Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns

Javier Gonzalez-Castillo[a,1,2], Colin W. Hoy[a,b,1], Daniel A. Handwerker[a], Meghan E. Robinson[a,c], Laura C. Buchanan[a], Ziad S. Saad[d], and Peter A. Bandettini[a,e]

[a]Section on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892; [b]Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720; [c]Veterans Affairs Boston Healthcare System, Boston, MA 02130; [d]Statistical and Scientific Computing Core, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892; and [e]Functional MRI Facility, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892

Edited by Russell A. Poldrack, University of Texas at Austin, Austin, TX, and accepted by the Editorial Board June 1, 2015 (received for review January 21, 2015)

Functional connectivity (FC) patterns in functional MRI exhibit dynamic behavior on the scale of seconds, with rich spatiotemporal structure and limited sets of whole-brain, quasi-stable FC configurations (FC states) recurring across time and subjects. Based on previous evidence linking various aspects of cognition to group-level, minute-to-minute FC changes in localized connections, we hypothesized that whole-brain FC states may reflect the global, orchestrated dynamics of cognitive processing on the scale of seconds. To test this hypothesis, subjects were continuously scanned as they engaged in and transitioned between mental states dictated by tasks. FC states computed within windows as short as 22.5 s permitted robust tracking of cognition in single subjects with near perfect accuracy. Accuracy dropped markedly for subjects with the lowest task performance. Spatially restricting FC information decreased accuracy at short time scales, emphasizing the distributed nature of whole-brain FC dynamics, beyond univariate magnitude changes, as valuable markers of cognition.

fMRI | connectivity dynamics | functional connectivity states | cognitive states | classification

Resting state functional MRI (rs-fMRI) focuses on spatial patterns of blood oxygenation level dependent (BOLD) signal cofluctuations recorded in the absence of externally driven tasks or stimulation. These patterns, known as functional connectivity (FC) patterns, are usually computed on the basis of an entire scan (often >6 min). Their cognitive significance (1) and long term reproducibility (2) are well established, and preliminary data suggest that they have potential clinical value (3). However, recent studies have shown that FC patterns are highly dynamic at shorter temporal scales (4) (i.e., tens of seconds), adding yet another challenge to developing fMRI-based protocols with sufficient single-subject specificity and sensitivity to inform clinical decisions.

FC patterns computed with 1- to 2-min portions of a scan can vary substantially around a mean FC pattern obtained using complete 6- to 20-min scans. This dynamic behavior has been observed in awake and sleeping humans (5–8), as well as in anesthetized animals (9, 10). Several studies involving simultaneous fMRI and electrophysiological recordings have suggested that FC dynamics may be driven by neurophysiological sources rather than noise (6, 11, 12). Furthermore, FC dynamics exhibit rich spatiotemporal structure. Connections between higher order cognitive regions are more variable than those between primary sensorymotor regions (13–15), and a limited set of whole-brain, quasi-stable FC configurations–known as FC states–reliably recur both within and across subjects at rest (13, 16).

Given that cognition is supported by highly dynamic brain processes, it has been hypothesized that FC states may reflect changes in ongoing cognitive states during rest (13). Initial task-based studies have been able to differentiate between a limited set of mental tasks (17, 18) and arousal levels (19) on the basis of localized changes in FC at the scale of 45 s to 2 min. However, all these studies examined only a reduced set of predefined connections expected to best differentiate states of interest. In doing so, these studies neglect the highly distributed and parallel nature of cognitive processing (20, 21) and consciousness itself (22), thus only partially validating the cognitive significance of FC states as originally defined: quasi-stable, global representations of FC across the whole brain.

In the following study, we used a continuous, multitask paradigm to study the relationship between ongoing cognition and dynamic changes in FC patterns derived from BOLD measurements at different temporal scales. We use the term "ongoing cognition" to refer to brain activity that is the result of experimentally constrained cognitive processes, in contrast to resting state, which includes unconstrained ongoing cognition of an unknown, internally driven nature (23). Although giving subjects freedom to transition between states at will may better reflect the self-directed nature of cognition during rest, such an experimental design would lack ground truth regarding the timing and nature of cognitive states, thus precluding detailed evaluation of their relationship with FC states. Consequently, our experimental design uses tasks to control these variables and monitor subject performance via behavioral responses. This framework provides the basis for rigorously testing the degree to which FC states reflect ongoing cognition at short temporal scales and, in that manner, informs the interpretation of BOLD connectivity dynamics during both task and rest.

Significance

Recently, it was shown that functional connectivity patterns exhibit complex spatiotemporal dynamics at the scale of tens of seconds. Of particular interest is the observation of a limited set of quasi-stable, whole-brain, recurring configurations—commonly referred to as functional connectivity states (FC states)—hypothesized to reflect the continuous flux of cognitive processes. Here, to test this hypothesis, subjects were continuously scanned as they engaged in and transitioned between mental states dictated by tasks. We demonstrate that there is a strong relationship between FC states and ongoing cognition that permits accurate tracking of mental states in individual subjects. We also demonstrate how informative changes in connectivity are not restricted solely to those regions with sustained elevations in activity during task performance.

Author contributions: J.G.-C., C.W.H., D.A.H., M.E.R., and P.A.B. designed research; J.G.-C., C.W.H., M.E.R., and L.C.B. performed research; Z.S.S. contributed new reagents/analytic tools; J.G.-C., C.W.H., D.A.H., M.E.R., and L.C.B. analyzed data; and J.G.-C., C.W.H., D.A.H., Z.S.S., and P.A.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. R.A.P. is a Guest Editor invited by the Editorial Board.

Data deposition: The MRI data have been deposited in Xnat Central, https://central.xnat.org (project ID: FCStateClassif).

[1]J.G.-C. and C.W.H. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: javier.gonzalez-castillo@nih.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1501242112/-/DCSupplemental.

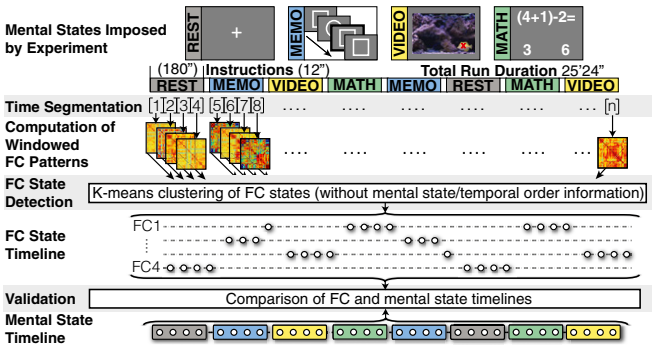


Fig. 1. Experimental paradigm and summary of analysis for multitask scans.

Participants were scanned continuously for ~25 min as they engaged and transitioned between four mental tasks: two-back working memory (memory), numerical computations (math), visual attention (video), and rest. Subjects engaged in each task for two nonconsecutive 3 min periods (Fig. 1). Responses and response times were obtained for the three active tasks. For each of 20 subjects, whole-brain FC patterns were computed for nonoverlapping windows of different durations (range 180 s to 22.5 s) and used to generate time lines of FC states. In this manuscript, the word FC state refers to a vector populated with connectivity values (i.e., Fisher's transform of the Pearson's correlation value) computed using only the portion of the data inside a given window. Because the Pearson's correlation removes the means and divides by the SDs of the input time series, windowed connectivity matrices computed this way are less prone to bias by shifts in activity levels across task blocks. Finally, these FC state time lines were then compared with mental state time lines defined by the timing of the experimental paradigm (see Fig. S1 for a detailed depiction of the analysis). We show that the structure imposed by these task-related demands, or ongoing cognition, can be recovered from short-term FC state dynamics at the single-subject level on the scale of tens of seconds with near perfect accuracy.

Results

Behavioral. Table 1 shows across-subject average values for percentage of correct responses ($P_{Correct}$), percentage of trials when a response was required but none was given ($P_{Missing}$), and reaction time (RT) for all task blocks. On average, subjects performed much better on the math and memory tasks ($P_{Correct} > 90\%$; $P_{Missing} < 20\%$) than they did on the video task ($P_{Correct} \approx 67\%$; $P_{Missing} \approx 30\%$). RTs were on average quite similar across blocks of the same task. Individualized measures of $P_{Correct}$, RT, and $P_{Missing}$ are reported in Fig. S2. This figure also shows the difference across same-task blocks for each of the metrics (i.e., $\Delta P_{Correct}$, ΔRT, and $\Delta P_{Missing}$), which informs us about performance consistency across the whole scan. Subjects S05, S12, and S14 had the worst overall performance in terms of $P_{Correct}$ (Fig. S2*A*) and $P_{Missing}$ (Fig. S2*E*), being the only ones with $P_{Missing} > 25\%$ and $P_{Correct} < 80\%$. They were also among those with the highest RTs (Fig. S2*C*), ΔRT (Fig. S2*D*), and $\Delta P_{Missing}$ (Fig. S2*F*). One of them in particular, S12, showed the highest levels of discrepancy for all three metrics, perhaps signaling an intermittent loss of concentration/awareness during the scan. Importantly for our discussion, these three subjects also had markedly degraded performance in terms of FC state-based classification (Figs. 2 and 3).

FC-Based Classification. Fig. 2 shows group-level, FC-based classification results in terms of the adjusted rand index (ARI). The ARI is a clustering validation metric that quantifies the level of agreement between a data-driven clustering (e.g., FC states + *k*-means) and existing knowledge of the underlying structure of the data (i.e., window groupings based on mental tasks) while also correcting for chance (24). Interpretation of the ARI is well established in the literature (25): ARI < 0.65 signals poor recovery of the underlying group structure of the data; 0.65 < ARI < 0.8 indicates moderate recovery; 0.8 < ARI < 0.9 indicates good recovery; and 0.9 < ARI < 1.0 indicates excellent recovery. Group-level ARI results are reported in Fig. 2 in terms of the median (white dot), 25–75% percentiles (gray box), and most extreme data points not considered outliers (dotted whiskers). In addition, outliers are marked with a (+) symbol in Fig. 2. For window length (WL) ≥ 30 s, the median ARI equals 1, signaling that we were able to perfectly group windows based on whole-brain FC snapshots. For WL = 22.5 s, the median ARI decreases slightly, but it is still within the excellent recovery range (green). Finally, only five subjects were marked as outliers with classification accuracies well below the rest of the group. Subjects S03 and S08 were outliers only for one of the various WLs, yet their ARI remained within the good recovery zone. Conversely, subjects S05, S12, and S14 were outliers for several WLs, and their ARI dropped in the moderate (subject S05) or poor recovery zone (subjects S12 and S14).

Fig. 3 shows classification results for six representative subjects at all WLs in the form of "classification staffs" (results for all 14 additional subjects can be found in Figs. S3 and S4). In each "staff," the *x* axis corresponds to time (in units of windows), and the *y* axis to FC states. Each time window is represented by a color-coded bar and a dot. The color of the bar signals the imposed mental state (gray, rest; blue, memory; green, math; yellow, video). The location of the dot on the *y* axis signals the FC state to which that window was assigned. Agreements between groupings based on mental state and FC state are marked with black dots, and errors are marked with red dots. In addition, for each subject, we report two measures of classification success (classification accuracy and ARI) to the right of the staff. $P_{Correct}$, RT, and $P_{Missing}$ for each task block and subject are reported in Table S1.

Fig. 3*A* shows results for subject S01, a representative non-outlier subject. No classification errors occurred for this subject. Fig. 3 *B*–*F* shows results for the five outliers reported above. Fig. 3*B* shows results for subject S03, outlier at WL = 30 s, due to two errors (first rest and seventh video windows). These two errors at WL = 30 s were sufficient to push the ARI down to the good recovery zone although accuracy remained above 95%. Two additional errors occurred for this subject at windows at the edge of task blocks (transition windows) for WL = 22.5 s. Fig. 3*D* shows results for subject S08, outlier at WL = 60 s, due to a single misclassification (last rest window). One more error occurred on the same rest window for WL = 30 s.

