and accuracy: higher accuracy was predicted by lower strength of connections across the network, but by higher relative strength of intra-modular connections, indicating a reorganization of network topology.

Fig. 12B shows that higher accuracy was accompanied by higher intra-modular strength z-score in all subcortical (reddish bars) and motor (greenish bars) ROIs, and by lower intra-modular strength z-score in some visual (yellowish bars) and high-level cognition (bluish bars) ROIs. The mean rank of subcortical areas was significantly higher than expected by chance (permutation test $\bar{x} = 10.82$, $p < 0.05$), whereas the mean rank of visual areas was significantly lower than expected by chance ($\bar{x} = 25.5$, $p < 0.01$). Other tests were not significant. These results are also visible from the regression weights of individual nodes, as shown in the left panel of Fig. 13. Note that although both positive (red-yellow) and negative (blue) weights are observed, the former tend to concentrate in subcortical and motor areas (at $y = -21$ mm, $y = -6$ mm and $y = 9$ mm), whereas the latter concentrate in visual ($y = -66$ mm and $y = -51$ mm, also the bottom of $y = -21$ mm), frontal areas (see top-lateral ROIs from $y = 9$ mm to $y = 39$ mm), and hippocampus (bottom of $y = -36$ mm and $y = -21$ mm).

From the response time analysis, the distribution of mean values across ROIs appears less straightforward (see Figs. 12 and 13). Permutation tests on mean ranks were not significant for any area type (all $p > 0.05$). The mean rank of visual areas was significantly lower in the analysis of accuracy than in the analysis of response time, $p < 0.01$, but other differences were not significant.

In sum, since the distribution of regression coefficients for both participation coefficient (Fig. 10A) and intra-modular strength z-score (Fig. 12A) showed a positive bias in the analysis of accuracy, higher accuracy was accompanied by nodes having a connection profile increasingly similar to that of connector hubs (Guimera and Amaral, 2005), which are nodes that allow communication between functionally segregated network communities. This was true for most ROIs, with the exception of visual and some high-level cognition areas (Fig. 12B), in which higher accuracy was accompanied by a connection profile increasingly similar to that of connector non-hubs (that is, higher participation coefficient, but lower intra-modular strength). The relationship between the role of a node in the network's community structure and speed in the task seemed idiosyncratic to each specific node and was not influenced by area type.

Betweenness centrality

Betweenness centrality showed a pattern opposite to that of the previous two predictors, with the distribution of regression coefficients

(Fig. 14A) being biased toward negative values in the analysis of response times, $p < 0.001$, but not biased in the analysis of accuracy, $p > 0.50$. The difference between distributions was significant, $p < 0.05$.

The distribution of mean coefficients across ROIs (Fig. 14B) reveals that only nine ROIs showed a slightly positive average relation between speed in the task and betweenness centrality. These results can also be observed at the level of individual nodes in Fig. 15. The right panel in this figure shows that individual regression coefficients for betweenness centrality were mostly negative (in blue), although some select regions included positive weights (red-yellow). The latter includes right hippocampus and inferotemporal cortex (see bottom-left ROIs in $y = -21$ mm), and left dorsal premotor cortex (top-right at $y = -21$ mm) and medial dorsal thalamus (right-middle at $y = -6$ mm). The mean ranks of all area types were not significantly different from what would be expected by chance.

The average relationship between accuracy and betweenness centrality depended on each specific ROI and did not seem to be consistent for any area type. Fig. 14B and the left panel of Fig. 15 show some trends; for example, many subcortical areas seem to show positive regression weights (see central areas from $y = -21$ mm to $y = 24$ mm; in particular, both right and left head of the caudate had relatively high positive weights) and many motor areas seem to show negative regression weights (see top-lateral cortical areas, in green, from $y = -36$ mm to $y = -6$ mm). However, none of these results were statistically reliable: all permutation tests on mean rank of different area types were not significant ($p > 0.10$). Tests on the difference in mean rank for each area type between accuracy and response times were also not significant ($p > 0.10$).

Head motion analysis

Several recent studies have shown that in-scanner head motion artificially modifies measures of functional connectivity in resting-state fMRI studies and can affect measures derived from network theory (Power et al., 2014; Power et al., 2012; Satterthwaite et al., 2012; Van Dijk et al., 2012). To account for a possible effect of head motion on our results, we repeated our analyses after removing the influence of head motion from the behavioral variables.

We computed the mean relative displacement (MRD; Satterthwaite et al., 2013) in each functional scan from estimates of head motion provided by MCFLIRT (Jenkinson et al., 2002). We used multilevel regression models with the MRDs as predictors to regress-out the influence of motion from the values of behavioral variables (accuracy and response time). Regressing-out summary statistics of motion in the group-level analysis has been proposed as a way to control for the effects of motion that persist after preprocessing at the individual level (Power et al., 2014; Satterthwaite et al., 2012). We then repeated all

