Title: Masked features of task states found in individual brain networks

Authors: Alexis Porter[1], Ashley Nielsen[2], Caterina Gratton*[1,3]
Affiliations: [1]Department of Psychology, Northwestern University, Evanston, IL, US, [2]Department of Neurology, Washington University in St. Louis, Saint Louis, MO, US, [3] Department of Neurology, Northwestern University, Evanston, IL, US

Correspondence: Caterina Gratton, cgratton@northwestern.edu

Acknowledgements: This work was supported by NIH grants R01MH118370 (CG), 2T32MH067564 (AP). Special thanks to Babatunde Adeyemo, Steven E. Petersen, and members of the Gratton, Dosenbach, and Greene Labs for feedback on various stages of this project.

Author Contributions: A.P, conceptualization, investigation, software, formal analysis, visualization, writing (original draft, review, and editing). A.N., conceptualization, methodology, writing (review and editing). C.G., conceptualization, methodology, investigation, writing (original draft, review, and editing), resources, supervision.

Competing interests: The authors have no competing interests

Citation Diversity Statement: Recently, the field of neuroscience has reported a bias in citation practices such that papers from minority groups are more often under-cited relative to the number of papers in the field[1]. The authors of this paper were proactive in consideration of selecting references that reflect diversity of the field in thought, contribution, and gender. Utilizing previously derived databases [1,2] we obtained the predicted gender of authors referenced in this manuscript. By this measure (and excluding self-citations to the first and last authors of our current paper), our references contain 13.87% woman(first)/woman(last), 23.3% man/woman, 23.3% woman/man, and 39.53% man/man. This method is limited in that a) names, pronouns, and social media profiles used to construct the databases may not, in every case, be indicative of gender identity and b) it cannot account for intersex, non-binary, or transgender people. Second, we obtained the predicted racial/ethnic category of the first and last author of each reference by databases that store the probability of a first and last name being carried by an author of color[3]. By this measure (and excluding self-citations), our references contain 10.83% author of color (first)/author of color(last), 10.64% white author/author of color, 23.55% author of color/white author, and 54.98% white author/white author. This method is limited in that a) names and Florida Voter Data to make the predictions may not be indicative of racial/ethnic identity, and b) it cannot account for Indigenous and mixed-race authors, or those who may face differential biases due to the ambiguous racialization or ethnicization of their names. We look forward to future work that could help us to better understand how to support equitable practices in science.

Abstract

Completing complex tasks requires flexible integration of functions across brain regions. While studies have shown that functional networks are altered across tasks, recent work highlights that brain networks exhibit substantial individual differences. Here we asked whether individual differences are important for predicting brain network interactions across cognitive states. We trained classifiers to decode state using data from single person “precision” fMRI datasets across 5 diverse cognitive states. Classifiers were then tested on either independent sessions from the same person or new individuals. Classifiers were able to decode task states in both the same and new participants above chance. However, classification performance was significantly higher within a person, a pattern consistent across model types, datasets, tasks, and feature subsets. This suggests that individualized approaches can uncover robust features of brain states, including features obscured in cross-subject analyses. Individualized approaches have potential to deepen our understanding of brain interactions during complex cognition.

Introduction

Achieving task goals requires the integration of functions associated with many different brain regions organized into large-scale brain networks [4,5]. Functional connectivity MRI (FC) provides the opportunity to examine how brain region interactions change in healthy humans when performing various tasks. FC involves tracking correlations between activity of brain regions over time. These correlated signals recapitulate functional systems seen in task activations, and are seen even in spontaneous activity at rest [6]. Relative to rest, FC is altered subtly but significantly across diverse task states[7–9] and researchers have been able to use machine learning on FC signals to classify task states with good accuracy [7,10–16], even during self-driven tasks [15,17]. These findings of significant and predictable changes in FC during complex tasks improve our understanding of how goals alter the ways that brain regions interact.

Most prior FC MRI studies combine FC patterns across individuals to overcome relatively low reliability seen with small amounts of individual subject data, inherently making the assumption that network organization has high correspondence across participants. Studies using machine learning to predict cognitive state [14,15,17–21] typically train classifiers on multi-subject data (e.g., N-1 participants) to predict network changes in unseen participants (e.g., participant N) . However, recent work has shown that there is substantial individual variability in the network organization of the human brain [22–26]. This cross-subject variation likely limits task state prediction by introducing variability and obscuring relevant features of the data which may be critical to informing our understanding of network changes during tasks.

One approach to measuring individualized brain function may be to shift the focus of data collection and analysis to single individuals through extended single subject data collection, an approach termed precision fMRI (pfMRI)[23,27]. This method produces reliable FC maps that are sensitive to individual features [9,22,23,26,28,29]. We recently used pfMRI data to demonstrate that brain networks are largely stable within an individual and subject to only subtle modulations from cognitive states [9]. These task-modulations appeared to vary across people [9], consistent with past reports that task and individual differences may have interactive influences on brain networks [9,30–32]. Past work has taken these results to suggest that measuring FC across different states can be a way to maximize individual identification (i.e., fingerprinting) across people. Here we take the complementary view that by studying task related changes at the level of single individuals, we may be able to provide a more accurate depiction of how cognitive state influences brain networks. That is, a precision fMRI approach may reveal novel components of how brain areas interact during complex tasks, which will necessarily be masked in approaches based on grouping data across people.

The goal of this paper is to determine the extent to which brain network properties that classify task states in a single person are unique. Specifically, we ask whether these individualized models generalize or show specificity by testing their ability to predict task state in new individuals. To address these questions, we used the Midnight Scan Club (MSC) pfMRI dataset – which contains over 10 hrs of fMRI data per participant - to build classifiers to discriminate brain states (task vs rest) from a single person's multi-session data. We then tested whether these classifiers could accurately determine task state in independent data from either the same person or a different person. Our results demonstrate that individualized classifiers can predict task state with high accuracy, exhibiting sensitivity to subject-specific features that

are obscured in cross-subject data. This work shines a light on previously masked – individually specific – characteristics of how functional brain networks are altered by cognitive state.

RESULTS

In this project, our goal was to improve our understanding of how brain areas interact during tasks by using an individual-level approach to test whether (and which) state changes in functional networks are specific to particular individuals. We measured functional brain networks across 4 task states and rest in highly sampled individuals with 10 separate fMRI sessions from the MSC dataset [23]. We used a machine learning approach, wherein we trained classifiers to predict task state based on data from a single individual. We then tested whether these classifiers could effectively classify state on independent data from the same individual or data from other individuals (**Fig. 1**). Eight individuals were examined, serving as 8 replications of our findings.