Fig. 3*C* shows results for subject S05, outlier for WL = 60 s, 45 s, 30 s, and 22.5 s. For all these windows, the ARI fell within the moderate recovery range. All but one misclassification involved grouping of rest and memory windows together (red line). This subject had the largest $P_{Missing}$ for the memory task (along with

Table 1. Average behavioral measures across all subjects

	Memory-B1	Memory-B2	Math-B1	Math-B2	Video-B1	Video-B2
$P_{Correct}$, %	91.92 ± 7.20	93.25 ± 6.65	95.69 ± 4.64	91.11 ± 9.30	66.25 ± 18.18	67.19 ± 21.92
$P_{Missing}$, %	12.75 ± 15.85	18.33 ± 24.27	0.83 ± 1.82	2.50 ± 4.84	30.62 ± 16.95	31.25 ± 21.65
RT, s	0.94 ± 0.37	1.05 ± 0.58	2.31 ± 0.45	2.55 ± 0.51	1.39 ± 0.17	1.37 ± 0.22

B, block.

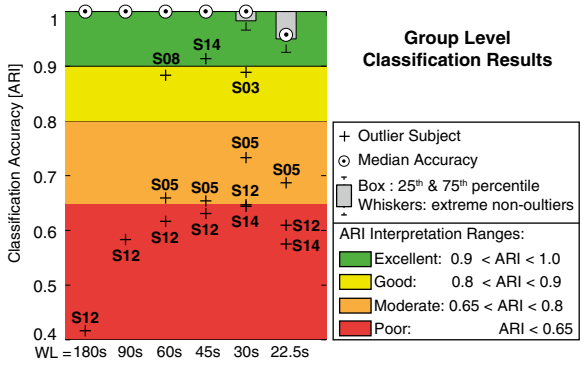


Fig. 2. Group-level FC state classification results.

subject S14) (Fig. S2*E*) and had the second worst $P_{Correct}$ for this task (Fig. S2*A*).

Fig. 3*F* shows results for subject S14, outlier for WL = 45 s, 30 s, and 22.5 s. At these WLs, all errors but one were related to confusion with the memory task, mostly (26 of 30) with rest windows (red line). Subject S14's ARI for WL = 30 s and 22.5 s lies in the poor recovery zone. Behaviorally, subject S14 had the largest $P_{Missing}$ (tied with subject S05) for the memory task.

Finally, Fig. 3*E* shows results for subject S12, outlier for all window lengths. Subject S12 had the worst classification of the group, with ARIs in the poor recovery zone for all WLs. According to all three behavioral metrics, this subject was the most inconsistent across blocks and was also among the four worst subjects in terms of task performance. Across all WLs, 70 of 73 misclassifications involved confusion with the video task (red lines). Subject S12's performance was low and variable during the video blocks, as evidenced by having the lowest $P_{Correct}$, largest $\Delta P_{Missing}$ and $\Delta P_{Correct}$, and second largest $P_{Missing}$ and ΔRT for this task.

FC-Based Classification Accuracy vs. Behavior. Scatter plots of classification accuracy (ARI) versus each of the six behavioral indices are shown in Fig. 4 for WL = 22.5 s. In each plot, subjects are represented as gray circles. A linear fit to the data is shown (dotted line), and correlation values and their significance (*P* value) are reported. We found significant correlations between ARI and all behavioral metrics for this window length, as well as for WL = 30 s, 45 s, and 60 s. When the three worst performers (subjects S05, S12, and S14) were excluded from this analysis, the correlations were no longer significant although this negative result may be partly due to ceiling effects on the ARI (all remaining subjects had ARI ≥ 0.88 and 9 of 17 had ARI = 1). This observation suggests that large deviations in task performance are required to produce substantial errors in FC classification.

Contribution of Task-Specific Regions to Classification. To determine the overlap between the connectivity changes driving the FC-based classification and the locations of activation-induced changes in BOLD magnitude, we sorted regions of interest (ROIs) according to how well they differentiated the tasks in terms of activation levels (see *SI Methods* for details). We subsequently attempted FC-based classification using progressively smaller sets of ROIs. In one analysis, we removed the most task-discriminative ROIs first. In a second parallel analysis, we removed ROIs in the opposite order (least task-discriminative first).

Fig. S5*A* shows activation maps for the contrasts between the three active tasks and rest. Although maps do differ, in all instances, activation foci were present in visual cortex, anterior insula, inferior

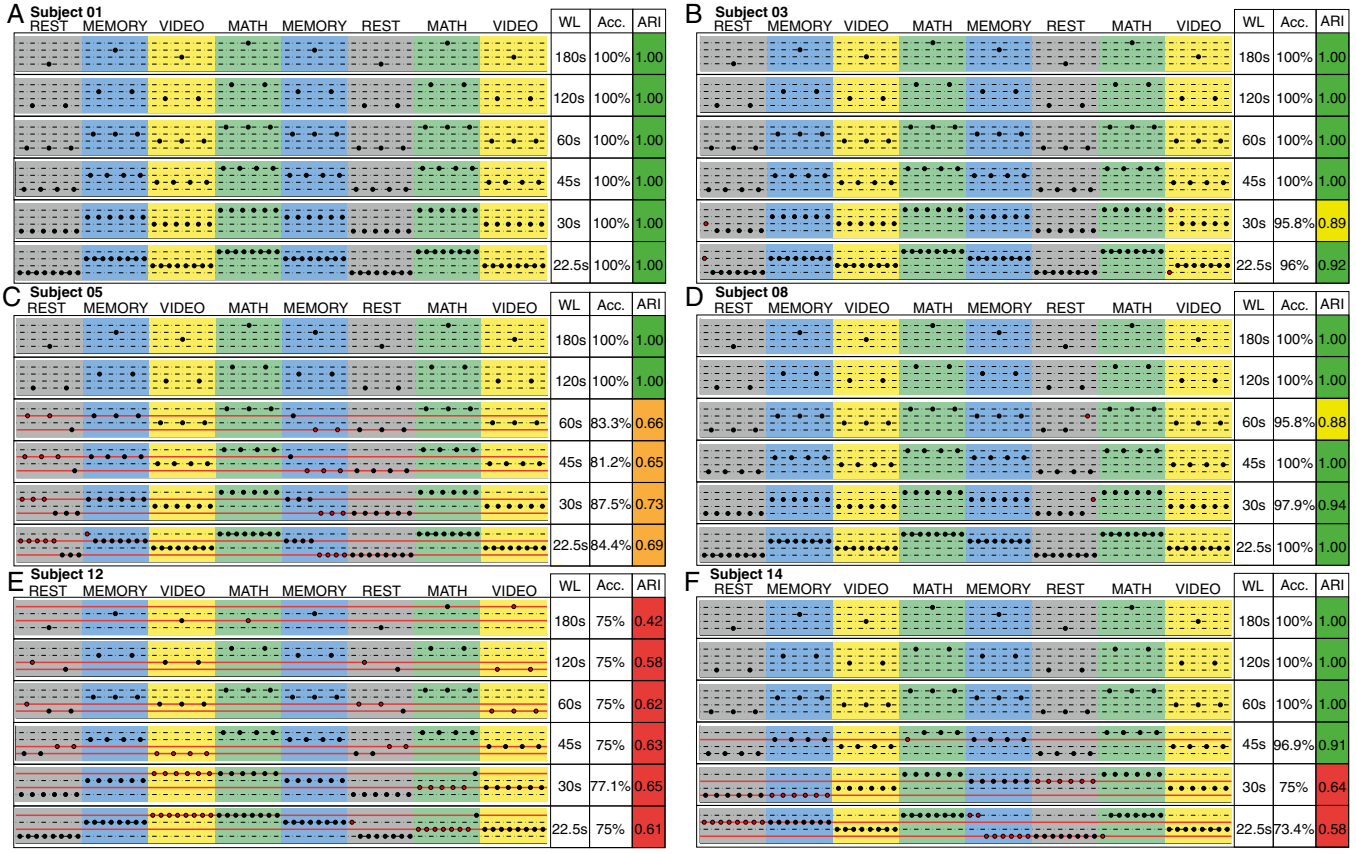


Fig. 3. Individual level FC state classification results. (*A*) Representative nonoutlier subject (S01). (*B–F*) Outlier subjects for one or more WLs.

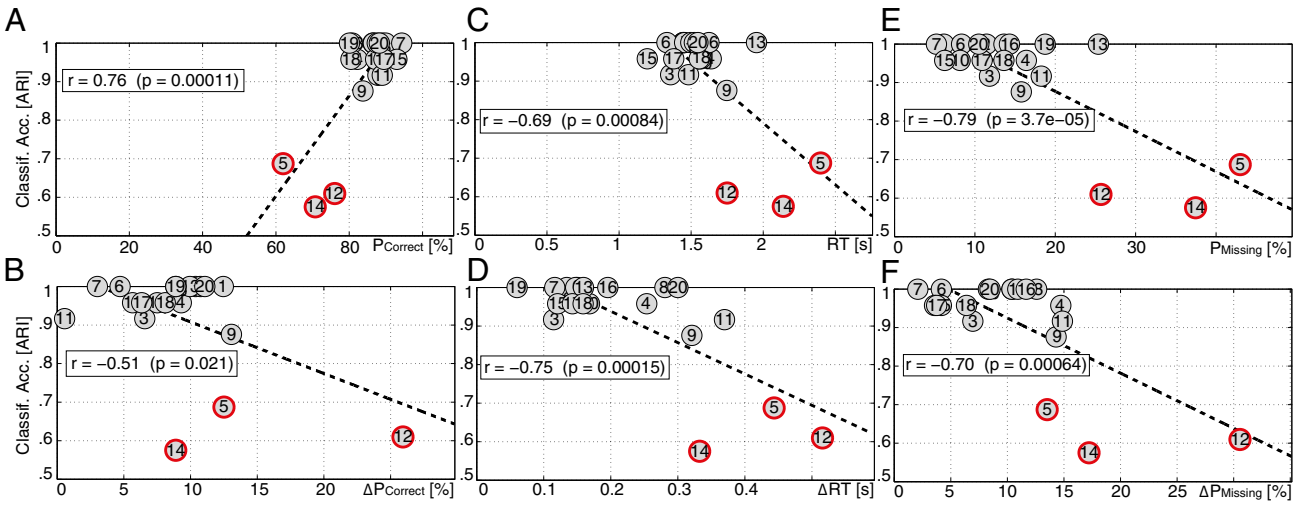


Fig. 4. Behavior versus FC classification (WL = 22.5 s). ARI outliers at multiple WLs marked in red. (*A*) ARI vs. $P_{Correct}$. (*B*) ARI vs. $\Delta P_{Correct}$. (*C*) ARI vs. RT. (*D*) ARI vs. ΔRT. (*E*) ARI vs. $P_{Missing}$. (*F*) ARI vs. $\Delta P_{Missing}$.

parietal lobule, thalamus, and both medial and lateral portions of the premotor cortex. Fig. S5*B* shows contrast maps between the active tasks. A more limited set of regions is observed for these higher order contrasts, with differences concentrated mainly in occipital cortex, posterior thalamus, and inferior frontal regions. The largest differences occurred for the memory vs. video contrast. Fig. S5*C* shows all 157 ROIs color-coded according to their activation-based discriminative power. Most discriminative ROIs are depicted in warmer colors (yellow to red); cooler colors (cyan to dark blue) are used for the least discriminative. All regions from the high order contrasts (Fig. S5*B*) and many regions from rest contrast (Fig. S5*A*) appear colored in green or warmer colors (high ranks) in the map.

Fig. 5 shows results after selective removal of ROIs for two representative windows (WL = 60 s, WL = 22.5 s). Independent of WL or exclusion order, classification accuracy decreases monotonically as the number of discarded ROIs increases. The rate of decrease in accuracy is faster when most discriminative ROIs are removed first, yet removal of a limited set of least discriminative ROIs can also degrade classification. This last observation is most apparent for shorter WLs. Classification remained above poor levels despite removal of 50% of ROIs. For WL = 60 s, classification was excellent even with the removal of 40% of the ROIs that best discriminate the tasks. This result suggests that an FC state is better described by the state of wide spread connections across the brain, rather than by the considerably smaller set of connections between regions whose overall activity changes the most across the tasks under study.