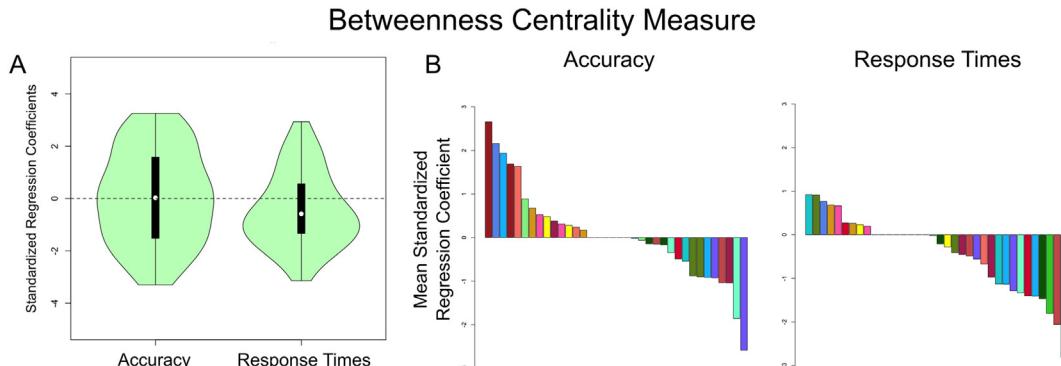


Fig. 14. Distributions of standardized regression coefficients for the betweenness centrality measure. (A) Violin plots with the overall distribution of regression coefficients across all predictive nodes. Predictive nodes are those for which variations in network measures could reliably predict variations in either accuracy or response times (according to our criteria, see "Materials and methods"). The contour of each violin plot represents an estimate of the density of nodes with particular regression coefficient values, the black bar at the center of each plot represents the interval containing the central 50% of the values in the distribution, and the white circle inside this bar represents the median. (B) Mean standardized regression coefficient for each region of interest, indicating whether betweenness centrality computed from nodes in that region was, on average, a positive (bars above dotted line) or negative (bars below dotted line) predictor of task performance. Bars of different color represent different regions of interest, following the color scheme presented in Fig. 2.

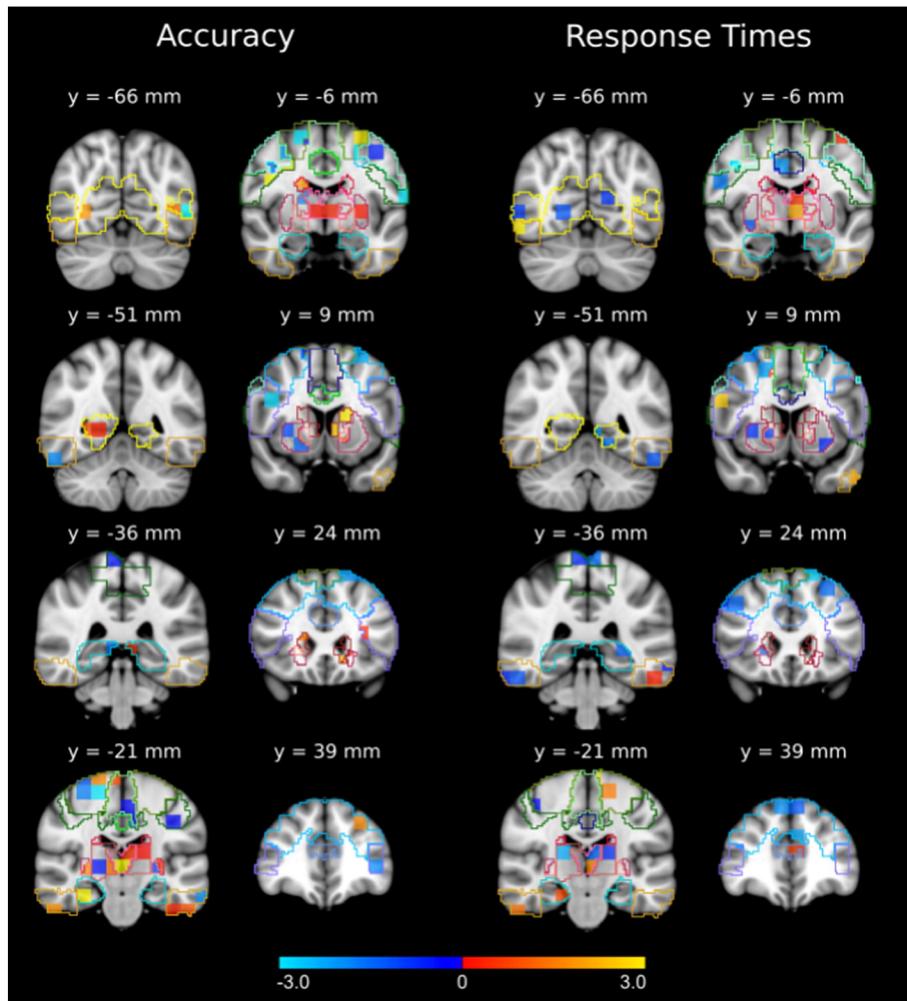


Fig. 15. Standardized regression coefficients for betweenness centrality, shown for individual selected nodes. Different colors are used to plot positive (red-yellow) and negative (blue) coefficients. Lines of different color represent ROI borders, following the color scheme presented in Fig. 2.

our analyses using these motion-corrected outcome variables. We found no indication that a motion artifact could drive the results reported here, which were reproduced in the analysis of motion-corrected data (for more details, see supplementary material).

Discussion

Our results demonstrate that it is possible to successfully explain variation in performance across training in a categorization task from the topological network properties of small regions in categorization-related brain areas. As expected, we found that the topological network properties of nodes in subcortical areas were particularly informative about changes in accuracy, but much less informative about changes in speed of correct responses. This observation is in line with neurocomputational models of automaticity in categorization (see Ashby et al., 2007), which predict that changes in accuracy are due to initial learning supported by basal ganglia, whereas changes in speed of correct responses are largely due to the development of direct connections between sensory and motor areas as skills become automatic.

The best-fitting models for most nodes did not include interaction terms, allowing a straightforward interpretation of the relationship between network predictors and behavior. Increments in accuracy were predicted by a higher clustering coefficient, participation coefficient and intra-modular strength z-score, and by lower strength across the network. The connection strength of nodes within visual areas, both

across the whole network and within modules, did not show the same relationship with accuracy as in other areas. On the other hand, speed of correct responding was predicted by higher clustering, higher strength and lower betweenness centrality.

The pattern of network changes related to increased speed of task performance is strikingly similar to that reported in a prior study of motor learning (Heitger et al., 2012), in which participants practiced a complex bimanual coordination task for four daily sessions and were scanned while performing the task before and after training. A comparison of functional networks before and after training revealed increments in clustering, degree, and strength, together with decrements in path length and betweenness centrality. The similarity of the present results and those found by Heitger et al. (2012) suggests that these changes in network properties may form a generalizable mechanism of behavioral change across all highly practiced procedural tasks.