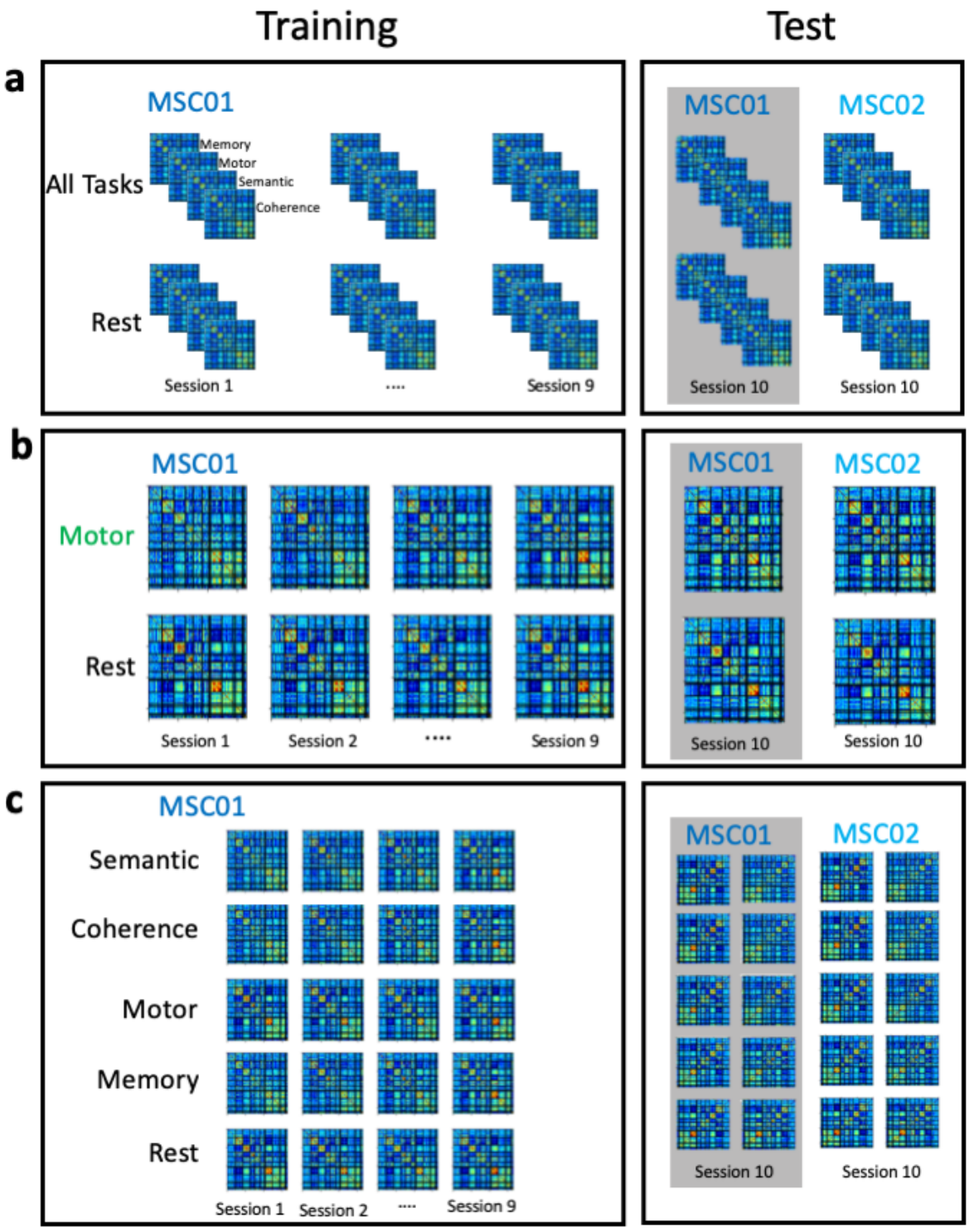

Figure 1. Overview of each analysis Classifiers were trained to distinguish task from rest functional connectivity using many sessions of data from a single participant with ridge regression. (**a**) For our first set of models, classifiers

were trained to discriminate between all tasks and rest (80 total samples per participant). **(b**) In the second group of models, classifiers were trained to distinguish a single task (vs. rest) (**c**) Finally, multiclass models were trained to discriminate among all task states at once In all cases, models were then tested on independent session data from either the same person (gray bar) or a different person; performance was contrasted across these tests.

Individualized classifiers can predict task state both within and between individuals

We first evaluated the performance of models trained in a single person to distinguish between task states and rest (binary - any task vs. rest) using independent data in the same individual. We began by training a classifier using ridge-regression with leave-one-session-out cross validation (80 samples per person, 72 training and 8 testing per fold; see *Methods*). In all participants, models performed well at discriminating task from rest for new data from that same person (Within-person test: M=.97+/-.01; p<.001in contrast to a permuted label null). Models also predicted task state significantly greater than chance when tested in other participants (Between-person test: M=.68+/-.11; p<.001 for all participants; **Fig. 2a;** see **Fig. S5** for results and p-values per person).

We then evaluated prediction performance for specific tasks. Again, classifiers were trained on a subset of a given participant's data and tested on data from an independent session from the same person or another person using leave-one-session-out cross validation. We saw consistent results across tasks and participants (**Fig. 2b**), with significant task state decoding when tested on data from both the same (Within-person test: Coherence M = .87+/-.06; Memory M = .99+/-.01; Motor M = .98+/-.02; Semantic M = .94+/-.03; all p<.001) and other individuals (Between-person test: Memory = .89+/-.1; Motor M = .78+/-.15; Semantic: M = .66+/-.14; and Coherence: M = .57+/-.1; all p < .001). These results demonstrate that functional networks carry task state information, which can be used to make accurate predictions about state in the same and other participants.

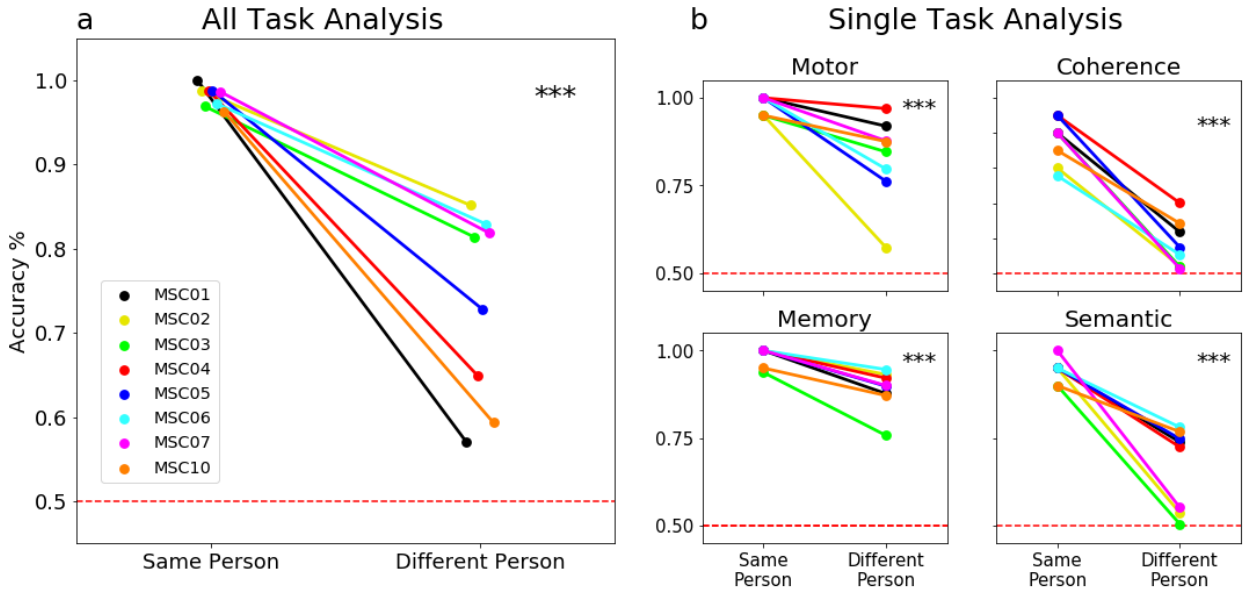

Figure 2. Differences in classifier accuracy within and between person (a) Performance of an individualized machine learning classifier for discriminating task from rest states when tested on independent data from either the same individual or different individuals (colored lines = different participants). Initial models contrasted all task states with equal samples of rest (N = 80 samples total per person). **(b)** Secondary analyses trained models to discriminate a single task from rest. On average, all models were able to predict task state significantly better than chance in the same person person ($p<0.001$ when compared to a permuted null with randomized task/rest labels) and in new people ($p<0.001$; red dashed lines = chance). However, model performance was significantly higher when tested on data from the same participant relative to other participants (*** $p<0.001$ compared to a permuted null on within/between person labels; All Tasks vs. Rest: effect size=.24; Motor vs. Rest=.15; Coherence vs. Rest=.30; Memory vs. Rest=.09; Semantic vs. Rest=.27;). Similar results were seen when tested with other machine learning algorithms (***Fig. S4***). For a more detailed breakdown of person to person performance see ***Fig. S5*.**

Task state prediction is higher within than between individuals

Although we were able to predict task state both within and between individuals, task prediction performance differed significantly in these two test sets. Classification was significantly more accurate when discriminating between all task states and rest within the same person vs. other individuals (mean effect size=.24, $p<0.001$ based on permutation, see *Methods*; **Fig. 2a**). These findings were robust to the number of samples included in the training set (**Fig. S1**, when testing on 16 - 80 samples): classification of task state generally improved for both the same and other people with more samples, but more dramatic improvements were seen when tested in the same person. In all cases, within person classification was higher than between person classification.

This effect remained consistent when predicting tasks separately (**Fig. 2b**). Within-person classification was significantly higher than between-person classification for all comparisons (Mean difference of within and between person performance: Motor=.15; Coherence=.30; Memory=.09; Semantic=.27, $p<.001$ for each separate comparison). Notably, the between-person test performance was similar to the performance seen for a more classic leave-one-*subject* (groupwise) classifier both for all tasks and single tasks (where training was performed on a session from many different participants and tested on a novel participant; **Fig. S2**). The significantly higher performance of classifiers tested on the same individual suggests that there are task network effects that are unique to individuals that do not generalize between people. These findings reveal that individualized within-person analyses unmask substantial information on how brain areas communicate during tasks.

To provide a more thorough investigation of classification accuracy we calculated task predictive value (TPV; correct labeling of task divided by any labeling of task) and rest predictive value (RPV; correct labeling of rest divided by any labeling of rest). When tested within the same individual, TPV and RPV scores were both consistently high (all tasks vs. rest, within person: TPV .97+/-.01 RPV .99+/-.006; single tasks vs. rest: see **Table S1**). More errors were seen for between person analyses (all tasks vs. rest, between person: TPV .69+/-.1 RPV .93+/-.07; single tasks vs. rest: see **Table S1**). The errors in state classification for new participants differed by classification procedure: in the all task vs. rest discrimination, rest was frequently mis-classified as a task, whereas in the single task analysis certain tasks (coherence, semantic) were frequently mis-classified as rest. Importantly, in all cases the within person analyses had both high TPV and RPV. These results suggest that classification across individuals can result

in biased errors dependent on the training set. Within person classification is less subject to these biases.