Influence of Analytical Decisions on the Results. Knowledge about ongoing mental states allowed us to evaluate our methods for FC-based classification. We tested the effect of atlas size, level of dimensionality reduction, and clustering algorithm. We performed FC classification with seven versions of the Craddock atlas (26) (range: 30–500 ROIs), five levels of dimensionality reduction (keep all, 97.5%, 95%, 90%, or 75% of the variance), and two clustering algorithms (*k*-means and hierarchical clustering). We found that methodological decisions can heavily influence classification results, especially for the shorter WLs. In particular we found the following: (*i*) When selecting the atlas, it is best to use a larger set of small ROIs rather than a smaller set of large ROIs; (*ii*) discarding a small amount of variability from the correlation matrix via principal component analysis (PCA) can help the clustering algorithm substantially, yet removing too much variance can be damaging; and (*iii*) *k*-means outperformed hierarchical clustering in all scenarios. Fig. S6 shows how ARI changed across all these conditions in detail.

Discussion

In this study, we scanned a group of subjects continuously as they engaged in and transitioned between a series of well-defined tasks that created distinct mental states. We demonstrate that there is a strong relationship between FC states and ongoing cognition that can be detected in individual subjects for windows ranging in duration from 180 s to 22.5 s. Moreover, for shorter windows (WL ≤ 60 s), we found significant correlations between classification performance and behavioral metrics of performance and variability. Finally, by selectively excluding ROIs from the classification based on their level of task-specific activation, we also show that informative changes in connectivity are widely distributed across the brain and not restricted to the few regions with sustained elevations in activity during the tasks. Importantly, this last point implies that inferences based solely on maps of relative magnitude changes may be missing valuable information embedded in a wider distribution of connectivity patterns that change without concomitant magnitude changes, thereby leaving our understanding of specific tasks and states incomplete. These results fit with prior studies showing how system-restricted (i.e., not whole-brain) connectivity-based classification of motor (27) and emotional

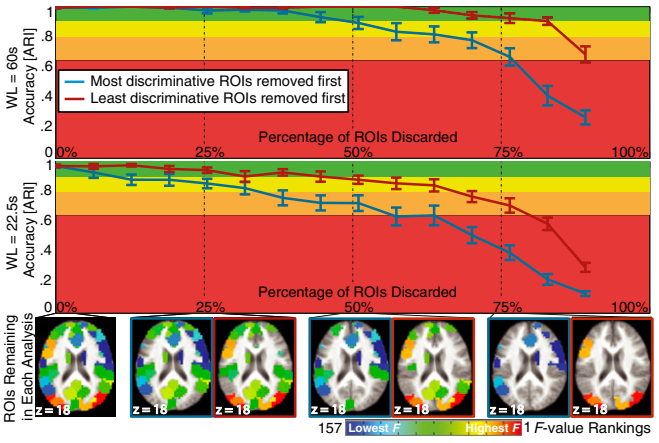


Fig. 5. Change in accuracy when ROIs are selectively removed from the analysis for WL = 60 s (*Top*) and WL = 22.5 s (*Bottom*). Below the plots we show ROIs entering analyses when 0%, 25%, 50%, or 75% of ROIs are removed. Blue frames denote ROIs entering the analysis when most discriminative ROIs are removed first, and red frames denote the complementary analyses.

NEUROSCIENCE

task (28) content outperforms activity-based classification, and extend this finding to whole-brain FC patterns. Overall, we show that short-term fluctuations in whole-brain fMRI connectivity patterns can be reliably used to track ongoing cognition, on an individual subject basis, despite the noisy and indirect nature of BOLD signals as a marker of neuronal activity.

Our results extend previous attempts at tracking cognition on the basis of short-term fMRI FC patterns (17–19, 29–31) in several ways: (*i*) We report higher accuracy for shorter windows, which suggests that FC states follow cognition at much shorter temporal scales than previously reported; (*ii*) we do so in individual subjects without a priori selection of informative connections or training datasets, which makes this approach suitable for studying a wide range of cognitive states and for single-subject clinical applications; (*iii*) we demonstrate that FC states are most identifiable when defined as global, whole-brain phenomena; and (*iv*) we report additional analyses aimed to devise the best methods for tracking cognition using FC states.

Best results from prior studies showed classification accuracies around 80% for durations of 60 s or more (17, 19). For shorter windows, when reported, accuracy dropped quickly. Here, we report accuracies well above 80% for windows as short as 22.5 s (Fig. 3), which approaches the upper limit of temporal scales (256 ms to 16 s) for which EEG microstate (32) sequences show scale-free behavior (33). Previous research already had limited success matching temporal sequences of EEG microstates to rs-fMRI networks, despite the contrast between their dynamic and stationary definitions (34–36). In addition to the better conceptual match between microstates and FC states as quasi-stable, global brain phenomena that are in flux over short time periods, our results show that FC states also share the functional utility of microstates in terms of tracking ongoing cognition at behaviorally relevant time scales. These findings suggest that examining these two phenomena in conjunction may help elucidate the relationship between dynamics in fMRI and electrophysiological signals.

The original observation of FC states by Allen et al. (13) relied on whole-brain connectivity data and *k*-means, a data-mining technique that does not require a preexisting training dataset. However, all attempts at validating the concept of FC states as a correlate of cognition (17–19, 29, 37) have relied upon supervised techniques that required a training dataset and/or some a priori selection of connections based on the mental states under scrutiny, both of which limit the generalizability of those approaches. To our knowledge, ours is the first study to show a strong correlation between FC and mental states without the need of any such priors. Unsupervised approaches such as *k*-means are preferable because they are independent of mental state and suitable for single-subject applications, but they have their own set of limitations. First, most of them require a priori selection of the number of states (*k*). We were able to develop robust classification procedures based on the ground truth for *k* established by the experimental paradigm, but obtaining the correct number of FC states in an unbiased, data-driven manner is an important challenge for the field. Second, unsupervised classification techniques simply group similar objects (e.g., whole-brain FC patterns) into sets. They cannot assign any label or meaning (e.g., mental computations) to the discovered groups. In other words, unsupervised methods can reveal when subjects are engaged in the same mental state, but not what state that is. Fortunately, once proven that FC states consistently represent cognitive states within subjects, one could envision a second step in which detected FC states are compared against dictionaries of FC patterns for which a label does exist. For such a dual-step approach to work, we first need to understand the conditions under which FC states faithfully represent cognitive states, particularly across subjects and for internally, self-induced mental states such as those present in rest. This topic is not directly addressed in this work, but previous studies have demonstrated successful group-level FC-based classification of more loosely controlled cognitive states without sensory stimulation or motor responses (17, 37). Although FC/cognitive state dictionaries are not yet available, equivalent dictionaries for activity-based patterns are being constructed today (38). We believe that better characterization of across-subject consistency in FC states, combined with parallel efforts to develop such FC-based dictionaries, may provide the means to track ongoing cognition during rest.

Several factors may have contributed to obtaining such high classification results for shorter temporal scales than previously reported. First, data were acquired on a high-field scanner, which translates into better signal-to-noise ratio. Second, an adaptive high-pass filtering scheme was used to avoid spurious fluctuations in correlation values (39). In fact, when such filtering is not in place, classification notably degraded for the shorter windows (Fig. S6*D*). Third, we used a finer grained parcellation of the brain than previous studies (200 ROIs vs. ∼100 ROIs). Reanalysis of the data with versions of the Craddock atlas (26) ranging in number of ROIs from 30 to 500 (Fig. S6*A*) suggests that best results are obtained with a minimum of 200 ROIs. Fourth, we used distinctive tasks in terms of their cognitive load, response patterns, and difficulty, which may have enhanced our ability to discern between cognitive states. Fifth, most previous studies used supervised classification algorithms and evaluated their results using cross-validation across subjects and/or runs. Here, because each subject was analyzed separately with unsupervised methods, an external evaluation metric (ARI) was used to evaluate each subject individually. Differences in the evaluation framework may also contribute to some differences in classification results. Finally, considering whole-brain connectivity patterns as opposed to just a subset of connections may have also provided an additional boost in accuracy for the shortest windows. Results from the "selective exclusion of ROIs" analyses suggest that valuable information for tracking cognition is spatially distributed across the whole brain. Particularly at shorter time windows, removing a limited set of regions/connections can cause a substantial reduction in classification accuracy (Fig. 5, *Bottom Plot*), even if those connections do not show sustained task activations. These findings reemphasize the original definition of FC states as whole-brain phenomena (13, 18) and fit current theories of consciousness that characterize mental states in terms of global access to widely distributed information (40) and highlight the irreducible nature of information integrated across large-scale networks (41).

Additional analyses were conducted to understand what drives the classification. First, to ensure that across-task changes in activation levels or signal variability were not responsible, we attempted classification based on average signal levels and SDs of ROI representative time series. Second, we phase-randomized each ROI time series before computing FC states to ensure that time-varying relationships were needed for accurate classification. Third, we also randomized FC state features (i.e., scrambled the strength of individual connections) to ensure classification was not driven simply by overall across-task changes in connectivity levels. For all these analyses, the ARI dropped into the poor recovery zone for all WLs in all subjects, except for intensity-based classification with WL = 180 s, for which the ARI fell into the moderate recovery zone but was still well below FC-based classification rates (Fig. S7 *B*, *C*, *E*, and *F*). Finally, to determine whether univariate task activations were necessary for classification, we regressed out the task effects as a nuisance variable in preprocessing and found that the ARI remained excellent (Fig. S7*D*). These control analyses suggest that classification is driven by orchestrated changes in connectivity across tasks for all window lengths, especially for WL < 180 s, and that first-order activations were neither necessary nor sufficient for robust classification at short time scales.

Additionally, a series of analyses were conducted to evaluate our preprocessing pipeline (Fig. S6). Our results suggest using a finer-grained atlas (more smaller ROIs is better than fewer larger ROIs), a conservative approach to discarding variance (keeping 97.5% of variance produced the best results), and *k*-means over hierarchical clustering. Most importantly, these results suggest that methodological decisions can influence the strength of measured relationships between FC and mental states, meaning that

caution should be exercised when interpreting results in situations with no ground truth (e.g., resting state).

The brain is an inherently dynamic and distributed system. Although valuable information can be obtained by studying its stable characteristics (e.g., anatomical or stationary rs-fMRI networks), a full understanding will require exploration of its dynamic behavior at different spatial and temporal scales. The strong correspondence between FC states and mental states reported here suggests that the detailed study of FC states may provide novel insights into system-level behaviors of the human brain. Moreover, the rich structure in these dynamics (e.g., number of states, dwell-time, etc.) may be a more sensitive marker for mental conditions than metrics about stable characteristics of the brain (16). Preliminary research has already revealed differences in dwell time between controls and both Alzheimer's disease patients (42) and schizophrenics (43). In two recent reviews on resting state dynamics (4, 16), it was acknowledged that a better understanding of the relationship between BOLD dynamics and behavior was still needed. The work summarized here represents an important step in that direction by showing how fluctuations in FC states, when computed appropriately, are directly related to ongoing cognition in individual subjects. We believe that future research on FC states in more experimental conditions and populations will help reveal the most cognitively and clinically meaningful ways to observe, describe, and quantify the dynamics of FC states.

Methods

Twenty-two participants were scanned continuously as they engaged in and transitioned between four mental tasks after giving informed consent in compliance with a protocol approved by the Institutional Review Board of the National Institute of Mental Health in Bethesda, MD. After data preprocessing, whole-brain FC patterns (based on Pearson's correlation matrices) were computed for nonoverlapping windows of different durations and used to generate time lines of FC states. These FC-state time lines were then compared with mental-state time lines defined by the timing of the experimental paradigm. Detailed methods are provided in *SI Methods*.

ACKNOWLEDGMENTS. This research was possible thanks to the support of the National Institute of Mental Health Intramural Research Program. Portions of this study used the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (biowulf.nih.gov). This study is part of NIH clinical protocol number NCT00001360 and protocol ID 93-M-0170.

