

Distributed representation of context by intrinsic subnetworks in prefrontal cortex

Michael L. Waskom^{a,b,1} and Anthony D. Wagner^{a,c}

^aDepartment of Psychology, Stanford University, Stanford, CA 94305; ^bCenter for Neural Science, New York University, New York, NY 10003; and ^cStanford Neurosciences Institute, Stanford University, Stanford, CA 94305

Edited by John Duncan, Medical Research Council, Cambridge, United Kingdom, and accepted by Editorial Board Member Leslie G. Ungerleider January 6, 2017 (received for review September 12, 2016)

Human prefrontal cortex supports goal-directed behavior by representing abstract information about task context. The organizational basis of these context representations, and of representations underlying other higher-order processes, is unknown. Here, we use multivariate decoding and analyses of spontaneous correlations to show that context representations are distributed across subnetworks within prefrontal cortex. Examining targeted prefrontal regions, we found that pairs of voxels with similar context preferences exhibited spontaneous correlations that were approximately twice as large as those between pairs with opposite context preferences. This subnetwork organization was stable across task-engaged and resting states, suggesting that abstract context representations are constrained by an intrinsic functional architecture. These results reveal a principle of fine-scaled functional organization in association cortex.

fMRI | resting-state | rule | cognitive control | functional organization

The cerebral cortex exhibits functional organization at multiple spatial scales. At a coarse scale, the cortex is parcellated into functional areas (1, 2) that coordinate as networks through long-range connections (3–5). These areas represent and compute information with functional circuitry that is organized more finely. Some fine-scaled, intrinsic principles have been described in detail, particularly for sensory cortex (6, 7). In contrast, the subregional organization of prefrontal cortex (PFC), which gives rise to higher-order processes, including attention, decision-making, and goal-directed action (8, 9), remains largely uncharacterized. Whether PFC representations are encoded within an equipotential system or constrained by an intrinsic functional architecture is currently unknown.

Sensory cortex can be mapped by parametrically varying stimulus attributes and measuring changes in the neural response, but complex and dynamic response properties limit the effectiveness of this approach for mapping PFC and other association regions. An alternate strategy is to leverage spontaneous variability in neural activity. Neural responses to repeated presentations of sensory stimuli, or in the absence of stimulation and explicit task demands, exhibit variability that is attributed to ongoing spontaneous activity (10–13). Traditional analyses consider spontaneous activity to be noise, but there is increasing evidence that shared spontaneous variability is a signature of functional organization (14). Analyses of spontaneous correlations have been used to identify multiple large-scale functional networks (4, 5, 15) and boundaries between functional areas (2, 16–18) in human and nonhuman primate association cortex. The spontaneous correlation structure also mirrors established fine-scaled principles of functional organization in visual cortex, such as preferences for retinotopic position (19) and stimulus orientation (20). We therefore leveraged spontaneous activity to examine the fine-scaled functional organization of human PFC.

We focused our investigation on prefrontal computations that enable goal-directed behavior. Across two experiments, we scanned subjects with high-resolution fMRI while they performed tasks that demanded selective integration of noisy sensory evidence according to a shifting decision rule, or context. During rule-based decision-making, distributed patterns of activation in lateral PFC

encode a representation of task context (21–23). We have previously reported that these context representations are strongly expressed in the inferior frontal sulcus (IFS), a prefrontal region defined through large-scale analyses of spontaneous correlations (24, 25). Here, we use multivariate decoding and analyses of subregional spontaneous correlations to show that context representations are organized across subnetworks within the IFS. Our results demonstrate that abstract cognitive representations emerge from an intrinsic functional organization in human PFC.