Individualized task state prediction is high across many tasks at once

Next we asked if individualized task state prediction could extend beyond a binary classification (task vs. rest) to classify varied task states at once. Using ridge regression we trained a multiclass model on data from a single individual to distinguish between the 5 different cognitive states (rest, motor, memory, semantic, coherence) and then tested this model on states from an independent session either from the same person or a new person (**Fig. 3**). In all participants, multiclass models were able to predict task state above chance for both within and between person test sets (Within person M=.83+/-.05, p<.001; Between person M=.53+/-.07, p<.001 relative to null permutation model). Again, however, we found that within-person classification significantly out-performed classification across individuals (mean effect size= 0.3, p<.001 based on permutation). An analysis of the errors in multiclass performance showed greater overall errors in between-person classification and a greater propensity to classify other tasks as rest (especially for the motor and semantic tasks, see **Fig. S6**). Thus, individualized classifiers can perform even fine-scale multiclass predictions with good accuracy, substantially out-performing more standard between-person approaches. These results suggest that individualized classifiers are able to capitalize on idiosyncratic task state information and may be useful in providing robust analyses of the commonalities and differences across task states.

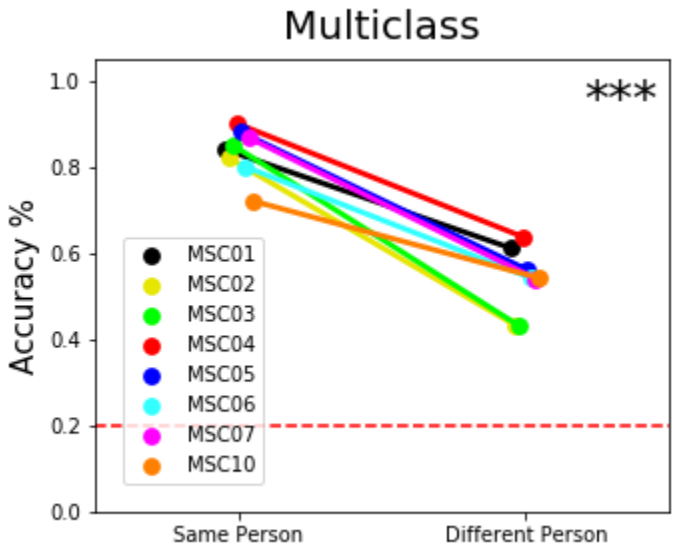

Figure 3. Accuracy using multiclass prediction to discriminate among five cognitive states. Average accuracy when testing the individualized classifier on new sessions from either the same participant or a different person. For a detailed breakdown of each individualized classifier see ***Fig. S7.*** Multiclass prediction was significant both in the same and new people (*red line = chance performance across 5 states)*, with higher performance in the same individual (**** p<0.001).*

Task state can be decoded from single networks

Next, we asked which brain network connections are most important for predicting cognitive states. We addressed this question through complementary analyses on feature weights and

feature selection procedures. First, we examined the anatomical distribution of the average feature weights for each brain region to better understand which features contributed the most to task state decoding. Feature weights were fairly consistent across folds for each individualized classifier (**Fig. 4a**). However, strong feature weights mapped on to different regions across people (**Fig. 4b**), likely underlying the variation in classification performance across people . Feature weights also varied somewhat across tasks in the separate single task classifiers (**Fig. S3**), indicating the presence of task-specific components to prediction and likely underlying the success of multi class prediction.

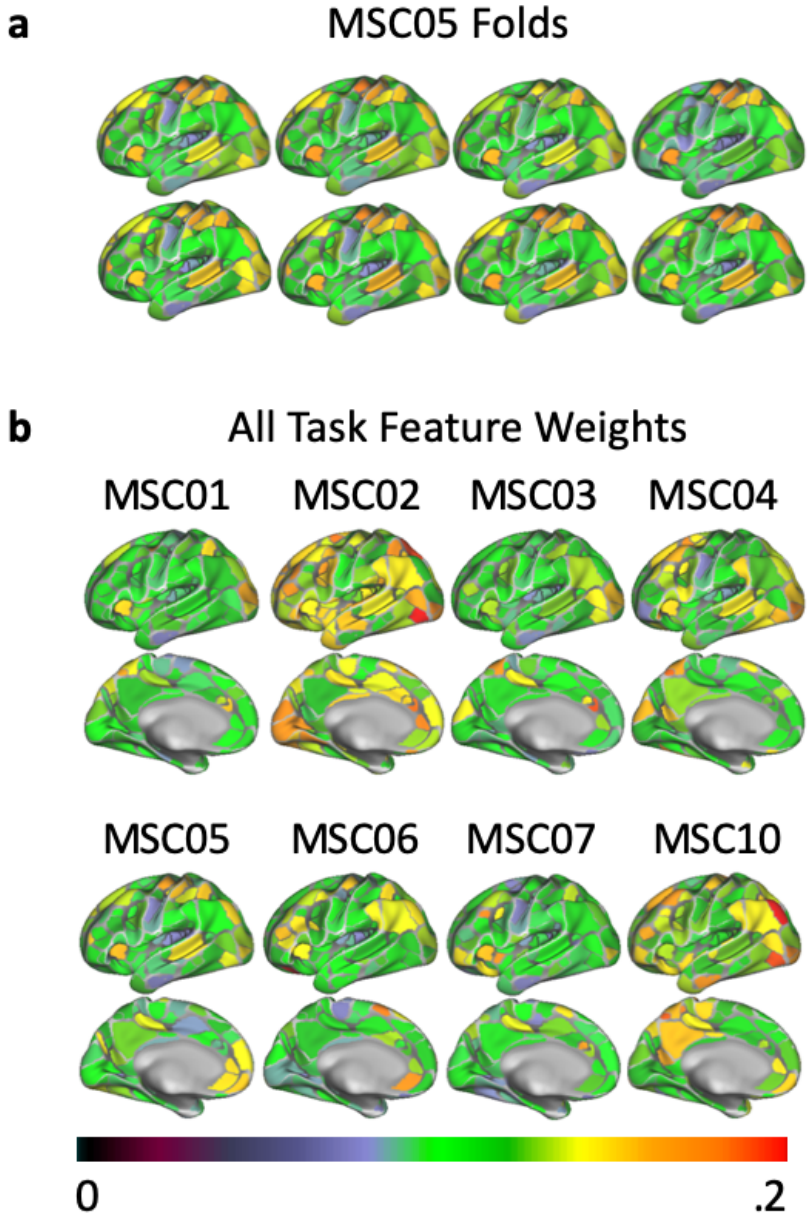

Figure 4. Feature Weight Analysis. Average absolute feature weights for each brain region for individualized classifiers built to discriminate between all task states and rest. (**a)** Feature weights for a single person (MCS05) for each fold. **(b)** Average feature weights for each individualized classifier. Note that while feature weights are fairly similar across folds within a person, they show variation across people.

Next we asked whether single networks would be sufficient for predicting task state in individualized classifiers. To address this question, we built classifiers restricted to features in specific blocks of network-to-network connections (**Fig. 5**). We found good performance for most classifiers based on single network blocks. For within-person analyses average block performance was 83%, (range 61%-95%); 99% of blocks were able to decode task state on their own above chance (SD=.008). In between-person analyses average block performance was 57% (range 51%-68%). 95% of blocks were able to decode task state above chance

(SD=.04).