1. 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.
2. Chou YH, Panych LP, Dickey CC, Petrella JR, Chen NK (2012) Investigation of long-term reproducibility of intrinsic connectivity network mapping: A resting-state fMRI study. *AJNR Am J Neuroradiol* 33(5):833–838.
3. Castellanos FX, Di Martino A, Craddock RC, Mehta AD, Milham MP (2013) Clinical applications of the functional connectome. *Neuroimage* 80:527–540.
4. Hutchison RM, et al. (2013) Dynamic functional connectivity: Promise, issues, and interpretations. *Neuroimage* 80:360–378.
5. Chang C, Glover GH (2010) Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage* 50(1):81–98.
6. Tagliazucchi E, von Wegner F, Morzelewski A, Brodbeck V, Laufs H (2012) Dynamic BOLD functional connectivity in humans and its electrophysiological correlates. *Front Hum Neurosci* 6(December):339.
7. Smith SM, et al. (2012) Temporally-independent functional modes of spontaneous brain activity. *Proc Natl Acad Sci USA* 109(8):3131–3136.
8. Handwerker DA, Roopchansingh V, Gonzalez-Castillo J, Bandettini PA (2012) Periodic changes in fMRI connectivity. *Neuroimage* 63(3):1712–1719.
9. Hutchison RM, Womelsdorf T, Gati JS, Everling S, Menon RS (2013) Resting-state networks show dynamic functional connectivity in awake humans and anesthetized macaques. *Hum Brain Mapp* 34(9):2154–2177.
10. Keilholz SD, Magnuson ME, Pan W-J, Willis M, Thompson GJ (2013) Dynamic properties of functional connectivity in the rodent. *Brain Connect* 3(1):31–40.
11. Chang C, Liu Z, Chen MC, Liu X, Duyn JH (2013) EEG correlates of time-varying BOLD functional connectivity. *Neuroimage* 72:227–236.
12. Thompson GJ, et al. (2013) Neural correlates of time-varying functional connectivity in the rat. *Neuroimage* 83:826–836.
13. Allen EA, et al. (2014) Tracking whole-brain connectivity dynamics in the resting state. *Cereb Cortex* 24(3):663–676.
14. Bassett DS, et al. (2013) Robust detection of dynamic community structure in networks. *Chaos* 23(1):013142.
15. Gonzalez-Castillo J, et al. (2014) The spatial structure of resting state connectivity stability on the scale of minutes. *Front Neurosci* 8(June):138.
16. Calhoun VD, Miller R, Pearlson G, Adalı T (2014) The chronnectome: Time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron* 84(2):262–274.
17. Shirer WR, Ryali S, Rykhlevskaia E, Menon V, Greicius MD (2012) Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cereb Cortex* 22(1):158–165.
18. Richiardi J, Eryilmaz H, Schwartz S, Vuilleumier P, Van De Ville D (2011) Decoding brain states from fMRI connectivity graphs. *Neuroimage* 56(2):616–626.
19. Tagliazucchi E, et al. (2012) Automatic sleep staging using fMRI functional connectivity data. *Neuroimage* 63(1):63–72.
20. Mesulam MM (1998) From sensation to cognition. *Brain* 121(Pt 6):1013–1052.
21. Gonzalez-Castillo J, et al. (2012) Whole-brain, time-locked activation with simple tasks revealed using massive averaging and model-free analysis. *Proc Natl Acad Sci USA* 109(14):5487–5492.
22. Tononi G (2010) Information integration: Its relevance to brain function and consciousness. *Arch Ital Biol* 148(3):299–322.
23. Delamillieure P, et al. (2010) The resting state questionnaire: An introspective questionnaire for evaluation of inner experience during the conscious resting state. *Brain Res Bull* 81(6):565–573.
24. Hubert L, Arabie P (1985) Comparing partitions. *J Classif* 2(1):193–218.
25. Steinley D (2004) Properties of the Hubert-Arabie adjusted Rand index. *Psychol Methods* 9(3):386–396.
26. Craddock RC, James GA, Holtzheimer PE, 3rd, Hu XP, Mayberg HS (2012) A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum Brain Mapp* 33(8):1914–1928.
27. Zilverstand A, Sorger B, Zimmermann J, Kaas A, Goebel R (2014) Windowed correlation: A suitable tool for providing dynamic fMRI-based functional connectivity neurofeedback on task difficulty. *PLoS ONE* 9(1):e85929.
28. Pantazatos SP, Talati A, Pavlidis P, Hirsch J (2012) Decoding unattended fearful faces with whole-brain correlations: An approach to identify condition-dependent large-scale functional connectivity. *PLOS Comput Biol* 8(3):e1002441.
29. Cribben I, Haraldsdottir R, Atlas LY, Wager TD, Lindquist MA (2012) Dynamic connectivity regression: Determining state-related changes in brain connectivity. *Neuroimage* 61(4):907–920.
30. Heinzle J, Wenzel MA, Haynes JD (2012) Visuomotor functional network topology predicts upcoming tasks. *J Neurosci* 32(29):9960–9968.
31. Schlegel A, et al. (2013) Network structure and dynamics of the mental workspace. *Proc Natl Acad Sci USA* 110(40):16277–16282.
32. Lehmann D, Strik WK, Henggeler B, Koenig T, Koukkou M (1998) Brain electric microstates and momentary conscious mind states as building blocks of spontaneous thinking: I. Visual imagery and abstract thoughts. *Int J Psychophysiol* 29(1):1–11.
33. Van de Ville D, Britz J, Michel CM (2010) EEG microstate sequences in healthy humans at rest reveal scale-free dynamics. *Proc Natl Acad Sci USA* 107(42):18179–18184.
34. Britz J, Van De Ville D, Michel CM (2010) BOLD correlates of EEG topography reveal rapid resting-state network dynamics. *Neuroimage* 52(4):1162–1170.
35. Yuan H, Zotev V, Phillips R, Drevets WC, Bodurka J (2012) Spatiotemporal dynamics of the brain at rest: Exploring EEG microstates as electrophysiological signatures of BOLD resting state networks. *Neuroimage* 60(4):2062–2072.
36. Musso F, Brinkmeyer J, Mobascher A, Warbrick T, Winterer G (2010) Spontaneous brain activity and EEG microstates: A novel EEG/fMRI analysis approach to explore resting-state networks. *Neuroimage* 52(4):1149–1161.
37. Milazzo AC, et al. (2014) Identification of mood-relevant brain connections using a continuous, subject-driven rumination paradigm. *Cereb Cortex* 2014:bhu255.
38. Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD (2011) Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods* 8(8):665–670.
39. Leonardi N, Van De Ville D (2015) On spurious and real fluctuations of dynamic functional connectivity during rest. *Neuroimage* 104:430–436.
40. Baars BJ (2002) The conscious access hypothesis: Origins and recent evidence. *Trends Cogn Sci* 6(1):47–52.
41. Tononi G, Edelman GM (1998) Consciousness and complexity. *Science* 282(5395):1846–1851.
42. Jones DT, et al. (2012) Non-stationarity in the "resting brain's" modular architecture. *PLoS ONE* 7(6):e39731.
43. Damaraju E, et al. (2014) Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. *Neuroimage Clin* 5(July):298–308.
44. Cox RW (1996) AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29(3):162–173.
45. Glover GH, Li TQ, Ress D (2000) Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magn Reson Med* 44(1):162–167.
46. Birn RM, Diamond JB, Smith MA, Bandettini PA (2006) Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. *Neuroimage* 31(4):1536–1548.
47. Chang C, Cunningham JP, Glover GH (2009) Influence of heart rate on the BOLD signal: The cardiac response function. *Neuroimage* 44(3):857–869.
48. Jo HJ, Saad ZS, Simmons WK, Milbury LA, Cox RW (2010) Mapping sources of correlation in resting state FMRI, with artifact detection and removal. *Neuroimage* 52(2):571–582.
49. Gonzalez-Castillo J, et al. (2013) Effects of image contrast on functional MRI image registration. *Neuroimage* 67:163–174.

Supporting Information

Gonzalez-Castillo et al. 10.1073/pnas.1501242112

SI Methods

Subjects. Twenty-two subjects (13 females; age, 27 ± 5 y old) completed these experiments after giving informed consent in compliance with a protocol approved by the Institutional Review Board of the National Institute of Mental Health in Bethesda, MD. The data from two subjects were discarded from the analysis due to excessive spatial distortions in the functional time series.

Experimental Paradigms. Subjects were scanned under two different experimental paradigms. The first paradigm consisted of a continuous ∼25-min-long scan during which subjects performed and transitioned between four different cognitive tasks: namely, rest, simple math, a two-back working memory task, and a visual attention task. We refer to this paradigm as the continuous multitask paradigm. The second paradigm consisted of a series of three separate scans (one per active task: math, memory, and visual attention) using a "20 s on/40 s off" block design paradigm. We refer to these three scans as functional localizer paradigms.

Continuous multitask paradigm. Subjects were scanned continuously for 25 min and 24 s while performing four different tasks (Fig. 1*A*). Each task appeared for two separate 180-s blocks, with each task block being preceded by a 12-s instruction period. The paradigm was the same for all subjects. The order of task blocks was randomized so that each task was always preceded and followed by a different task. The four tasks were as follows:

- Rest: Subjects were instructed to fixate on a crosshair in the center of the screen and let their mind wander.
- Math: Subjects were presented with simple arithmetic involving three numbers between 1 and 10 and two operands (only addition and subtraction). These problems were presented at the top of the screen (e.g., "(2 + 3) − 1 ="), and two answers (one correct and one incorrect) were presented at the bottom of the screen. Subjects used an MRI-compatible response box to select the correct response. Operations remained on the screen for 4 s, and a blank screen appeared for 1 s between successive trials. This timing resulted in a total of 36 operations per 180-s block. Problems were different across all trials, and the left/right position of correct responses was randomized.
- Memory: Subjects were presented with a continuous sequence of individual geometric shapes (triangles, squares, rhomboids, circles, and pentagons) that appeared at the center of the screen every 3 s (shapes appeared on the screen for 2.6 s, followed by a blank screen for 400 ms). Subjects were instructed to press a button on an MRI-compatible response box whenever the shape currently on the screen was the same as two shapes before in the sequence. The sequence was designed such that a response was required at least every seven shapes. Different sequences were presented during the two 180-s blocks.
- Visual attention (video): Subjects were instructed to watch a video of a real fish tank. The video shows different types of fish swimming in and out of view from a single, stationary point of view. Subjects were instructed to monitor the fish for a red crosshair target that would appear at random times over a single fish for a duration of 200 ms, and to use an MRI-compatible response box to signal whether the target appeared on top of a clown fish, or any other type of fish. A total of 16 targets appeared during each 180-s block, and the same video was used for both blocks. For one subject, the video during the second visual attention block flickered. The subject's subjective report and behavioral data confirmed that the subject was still able to perform the task at a comparable level.

Both responses and reaction times were recorded for the math, memory, and video tasks. Subjects were instructed to respond as quickly and accurately as possible, and only once per question, regardless of potential errors. To get subjects accustomed to all of the tasks, subjects were presented with a shorter training version of the paradigm outside the scanner with a different list of trials for all tasks. Before doing the training, subjects were instructed to use this opportunity to come up with a cognitive strategy for each of the tasks. Subjects were advised about the importance of keeping the cognitive strategy consistent within and across experimental blocks. All subjects confirmed that they had come up with strategies for each task before entering the scanner.

Functional localizer scans. Three additional functional scans were acquired in a subset of 18 subjects after the continuous multitask scan. All three functional localizer scans had the same organization. An initial 18-s rest period was followed by five repetitions of a 21-s task block followed by a 39-s rest block. An additional 10.5 s of rest were added at the end of each scan, resulting in 328.5-s runs. During the rest periods, subjects were instructed to remain still and focus their attention on a fixation crosshair on the center of the screen. During the task block periods, subjects performed one of the three active tasks previously described (e.g., math, memory, or visual attention). The same task was performed in all five blocks of a scan. Therefore, three different scans, one per task, were acquired in each subject.