Results

In experiment 1, subjects viewed a bivalent random dot stimulus and were cued to discriminate either the direction of coherent motion or the more prevalent color (Fig. S1A). We first sought to identify distributed patterns representing the task context (i.e., the motion vs. color rule). Guided by our previous findings (24), we applied linear classifiers to patterns of activation within the bilateral IFS (Fig. 1A). Cross-validated decoding accuracy exceeded chance at both the group (mean = 72.2%, $t_{13} = 11.22$, $P < 0.001$; Fig. 1B) and subject (13/14 subjects $P < 0.05$, permutation test) levels. We also evaluated a region in medial frontal cortex (MFC) (Fig. 1A) that is widely implicated in cognitive control processes (26). Decoding performance in MFC exceeded chance at the group level (mean = 57.3%, $t_{13} = 3.89$, $P = 0.001$) and in some individual subjects (6/14 subjects $P < 0.05$, permutation test) but was significantly weaker than in the IFS (paired $t_{13} = 7.09$, $P < 0.001$; Fig. 1B). Because we could reliably decode task context representations from the IFS in nearly all subjects, we focus on this region in subsequent analyses.

Significance

Information is represented in the brain by distributed patterns of cortical activity. In sensory cortex, these patterns are expressed across circuits with an intrinsic functional architecture that is organized along relevant stimulus dimensions. However, it is unknown whether similar organizational principles underlie distributed representations of more abstract information, such as rules or goals. We analyzed correlations in spontaneous activity to identify fine-scaled subnetworks in human prefrontal cortex. These subnetworks were differentially engaged when subjects followed rules in a complex decision-making task. Our results show how the abstract representations that support goal-directed cognition are constrained by an intrinsic functional architecture and prompt new models of information representation in association cortex.

Author contributions: M.L.W. and A.D.W. designed research; M.L.W. performed research; M.L.W. analyzed data; and M.L.W. and A.D.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. J.D. is a Guest Editor invited by the Editorial Board.

¹To whom correspondence should be addressed. Email: mwaskom@nyu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1615269114/-DCSupplemental.

representations in human PFC. Despite this qualitative agreement, our spatial clustering analyses indicate that context preferences are shared across larger spatial extents than can be accounted for by the stripe sizes previously observed in nonhuman primates (40). This may represent a difference between species or the effect of blurring within voxels. Quantitative between-species comparisons will require functional imaging data with a higher spatial resolution.

Multivariate decoding approaches are increasingly used to test theories about higher-order representations (42), but the organizational principles that underlie successful decoding in association cortex have remained unclear. Our data indicate that, for context decoding in lateral PFC, information is expressed across patches on the order of tens of millimeters and that multiple patches comprise functional subnetworks. It remains to be determined whether subnetwork organization is specific to the representation of context in lateral PFC or if it is a general principle in association cortex. We note that large-scale networks exhibit consistent organization across the brain (4, 5) and suggest that fine-scaled subnetworks in other regions may also provide structure for high-level cognitive representations.

Analyses of spontaneous correlations identify highly reproducible large-scale networks (4, 5) that are recruited in the performance of functional tasks (43–45). Despite this, the relevance of spontaneous correlations to the finer scales of organization that underlie cognitive representations has not been determined. Some aspects of fine-scaled organization in sensory cortex are reflected in spontaneous correlation structure, including retinotopic eccentricity biases (19) and orientation maps (20). Analyses that identify discontinuities in spontaneous correlation structure have also been applied in PFC to characterize areal boundaries (16, 18). Our results extend these previous findings in two important ways. First, they show that the fine-scale spontaneous correlation structure in PFC underlies task-evoked representations of abstract context information. Second, they indicate that subregional organization is characterized not just by discrete areas but by distributed subnetworks. These findings demonstrate the utility of spontaneous correlation analyses as a tool for studying the fine-scaled functional organization of association cortex and clarify the importance of functionally specific network organization in the human brain at multiple spatial scales.

Methods

Subjects. Subjects were healthy right-handed members of the Stanford community and gave written informed consent before participating. The study was approved by the Stanford University Institutional Review Board. All subjects had normal or corrected-to-normal visual acuity and normal color vision. Subjects received \$10/h for behavioral sessions and \$20/h for scanning, and they additionally received a monetary bonus based on their performance during the imaging sessions.