With the exception of the clustering coefficient, we observed a dissociation between accuracy and response times regarding which network properties could consistently explain behavioral changes. For example, higher strength predicted lower accuracy but higher speed of correct responses. This dissociation is in line with the idea that different learning mechanisms mediate changes of accuracy and speed throughout categorization training. However, the specific relationships found between graph statistics and behavior are not predicted by neurocomputational theories of categorization and are easier to interpret within the more general framework of network science.

Effortful processing, automaticity, and network topology

Recent work has linked effortful information processing during working memory (Kitzbichler et al., 2011) and preparatory attention (Ekman et al., 2012) to higher integration and lower segregation in functional networks. Neurocomputational models (e.g., Ashby et al., 1998) also suggest higher integration during initial category learning in our task, but orchestrated by the basal ganglia and related subcortical structures. The results both support and qualify this hypothesis: learning is predicted by a form of integration that accounts for the network's community structure. Higher accuracy was predicted by higher integration within and across network modules – as measured by the intra-modular strength and participation coefficient, respectively. These changes at the meso-scale of network communities were accompanied by lower integration (strength) and higher segregation (clustering) at the global scale of the whole network.

Nodes with a profile increasingly similar to connector hubs predicted higher accuracy. Connector hubs allow communication between functionally segregated network communities and their presence in prefrontal cortex has been previously linked to performance of a simple visually guided finger-tapping task (Bassett et al., 2006). It is intuitively plausible that the increased presence of connector hubs enables flexible reconfiguration of brain networks, thereby facilitating learning (Bassett et al., 2013b, 2011, 2015). Here, an increase in the “connector hubness” of nodes in subcortical areas was a particularly good predictor of task performance, which is consistent with the nature of our task (Ashby and Ennis, 2006; see Ashby and Spiering, 2004).

Taking speed of correct responding as an index of automaticity, we found that automaticity is related to increased segregation, as indexed by the clustering coefficient, a relationship that was particularly strong in subcortical areas. This supports our hypothesis that the basal ganglia and related subcortical areas are involved in integrative processing early in category learning but are released from such function with overtraining. Previous findings of decrements in dynamic centrality with learning of a motor task have also been interpreted as arising from a lower requirement for integration as a skill becomes automated (Mantzaris et al., 2013). Furthermore, the rate of decreased integration between motor and visual modules in a finger sequencing task (in which participants generated responses to sequences of visual stimuli corresponding to fingers of their right hand) practiced over 6 weeks predicted individual differences in learning, further highlighting the utility of network segregation for automatic processing (Bassett et al., 2015).

Although automaticity learning was also predicted by an increase in strength in many ROIs, this effect appeared stronger in visual and motor areas than in subcortical areas (Fig. 5B). Strength is a measure of the connectivity of pairs of regions, which is expected to increase in cortical areas as automaticity develops. On the other hand, integration involving more than two regions, as measured by betweenness centrality, did decrease with automaticity.

Focal points of functional network reconfiguration are task-specific

Several recent papers have shown that multimodal association areas—including frontal, temporal, and parietal cortex—are focal points of functional network reconfiguration during learning and the performance of cognitive tasks (Bassett et al., 2015, 2013b, 2011, 2009, 2006; Mantzaris et al., 2013; Cole et al., 2013; Ekman et al., 2012; Fornito et al., 2012; Braun et al., 2015). Here, the network topology of frontal and medial temporal nodes played only a minor role in explaining behavioral changes, whereas subcortical areas played a major role. One likely explanation for this departure from previous studies is that they used tasks that engage explicit processes, such as working memory (Bassett et al., 2009; Braun et al., 2015), preparatory attention (Ekman et al., 2012), memory recollection (Fornito et al., 2012), and rule use (Cole et al., 2013). On the other hand, we used a categorization task known to engage implicit procedural learning (Ashby

et al., 2003; Maddox et al., 2004) and mediated by the basal ganglia (Ashby and Ennis, 2006; Ashby and Spiering, 2004).

It is unlikely that a single set of brain regions drives functional network reconfigurations across all tasks. However, some general principles might still be at work. For example, areas with structural connections to a variety of other areas across the brain are more likely to play a role in integrative function (Sporns, 2014; Sporns, 2013; van den Heuvel and Sporns, 2013; Gu et al., 2015). Both multimodal association areas in cortex and the basal ganglia fit this description. The basal ganglia in particular receive projections from almost all of cortex and project back to it through the thalamus (Haber and Johnson Gdowski, 2004), which might explain their strong functional connections to key cortical hubs (Zuo et al., 2012) and their role in learning and cognitive function (e.g., Ashby and Ennis, 2006; Ding and Gold, 2013; Doyon et al., 2009; Packard and Knowlton, 2002). Indeed, recent theoretical work underscores the predicted relationship between a brain region's pattern of structural connectivity and the role that brain region plays in controlling or constraining brain dynamics leading to adaptive function such as learning (Gu et al., 2015).

Relation to previous analyses of this dataset

Two other analyses of the data used here have been previously reported, one involving traditional univariate statistics (Waldschmidt and Ashby, 2011) and one involving multivoxel pattern analyses (Soto et al., 2013). Several results from those studies are in line with the hypothesis that early learning of the information-integration categorization task depends on the basal ganglia and associated subcortical areas, whereas extended practice is associated with transfer of control from the basal ganglia to cortex, as predicted by the neurocomputational theory of Ashby and colleagues (Ashby et al., 2007). Univariate analyses (Waldschmidt and Ashby, 2011) found that activity in putamen correlated with task performance early in training. After extensive categorization training, putamen activity was no longer related to performance in the task, the only areas in which activity correlated with accuracy were cortical (SMA and pre-SMA). There was no evidence that the basal ganglia made a meaningful contribution to categorization performance after overtraining.