Data Acquisition. Imaging was performed on a Siemens 7 Tesla MRI scanner equipped with a 32-element receive coil (Nova Medical). Functional runs were obtained using a gradient recalled, single shot, echo planar imaging (gre-EPI) sequence {repetition time [TR], 1.5 s; echo time [TE], 25 ms; flip angle [FA], 50°; 36 interleaved slices; slice thickness, 2 mm; in-plane resolution, 2 × 2 mm; field-of-view [FOV], 192 mm; right-to-left phase encoding, integrated parallel imaging technique [iPAT] [generalized autocalibrating partially parallel acquisition (GRAPPA)], 2}. To accommodate differences in head size, acquisition angle and slice spacing (0.1–0.4 mm) were varied across subjects. We acquired a total of 1,017 volumes while subjects performed the continuous-task paradigm, and we acquired 219 volumes during each functional localizer paradigm. For one subject, functional data were collected with a slightly different set of parameters [TR, 1.5 s; TE, 25 ms; FA, 70°; 1,021 volumes; 35 oblique, interleaved slices; slice thickness, 2 mm with 0.3-mm gap; in-plane resolution, 1.4 × 1.4 mm; field-of-view (FOV), 200 mm; anterior-to-posterior phase encoding, iPAT (GRAPPA), 4].

In addition, T1-weighted magnetization-prepared rapid gradient-echo and proton density (PD) sequences were acquired for presentation and alignment purposes (axial prescription, number of slices per slab, 192; slice thickness, 1 mm; square FOV, 256 mm; image matrix, 256 × 256).

Behavioral Data Analysis. Here, we describe computations of percent correct responses, percent missing responses, and average reaction times for each of the three active tasks.

Memory task. Percent correct responses for the memory blocks was computed as follows:

$$P_{Correct,\,Mem} = 100x\frac{(N_R - N_{Missing}) + (N_{NR} - N_{Misfires})}{N_R + N_{NR}} \quad \textbf{[S1]}$$

where N_{NR} is the number of trials that required no response, $N_{Misfires}$ is the number of times a subject pressed the button for

a trial that required no answer, N_R is the number of trials that required a response, $N_{Missing}$ is the number of times a subject did not press the button when he/she was required to do so.

Percent trials missing responses for the memory blocks was computed as follows:

$$P_{Missing} = 100x\frac{N_{Missing}}{N_R} \quad \textbf{[S2]}$$

Average reaction time for memory blocks was computed using reaction times for all trials in which subjects responded. However, this information did not take into account trials requiring a response but missing one (e.g., if subject went to sleep). To take into account such inconsistencies in performance, which may indicate lapses in cognitive processing that would result in poor classification of FC states, trials missing responses were penalized by counting them as a reaction time equal to the duration of a trial (3 s) in the average reaction time. In this manner, the average reaction time for subjects that may have fallen asleep during a block will increase in parallel with the number of missing responses.

Math task. Percent correct responses for math blocks ($P_{Correct,Math}$) was computed as the number of correct answers divided by the total number of trials. Trials with no response counted as incorrect answers.

Percent trials missing responses for the math blocks was computed as the number of trials for which there was no response divided by the total number of trials in a block (36 trials).

Using the same reasoning used to calculate average reaction time in the memory task, average reaction time for math blocks was computed by averaging together reaction times from trials in which subjects responded and penalty reaction times equal to the duration of a trial (5 s) from trials missing responses.

Visual attention/video task. Percent correct responses for the video task was computed as follows:

$$P_{Correct,Vid} = 100x\frac{N_{CR}}{N_{Trials}} \quad \textbf{[S3]}$$

where N_{CR} is the number of correct responses (crosshair appeared on screen and subject selected the correct fish type) and N_{Trials} is the number of times a crosshair appeared in the screen.

Percent trials missing responses for the video blocks was computed as the number of trials for which there was no response divided by the total number of trials in a block (16 trials).

Using the same reasoning used to calculate average reaction time in the memory task, average reaction time for video blocks was computed by averaging together reaction times from trials in which subjects responded and penalty reaction times from trials missing responses that were equal to the average plus two SDs of the reaction time across all subjects during the video task (1.82 s).

Continuous-Task fMRI Data Processing. Fig. S1*C* contains a flowchart that summarizes all of the main processing steps in our connectivity-based decoding pipeline. Details about the implementation of each of these processing steps are provided below.

Preprocessing. The Analysis of Functional NeuroImages (AFNI) software (44) was used for data preprocessing. For individual EPI runs, preprocessing included the following: (*i*) despiking (AFNI program *3dDespike*); (*ii*) physiological noise correction (in all but four subjects), including regressors for the retrospective image correction (RETROICOR) (45), respiration volume per time (RVT) (46), and heart rate (47) models; (*iii*) slice time correction (AFNI program *3dTshift*); and (*iv*) head motion correction (AFNI program *3dvolreg*). In addition, mean, slow signal trends modeled with legendre polynomials up to seventh order, signal from eroded local white matter, signal from the lateral ventricles (cerebrospinal fluid), motion estimates, and the first derivatives of motion were regressed out in a single regression step (AFNI program *3dTfitter*) to account for potential hardware instabilities and remaining physiological noise [anatomy-based image correction (ANATICOR)] (48). Finally, time series were converted to signal percent change and bandpass filtered. The high end of the bandpass filter was set to 0.18 Hz. The lower end of the bandpass filter was chosen to match the inverse of the window duration used in subsequent analyses. Consequently, a different frequency band was used per window duration: for 180-s window analysis, data were filtered to 0.006–0.18 Hz; for 90-s windows, to 0.012–0.18 Hz; for 60-s windows, to 0.017–0.18 Hz; for 45-s windows, to 0.023–0.18 Hz; for 30-s windows, to 0.034–0.18 Hz; and, finally, for 22.5-s windows, to 0.045–0.18 Hz. This step is necessary to avoid inducing spurious fluctuations in correlation calculations in successive steps (39). Window-specific time series were then spatially smoothed [full width at half maximum (FWHM), 4 mm; AFNI program *3dBlurInMask*).

Spatial transformation matrices to go from EPI native space to Montreal Neurological Institute (MNI) space were computed for all subjects using the magnetization-prepared rapid gradient echo (MP-RAGE) and PD scans with AFNI program *align_epi_anat.py* following procedures previously described in ref. 49. These matrices were then used to bring publicly available regions of interest (ROI) definitions (see *Brain Parcellation* below for more details) from MNI space into each subject's EPI native space.

Brain parcellation. Two hundred ROIs covering the whole brain were obtained from the human brain atlas provided by Craddock et al. (26). These ROIs are spatially contiguous, similar in size, and represent functionally homogeneous brain regions. Publicly available atlases were transformed from MNI to subject space for each individual. Any ROIs that did not contain at least 10 voxels in all of a subject's scanning field of view (FOV) were removed from subsequent analyses. This process resulted in the exclusion of 43 ROIs located primarily in cerebellar, inferior temporal, and orbitofrontal regions.

For comparison purposes, we also conducted the classification using the 30, 50, 70, 100, 150, and 500 ROI atlases provided by Craddock et al. (26). Results for these alternative analyses pipeline are shown in Fig. S6*A*.

ROI time series extraction. For each ROI, a representative time series consisting of the principal singular vector was obtained using the AFNI program *3dmaskSVD*. Voxels with a temporal SD greater than 7 were discarded to minimize contributions from large vasculature.

Dimensionality reduction. To reduce the dimensionality of the input feature space before the clustering/classification step, we used principal component analysis (PCA) and kept all necessary components to account for 97.5% of the variance. On average, this procedure permitted us to reduce the dimensionality of the input feature space from 12,246 to 2,556. Those numbers correspond to the number of unique pairwise connections associated with a 157 × 157 connectivity matrix (before the PCA) and a 72 × 72 matrix (after the PCA step), respectively.

For comparison purposes, we also conducted the classification after applying PCA but keeping all components (no dimensionality reduction), and also after applying PCA and keeping smaller amounts of variance (95%, 90%, and 75%). Results for these alternative analysis pipelines are shown in Fig. S6*B*.

Windowed connectivity snapshots. Remaining PCA time series were subsequently segmented in time using nonoverlapping windows of 180, 90, 60, 45, 30, and 22.5 s that match the experimental paradigm timing. Instruction periods were discarded. For each window, we computed all pairwise correlations between PCA time series and put them in vector form. We then transformed these Pearson's correlation values into Z-scores using the Fisher transformation. We refer to the Fisher-transformed vectors as

connectivity snapshots throughout this manuscript. They can be regarded as a picture of covariance across the brain in a given window of time during the experiment. These connectivity snapshots are the input to the subsequent classification step.

Connectivity-based classification/clustering. Connectivity snapshots were input to the *k*-means clustering algorithm in MATLAB. This algorithm sorts the connectivity snapshots into k groups by maximizing within-cluster similarity and between-cluster dissimilarity, using correlation as a distance metric between snapshots. We selected $k = 4$ clusters for all analyses, given that subjects were asked to engage in four different cognitive tasks during the scan. The algorithm ran with a maximum of 1,000 iterations to ensure convergence to the optimal clustering solution. This analysis was performed separately for each subject and window length. No information about the timing of each snapshot relative to the paradigm or tasks was provided to the clustering algorithm.

Quantitative clustering validation. To quantitatively evaluate how successful we were at recovering the periods during which subjects were performing the same mental task, we used the adjusted rand index (ARI) (24). As an external clustering validation technique, its computation requires knowledge about the real structure of the data (i.e., the real groupings of elements into clusters that should be recovered by the clustering algorithm). In our particular case, this ground truth comes from the experimental paradigm (e.g., what task subjects were engaged in during a given window). The ARI ranges from 1 to below 0, with 1.00 indicating perfect recovery of the known clusters, 0 indicating chance level clustering performance, and < 0 indicating worse than chance. Ranges established in the literature describe an ARI > 0.9 as excellent, $0.9 >$ ARI > 0.8 as good, $0.8 >$ ARI > 0.65 as moderate, and ARI < 0.65 as poor recovery, respectively (25). In this manner, the ARI enables us to evaluate the agreement between grouping of windows based on mental state and those based on connectivity snapshots, while also adjusting for potential agreements just by chance.

The ARI was calculated separately for each subject and window length. ARI values were then averaged across subjects to provide summary results at each window length.

Correlation with behavioral metrics. To determine whether clustering results were contingent on consistent task performance, we computed three behavioral indices based on responses recorded inside the scanner. Each index was computed separately for each subject. For each metric, we also calculated the discrepancy between task blocks. The three indices are as follows:

- Average percent correct responses [$P_{Correct}$]: Percentage of correct responses averaged across all six active task blocks.
- Average percent missing responses [$P_{Missing}$]: Percentage of trials that required a response for which subjects did not provide one. This metric includes all math trials, memory trials that required a button press, and each appearance of the target crosshair during the visual attention task.
- Average reaction time [RT]: Average reaction times across all six active task blocks were averaged into one representative value.
- Across-block discrepancies [$\Delta P_{Correct}$, $\Delta P_{Missing}$, and ΔRT]: For each index, metrics were calculated as described above within each task block. For each metric, the differences between blocks of the same task were found and then averaged to arrive at a single value representing how each index varied across task blocks for each subject.

We plotted each of the six behavioral indices versus the ARI for each window length. We then tested for significant correlations between the ARI and each of these behavioral indices using MATLAB function *corrcoef*.