In experiment 1, 20 subjects were recruited. Of this group, one was excluded early in the training session for poor color vision, one was excluded after the training session for performing at chance, and three ended the scanning session early due to fatigue or illness. One additional subject was excluded after initial data analysis suggested experimenter error during data acquisition; the decision to exclude this subject was made before performing the present analyses. The analyses reported here reflect data from the 14 remaining subjects (19 y to 29 y old; seven females). Each subject participated in one behavioral training session and one imaging session, for ~3.5 total hours participation.

In experiment 2, eight subjects were recruited; all had extensive experience being scanned, one subject was author M.L.W., and one subject had participated in experiment 1. Of this group, one subject was excluded for early exit from the scanning session, and one subject was excluded due to data loss during image reconstruction. The analyses reported here reflect data from the remaining six subjects (24 y to 29 y old; three females). Each subject participated in two behavioral training sessions and two imaging sessions for ~5 total hours participation.

Stimuli and Experimental Design.

Experiment 1. The stimuli and experimental design for experiment 1 have been previously reported in detail (25). Briefly, subjects viewed a bivalent random dot stimulus and were cued to make either a motion or a color

discrimination on each trial (Fig. S1A). Motion information was controlled by displacing a selected proportion of the dots coherently either up or down on each screen refresh, whereas the remainder of the dots were redrawn at a random location. Color information was controlled by drawing a selected proportion of the dots in either red or green on each screen refresh. The difficulties of the color and motion discriminations were set independently for each subject using a staircase procedure in a separate training session before scanning. On each trial, the relevant dimension (the context) was cued by the pattern of the frame surrounding the stimulus. Two distinct patterns cued a motion trial, and two distinct patterns cued a color trial. The cue appeared 1 s before the stimulus on one third of trials, it appeared concurrent with stimulus onset on one third of trials, and one third of trials were “cue-only” trials where the frame was presented for 1 s but was not followed by a stimulus. On trials with a stimulus, it was shown for 2 s. Subjects were instructed to respond as soon as they had made a decision, and they indicated their response with a button box held in the right hand. No feedback was presented during scanning. The experiment was structured into epochs with different proportions of motion and color trials (see ref. 25 for details). In total, each subject performed 900 trials (600 with a stimulus), evenly divided between motion and color contexts, across 12 scanner runs. Each run had a duration of 460 s (230 volumes).

Experiment 2. The stimuli and experimental design for experiment 2 are reported in detail in the *SI Methods*. Briefly, subjects performed a context-dependent perceptual decision-making task that was similar to the task in experiment 1. Subjects viewed a bivalent stimulus that comprised a field of small sticks and were cued to make either an orientation or a color discrimination on each trial (Fig. S1B). Orientation and color information was manipulated by controlling the proportion of sticks drawn at either 45° or 135° and in either red or green. The difficulties of the orientation and color discriminations were set independently for each subject, using performance in two separate training sessions before scanning. In contrast to experiment 1, two different difficulty levels were used for each dimension during scanning. On each trial, the relevant dimension (the context) was cued by the shape of a polygon drawn at fixation. Two distinct polygons cued an orientation trial, and two distinct polygons cued a color trial. In contrast to experiment 1, the cue appeared concurrently with stimulus onset on every trial, the stimulus disappeared when subjects made their first button press response, and feedback was provided by blinking the fixation point after error trials. The relevant dimension on each trial was chosen randomly from a balanced distribution. In total, each subject performed 768 trials, evenly divided between orientation and color contexts, across 12 scanner runs in two separate scanning sessions. Each run had a duration of 370.8 s (515 volumes).