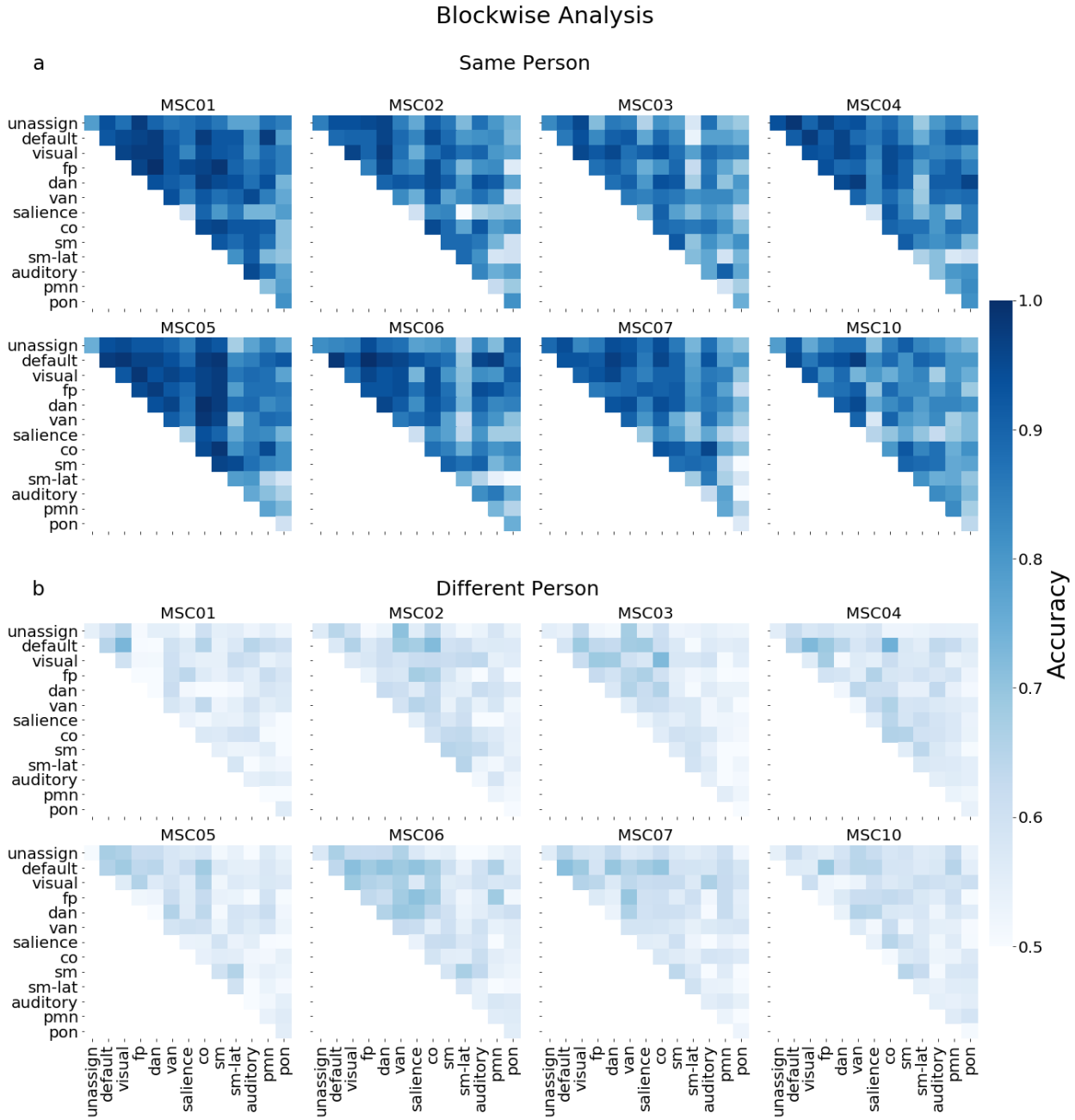

Figure 5. Single network block performance in decoding task state. Here we trained classifiers to decode all tasks from rest using subsets of network to network connections (each block was an independent classifier). We then tested the classifier on new data from either (a) the same person or (b) a new person. Due to low feature numbers prior to training the classifiers we standardized features (see Methods). Task state information was distributed across many networks of the brain: in within person analyses 99% of blocks could decode task state; in between person analyses 95%. Within person performance was consistently higher than between person classification.

This sufficiency test indicates that task state classification features can be found in almost all networks. While performance levels varied by network block, these network blocks also differed substantially in size (different numbers of regions in each network lead to different numbers of

features), and better performance was generally found for larger networks. To systematically test whether specific networks or simply higher feature numbers were important for classification performance, we contrasted performance of features extracted from single networks (all connections associated with regions from a single network) relative to feature sets that were randomly selected from the full connectome. Random feature sets numbering 10-50,000 features were tested. In all cases, classifier performance improved with more features. Moreover, at all feature numbers, within-person classifier performance outperformed between person classifier performance (see **Fig. 6**). Interestingly, there were no cases where single networks performed significantly better than random feature selection when feature numbers were matched. Instead, in the majority of cases, single networks performed similar or worse than randomly selected features. This suggests that task state information, including individual-specific aspects of that information, is distributed across multiple networks that a classifier can utilize in prediction. We hypothesize that prediction is better with random features because random feature selection can utilize more independent sources of information from different networks (within network features are more likely to share correlated information, as has previously been found for age prediction performance[33]).

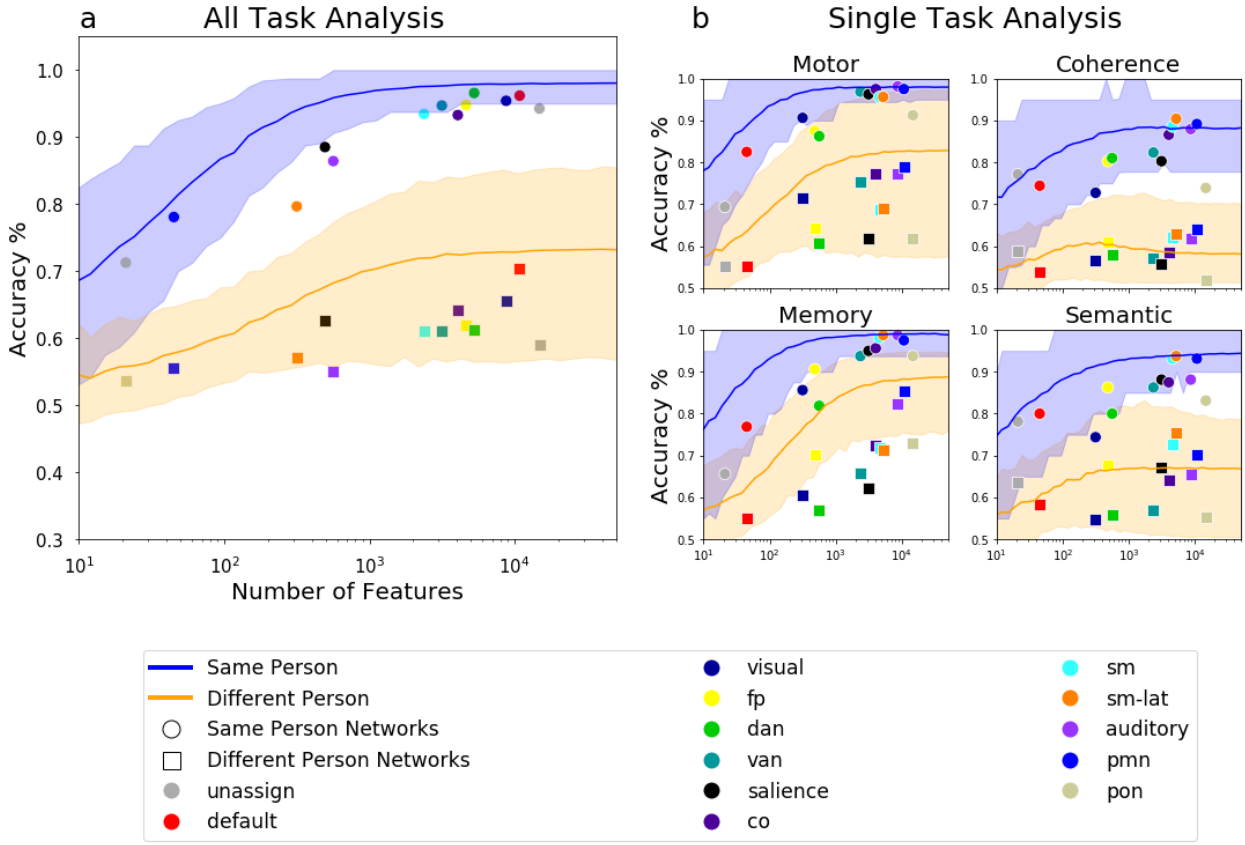

Figure 6. **Overview of feature selection.** Model classification performance when classifiers were trained on randomly selected features of different numbers, or features associated with a given network. Random feature number varied from 10-50,000, randomly indexed across all unique FC edges. Networks have been plotted on top at the position equivalent to their feature size. Accuracy is shown both for models tested on the same person (circle markers, blue line) or a different person (square markers, orange line). Error bars on the random feature selection line represent the 5th and 95th percentile across iterations. (**a**) Performance when training and testing to discriminate

all tasks from rest. (**b)** Performance when training and testing to discriminate a single task from rest. For a breakdown of participants see **Fig. S8**.

A within person advantage is seen even on individually-mapped network regions

Most past machine learning studies rely on between person analyses using a common standardized set of brain regions defined for group comparisons, which we demonstrate systematically underperform within-person models. This observation naturally raises questions about the source(s) of improved classification within-person. One possibility is that within person classification is driven by differences in the underlying spatial positions of networks across individuals; within a person, these are likely consistent over time [9,23,26] and may confer an advantage in task state decoding for that same person relative to other individuals who differ in their spatial network arrangement. Another possibility is that the magnitude of functional connectivity changes during tasks, rather than the spatial position, varies systematically across people.

To test whether variation in the spatial locations of networks across individuals drives differences in classifier performance, we shifted our analysis strategy. Instead of relying on common group parcellations and network definitions, we used individualized parcels and network definitions from Gordon et al[34] for each participant, derived using data-driven community detection algorithms specifically for that person. We then estimated the average functional connectivity within and between networks for each state in each person, creating a set of 14x14 network correlation matrices (while differing in specific topography, 14 common networks were consistently identified across individuals - see *Methods* and **Fig. 7a**). Classifiers were trained to distinguish all tasks from rest in each participant, and then tested on independent data from either that same participant or other participants. As with the other analyses reported above, individualized classifiers were able to significantly decode task from rest in both cases (Within-person test: M=.91+/-.04; Between-person test: M=.70+/-.05). Importantly, however, classifiers still performed significantly better on the same person than other people, even after matching for individualized network locations (mean effect size=.19; p<.001, **Fig. 7b**). This suggests that the individualized advantage for task state decoding is not driven only by differences in the spatial layout of networks, but may also be associated with differences in the magnitude of functional connectivity changes across people.