Classification under control conditions. To confirm that classification results were driven primarily by meaningful changes in connectivity across the brain, we conducted a series of four additional analyses, in which the input to the classification algorithm was altered as follows:

- ROI intensity-based features: Each window was characterized by the average intensity of each ROI representative time series within that window. Consequently, in this analysis, the dimensionality of the feature vectors was 157 (the number of ROIs). No dimensionality reduction step is performed in this analysis. The purpose of this analysis is to rule out the possibility that high classification accuracy results primarily from differences in sustained changes in activity levels across tasks. If connectivity changes are the primary source of information driving the classification algorithm, this analysis should lead to poor recovery levels of the original mental states.
- ROI variability-based features: Each window was characterized by the SD across time of each ROI representative time series within that window. As in the previous case, the dimensionality of the feature vectors equals the number of ROIs, and no dimensionality reduction step was performed. The goal of this analysis is to rule out the possibility that high classification accuracy results primarily from differences in the amount of signal fluctuations across tasks. Once more, if connectivity changes are the primary source of information driving the classification algorithm, this analysis should lead to poor recovery levels of the original mental states.
- Connectivity-based classification after task regression: Three regressors corresponding to each task were generated by convolving the hemodynamic response with boxcar functions with ones during the two task blocks and zeros elsewhere. These task regressors were included as nuisance regressors in the preprocessing step when slow trends, white matter, ventricle, and motion signals were regressed out of the data. The rest of the analysis was the same as in the main experiment.
- Randomization of connectivity snapshots: The order of the elements in the connectivity snapshots was randomized independently for each window entering the analysis. After this step, a given position in a connectivity snapshot no longer corresponds to connectivity between the same two PCA components across all windows. Because the temporal evolution of connections across snapshots is lost, this analysis should lead to poor recovery of the original mental states.
- ROI time series phase randomization: We phase randomized the ROI representative time series before the dimensionality reduction step. Phase randomization produces surrogate time series with identical autoregressive properties, yet the precise timing of signal fluctuations is destroyed. Because correlations depend heavily on phase alignment between time series, this analysis should lead to poor recovery of the original mental states.

Results from these analyses are presented in Fig. S7.

Functional Localizer Scan Processing. Functional localizer scans were also processed with the AFNI software (44). Preprocessing steps for each functional scan included the following: (*i*) despiking (AFNI program *3dDespike*); (*ii*) slice time correction (AFNI program *3dTshift*); (*iii*) estimation of head motion parameters (AFNI program *3dvolreg*); (*iv*) head motion correction and transformation into MNI space through a single interpolation step (AFNI programs *align_epi_anat.py* and *3dAllineate*); and (*v*) spatial smoothing (FWHM, 4 mm; AFNI program *3dBlurInMask*). Individual subject levels of activation were subsequently computed separately for each individual localizer scan using AFNI program *3dDeconvolve*. Head motion estimates and their first derivative were incorporated into the analysis as covariates.

To generate group activation maps, we input individual subject activation levels for all three tasks into a single two-way, mixed-effects ANOVA [factor A, task, fixed; factor B, subject, random].

Using AFNI program *3dANOVA2*, we computed statistical maps of activation for the following contrasts: (*i*) math vs. rest, (*ii*) memory vs. rest, (*iii*) video vs. rest, (*iv*) math vs. memory, (*v*) math vs. video, and (*vi*) memory vs. video. In addition, we also computed a map of *F* statistics (*F* map) for factor A, the task effect.

Contribution of task-specific regions to classification. The purpose of these analyses was to evaluate the spatial distribution and compactness of the information driving the classification algorithm. In particular, we wanted to evaluate whether or not connectivity levels in areas outside those identified as active during the tasks by traditional univariate analyses were providing valuable information to the classification algorithm. For this purpose, we first ranked ROIs according to the *F* map of task effect obtained from the functional localizer scans. We then performed classification using a decreasing number of ROIs. ROIs were removed in two opposite ways as described below.

Ranking of ROIs in relationship to task set. For each of the 157 ROIs entering the final steps of the analysis, we computed the average *F* statistic across all voxels in the ROI. We then ranked all 157 ROIs according to this average *F* value. In this manner, ROIs were ranked according to the effect size for the task factor. In other words, ROIs were ranked according to how well they differentiated the tasks based on activity levels.

Leave-out-high-F-ROIs analysis. Here, we attempted classification with a variable number of ROIs. The number of ROIs entering the analysis ranged from 157 (all ROIs) to only 10, in decrements of 10 ROIs. ROIs with the highest ranks (highest average *F* statistic) were removed first. This analysis was repeated for all window lengths used in the original analysis, and dimensionality was reduced to keep 97.5% of the variance.

Leave-out-low-F-ROIs analysis. This analysis is equivalent to the one described above, except that ROIs with the lowest ranks (lowest average *F* statistic) were removed first.

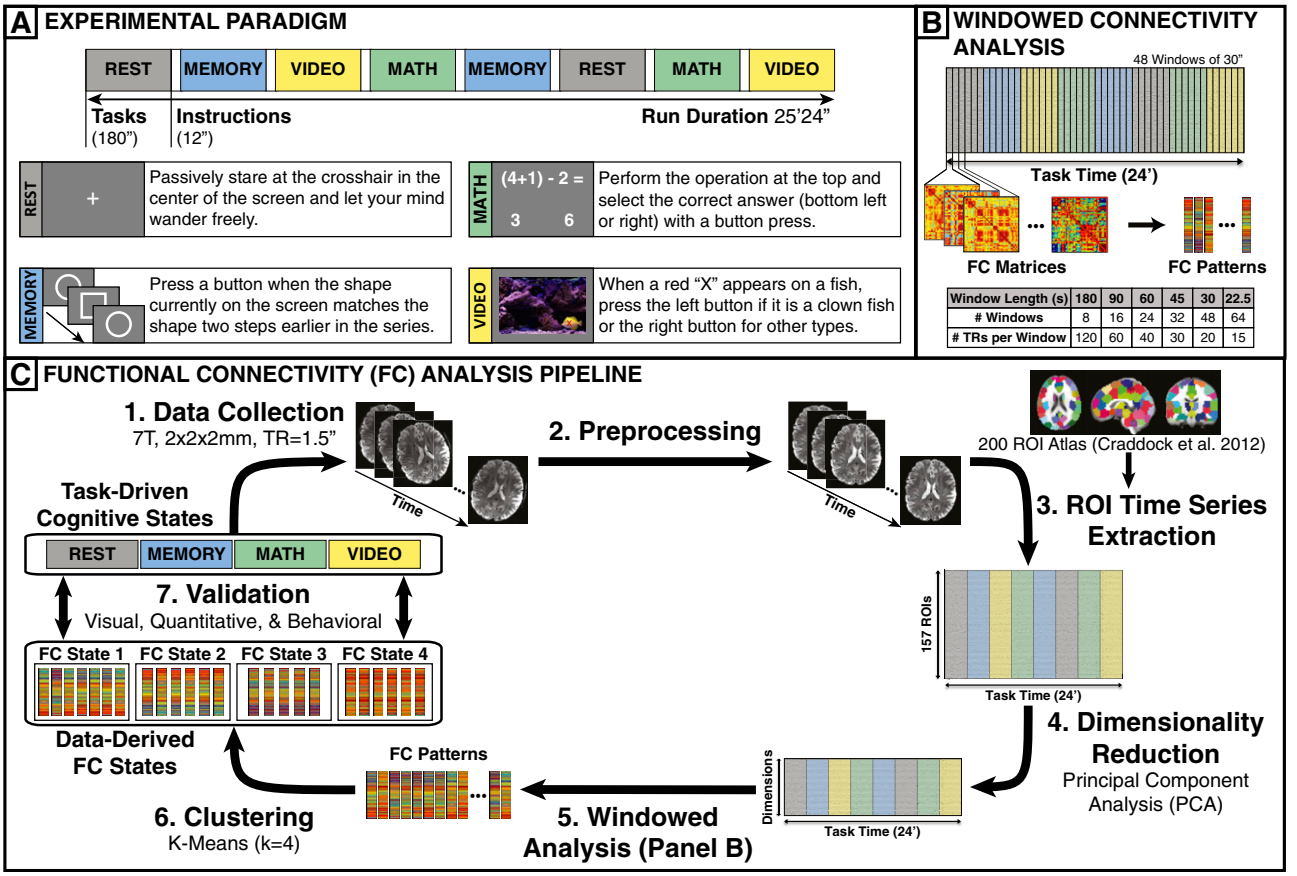


Fig. S1. Detailed schematic of experimental paradigm and analysis pipeline. (*A*) Timing of experimental paradigm for the continuous multitask experiments, as well as descriptions of the visuals presented to the subjects during each task. (*B*) Detailed depiction of how window-based FC connectivity patterns entering the classification analysis were computed. (*C*) Step-by-step diagram of the analysis pipeline used for the multitask paradigm. Data collected as subjects engaged in and transition between the different tasks were first preprocessed. We then extracted representative time series for the ROIs. These whole-length representative time series entered a PCA analysis used to reduce the dimensionality of the data. Selected PCA time series were then used to construct window-based connectivity patterns that entered the final clustering step. Groupings of windows based on connectivity patterns were finally compared against groupings of windows based on the mental task being performed.

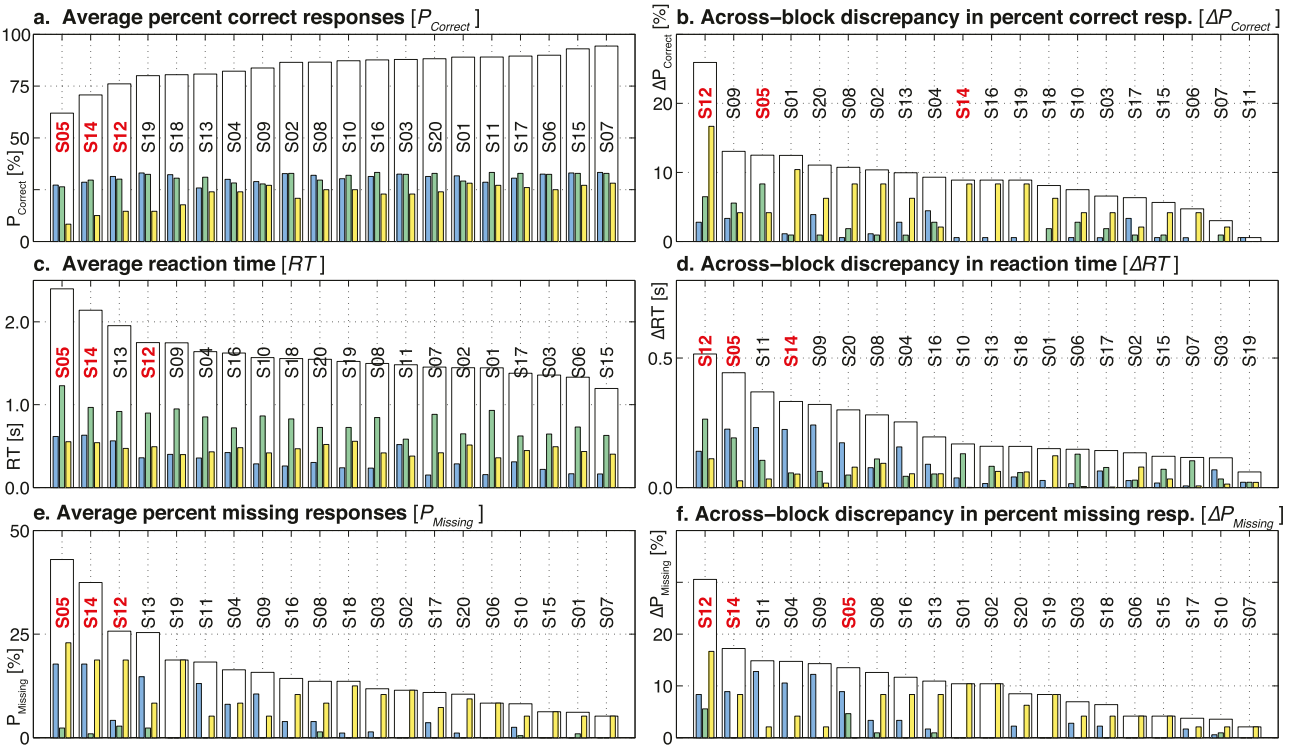


Fig. S2. Single-subject behavioral results. (*A*) Percentage of correct responses ($P_{Correct}$). (*B*) Discrepancy in $P_{Correct}$ across blocks of the same task (a.b.s.t.). (*C*) Reaction time (RT). (*D*) Discrepancy in RT a.b.s.t. (*E*) Percentage of trials with missing responses ($P_{Missing}$). (*F*) Discrepancy in $P_{Missing}$ a.b.s.t. In all panels, each subject is represented by four bars: transparent bar, overall value across all tasks; blue, memory; green, math; yellow, video. Subjects are sorted according to the panel's specific metric. Outliers for more than one WL are marked in red.