Resting-state scans. During resting-state scans, a black fixation cross was displayed on a gray background drawn at 30% luminance. Subjects were instructed to fixate on the cross and to let their minds wander. Fixation and wakefulness were monitored with an eye-tracking camera. We collected eight separate resting-state scans from each subject in experiment 2. Each resting-state scan had a duration of 367.2 s (510 volumes). We aimed to collect an amount of data that has been previously shown to produce reliable estimates of spontaneous correlations at the individual subject level (46). Both scanning sessions began with a resting-state scan, and the remainder were interleaved with the task runs.

MRI Acquisition and Preprocessing. Imaging data acquisition and preprocessing methods are described in detail in the *SI Methods*. Briefly, functional data were acquired with high spatial resolution in experiment 1 using a partial brain acquisition [$2 \times 2 \times 2.3$ -mm voxels; 2,000-ms temporal resolution (TR)] and high spatiotemporal resolution in experiment 2 using a whole-brain multiband acquisition (47) ($2 \times 2 \times 2$ mm; 720-ms TR). Functional data were distortion corrected (48) (only in experiment 2), spatially realigned, temporally interpolated (only in experiment 1), and high-pass temporally filtered. No spatial filtering or low-pass temporal filtering was applied. All analyses were performed in subject-specific space; no spatial normalization was applied. Functional time series data were denoised using estimates of head movements, nuisance signals derived from the white matter (49), and indicator vectors to remove frames with signal artifacts. Anatomical data were processed to construct a tessellated mesh model of the cortical surface (50), and functional data were placed in register by aligning each run's mean volume with the anatomy (51).

Region of Interest Definition. Data were analyzed within a priori regions of interest (ROIs) derived from a population atlas of large-scale resting-state networks (4). These regions are defined on the Freesurfer average cortical surface mesh. To obtain ROI masks in native functional space, we first reverse-normalized the ROI labels using the spherical registration parameters (52) and then transformed the vertex coordinates into the space of the first run using the inverse of the functional-to-anatomical registration. Voxels intersecting the

midpoint between the gray–white and gray–pial boundaries were included in the analyses. This produced ROIs that respected the underlying 2D topology of the cortical surface and minimized the contributions of voxels lying outside of gray matter. To estimate distances between voxels along the cortical surface, we established a one-to-one mapping between voxels and surface vertices. A small number of vertices were originally mapped to multiple voxels; in these cases, we used the voxel whose center was closest to the vertex coordinate. We then constructed a distance matrix between each voxel center, where distances were measured with Dijkstra's algorithm on the midthickness cortex mesh (53).

Decoding Analyses and Context Preference Estimation. We used linear classifiers to decode information about task context and estimate each voxel's context preference. Decoding was implemented in three steps. We first used voxelwise general linear models to deconvolve the amplitude of the evoked response in different experimental conditions. The estimated response amplitudes were then used as samples in the multivariate analyses. Finally, we inverted the resulting decoding models to obtain estimates of each voxel's context preference.

For experiment 1, the deconvolution model included 12 regressors reflecting a crossing between context (color or motion), trial type (early cue, concurrent cue, or cue only), and cue pattern. For experiment 2, the deconvolution model included 16 regressors reflecting a crossing between context (color or orientation), cue shape, color stimulus strength, and orientation stimulus strength. These regressors were dummy-coded such that, for each trial, there was only one regressor event. Regressors were defined as boxcars onsetting at the time of the cue and lasting for a duration equal to the sum of the cue duration and mean reaction time (RT), where applicable. To control for potentially confounding RT differences between contexts (54), we also included parametric regressors where the amplitude of the boxcar was determined by the z-scored RT on each trial. RT effects were modeled separately for each context and, in experiment 2, each stimulus strength. Condition regressors were convolved with a canonical difference-of-gammas model of the hemodynamic response function (HRF). We additionally included a set of identical regressors that were convolved with the temporal derivative of the HRF. After assembling and high-pass filtering the design matrix, the confound model that had been used to denoise the data was regressed out. The final deconvolution model was fit to the preprocessed time series from each voxel using ordinary least squares, separately for each run, producing a matrix of parameter estimates across conditions and voxels.