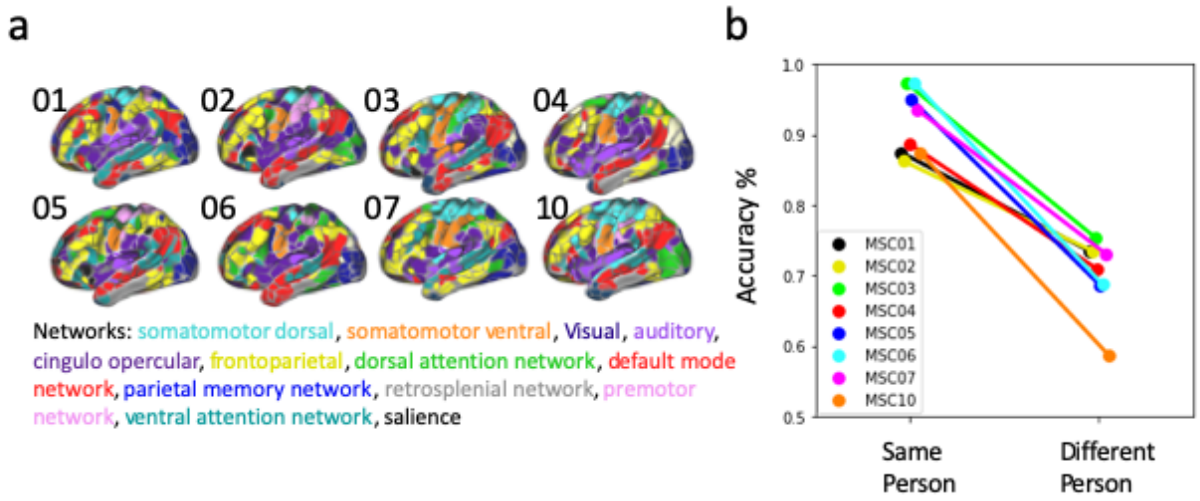

Figure 7. Classification performance built on individualized network parcellations. Comparison of performance of individualized classifiers on networks derived from individual parcellations. (**a**) Individual person parcellations and network definitions. FC matrices were calculated at the network level (network x network) using each participant's network definitions in order to align across people. These FC matrices are then used to train a classifier to decode all task states from rest as in previous analyses. **(b)** Individualized classifiers based on individualized networks were significantly able to decode task from rest in both the same and new participants, but performed consistently better for the same person ($p<.001$).

Discussion

A central question in cognitive neuroscience is how brain regions coordinate their actions during complex tasks. In this study, we applied machine learning to precision fMRI data to determine whether individual-specific classifiers can predict task state, and how these effects generalize to new people. We found that high quality precision fMRI data led to robust estimates of state differences in brain network interactions, permitting individualized classifiers to predict states across a variety of tasks, within and across individuals. However, classifiers tested within the same individual had significantly higher accuracy than those tested across individuals, suggesting the presence of idiosyncratic features that are obscured in cross-subject analysis. While single brain networks were able to predict task state alone, features for predicting task state were distributed across multiple systems. Based on these findings, we suggest that cognitive state leads to systematic and idiosyncratic changes in brain networks that are often masked in cross-subject analyses. Using an individualized, high data approach can help capture these features and improve our understanding of how brain networks are altered during tasks.

Task states alter networks in predictable ways

Task states have only relatively subtle effects on the intrinsic functional connectivity architecture [5,8]. However, prior work has consistently demonstrated that sensitive machine learning methods can effectively be used to predict task state from whole brain functional connectivity data[11,15,18,21,35–37]. In fact, it is possible to predict tasks designed to measure a variety of cognitive processes including attention [10,11,21,35–37], memory [11,15,17,20,38], language [17,20], and math processing [11] with accuracy levels well above chance. These investigations have typically been conducted

using cross-subject prediction, with classifiers trained on a subset of participants and then tested on left out subjects. This approach inherently assumes that task states will induce similar network modulations across participants.

Our work builds on these findings by showing that you can predict task state even using classifiers trained on a single person's data. These individualized classifiers were able to predict task state above chance even in other individuals (at similar levels to that seen with more typical cross-subject classifiers, see **Fig. S2**). These findings underscore that there are common, generalizable aspects to how brain networks are affected by task state. However, our findings also showed a significant improvement in prediction when classifiers were used to predict state within the same person. This suggests that an individualized approach may help to uncover additional novel features associated with brain states that are masked in typical cross-person analyses.

Cognitive effects on brain networks are strongly influenced by individual characteristics

Here, through a number of diverse analyses, we show a consistent benefit for within person classification in the prediction of task states. Classifiers showed significantly higher accuracy for predicting task state from data from the same person relative to new people. This finding was seen in each of the eight precision datasets we measured and was consistent across different tasks (**Fig. 2b**), networks (**Fig. 5**), subsets of samples (**Fig. S1**), and versions of the classifier (all task **Fig. 2a**, single task **Fig. 2b**, multiclass **Fig. 3**) and classifier implementations (ridge regression, logistic regression, support vector machines, **Fig. S4**). Consistent with past results using cross-subject approaches [10,11,15,17,20,21,35,36], individualized classifiers tested on new people were able to perform significantly above chance (and similar to that seen with classic cross-subject classifiers, **Fig. S2**; see previous section), indicating that our findings were not driven by general poor performance of the individualized classifiers. Rather, the large enhancement in within-person classification suggests that individualized classifiers have the potential to reveal novel, robust features of task states that are obscured in typical cross-person designs.

These findings align with past results indicating large differences in functional brain networks across people [22–26,30,34,39,40], as well as interactions between individual differences and changes in task states [9,32,41]. Here, we provide evidence that these individual differences have substantive impact on classification of task state. These results concord with past evidence using task evoked fMRI signals for classification, which also shows a within-person advantage [42–45]. Within-person classification may be necessary to reveal robust, impactful, but idiosyncratic features of brain network states-- observations that are likely to be central to enhancing our understanding of cognitive neuroscience. The growth of precision fMRI datasets sampling a variety of task conditions such as the Natural Scenes Database[46], the Individual Brain Charting dataset[47], and StudyForrest[48,49] will be instrumental in this regard.

What are the sources of these individual differences in classification? One possibility is that individual features may be driven by differences in the functional-anatomical topography of brain networks. For many of the analyses in this manuscript, we used a group parcellation [34].

Disruptions in classification could be driven to the extent that any individual showed deviations from this typical group architecture[23], consistent with suggestions that spatial topography is a major form of individual differences in brain networks[40]. Deviations in spatial topography may be important limitations in most past task state classification relying on cross-subject correspondence, but could be addressed in the future through the use of functional alignment procedures, such as hyperalignment, a method that has been successful at increasing model performance across participants [41–45]. However, we found that a within-person advantage persisted even after functionally aligning across people based on their own intrinsic network architecture (**Fig. 7**). This suggests that other sources of deviations, such as individual differences in the magnitude of task modulations may also contribute substantially to task network effects.

Individual task effects may be associated with differences in strategy for completing tasks [6,50]. As the tasks from the Midnight Scan Club were relatively simple and each participant was at ceiling performance, this hypothesis would require that even minor differences in strategies during a simple task could have measurable effects on brain networks. One exciting avenue of future work will be to evaluate how precision data can generalize to predict novel task states and diverse strategies. Studying individualized task network features across a variety of contexts may help untangle these possibilities and piece apart the functional relevance associated with different states.

Task state information is distributed across many large-scale systems

One question that arises from the current results is whether specific brain regions or connections are particularly influenced by task states. Prior work has suggested that tasks influence a range of within and between network connections[8], and focused on highlighting specific connections associated with various tasks [5,8,37,51]. Like with past studies, we also found that both within and between network connections contributed to decoding of task state. Interestingly, however, we found that task state information was very distributed across the connectome: almost every block of network to network connections was sufficient to decode task state (within a person), but these single networks didn't outperform random distributed feature selection when matched on feature number. These results are consistent with past work[33] which showed that distributed brain network features outperform features from single networks in predicting age, likely driven by the more diverse information that can be gained by sampling from many different systems. Our findings suggest that completing various tasks optimally recruits distributed systems rather than relying solely on any one network.

Importantly, regardless of the feature selection approach (single network, block, or random feature), we found that within-person tests consistently outperformed between-person tests. This finding not only underscores the robustness of our original result, but also suggests that idiosyncratic task network features are distributed throughout different brain regions. Indeed, analyses of the feature weights used in our machine learning models showed consistency of distributed network features across folds of a single person, but substantial deviations across people. Future deep analyses of the consistent and idiosyncratic features used across a variety

of task contexts has substantial potential to help clarify the contributions of diverse systems to completing complex tasks.