a. Subject 02

Window Duration	Accuracy	ARI
180s	100%	1.00
120s	100%	1.00
60s	100%	1.00
45s	100%	1.00
30s	100%	1.00
22.5s	100%	1.00

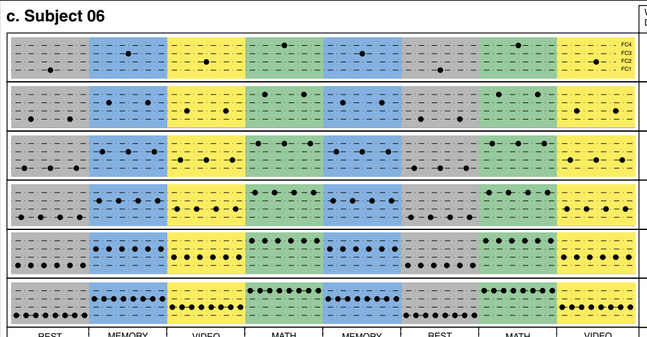
REST	MEMORY	VIDEO	MATH	MEMORY	REST	MATH	VIDEO	
N/A	96.67%	75%	100%	100%	N/A	97.2%	50%	Percent Correct
N/A	0.0%	18.75%	0%	0.0%	N/A	0%	50%	Percent Missing Responses
N/A	0.82s	1.42s	1.90s	0.90s	N/A	1.99s	1.66s	Average RT

b. Subject 04

Window Duration	Accuracy	ARI
180s	100%	1.00
120s	100%	1.00
60s	100%	1.00
45s	100%	1.00
30s	100%	1.00
22.5s	98.4%	0.96

REST	MEMORY	VIDEO	MATH	MEMORY	REST	MATH	VIDEO	
N/A	83.33%	68.75%	88.89%	96.67%	N/A	80.56%	75%	Percent Correct
N/A	40.0%	31.25%	0%	8.33%	N/A	0%	18.75%	Percent Missing Responses
N/A	1.30s	1.37s	2.49s	0.83s	N/A	2.62s	1.21s	Average RT

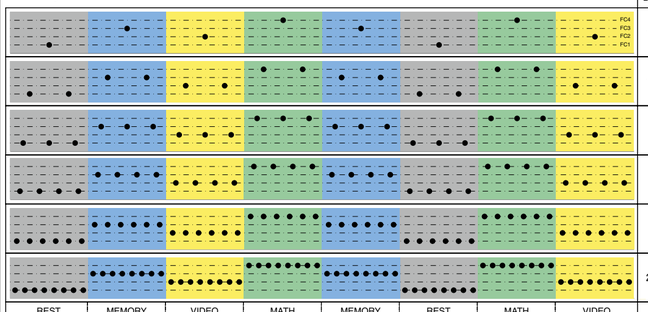
c. Subject 06



Window Duration	Accuracy	ARI
180s	100%	1.00
120s	100%	1.00
60s	100%	1.00
45s	100%	1.00
30s	100%	1.00
22.5s	100%	1.00

REST	MEMORY	VIDEO	MATH	MEMORY	REST	MATH	VIDEO	
N/A	96.67%	81.25%	97.22%	98.33%	N/A	97.22%	68.75%	Percent Correct
N/A	0%	18.75%	0%	0%	N/A	0%	31.25%	Percent Missing Responses
N/A	0.52s	1.31s	2.00s	0.48s	N/A	2.38s	1.30s	Average RT

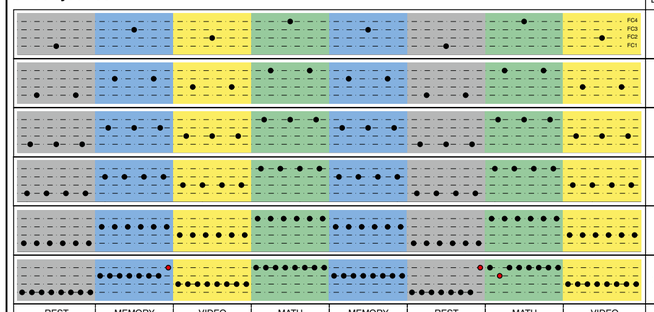
d. Subject 07



Window Duration	Accuracy	ARI
180s	100%	1.00
120s	100%	1.00
60s	100%	1.00
45s	100%	1.00
30s	100%	1.00
22.5s	100%	1.00

REST	MEMORY	VIDEO	MATH	MEMORY	REST	MATH	VIDEO	
N/A	100%	87.50%	100%	100%	N/A	97.22%	97.22%	Percent Correct
N/A	0%	12.50%	0%	0%	N/A	0%	0%	Percent Missing Responses
N/A	0.46s	1.25s	2.50s	0.44s	N/A	2.81s	1.27s	Average RT

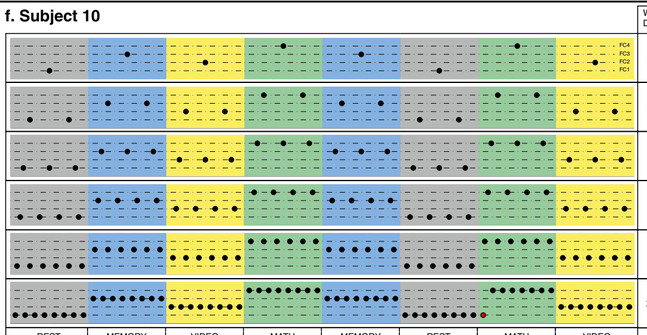
e. Subject 09



Window Duration	Accuracy	ARI
180s	100%	1.00
120s	100%	1.00
60s	100%	1.00
45s	100%	1.00
30s	100%	1.00
22.5s	95.3%	0.88

REST	MEMORY	VIDEO	MATH	MEMORY	REST	MATH	VIDEO	
N/A	91.67%	75%	91.67%	81.67%	N/A	75%	87.5%	Percent Correct
N/A	13.33%	18.75%	0%	50%	N/A	0%	12.5%	Percent Missing Responses
N/A	0.84s	1.17s	2.75s	1.56s	N/A	2.94s	1.22s	Average RT

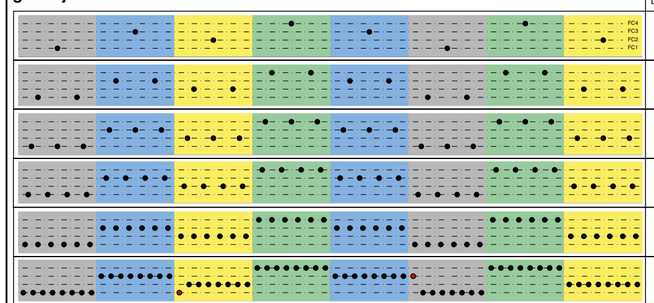
f. Subject 10



Window Duration	Accuracy	ARI
180s	100%	1.00
120s	100%	1.00
60s	100%	1.00
45s	100%	1.00
30s	100%	1.00
22.5s	98.4%	0.96

REST	MEMORY	VIDEO	MATH	MEMORY	REST	MATH	VIDEO	
N/A	90%	81.25%	100%	91.67%	N/A	91.67%	68.75%	Percent Correct
N/A	6.67%	12.50%	0%	8.33%	N/A	2.78%	18.75%	Percent Missing Responses
N/A	0.80s	1.25s	2.40s	0.91s	N/A	2.79s	1.25s	Average RT

g. Subject 11



Window Duration	Accuracy	ARI
180s	100%	1.00
120s	100%	1.00
60s	100%	1.00
45s	100%	1.00
30s	100%	1.00
22.5s	96.9%	0.92

REST	MEMORY	VIDEO	MATH	MEMORY	REST	MATH	VIDEO	
N/A	86.67%	81.25%	100%	85%	N/A	100%	81.25%	Percent Correct
N/A	20%	12.5%	0%	58.33%	N/A	0%	18.75%	Percent Missing Responses
N/A	1.21s	1.09s	1.59s	1.90s	N/A	1.91s	1.19s	Average RT

h. Subject 13

Window Duration	Accuracy	ARI
180s	100%	1.00
120s	100%	1.00
60s	100%	1.00
45s	100%	1.00
30s	100%	1.00
22.5s	100%	1.00

REST	MEMORY	VIDEO	MATH	MEMORY	REST	MATH	VIDEO	
N/A	73.33%	81.25%	94.44%	81.67%	N/A	91.67%	62.5%	Percent Correct
N/A	46.67%	12.50%	5.56%	41.67%	N/A	8.33%	37.5%	Percent Missing Responses
N/A	1.71s	1.32s	2.63s	1.67s	N/A	2.88s	1.51s	Average RT

Fig. S3. Individual subject classification results (part 1). Classification results for nonoutlier subjects 2 (*A*), 4 (*B*), 6 (*C*), 7 (*D*), 9 (*E*), 10 (*F*), 11 (*G*), and 13 (*H*) are shown here. For each subject, we show classification results for WL = 180 s, 90 s, 45 s, 60 s, 30 s, and 22.5 s in the form of classification staffs. Correctly classified windows are marked with black dots whereas incorrectly classified windows are marked in red. To the right of the classification staffs, we report quantitative measures of classification in terms of classification accuracy and adjusted rand index (ARI). Below the classification staffs, we report values of percent correct responses, percent of missing trials, and response time for each active task block.

Fig. S4. Individual subject classification results (part 2). Classification results for the remaining nonoutlier subjects: 15 (*A*), 16 (*B*), 17 (*C*), 18 (*D*), 19 (*E*), and 20 (*F*). The organization of results within each panel is the same as in Fig. S3.

Fig. S4. Individual subject classification results (part 2). Classification results for the remaining nonoutlier subjects: 15 (*A*), 16 (*B*), 17 (*C*), 18 (*D*), 19 (*E*), and 20 (*F*). The organization of results within each panel is the same as in Fig. S3.

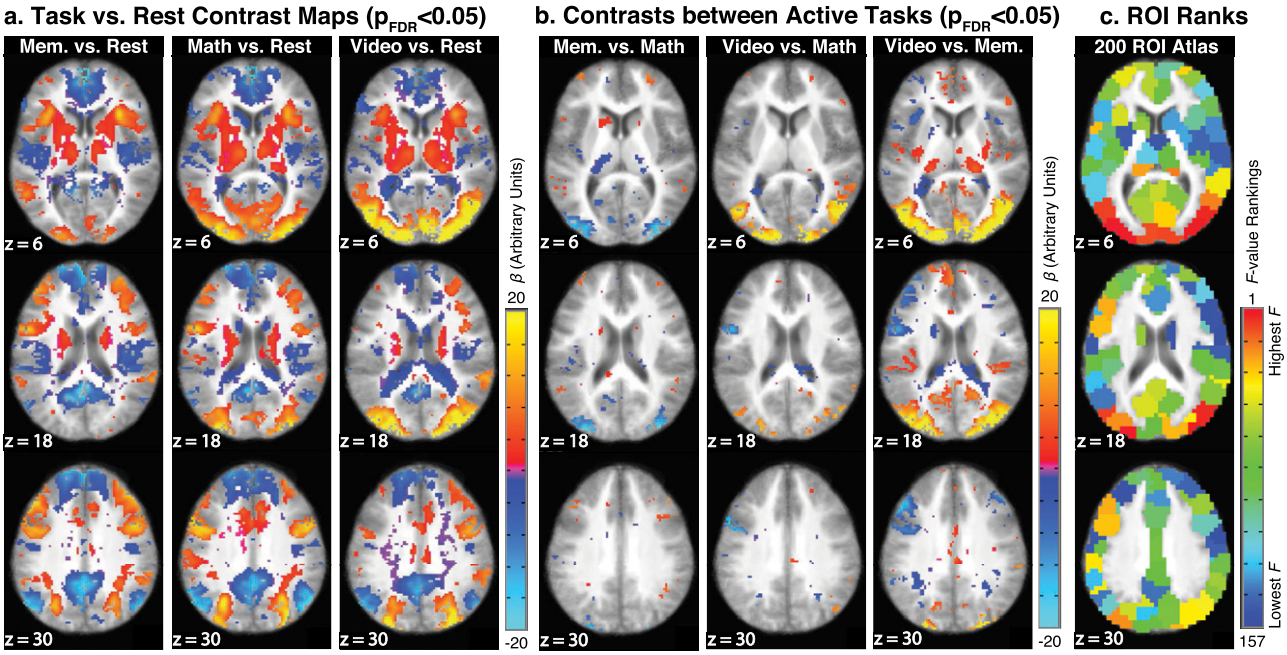


Fig. S5. Localizer scan analyses. (*A*) Low order contrast maps for all active tasks (memory vs. rest; math vs. rest; and video vs. rest). Maps are thresholded at $P_{FDR} < 0.05$. (*B*) High order contrast maps (task vs. task) for all possible active task pairs (memory vs. math; video vs. math; and video vs. memory) also thresholded at $P_{FDR} < 0.05$. (*C*) Maps of ROIs ranked according to their activity-based discriminatory power across tasks, as determined by the *F* statistic for the task vs. task contrasts. Cooler colors are used for ROIs with the highest rank (lowest *F* and lowest discriminative power across tasks) whereas warmer colors are used for the ROIs with the lowest rank (highest *F* and highest discriminative power across tasks)