Previous multivariate analyses of the data (Soto et al., 2013) compared patterns of activity in categorization-related ROIs across three different categorization tasks, including the information-integration task studied here. The results of those analyses revealed to what extent the activity patterns observed in the information-integration task were similar or dissimilar to those observed in rule-based tasks. The results suggested that, early in training, the information-integration task could be distinguished from rule-based tasks based on activity patterns from a variety of motor and subcortical areas, including caudate and thalamus. This is in line with the hypothesis that early category learning depends on different neural mechanisms for information-integration and rule-based tasks. After extensive training, the information-integration task could be distinguished from rule-based tasks based on activity patterns from only a handful of regions. That is, patterns across the brain became increasingly similar across tasks after overtraining, which is what would be expected if the development of automaticity in such tasks involves gradual transfer of control from originally dissociable learning systems to common sensory-motor cortical networks.

As discussed previously, the present results are also in line with the hypothesis of an early engagement of basal ganglia and related subcortical structures in category learning, which decreases with extensive training in a categorization task. However, the present study also extends beyond the scope of previous analyses in two main ways. First, this is the first study to directly compare how accuracy and response times are related to the same measures of neural function across categorization training, taking advantage of the fact that changes in these two behavioral measures of performance should accompany

different learning processes according to the neurocomputational theory of Ashby and colleagues (Ashby et al., 2007). Second, the present study focused on characterizing the pattern of connectivity of small regions in the brain network related to categorization, instead of characterizing the pattern of localized activity in such regions, as previous studies did. Our results suggest that subcortical areas not only show stronger and more unique patterns of activity during early category learning, but that changes in their topological properties are predictive of changes in task accuracy. On the other hand, as predicted by the neurocomputational theory, the same topological properties are much less informative about changes in speed of performance throughout training.

Methodological considerations

Although the number of participants in our study is small, the number of scans is much larger than is common in traditional fMRI studies: each participant was trained in the same task for 20 days, and scanned during 4 of these days. This large longitudinal sample size is critical for regression analysis used to understand how changes in network topology could explain changes in behavior *across training*. Indeed, each participant completed more than 11,520 categorization trials and the modeling was based on 24 blocks of data (i.e., 80 trials each) per participant. In total, each model was fit to 237 data points. As we discuss more extensively in the supplementary material, both this larger sample of data points and the number of participants is above the level recommended in the multilevel regression literature. Furthermore, previous work in the field of multilevel regression has shown that the main effect of a small number of participants should be a loss of power in statistical tests, but estimates from regression analyses should have good statistical properties. Thus, the statistically significant results reported here should be generalizable and the regression estimates precise, although we may have missed true effects due to a lack of statistical power.

Computation of intra-modular strength and participation coefficient depends on how the network is partitioned into modules, and thus it could be affected by such partition. The intra-modular strength is standardized to reflect the strength distribution within each module; it reflects the strength of connections of a node in relation to all nodes in the same module. Changes in module size should not affect this measure. Changes in the node's module affiliation could. Thus, our interpretation of results related to intra-modular strength should be informed by the fact that an increase or decrease in the measure could be due to a change in module affiliation instead of a change in the connectivity of the node within the same module. In either case, though, an increase in intra-modular strength can be interpreted as an increase in the "hubness" of the node relative to its own module.

The situation is more complicated for the participation coefficient. The scale of this measure changes with the number of modules in the network. For example, the maximum value (the case in which connection strength is evenly distributed across modules) for a network with two modules is 0.5, whereas for a network with 5 modules is 0.8. Thus, the number of modules affects the value of this measure independently of the node's connectivity. If a particular node does not change its connectivity to other nodes at all, but the number of modules increases, then the PC of that node should increase. Although the results of our analysis of modularity suggest that the number of modules was relatively stable (between 4 and 5 on average, without statistically significant changes across blocks; see supplementary material), we should note that it is possible that what predicts increases in accuracy is not a higher connectivity across modules, but a larger number of modules. Note, however, that a change in global network properties like this could not explain the distribution of regression weights across ROIs depicted in Fig. 10, where increments in participation coefficient were particularly predictive of changes in accuracy in subcortical areas (a result confirmed by a permutation test). Still, we performed a regression analysis (see supplementary material) to determine whether variation

in the number of modules across blocks and participants could explain changes in accuracy or response times. The resulting regression models could not predict either behavioral measure better than a null model, indicating that it is unlikely that the results shown in Fig. 10 could be the result of changes of scale in the participation coefficient due to changes in the number of modules in the network.

While a larger graph statistic battery may be relevant for some scientific questions (Ekman et al., 2012), here we use a small number of measures to (i) reduce collinearity issues (Dormann et al., 2013; see supplementary material) and (ii) maximize interpretability of the relationship between graph measures and behavior, as revealed by regression coefficients.

Here we study networks built using functional connectivity, which is a measure of the statistical association between signals in pairs of nodes. Measures of functional connectivity may or may not reflect the strength of information transmission between two areas, as the former can change without changes in the latter (Friston, 2011; Horwitz, 2003). However, the interpretation of our results does not change significantly when we take these issues into consideration. For example, correlated signals suggest that two areas, if not directly connected, at least process similar information or process information in a similar way. Network measures can be interpreted in terms of integration and segregation of function, and integration may be more costly than segregation if it utilizes redundant information processing. On the other hand, an area may change functional connectivity without any underlying change in effective connectivity (Friston, 2011), meaning that localization of an effect in a particular ROI suggests, but does not imply, localization of the underlying mechanisms in that ROI. Furthermore, relations between functional connectivity changes and behavioral changes should not be interpreted as causal, since both are likely the product of latent changes in effective connectivity.