Multivariate analyses were performed using L2 penalized logistic regression models (55, 56) trained in a binary classification problem to predict task context (motion vs. color trials in experiment 1 and orientation vs. color trials in experiment 2). Specifically, this involved minimizing the following cost function:

$$\min_{\mathbf{w}, c} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \log \left\{ \exp[-y_i (\mathbf{X}_i^T \mathbf{w} + c)] + 1 \right\},$$

where y_i is the binary label for condition i , \mathbf{X}_i is the sample vector of signal amplitude across the n voxels in the ROI, and C is a penalty parameter, which we set to the default value of 1.

We evaluated the accuracy of the decoding models using leave-one-run-out cross-validation. The cross-validation procedure was as follows: On each fold, we z-scored the training and test sets, regressed out any residual relationship with RT, fit the classifier to the training set, and evaluated its performance on the test set. We used the mean and SD of each feature in the training set to z-score the features in both the training and testing sets. RT effects were removed by averaging the RT across trials corresponding to each classifier sample. These vectors were then z-scored using the mean and SD in the training set. We regressed each feature in the training set onto the training set RT vector and used the resulting parameter estimates to residualize both the training and test set data with respect to RT.

We assessed the significance of the classification accuracy scores in a permutation framework where we randomly shuffled the context labels, within run, and then retrained and tested the classifier on each of 100 iterations to establish a distribution of accuracy scores under the null hypothesis (57). For group inferences, we subtracted the mean of each subject's null distribution from the observed accuracy value and then performed a one-sample t test against 0. For subject inferences, we used the percentile of the observed value in the null distribution to obtain a P value.

The weights that are estimated when training the decoding model are not directly interpretable, but they can be made interpretable by multiplying the weight vector \mathbf{w} with the data covariance matrix, Σ_X , to produce a vector of context preferences (58), $\mathbf{a} = \Sigma_X \mathbf{w}$. The resulting vector, \mathbf{a} , is a distributed pattern corresponding to each voxel's relative preference for the two contexts. To identify voxels with strong context preferences, we compared each voxel's preference value, a_j , to the vector of preferences assigned to voxel j in the permutation test. Voxels where the observed value fell below the

10th percentile or above the 90th percentile in the null distribution were used in subsequent analyses. Note that this procedure uses a liberal threshold, as our goal was not to reject a null hypothesis in any individual voxel but rather to identify a suitably large population of voxels that expressed relatively strong preferences for each of the contexts. Our results did not depend on the specific threshold chosen.

To evaluate the spatial organization of the context representations, we determined how well we could predict the preference of a given voxel from the preferences of its neighbors at a given radius. Specifically, we quantified the error at radius i as the mean-squared difference between the preference in each of the n ROI voxels and the mean of the m voxels that were separated from it by a distance within 2 mm of the specified radius,

$$e_i = \frac{1}{n} \sum_{j=1}^n \left(a_j - \frac{1}{m} \sum_{k=1}^m a_k \right)^2.$$

We then compared the observed error value to a null distribution of errors obtained by randomly shuffling the voxel preferences. To quantify the clustering, we estimated an upper bound for the neighborhood size as the first point at which the observed error values were greater than the mean of the null distribution at that distance.

Estimation of Spontaneous Correlation Strength. We estimated the magnitude of spontaneous correlations in both task-engaged and resting states. Spontaneous activity in a task-engaged state was estimated using residual data after regressing out a model of task-evoked effects (12, 27). Specifically, the task models were structured identically to the models used during the deconvolution step of the decoding analyses, with the difference that, rather than convolving regressors with a canonical HRF model, we used a finite-impulse-response (FIR) basis to more flexibly fit and remove task-evoked responses. In experiment 1, where the design was oversampled relative to the data acquisition, we first upsampled the time series data to 1-s resolution using cubic spline interpolation. The sampling rate of the design of experiment 2 was matched to that of the data acquisition and so no resampling was applied. In each case, we used 24 FIR regressors per condition, onsetting two TRs before the cue. These models were fit using ordinary least squares to denoised time series data that had been concatenated across runs. Analyses of spontaneous correlations in the resting state were performed on the entire time series of denoised data. Spontaneous correlation matrices were created for both datasets by estimating the Pearson correlation between the time series from each voxel, separately within each run, and then averaging correlation values over runs.