Practical considerations for individualize task state prediction

By necessity, individualized classifiers were built on a relatively small number of samples (80) and sessions per participant (10). However, the presence of eight participants allowed for eight independent models to replicate the individualized classifier results. These individualized classifiers showed many consistent results, including significant ability to decode task state within and across people, across tasks, and across feature subsets. Prior work has shed light on the effects of data quantity and reliability in FC matrices [52,53]. To study how data quantity affects individualized classifiers, we iteratively increased the number of samples (16-80 task/rest pairs) used in our analysis. We found that increasing the amount of samples in the training set influenced the models ability to accurately decode tasks from rest within a person, plateauing around 25 sample pairs (50 samples). However, the number of samples only modestly improved the model's ability to predict state in a new person. Similarly, we found that increasing numbers of features improved classification, even when drawn randomly from different positions in the functional connectome. This was particularly true for within-person classification, which showed substantial performance improvements plateauing at ~1000 features. This finding suggests that caution should be warranted in interpreting differences in classification between networks of different sizes. Notably, these results provide guidelines for the practical application of building individualized classifiers in future studies seeking to maximize within and between person prediction.

Applications of personalized state classifiers

Our current findings have focused on predicting cognitive states. One interesting question for future work will be to determine whether individualized classifiers will also be useful in predicting other types of states a person experiences such as arousal [51,54] and sleep [55,56], as well as clinically relevant mood states such as bipolar disorder, depression episodes, and hallucination states in schizophrenia. Given the highly replicated finding of individual variability in functional network topography across individuals [23,26,28,30,34,39] including in large community-based samples [24,25,40,57], it is likely that individualized classifiers may have unique advantages across a variety of contexts. If it is broadly true that idiosyncratic features are important for classifying states, then person-specific precision analyses such as the ones conducted in this study may uncover new important features for future study. Clinical applications may ultimately require a balance between precision applications which may offer unique advantages in identifying idiosyncratic features of diverse states and classifiers built to maximize cross-person generalization (e.g., large datasets with less precise measurements which can be used to capture large scale generalizable features across a population[58]). One exciting avenue for future research will be to understand whether there are common subsets of individuals that share idiosyncratic features [57], that can be used to improve cross-subject decoding while preserving sensitivity to individual differences. Uniting individualized approaches and targeted cross-subject studies may help to maximize our knowledge and prediction power.

Conclusion

In this study, we asked whether people differ in how their brain interactions are altered by cognitive state. Using a machine learning approach, we find that it is possible to classify task state from a classifier trained on a single individual's multi session data. Classification was successful both for independent data from that person as well as for new people, but a big boost in accuracy was seen for classification within-person. This result was consistent across datasets, tasks, and classifier approaches. Feature selection analyses demonstrated that task state information is distributed across large-scale systems, with single networks performing well at classifying task state but not out-performing randomly selected features. These findings have important implications for understanding how cognitive state influences brain networks, emphasizing the importance of considering individual differences to reveal the full magnitude of variations that are present.

METHODS

Overview

Our goal in this project was to investigate whether there are individual differences in how brain networks integrate task states. We used the precision fMRI Midnight Scan Club Dataset (MSC) dataset for this analysis [23]. This dataset and its processing have been described extensively in [9,23], and we include a summary of the components relevant to this study below. We created functional connectivity matrices from task and rest data for each participant, task, session in the MSC. We then used a machine learning approach to train classifiers to distinguish rest from task on a subset of sessions and then tested these classifiers on held out data from either the same participant or a new participant. Variations on this analysis were conducted with different classifiers, tasks, and feature subsets, as described below.

Dataset

The MSC dataset includes data from 10 individuals (5 females, ages 24-34) with ten fMRI sessions each. Each session occurred on a separate day, beginning at midnight. Sessions were completed within 7 weeks for all participants. Each session included fMRI data for a resting-state scan and 4 tasks (see Task Designs and Analysis). Participants provided informed consent and procedures were approved by Washington University Institutional Review Board and School of Medicine Human Studies Committee. Participant MSC08 was excluded from the study due to high levels of head motion and sleep (more details in [23]). MSC09 was also excluded due to insufficient samples in the motor task (only 5 of the sessions had sufficient low motion data, see Functional Connectivity Processing). This resulted in eight participants used in the final analyses.

MRI Acquisition

MRI data were acquired on a 3T Siemens Trio. Four T1-weighted images (sagittal, 224 slices, 0.8 mm isotropic resolution, TE = 3.74 ms, TR = 2.4s, TI = 1.0s, flip angle = 8 degrees), four T2-weighted images (sagittal, 224 slices, 0.8 mm isotropic resolution, TE = 479 ms, TR = 3.2s) were collected for each participant across two separate days. Functional MRI data was

collected using a gradient-echo EPI BOLD sequence (TE =27ms, TR = 2.2 s, flip angle = 90, voxels =isotropic 4mm^3, 36 axial slices) during each of 10 sessions. The same sequence was used for task and resting-state data. The same parameters were also used to collect a gradient echo field map acquired during each session for de-warping of the functional data.

Task Designs and Analysis

Functional MRI data were collected during five conditions described briefly below (for a more thorough description see [9,23]). Task activations were modeled with a generalized linear model (GLM) using in-house software written in IDL (Research Systems, Inc.). GLM residuals were used for time-series correlations, following a background connectivity approach [59,60].

Resting-state: Each session started with a 30 min resting state scan, wherein participants were asked to stare at a fixation on the center of a black screen.

Motor Task: The motor task was adapted from the Human Connectome Project [61] and consisted of a block design in which participants are told to move either left or right hand, left or right feet, or tongue based on visual cues. This included two runs in each session (7.8 min). Each block began with a 2.2s cue, followed by a fixation caret flashing every 1.1s to signal a movement. Each run included two blocks for each type of movement and three fixation blocks (15.4s). For this GLM each motor condition was modeled separately with a block regressor convolved with a hemodynamic response function.

Semantic Task: Each session included two runs (14.2 min) of a mixed block/event related design modeled on tasks in [62]. This included four blocks per run, two for semantic, two for coherence (see below for coherence). The semantic task was a verbal discrimination task where participants were visually presented with short words and asked to identify if they were nouns or verbs (50% nouns, 50% verbs). Task blocks began with a 2.2s cue indicating which task was to be conducted in the following block; after this 30 individual trials were presented. Trials consisted of words presented for 0.5s with jittered 1.7-8.3s intervals. Participants responded to whether each word was a verb or a noun. After, participants were presented with a cue (2.2s) indicating the end of a block, with 44s fixation periods separating each block. The semantic and coherence tasks were modeled together in a single mixed block/event-related GLM. Separate regressors were included for each task block and for events (start and end cues in each task, correct and incorrect trials of different types). Events were modeled with delta functions for 8 separate time-points to model the time course of responses using an FIR approach [63].

Coherence Task: Coherence task blocks were interleaved with the semantic task in the same mixed runs and followed the same timing structure and analysis as the semantic task. Individual trials consisted of arrays of Glass-like patterns, with white dots on a black screen that varied in arrangement (0% or 50% coherence to a concentric arrangement, both presented with 50% frequency) [64]. Participants were told to identify dot patterns as concentric or random.

Memory Task: The incidental memory task employed an event-related design with 3 runs per session (15min). A single run was collected for each type of stimulus (faces, scenes, words).

During trials of the run, participants made binary decisions to categorize stimuli (male/female for faces, indoor/outdoor for scenes, abstract/concrete for words). In each run participants viewed 24 images repeated 3 times. Images were presented for 1.7s, with jittered 0.5 – 4.9s intervals. For the memory GLM, trials were modeled based on stimulus type and number of repetitions with delta functions across 8 time-points.

Structural MRI Processing and Surface Registration

The procedure used to process the structural MRI data and align with the surface can be found in extensive detail elsewhere [23,52]. The high-resolution structural T1 images were first aligned and averaged together, then registered to a volumetric Talairach atlas using an affine transform. These averaged template T1s were used to generate a cortical surface in Freesurfer [65], which was hand-edited to improve surface accuracy. The native freesurfer surface was registered to fs_LR_32k space using a similar procedure described in [66].

Functional MRI Pre-processing

All fMRI data first underwent pre-processing in the volume to correct for artifacts and align data. This included slice-timing correction, frame-to-frame alignment for motion correction, and intensity normalization to mode 1000. Functional data was registered to the T2 image, which was registered to the high-resolution T1 that had been registered to template space. Functional data underwent distortion correction (see [23] for more details). Registration, atlas transformation, resampling to 3mm isotropic resolution, and distortion correction were combined and applied in a single transformation [67]. All remaining operations were completed on the atlas transformed and resampled data.

Task fMRI data were processed in the volume using a GLM, with the approach and regressors described above (*Task Designs and Analysis*). The residuals from this model were used to compute task functional connectivity (as in [9]), following the same steps as rest (described below).