Finally, we assumed that changes in functional network measures across training were related to learning of the categorization task and not to other changes occurring throughout the study such as habituation to the scanner environment. Two features of our results support this assumption: (i) we found a dissociation between graph statistics that could predict changes in response time and changes in accuracy, and (ii) we confirmed the prediction that subcortical areas should be more important to predict accuracy changes than response time changes. At least two independent confounds are required to explain these results, which should follow very specific trends throughout training, and have different effects on different network measures and brain areas.

Conclusions

Using network science and multilevel regression, we showed that changes in the connectivity of fine-scale regions can predict behavioral changes during training in a visual categorization task. As expected, we found that changes in the connectivity of subcortical areas were particularly critical for predicting changes in accuracy reflecting initial learning. Such initial learning was predicted by increasingly efficient integrative processing in subcortical areas, with higher functional specialization, more efficient integration across modules, but at a lower cost in terms of redundancy of information processing. Changes in the speed of correct responding, reflecting the development of automaticity, were predicted by lower clustering (particularly in subcortical areas), higher strength (highest in cortical areas), and higher betweenness centrality.

Our work shows that two different recent developments in computational cognitive neuroscience—neurocomputational theories of category learning (Ashby et al., 2007; Ashby et al., 1998) and the application of network science to understanding brain connectivity (Bullmore and Sporns, 2009; Bullmore and Bassett, 2011; Sporns, 2014)—can be used in a complementary fashion to reach a better understanding of brain function. Neurocomputational models of category learning and automaticity (Ashby et al., 2007) make predictions about

changes in functional connectivity between a few specific brain regions, some of which have been supported by previous work (DeGutis and D'Esposito, 2009; Seger et al., 2010). Here, applying network science allowed us to explore global changes in functional connectivity as a function of learning, as opposed to changes in specific connections. The measures afforded by graph theory allowed us to study the distributed shifts in segregation and integration that accompany category learning and development of automaticity, providing a more complete characterization of changes in functional networks than is possible in the study of single connections. On the other hand, neurocomputational theories of categorization provided novel hypotheses about what brain areas should be particularly important focal points of functional network reconfiguration during learning, which have been supported by our results.

Acknowledgments

This work was supported in part by grant no. W911NF-07-1-0072 from the U.S. Army Research Office through the Institute for Collaborative Biotechnologies. DSB was supported by the John D. and Catherine T. MacArthur Foundation, Alfred P. Sloan Foundation, the Army Research Laboratory through contract no. W911NF-10-2-0022 from the U.S. Army Research Office, the U.S. Army Research Office through contract no. W911NF-14-1-0679, the Office of Naval Research through the Young Investigator Program, the National Institutes of Health through 1-R01-HD086888-01 and 2-R01-DC-009209-11, and awards #BCS-1441502 and #BCS-1430087 from the National Science Foundation. The authors declare no competing financial interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2016.07.032>.