We used the data from experiment 2, where we collected both task-based and resting-state data from the same subjects, to evaluate the similarity of the spontaneous correlation structure under these two conditions. Specifically, we vectorized the upper triangles of the two correlation matrices and then computed a Pearson correlation between the two vectors. We tested the significance of this value with the Mantel test, where, on each of 100 iterations, we applied the same permutation to the rows and columns of the resting-state correlation matrix and then recomputed the similarity to the task-engaged correlation matrix to generate a null distribution of similarity values.

Relating Context Preferences to Spontaneous Correlations. To visualize the relationship between context preference and spontaneous correlation, we submitted the spontaneous correlation matrix to metric MDS, which embeds the similarity matrix in a 2D space. We then created scatterplots where each point corresponded to a voxel. The position of the points was determined by the MDS embedding coordinates, and the color of the points was determined by their context preferences. Because the axes returned by MDS are arbitrary, we computed the matrix–vector product between the MDS coordinates and the voxel preferences and then applied a rotation so that the resulting coordinate fell on the positive side of the x axis. We determined the amount of variance explained by the low-dimensional projection by computing the squared Pearson correlation between the pairwise correlation distance measurements submitted to the MDS algorithm and the pairwise euclidean distance of points in the resulting embedding space.

To quantify the relationship between spontaneous correlations and context preferences, we computed the mean correlation between voxels with the same context preference and the mean correlation between voxels with different context preferences. We then evaluated the significance of this value with a permutation test. On each of 100 iterations, we randomly shuffled the context preference labels and recomputed the two measures. We then compared the observed difference in correlations to the distribution of differences under the null hypothesis.

To determine the spatial extent of the functionally selective spontaneous correlations, we recomputed the above measures while systematically excluding correlations between voxels that were separated by less than a specified distance threshold. Note that this process is similar to how we evaluated the spatial clustering of the voxel preferences, but here we used all voxels separated by a distance larger than the threshold rather than voxels situated at that specific distance.

Code Availability. Data were analyzed using published open-source software and custom code written in Python. Imaging data were processed with a workflow of FSL 5.0.8 (59) and FreeSurfer 5.3.0 (60) tools implemented in Nipype 0.11.0 (61). The Python code used a number of libraries including

- Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1(1):1–47.
- Glasser MF, et al. (2016) A multi-modal parcellation of human cerebral cortex. *Nature* 536(7615):171–178.
- Goldman-Rakic PS (1988) Topography of cognition: Parallel distributed networks in primate association cortex. *Annu Rev Neurosci* 11:137–156.
- Yeo BTT, et al. (2011) The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol* 106(3):1125–1165.
- Power JD, et al. (2011) Functional network organization of the human brain. *Neuron* 72(4):665–678.
- Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195(1):215–243.
- Wandell BA, Winawer J (2011) Imaging retinotopic maps in the human brain. *Vision Res* 51(7):718–737.
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202.
- Duncan J (2013) The structure of cognition: Attentional episodes in mind and brain. *Neuron* 80(1):35–50.
- Arieli A, Sterkin A, Grinvald A, Aertens A (1996) Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science* 273(5283):1868–1871.
- Ringach DL (2009) Spontaneous and driven cortical activity: Implications for computation. *Curr Opin Neurobiol* 19(4):439–444.
- Cole MW, Bassett DS, Power JD, Braver TS, Petersen SE (2014) Intrinsic and task-evoked network architectures of the human brain. *Neuron* 83(1):238–251.
- Power JD, Schlaggar BL, Petersen SE (2014) Studying brain organization via spontaneous fMRI signal. *Neuron* 84(4):681–696.
- Foster BL, et al. (2016) Spontaneous neural dynamics and multi-scale network organization. *Front Syst Neurosci* 10:7.
- Vincent JL, et al. (2007) Intrinsic functional architecture in the anaesthetized monkey brain. *Nature* 447(7140):83–86.
- Wig GS, Laumann TO, Petersen SE (2014) An approach for parcellating human cortical areas using resting-state correlations. *Neuroimage* 93(Pt 2):276–291.
- Kiani R, et al. (2015) Natural grouping of neural responses reveals spatially segregated clusters in prearcuate cortex. *Neuron* 85(6):1359–1373.
- Gordon EM, et al. (2016) Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb Cortex* 26(1):288–303.
- Arcaro MJ, Honey CJ, Mruczek REB, Kastner S, Hasson U (2015) Widespread correlation patterns of fMRI signal across visual cortex reflect eccentricity organization. *eLife* 4:e03952.
- Kenet T, Bibitchkov D, Tsodyks M, Grinvald A, Arieli A (2003) Spontaneously emerging cortical representations of visual attributes. *Nature* 425(6961):954–956.
- Cole MW, Etzel JA, Zacks JM, Schneider W, Braver TS (2011) Rapid transfer of abstract rules to novel contexts in human lateral prefrontal cortex. *Front Hum Neurosci* 5:142.
- Woolgar A, Hampshire A, Thompson R, Duncan J (2011) Adaptive coding of task-relevant information in human frontoparietal cortex. *J Neurosci* 31(41):14592–14599.
- Zhang J, Kriegeskorte N, Carlin JD, Rowe JB (2013) Choosing the rules: Distinct and overlapping frontoparietal representations of task rules for perceptual decisions. *J Neurosci* 33(29):11852–11862.
- Waskom ML, Kumaran D, Gordon AM, Rissman J, Wagner AD (2014) Frontoparietal representations of task context support the flexible control of goal-directed cognition. *J Neurosci* 34(32):10743–10755.
- Waskom ML, Frank MC, Wagner AD (January 4, 2016) Adaptive engagement of cognitive control in context-dependent decision making. *Cereb Cortex*, 10.1093/cercor/bhv333.
- Shenhav A, Botvinick MM, Cohen JD (2013) The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron* 79(2):217–240.
- Al-Aidroos N, Said CP, Turk-Browne NB (2012) Top-down attention switches coupling between low-level and high-level areas of human visual cortex. *Proc Natl Acad Sci USA* 109(36):14675–14680.
- Buckner RL, Krienen FM, Yeo BTT (2013) Opportunities and limitations of intrinsic functional connectivity MRI. *Nat Neurosci* 16(7):832–837.
- Rougier NP, Noelle DC, Braver TS, Cohen JD, O'Reilly RC (2005) Prefrontal cortex and flexible cognitive control: Rules without symbols. *Proc Natl Acad Sci USA* 102(20):7338–7343.
- Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370(6485):140–143.
- numpy, scipy, matplotlib, seaborn, pandas, and jupyter. Cortical surface visualizations were created using pysurfer. All custom code will be made available at https://github.com/WagnerLabPapers/Waskom_PNAS_2017.