Function Connectivity Processing

To reduce artifacts, data underwent initial functional connectivity (FC) processing in the volume, with methods described in detail in [23,68]. This involved demeaning and detrending the data, nuisance regression of motion parameters and signals from white matter, cerebrospinal fluid, and global signal, as well as their derivatives. High motion frames (framewise displacement, [69] FD > .2mm along with sequences containing less than 5 contiguous low motion frames, the first 30s of each run, and runs with <50 low motion frames) were censored and replaced with power spectral matched data. For participants, MSC03 and MSC10 motion parameters were low-pass filtered (<0.1Hz) before FD calculation. Following motion censoring, bandpass filtering (.009Hz-.08Hz) was applied. Cortical functional data was then registered to the surface and smoothed (Gaussian kernel, sigma=2.55mm) with 2D geodesic smoothing on the surface and 3D Euclidean smoothing for subcortical volumetric data (see [9] for more details). Finally, interpolated high motion frames were removed before functional connectivity analysis. To be included, each run was required to include at least 25 low-motion volumes and each session

required at least 50 volumes total per task; we removed 5 sessions based on these criteria. Task data was limited to task periods for each run (excluding fixation frames).

Functional Networks

Regions and Systems: This study examined functional connectivity among 333 cortical parcels defined based on boundary-mapping techniques in a large group of independent participants [34]. These 333 parcels are divided into 12 functional systems: somatomotor (SM), somatomotor lateral (SM-lat), visual (Vis), auditory (Aud), cingulo opercular (CO), salience (Sal), frontoparietal (FP), dorsal attention (DAN), ventral attention (VAN), default mode (DMN), parietal memory (PMN), and retrosplenial (RSP). Low signal regions that grouped poorly into a system were put in an ‘‘unassigned’’ group.

Functional Connectivity (FC): FC was computed by averaging the BOLD time course within each parcel, after removing censored and interpolated frames, and computing linear correlations between the time-series of each pair of parcels. Task data were limited to task periods within each run (i.e., excluding fixation periods). FC values were Fisher transformed for normality. FC was represented with a parcel by parcel functional network matrix, sorted by system; edges (FC values for a particular region-to-region pair) along the diagonal blocks represent within-system correlations, and edges in the off-diagonal blocks represent between-system correlations. Edges taken from the upper triangle of this matrix represent all unique pairwise parcel relationships resulting in 55,278 values for each task of a given session; combining all tasks and rest sessions for a given participant provided 80x55,278 values. Unique edges were used as features in the machine learning classifier.

Machine Learning

We employed a machine learning approach at the individual level where models were trained to classify task state using data from a single participant, which we call individualized classifiers. Models were trained and tested on independent data using a leave-one-session-out cross-validation framework, iteratively leaving out data (both task and rest) from an entire session for training and then testing on either this left out data or data from a single session in another person (for schematic see **Fig. 1**). Classification was performed using different classifiers, subsets of tasks, and subsets of features as specified below.

Ridge Regression Classification: We used ridge regression for classification due to the high collinearity among FC edges and the relatively small number of samples. Ridge regression implements a shrinkage estimator that can improve performance in the presence of multicollinearity [70]. This pipeline was implemented using Python’s Scikit-learn package [71]. As a control, we also conducted analyses using support vector classification and logistic regression and found that all models resulted in similar patterns of performance across conditions. (see **Fig. S1**).

Binary classification of task state across people: Our first goal was to determine the extent to which a model trained on one participant would be able to differentiate task from rest both in that person and in new people.

Training and testing procedure: We conducted two versions of binary classification to examine task state prediction from FC data. The "*All Task*" classifier was trained to distinguish all tasks (memory, motor, semantic, coherence) from rest in a single binary classifier. This classifier was used as a higher sample test of the ability for classifiers to distinguish task states from rest. This classifier had 80 total samples per participant: 40 task samples (10 sessions of motor, coherence, semantic, and memory tasks) paired with 40 rest samples (rest split into four 7 min increments for each of 10 sessions; we split rest samples in order to balance our training set and avoid biasing the model towards the task samples). For each individual, we then trained a classifier to predict task state (task vs rest) based on FC using a leave one session out cross-validation scheme as described above (90% training data, 10% testing per fold). Each model was also tested on a single session from other participants (resulting in 7 "other person" tests per trained model). This training/testing process was repeated for all folds and all participants.

The second "*Single Task*" binary classifier classified a single task at a time (e.g., memory vs. rest). This classifier allowed us to determine whether the All Task results were consistent for each specific task condition. The same procedure was used as for the "All Task" classifier, except in this case each classifier was based on a total of 20 samples per participant (10 from a single task, 10 from rest). Models were again trained on 90% of the data (9 sessions) and tested on either left out data from that same participant (1 session) or another participant. The procedure of training/testing an individual classifier was repeated for all 4 tasks (motor, memory, coherence, semantic).

Model evaluation: Classifier accuracy (total percentage of test samples classified correctly), task predictive value (TPV; correct labeling of task divided by any labeling of task), and rest predictive value (RPV; correct labeling of rest divided by any labeling of rest) were used to evaluate model performance for each individualized classifier. For each classifier, test performance was averaged over folds. Data is presented separately for each participant (each individualized classifier), representing an 8-fold replication of model training and testing on independent people. Performance values were contrasted with permuted null models (see below) to determine significance.

Model significance: First, we tested whether the performance of each classifier was significantly greater than chance when compared to a permuted null model. For each model, we trained a null classifier using the same FC data but permuted how samples were labeled as task or rest. The null classifiers were subsequently tested on held out data using the same cross-validation framework described above for each type of classifier. We repeated this process 1,000 times for each task, averaging permuted and true accuracies across all individualized classifiers to compute an omnibus statistic. We then calculated permutation-based p-values by comparing the average true classification accuracy relative to the distribution of classifier accuracy of the null classifiers (number of permutations that exceeded true model performance/1000). Additional tests were run per individualized classifier, comparing that classifier to a permuted null sample.

Second, we tested whether the performance of a classifier significantly differed when applied to independent data from the same individual vs. a different individual. We again used a permutation approach, in this case randomly shuffling the same-person or different-person labels of the test sets, then calculating the difference in accuracy scores. We repeated this process 1,000 times to create a null distribution, after which the true difference was contrasted with the null distribution and used to estimate a permutation-based p-value.

Groupwise cross-subject models: To evaluate how these measures compare to a standard machine learning approach we conducted a leave-one-*subject*-out analysis in which we trained the classifier using a single session of data across all participants in the dataset and tested on the left out participant (training 7 participants, testing 1 participant on the same session). We then repeated this process across all participants and across all sessions. We conducted cross-subject models both for decoding all tasks from rest (as in the All Task Analysis) and models for discriminating a single task from rest (as the Single Task Analysis). Model performance is reported as the average accuracy across all folds.

Multiclass classification of task state across people: We next asked if finer-scale discriminations of task state could also be decoded from FC. To address this question we trained a *Multiclass* classifier to discriminate among all 5 of the separate states (rest, motor, memory, semantic, coherence; chance = 20%) . We then repeated the same procedure as above for cross-validation (leave-one-session-out for a given participant, then test on the same or a different participant) and model evaluation and significance. For this analysis we also created confusion matrices of the test performance to better assess errors .

Classification of models built on individualized network maps: We conducted an additional set of classification analyses to evaluate whether differences in participant performance could be attributed to differences in underlying spatial layout of large-scale networks. To address this question, we reconducted our classification based on individually defined network layouts. Previously published [23] individually defined parcels and network assignments for the MSC dataset were used for this analysis. These parcels and networks were defined independently for each participant using data-driven procedures. Importantly, previous results show that these individual network assignments reach high reliability with the precision fMRI data employed here[23]. We then calculated functional connectivity among each of these individually defined regions for each state. Finally, we averaged across each of 14 networks that could be identified consistently in each individual (default mode network, visual, frontoparietal, dorsal attention network, premotor network, ventral attention network, salience, cingulo opercular network, somatomotor dorsal, somatomotor ventral, auditory, parietal memory network, and retrosplenial network; as in [9] Supp. Fig. S7). This resulted in FC matrices for each person, task, and session with a 14 x 14 network dimensionality - notably while the dimensionality and network labeling was matched across participants, each person's networks were defined separately to optimize accounting for that person's specific spatial layout of networks. We then extracted the features from these matrices and then conducted the same set of individualized model training and testing as described in the binary classification section above.

Feature Analyses

Next we asked if the classification of task state was dependent on specific features or networks. We took two distinct approaches to address this question.

Feature weight analysis: First, we examined the feature weights from our individualized classifiers. To do this we extracted the feature weights for the *All Task* binary classifier (see above) averaged for each region (row of the correlation matrix; averages were taken on the absolute values of the original weights). We qualitatively examined the variability of average feature weights across folds to look at consistency within a person. We then examined the variability of average feature weights across individuals (in this case averaged over folds) to look at consistency of feature weights across people. We also conducted the same procedure when analyzing feature weights at the single task level.