References

- Achard, S., Salvador, R., Whitcher, B., Suckling, J., Bullmore, E., 2006. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.* 26, 63–72. <http://dx.doi.org/10.1523/JNEUROSCI.3874-05.2006>.
- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19, 716–723.
- Alexander-Bloch, A., Lambiotte, R., Roberts, B., Giedd, J., Gogtay, N., Bullmore, E., 2012. The discovery of population differences in network community structure: New methods and applications to brain functional networks in schizophrenia. *NeuroImage* 59, 3889–3900. <http://dx.doi.org/10.1016/j.neuroimage.2011.11.035>.
- Andersson, J.L.R., Jenkins, M., Smith, S., 2007. Non-linear registration, aka spatial normalisation (FMRIB technical report No. TR07JA2). FMRIB Analysis Group of the University of Oxford, Oxford, UK.
- Ashby, F.G., Maddox, W.T., 2005. Human category learning. *Annu. Rev. Psychol.* 56, 149–178.
- Ashby, F.G., Ennis, J.M., 2006. The role of the basal ganglia in category learning. In: Ross, B.H. (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory*. Elsevier, New York, NY, pp. 1–36.
- Ashby, F.G., Spiering, B.J., 2004. The neurobiology of category learning. *Behav. Cogn. Neurosci. Rev.* 3, 101–113. <http://dx.doi.org/10.1177/1534582304270782>.
- Ashby, F.G., Paul, E.J., Maddox, W.T., 2011. COVIS. In: Pothos, E.M., Wills, A.J. (Eds.), *Formal approaches in categorization*. Cambridge University Press, New York, pp. 65–87.
- Ashby, F.G., Alfonso-Reese, L.A., Turken, A.U., Waldron, E.M., 1998. A neuropsychological theory of multiple systems in category learning. *Psychol. Rev.* 105, 442–481.
- Ashby, F.G., Ell, S.W., Waldron, E.M., 2003. Procedural learning in perceptual categorization. *Mem. Cogn.* 31, 1114–1125.
- Ashby, F.G., Ennis, J.M., Spiering, B.J., 2007. A neurobiological theory of automaticity in perceptual categorization. *Psychol. Rev.* 114, 632–656.
- Ashby, F.G., Turner, B.O., Horvitz, J.C., 2010. Cortical and basal ganglia contributions to habit learning and automaticity. *Trends Cogn. Sci.* 14 (5), 191–232.
- Barrat, A., Barthélémy, M., Pastor-Satorras, R., Vespignani, A., 2004. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. U. S. A.* 101, 3747–3752. <http://dx.doi.org/10.1073/pnas.0400087101>.
- Bassett, D.S., Meyer-Lindenberg, A., Achard, S., Duke, T., Bullmore, E., 2006. Adaptive re-configuration of fractal small-world human brain functional networks. *Proc. Natl. Acad. Sci.* 103, 19518.
- Bassett, D.S., Bullmore, E., Meyer-Lindenberg, A., Apud, J.A., Weinberger, D.R., Coppola, R., 2009. Cognitive fitness of cost-efficient brain functional networks. *Proc. Natl. Acad. Sci.* 106, 11747–11752.
- Bassett, D.S., Wymbs, N.F., Porter, M.A., Mucha, P.J., Carlson, J.M., Grafton, S.T., 2011. Dynamic reconfiguration of human brain networks during learning. *Proc. Natl. Acad. Sci.* 108, 7641–7646. <http://dx.doi.org/10.1073/pnas.1018985108>.
- Bassett, D.S., Nelson, B.G., Mueller, B.A., Camchong, J., Lim, K.O., 2012. Altered resting state complexity in schizophrenia. *NeuroImage* 59, 2196–2207. <http://dx.doi.org/10.1016/j.neuroimage.2011.10.002>.
- Bassett, D.S., Porter, M.A., Wymbs, N.F., Grafton, S.T., Carlson, J.M., Mucha, P.J., 2013a. Robust detection of dynamic community structure in networks. *Chaos Interdiscip. J. Nonlinear Sci.* 23, 013142. <http://dx.doi.org/10.1063/1.4790830>.
- Bassett, D.S., Wymbs, N.F., Rombach, M.P., Porter, M.A., Mucha, P.J., Grafton, S.T., 2013b. Task-based core-periphery organization of human brain dynamics. *PLoS Comput. Biol.* 9, e1003171. <http://dx.doi.org/10.1371/journal.pcbi.1003171>.
- Bassett, D.S., Yang, M., Wymbs, N.F., Grafton, S.T., 2015. Learning-induced autonomy of sensorimotor systems. *Nat. Neurosci.* 18, 744–751.
- Bates, D.M., Maechler, M., Bolker, B., Walker, S., 2014. *lme4: Linear mixed-effects models using Eigen and S4*.
- Bell, B.A., Morgan, G.B., Schoeneberger, J.A., Kromrey, J.D., Ferron, J.M., 2014. How low can you go? An investigation of the influence of sample size and model complexity on point and interval estimates in two-level linear models. *Methodol. Eur. J. Res. Methods Behav. Soc. Sci.* 10, 1–11.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300.
- Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. <http://dx.doi.org/10.1088/1742-5468/2008/10/P10008> (arXiv:0803.0476).
- Brainard, D.H., 1997. The psychophysics toolbox. *Spat. Vis.* 10, 433–436.
- Braun, U., Schäfer, A., Walter, H., Erk, S., Romanczuk-Seifert, N., Haddad, L., Schweiger, J.I., Grimm, O., Heinz, A., Tost, H., Meyer-Lindenberg, A., Bassett, D.S., 2015. Dynamic re-configuration of frontal brain networks during executive cognition in humans. *Proc. Natl. Acad. Sci.* 112, 11678–11683.
- Bullmore, E., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198. <http://dx.doi.org/10.1038/nrn2575>.
- Bullmore, E.T., Bassett, D.S., 2011. Brain graphs: graphical models of the human brain connectome. *AR Clinical Psychology* 7, 113–140.
- Burnham, K.P., Anderson, D.R., 2004. Multimodel inference understanding AIC and BIC in model selection. *Social. Methods Res.* 33, 261–304.
- Chen, Z.J., He, Y., Rosa-Neto, P., Germann, J., Evans, A.C., 2008. Revealing modular architecture of human brain structural networks by using cortical thickness from MRI. *Cereb. Cortex* 18, 2374–2381. <http://dx.doi.org/10.1093/cercor/bhn003>.
- Cole, M.W., Reynolds, J.R., Power, J.D., Repovs, G., Anticevic, A., Braver, T.S., 2013. Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat. Neurosci.* 16, 1348–1355. <http://dx.doi.org/10.1038/nn.3470>.
- Cole, M.W., Bassett, D.S., Power, J.D., Braver, T.S., Petersen, S.E., Cole, M.W., 2014. Intrinsic and task-evoked network architectures of the human brain. *Neuron* 83, 238–251.
- Cordes, D., Haughton, V.M., Arfanakis, K., Carew, J.D., Turski, P.A., Moritz, C.H., Quigley, M.A., Meyerand, M.E., 2001. Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data. *Am. J. Neuroradiol.* 22, 1326–1333.
- Cornish, C., 2006. *WMTSA Wavelet Toolkit for MATLAB*.
- Costa, L.F., Rodrigues, F.A., Travieso, G., Villas Boas, P.R., 2007. Characterization of complex networks: a survey of measurements. *Adv. Phys.* 56, 167–242. <http://dx.doi.org/10.1080/00018730601170527>.
- Crossman, E., 1959. A theory of the acquisition of speed-skill. *Ergonomics* 2, 153–166. <http://dx.doi.org/10.1080/00140135908930419>.
- DeGutis, J., D'Esposito, M., 2009. Network changes in the transition from initial learning to well-practiced visual categorization. *Front. Hum. Neurosci.* 3, 44.
- Ding, L., Gold, J.I., 2013. The basal ganglia's contributions to perceptual decision making. *Neuron* 79 (4), 640–649.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *EcoGraphy* 36, 027–046.
- Doyon, J., Bellec, P., Amsel, R., Penhune, V., Monchi, O., Carrier, J., 2009. Contributions of the basal ganglia and functionally related brain structures to motor learning. *Behav. Brain Res.* 199 (1), 61–75.
- Ekman, M., Derrfuss, J., Tittgemeyer, M., Fiebach, C.J., 2012. Predicting errors from reconfiguration patterns in human brain networks. *Proc. Natl. Acad. Sci.* 109, 16714–16719. <http://dx.doi.org/10.1073/pnas.1207523109>.
- Fornito, A., Harrison, B.J., Zalesky, A., Simons, J.S., 2012. Competitive and cooperative dynamics of large-scale brain functional networks supporting recollection. *Proc. Natl. Acad. Sci.* 109, 12788–12793. <http://dx.doi.org/10.1073/pnas.1204185109>.
- Freeman, L.C., 1978. Centrality in social networks conceptual clarification. *Soc. Networks* 1, 215–239. [http://dx.doi.org/10.1016/0378-8733\(78\)90021-7](http://dx.doi.org/10.1016/0378-8733(78)90021-7).
- Friston, K.J., 2011. Functional and effective connectivity: a review. *Brain Connect.* 1, 13–36. <http://dx.doi.org/10.1089/brain.2011.0008>.
- Gelman, A., Hill, J., 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York, NY.
- Good, B.H., de Montjoye, Y.A., Clauset, A., 2010. Performance of modularity maximization in practical contexts. *Phys. Rev. E* 81, 046106.
- Gu, S., Pasqualetti, F., Cieslik, M., Telesford, Q.K., Yu, A.B., Kahn, A.E., Medaglia, J.D., Vettel, J.M., Miller, M.B., Grafton, S.T., Bassett, D.S., 2015. Controllability of structural brain networks. *Nat. Commun.* 6, 8414.
- Guimera, R., Amaral, L.A., 2005. Functional cartography of complex metabolic networks. *Nature* 433, 895–900. <http://dx.doi.org/10.1038/nature03288>.
- Haber, S.N., Johnson-Gdwowski, M., 2004. The basal ganglia. In: Paxinos, G., Mai, J.K. (Eds.), *The Human Nervous System*. Academic Press, San Diego, CA, pp. 676–719.