- ACKNOWLEDGMENTS.** We thank Serra Favila and Karen LaRocque for assistance with data collection and helpful conversations, Bob Dougherty for assistance with the high-resolution pulse sequence, and Roozbeh Kiani, Brett Foster, and Thomas Yeo for comments on an earlier version of the manuscript. This work was supported by the Stanford Center for Cognitive and Neurobiological Imaging and the Knut and Alice Wallenberg Foundation (Network Initiative on Culture, Brain, and Learning). M.L.W. was supported by a National Science Foundation Integrative Graduate Education and Research Traineeship under Award 0801700.
- Kohn A, Coen-Cagli R, Kanitscheider I, Pouget A (2016) Correlations and neuronal population information. *Annu Rev Neurosci* 39:237–256.
- Cowan N (2001) The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behav Brain Sci* 24(1):87–114, discussion 114–185.
- Engle RW, Tuoholski SW, Laughlin JE, Conway AR (1999) Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *J Exp Psychol Gen* 128(3):309–331.
- Collins A, Koechlin E (2012) Reasoning, learning, and creativity: Frontal lobe function and human decision-making. *PLoS Biol* 10(3):e1001293.
- Gratton C, Laumann TO, Gordon EM, Adeyemo B, Petersen SE (2016) Evidence for two independent factors that modify brain networks to meet task goals. *Cell Reports* 17(5):1276–1288.
- Gonzalez-Castillo J, et al. (2015) Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. *Proc Natl Acad Sci USA* 112(28):8762–8767.
- Sakai K, Passingham RE (2003) Prefrontal interactions reflect future task operations. *Nat Neurosci* 6(1):75–81.
- Heinzle J, Wenzel MA, Haynes J-D (2012) Visuomotor functional network topology predicts upcoming tasks. *J Neurosci* 32(29):9960–9968.
- Cole MW, et al. (2013) Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat Neurosci* 16(9):1348–1355.
- Pucak ML, Levitt JB, Lund JS, Lewis DA (1996) Patterns of intrinsic and associational circuitry in monkey prefrontal cortex. *J Comp Neurol* 376(4):614–630.
- Frank MJ, Loughry B, O'Reilly RC (2001) Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cogn Affect Behav Neurosci* 1(2):137–160.
- Haynes J-D (2015) A primer on pattern-based approaches to fMRI: Principles, pitfalls, and perspectives. *Neuron* 87(2):257–270.
- Smith SM, et al. (2009) Correspondence of the brain's functional architecture during activation and rest. *Proc Natl Acad Sci USA* 106(31):13040–13045.
- Tavor I, et al. (2016) Task-free MRI predicts individual differences in brain activity during task performance. *Science* 352(6282):216–220.
- Cole MW, Ito T, Bassett DS, Schultz DH (2016) Activity flow over resting-state networks shapes cognitive task activations. *Nat Neurosci* 19(12):1718–1726.
- Laumann TO, et al. (2015) Functional system and areal organization of a highly sampled individual human brain. *Neuron* 87(3):657–670.
- Uğurbil K, et al.; WU-Minn HCP Consortium (2013) Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *Neuroimage* 80:80–104.
- Andersson JLR, Skare S, Ashburner J (2003) How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *Neuroimage* 20(2):870–888.
- Behzadi Y, Restom K, Liu J, Liu TT (2007) A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* 37(1):90–101.
- Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9(2):179–194.
- Greve DN, Fischl B (2009) Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* 48(1):63–72.
- Fischl B, Sereno MI, Tootell RB, Dale AM (1999) High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp* 8(4):272–284.
- Oosterhof NN, Wiestler T, Downing PE, Diedrichsen J (2011) A comparison of volume-based and surface-based multi-voxel pattern analysis. *Neuroimage* 56(2):593–600.
- Todd MT, Nystrom LE, Cohen JD (2013) Confounds in multivariate pattern analysis: Theory and rule representation case study. *Neuroimage* 77:157–165.
- Fan R-E, Chang K-W, Hsieh C-J, Wang X-R, Lin C-J (2008) LIBLINEAR: A library for large linear classification. *J Mach Learn Res* 9:1871–1874.
- Pedregosa F, et al. (2011) Scikit-learn: Machine learning in Python. *J Mach Learn Res* 12:2825–2830.
- Ojala M, Garriga G (2010) Permutation tests for studying classifier performance. *J Mach Learn Res* 11:1833–1863.
- Haufe S, et al. (2014) On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* 87:96–110.
- Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012) FSL. *Neuroimage* 62(2):782–790.
- Fischl B (2012) FreeSurfer. *Neuroimage* 62(2):774–781.
- Gorgolewski K, et al. (2011) Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Front Neuroinform* 5:13.