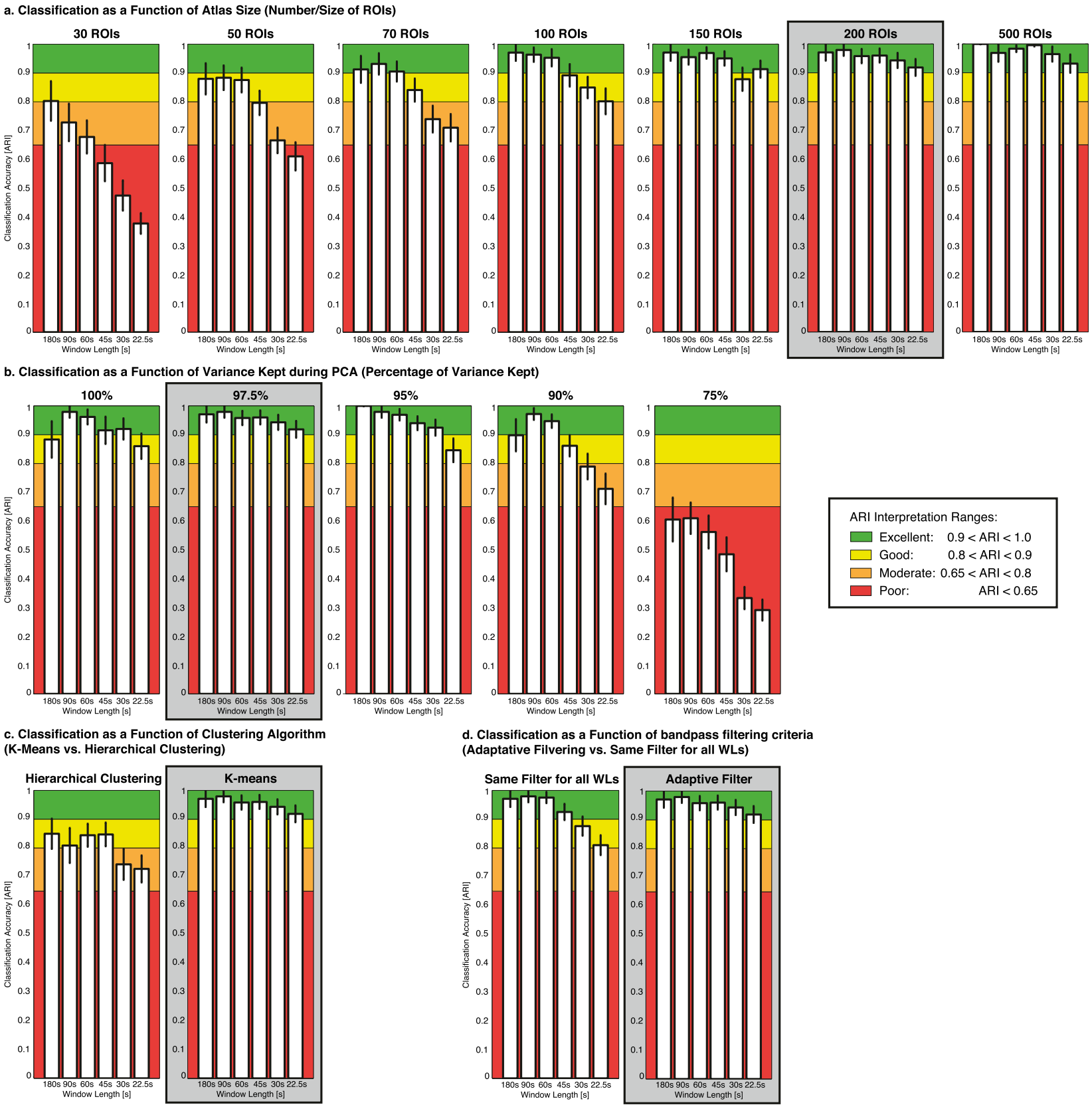


Fig. S6. Group-level classification results for additional analyses with different atlases, levels of kept variance, clustering algorithms, and band-pass filtering criteria. In all panels, the combination of parameters reported in the main analysis is highlighted with a gray background. Bars represent average ARI across subjects; error bars represent SE. (*A*) Average group-level ARI for all window lengths when the number of ROIs in the atlas changes. Results are shown for versions of the Craddock atlas (26) with 30, 50, 70, 100, 150, 200 and 500 ROIs. (*B*) Average group-level ARI for all window lengths for the 200 ROI atlas when different levels of variance are kept in the PCA step (100%, 97.5%, 95%, 90%, and 75%). (*C*) Average group-level ARI for all window lengths for the 200 ROI atlas for two different clustering algorithms: k-means and hierarchical clustering. (*D*) Average group-level ARI for all window lengths for the 200 ROI atlas for two different band-pass filtering criteria: adaptive filtering based on WL and same filtering for all WLs.

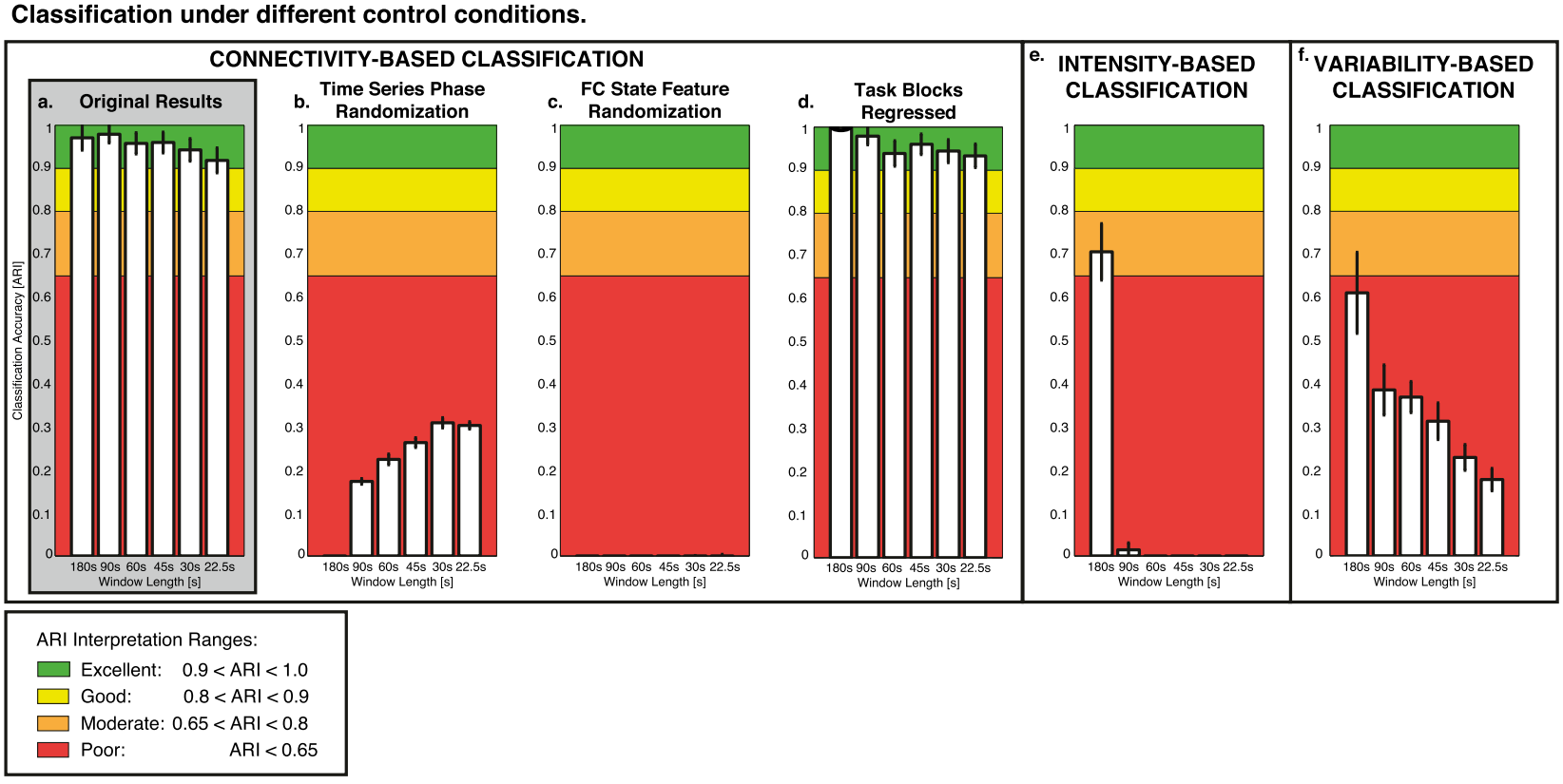


Fig. S7. Group-level classification results for all control analyses. Bars represent average ARI across subjects, and error bars represent SE. (*A*) Main analysis results for comparison purposes. (*B*) Average group-level ARI when classification is attempted after phase randomization of all ROI representative time series. (*C*) Average group-level ARI when classification is attempted after feature randomization. (*D*) Average group-level ARI after regressing out the task blocks. (*E*) Average group-level ARI when the features entering the classification are the ROI's average signal level, instead of connectivity measures. (*F*) Average group-level ARI when the features entering the classification are the ROI's SDs across time, instead of connectivity measures.

Table S1. Behavioral metrics for the subjects reported in Fig. 4

Subject	Memory-B1	Memory-B2	Math-B1	Math-B2	Video-B1	Video-B2
Subject 1						
$P_{Correct}$, %	93.33	68.75	88.89	96.67	86.11	100
$P_{Missing}$, %	0	31.25	2.78	0	2.78	0
RT, s	0.51	1.26	2.79	0.43	2.79	0.89
Subject 3						
$P_{Correct}$, %	96.67	62.50	94.44	98.33	100	75
$P_{Missing}$, %	8.33	37.50	0	0	0	25
RT, s	0.76	1.50	1.89	0.56	1.99	1.46
Subject 5						
$P_{Correct}$, %	81.67	18.75	91.67	81.67	66.67	31.25
$P_{Missing}$, %	40	68.74	0	66.67	13.89	68.75
RT, s	1.51	1.62	3.40	2.19	3.98	1.70
Subject 8						
$P_{Correct}$, %	96.67	62.50	91.67	95	86.11	87.50
$P_{Missing}$, %	6.67	37.50	5.56	16.67	2.78	12.50
RT, s	0.59	1.39	2.37	0.82	2.70	1.11
Subject 12						
$P_{Correct}$, %	98.33	68.75	100	90	80.56	18.75
$P_{Missing}$, %	0	31.25	0	25	16.67	81.25
RT, s	0.87	1.31	2.30	1.29	3.09	1.64
Subject 14						
$P_{Correct}$, %	86.67	50	88.89	85	88.89	25
$P_{Missing}$, %	40	43.75	2.78	66.67	2.78	68.75
RT, s	1.56	1.55	2.82	2.23	2.99	1.70