- Heitger, M.H., Ronsse, R., Dhollander, T., Dupont, P., Caeyenberghs, K., Swinnen, S.P., 2012. Motor learning-induced changes in functional brain connectivity as revealed by means of graph-theoretical network analysis. *NeuroImage* 61, 633–650. <http://dx.doi.org/10.1016/j.neuroimage.2012.03.067>.
- Heitz, R.P., 2014. The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Front. Neurosci.* 8, 150.
- Helie, S., Waldschmidt, J.G., Ashby, F.G., 2010. Automaticity in rule-based and information-integration categorization. *Atten. Percept. Psychophys.* 72, 1013–1031. <http://dx.doi.org/10.3758/APP.72.4.1013>.
- Helie, S., Ell, S.W., Ashby, F.G., 2015. Learning robust cortico-cortical associations with the basal ganglia: an integrative review. *Cortex* 64, 123–135.
- Horwitz, B., 2003. The elusive concept of brain connectivity. *NeuroImage* 19, 466–470. [http://dx.doi.org/10.1016/S1053-8119\(03\)00112-5](http://dx.doi.org/10.1016/S1053-8119(03)00112-5).
- Hurvich, C.M., Tsai, C.L., 1989. Regression and time series model selection in small samples. *Biometrika* 76, 297–307. <http://dx.doi.org/10.1093/biomet/76.2.297>.
- Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–156. [http://dx.doi.org/10.1016/S1361-8415\(01\)00036-6](http://dx.doi.org/10.1016/S1361-8415(01)00036-6).
- Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* 17, 825–841 (doi:06/nimg.2002.1132).
- Kitzbichler, M.G., Henson, R.N.A., Smith, M.L., Nathan, P.J., Bullmore, E., 2011. Cognitive effort drives workspace configuration of human brain functional networks. *J. Neurosci.* 31, 8259–8270. <http://dx.doi.org/10.1523/JNEUROSCI.0440-11.2011>.
- Lohse, C., Bassett, D.S., Lim, K.O., Carlson, J.M., 2013. Resolving structure in human brain organization: identifying mesoscale organization in weighted network representations (ArXiv13126070 Q-Bio).
- Maddox, W.T., Bohil, C.J., Ing, A.D., 2004. Evidence for a procedural-learning-based system in perceptual category learning. *Psychon. Bull. Rev.* 11, 945–952.
- Mantzaris, A.V., Bassett, D.S., Wymbs, N.F., Estrada, E., Porter, M.A., Mucha, P.J., Grafton, S.T., Higham, D.J., 2013. Dynamic network centrality summarizes learning in the human brain. *J. Complex Netw.* 1, 83–92. <http://dx.doi.org/10.1093/comnet/cnt001>.
- Martin, J.H., 2003. *Neuroanatomy: text and atlas*. McGraw-Hill Professional.
- Maas, C.J.M., Hox, J.J., 2005. Sufficient sample sizes for multilevel modeling. *Methodol. Eur. J. Res. Methods Behav. Soc. Sci.* 1, 86–92.
- Meunier, D., Achard, S., Morcom, A., Bullmore, E., 2009. Age-related changes in modular organization of human brain functional networks. *NeuroImage* 44, 715–723.
- Moussa, M.N., Vechlekar, C.D., Burdette, J.H., Steen, M.R., Hugenschmidt, C.E., Laurienti, P.J., 2011. Changes in cognitive state alter human functional brain networks. *Front. Hum. Neurosci.* 5, 83. <http://dx.doi.org/10.3389/fnhum.2011.00083>.
- Newman, M.E.J., 2004. Analysis of weighted networks. *Phys. Rev. E* 70, 056131. <http://dx.doi.org/10.1103/PhysRevE.70.056131>.
- Newman, M.E.J., 2010. Networks: An Introduction. Oxford University Press, Oxford, UK.
- Nolte, J., 2008. The Human Brain: An Introduction to Its Functional Anatomy. sixth ed. Mosby, Philadelphia.
- Nomura, E.M., Reber, P.J., 2008. A review of medial temporal lobe and caudate contributions to visual category learning. *Neurosci. Biobehav. Rev.* 32 (2), 279–291.
- Onnela, J.P., Saramäki, J., Kertész, J., Kaski, K., 2005. Intensity and coherence of motifs in weighted complex networks. *Phys. Rev. E* 71, 065103. <http://dx.doi.org/10.1103/PhysRevE.71.065103>.
- Packard, M.G., Knowlton, B.J., 2012. Learning and memory functions of the basal ganglia. *Annu. Rev. Neurosci.* 25 (1), 563–593.
- Percival, D.B., Walden, A.T., 2000. Wavelet methods for time series analysis. Cambridge University Press, Cambridge, UK.
- Petrides, M., Pandya, D.N., 2004. The frontal cortex. In: Paxinos, G., Mai, J.K. (Eds.), *The Human Nervous System*. Academic Press, San Diego, CA, pp. 950–972.
- Picard, N., Strick, P.L., 2001. Imaging the premotor areas. *Curr. Opin. Neurobiol.* 11, 663–672. [http://dx.doi.org/10.1016/S0959-4388\(01\)00266-5](http://dx.doi.org/10.1016/S0959-4388(01)00266-5).
- Poldrack, R.A., Foerde, K., 2008. Category learning and the memory systems debate. *Neurosci. Biobehav. Rev.* 32 (2), 197–205.
- 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., Petersen, S.E., 2011. Functional network organization of the human brain. *Neuron* 72, 665–678. <http://dx.doi.org/10.1016/j.neuron.2011.09.006>.
- Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage* 59, 2142–2154.
- Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2014. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage* 84, 320–341. <http://dx.doi.org/10.1016/j.neuroimage.2013.08.048>.
- R Development Core Team, 2014. R: A Language and Environment for Statistical Computing. Vienna, Austria.
- Rawlings, J.O., Pantula, S.G., Dickey, D.A., 1998. Applied regression analysis: a research tool. Springer, New York, NY.
- Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P., Van De Ville, D., 2011. Decoding brain states from fMRI connectivity graphs. *NeuroImage*, Multivar. Decoding Brain Read. 56, 616–626. <http://dx.doi.org/10.1016/j.neuroimage.2010.05.081>.
- Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage* 52, 1059–1069. <http://dx.doi.org/10.1016/j.neuroimage.2009.10.003>.
- Rubinov, M., Sporns, O., 2011. Weight-conserving characterization of complex functional brain networks. *NeuroImage* 56, 2068–2079. <http://dx.doi.org/10.1016/j.neuroimage.2011.03.069>.
- Rzucidlo, J.K., Roseman, P.L., Laurienti, P.J., Dagenbach, D., 2013. Stability of whole brain and regional network topology within and between resting and cognitive states. *PLoS ONE* 8, e70275. <http://dx.doi.org/10.1371/journal.pone.0070275>.
- Satterthwaite, T.D., Wolf, D.H., Loughead, J., Ruparel, K., Elliott, M.A., Hakonarson, H., Gur, R.C., Gur, R.E., 2012. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. *NeuroImage* 60, 623–632. <http://dx.doi.org/10.1016/j.neuroimage.2011.12.063>.
- Satterthwaite, T.D., Elliott, M.A., Gerraty, R.T., Ruparel, K., Loughead, J., Calkins, M.E., Eickhoff, S.B., Hakonarson, H., Gur, R.C., Gur, R.E., Wolf, D.H., 2013. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage* 64, 240–256. <http://dx.doi.org/10.1016/j.neuroimage.2012.08.052>.
- Seger, C.A., Peterson, E.J., Cincotta, C.M., Lopez-Paniagua, D., Anderson, C.W., 2010. Dissociating the contributions of independent corticostratal systems to visual categorization learning through the use of reinforcement learning modeling and Granger causality modeling. *NeuroImage* 50, 644–656.
- Snijders, T.A.B., Bosker, R.J., 2012. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. 2nd ed. Thousand Oaks, SAGE, CA.
- Soto, F.A., Waldschmidt, J.G., Helie, S., Ashby, F.G., 2013. Brain activity across the development of automatic categorization: a comparison of categorization tasks using multi-voxel pattern analysis. *NeuroImage* 71, 284–297. <http://dx.doi.org/10.1016/j.neuroimage.2013.01.008>.
- Sporns, O., 2013. Network attributes for segregation and integration in the human brain. *Curr. Opin. Neurobiol.* 23, 162–171. <http://dx.doi.org/10.1016/j.conb.2012.11.015>.
- Sporns, O., 2014. Contributions and challenges for network models in cognitive neuroscience. *Nat. Neurosci.* 17, 652–660. <http://dx.doi.org/10.1038/nn.3690>.
- Sun, F.T., Miller, L.M., D'Esposito, M., 2004. Measuring interregional functional connectivity using coherence and partial coherence analyses of fMRI data. *NeuroImage* 21, 647–658. <http://dx.doi.org/10.1016/j.neuroimage.2003.09.056>.
- van den Heuvel, M.P., Sporns, O., 2013. Network hubs in the human brain. *Trends Cogn. Sci.* 17, 683–696. <http://dx.doi.org/10.1016/j.tics.2013.09.012> (Special Issue: The Connectome).
- Van Dijk, K.R.A., Sabuncu, M.R., Buckner, R.L., 2012. The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage* 59, 431–438.
- Varoquaux, G., Craddock, R.C., 2013. Learning and comparing functional connectomes across subjects. *NeuroImage* 80, 405–415. <http://dx.doi.org/10.1016/j.neuroimage.2013.04.007>.
- Vogt, B.A., Hof, P.R., Vogt, L.J., 2004. Cingulate gyrus. In: Paxinos, G., Mai, J.K. (Eds.), *The Human Nervous System*. Academic Press, San Diego, pp. 915–949.
- Waldschmidt, J.G., Ashby, F.G., 2011. Cortical and striatal contributions to automaticity in information-integration categorization. *NeuroImage* 56, 1791–1802. <http://dx.doi.org/10.1016/j.neuroimage.2011.02.011>.
- Zuo, X.N., Ehmke, R., Mennes, M., Imperati, D., Castellanos, F.X., Sporns, O., Milham, M.P., 2012. Network centrality in the human functional connectome. *Cereb. Cortex* 22, 1862–1875. <http://dx.doi.org/10.1093/cercor/bhr269>.