Feature Selection Analysis: Secondly, we conducted additional quantitative classification analyses with only subsets of features to determine which features were sufficient for classification. Three forms of feature selection were performed: (1) features were selected from specific blocks of network to network connections (e.g., all of the features associated with connections between the frontoparietal network and the default mode network). Due to low feature size associated with specific blocks we standardized the data by removing the mean and scaling to unit variance of the training set. We then transformed all test sets using the mean and standard deviation from the training set [71]. (2) To examine properties for full networks, features were also selected for entire network rows (e.g., FC in all matrix rows associated with default mode network regions), and finally (3) as a comparison, random features numbering 10 – 50000 were randomly selected from the full matrix (this process was repeated 1,000 times). Model training and testing was then conducted as described in the binary classification section for both the *All Task* and *Single Task* analyses. This process was repeated independently for each feature subset and for each separate participant.

Data Quantity Dependence

Since data quantity has been demonstrated in the past to affect the reliability of FC matrices [53,72] we examined whether model performance depended on sample number (i.e., the number of FC matrix pairs which were included in the training data). To determine how the number of samples in the training set influenced model performance, we re-built the *All Tasks* classifier described above using varying amounts of sessions from a single individual. Beginning with 80 samples of FC matrices (40 task and 40 rest) for each person, we then sub-selected between 16 – 80 samples (in matched task/rest pairs) and repeated model training and classification, using a leave-one-session-out cross validation approach. We also tested the model on each of the other participants' task and rest data. This training/testing process was repeated for all folds and all participants. This process was repeated 1,000 times for different random pairs of samples per individualized classifier.

Data and Code Availability

Data is publicly available (https://openneuro.org/datasets/ds000224). Code for analysis related to MSC processing can be found at https://github.com/MidnightScanClub. Code related to the analysis in this paper will be located at https://github.com/GrattonLab/Porteretal_taskprediction.

REFERENCES

1. Dworkin, J. D. *et al.* The extent and drivers of gender imbalance in neuroscience reference lists. *Nature Neuroscience* **23**, 918–926 (2020).
2. Zhou, D. *et al. dalejn/cleanBib: v1.1*. (Zenodo, 2020). doi:10.5281/ZENODO.4062888.
3. Ambekar, A., Ward, C., Mohammed, J., Male, S. & Skiena, S. Name-ethnicity classification from open sources. in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09* 49 (ACM Press, 2009). doi:10.1145/1557019.1557032.
4. Bertolero, M. A., Yeo, B. T. T. & D'Esposito, M. The modular and integrative functional architecture of the human brain. *PNAS* (2015) doi:10.1073/pnas.1510619112.
5. Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S. & Petersen, S. E. Intrinsic and task-evoked network architectures of the human brain. *Neuron* **83**, 238–251 (2014).
6. Fox, M. D. & Raichle, M. E. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience* **8**, 700–711 (2007).
7. Cole, M. W. *et al.* Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat Neurosci* **16**, 1348–1355 (2013).
8. Gratton, C., Laumann, T. O., Gordon, E. M., Adeyemo, B. & Petersen, S. E. Evidence for Two Independent Factors that Modify Brain Networks to Meet Task Goals. *Cell Rep* **17**, 1276–1288 (2016).
9. Gratton, C. *et al.* Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. *Neuron* **98**, 439-452.e5 (2018).
10. Alnæs, D. *et al.* Attentional load modulates large-scale functional brain connectivity beyond the core attention networks. *Neuroimage* **109**, 260–272 (2015).

11. Gonzalez-Castillo, J. *et al.* Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. *PNAS* **112**, 8762–8767 (2015).
12. Greene, A. S., Gao, S., Scheinost, D. & Constable, R. T. Task-induced brain state manipulation improves prediction of individual traits. *Nature Communications* **9**, 2807 (2018).
13. Greene, A. S., Gao, S., Noble, S., Scheinost, D. & Constable, R. T. How Tasks Change Whole-Brain Functional Organization to Reveal Brain-Phenotype Relationships. *Cell Rep* **32**, 108066 (2020).
14. Rosenberg, M. D. *et al.* Functional connectivity predicts changes in attention observed across minutes, days, and months. *PNAS* **117**, 3797–3807 (2020).
15. Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V. & Greicius, M. D. Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cereb Cortex* **22**, 158–165 (2012).
16. Wu, E. X. W. *et al.* Overlapping attentional networks yield divergent behavioral predictions across tasks: Neuromarkers for diffuse and focused attention? *Neuroimage* **209**, 116535 (2020).
17. Krienen, F. M., Yeo, B. T. T. & Buckner, R. L. Reconfigurable task-dependent functional coupling modes cluster around a core functional architecture. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130526 (2014).
18. Chen, J. *et al.* Shared and unique brain network features predict cognition, personality and mental health in childhood. *bioRxiv* 2020.06.24.168724 (2020) doi:10.1101/2020.06.24.168724.
19. Elliott, M. L. *et al.* General functional connectivity: Shared features of resting-state and task fMRI drive reliable and heritable individual differences in functional brain networks. *Neuroimage* **189**, 516–532 (2019).
20. Jiang, R. *et al.* Task-induced brain connectivity promotes the detection of individual

differences in brain-behavior relationships. *NeuroImage* **207**, 116370 (2020).

21. Rosenberg, M. D. *et al.* A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience* **19**, 165–171 (2016).
22. Braga, R. M. & Buckner, R. L. Parallel Interdigitated Distributed Networks within the Individual Estimated by Intrinsic Functional Connectivity. *Neuron* **95**, 457-471.e5 (2017).
23. Gordon, E. M. *et al.* Precision Functional Mapping of Individual Human Brains. *Neuron* **95**, 791-807.e7 (2017).
24. Kong, R. *et al.* Spatial Topography of Individual-Specific Cortical Networks Predicts Human Cognition, Personality, and Emotion. *Cerebral Cortex* **29**, 2533–2551 (2019).
25. Mueller, S. *et al.* Individual variability in functional connectivity architecture of the human brain. *Neuron* **77**, 586–595 (2013).
26. Seitzman, B. A. *et al.* Trait-like variants in human functional brain networks. *PNAS* **116**, 22851–22861 (2019).
27. Gratton, C. *et al.* Defining Individual-Specific Functional Neuroanatomy for Precision Psychiatry. *Biological Psychiatry* **88**, 28–39 (2020).
28. Braga, R. M., Van Dijk, K. R. A., Polimeni, J. R., Eldaief, M. C. & Buckner, R. L. Parallel distributed networks resolved at high resolution reveal close juxtaposition of distinct regions. *Journal of Neurophysiology* **121**, 1513–1534 (2019).
29. Braga, R. M., Van Dijk, K. R. A., Polimeni, J. R., Eldaief, M. C. & Buckner, R. L. Parallel distributed networks resolved at high resolution reveal close juxtaposition of distinct regions. *Journal of Neurophysiology* **121**, 1513–1534 (2019).
30. Finn, E. S. *et al.* Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience* **18**, 1664–1671 (2015).
31. Salehi, M., Karbasi, A., Barron, D. S., Scheinost, D. & Constable, R. T. Individualized functional networks reconfigure with cognitive state. *Neuroimage* **206**, 116233 (2020).
32. Xie, H. *et al.* Whole-brain connectivity dynamics reflect both task-specific and

individual-specific modulation: A multitask study. *Neuroimage* **180**, 495–504 (2018).

33. Nielsen, A. N., Barch, D. M., Petersen, S. E., Schlaggar, B. L. & Greene, D. J. Machine Learning With Neuroimaging: Evaluating Its Applications in Psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **5**, 791–798 (2020).
34. Gordon, E. M. *et al.* Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cerebral Cortex* **26**, 288–303 (2016).
35. Fong, A. H. C. *et al.* Dynamic functional connectivity during task performance and rest predicts individual differences in attention across studies. *Neuroimage* **188**, 14–25 (2019).
36. Madhyastha, T. M., Askren, M. K., Boord, P. & Grabowski, T. J. Dynamic connectivity at rest predicts attention task performance. *Brain Connect* **5**, 45–59 (2015).
37. Rosenberg, M. D., Hsu, W.-T., Scheinost, D., Todd Constable, R. & Chun, M. M. Connectome-based Models Predict Separable Components of Attention in Novel Individuals. *J Cogn Neurosci* **30**, 160–173 (2018).
38. Braun, U. *et al.* Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proc Natl Acad Sci USA* **112**, 11678–11683 (2015).
39. Gordon, E. M. *et al.* Individual-specific features of brain systems identified with resting state functional correlations. *Neuroimage* **146**, 918–939 (2017).
40. Bijsterbosch, J. D. *et al.* The relationship between spatial configuration and functional connectivity of brain regions. *Elife* **7**, (2018).
41. Geerligs, L., Rubinov, M., Cam-CAN & Henson, R. N. State and Trait Components of Functional Connectivity: Individual Differences Vary with Mental State. *J. Neurosci.* **35**, 13949–13961 (2015).
42. Haxby, J. V. *et al.* A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* **72**, 404–416 (2011).
43. Haxby, J. V., Guntupalli, J. S., Nastase, S. A. & Feilong, M. Hyperalignment: Modeling shared information encoded in idiosyncratic cortical topographies. *eLife* **9**, e56601 (2020).

44. Guntupalli, J. S., Feilong, M. & Haxby, J. V. A computational model of shared fine-scale structure in the human connectome. *PLoS Comput Biol* **14**, e1006120 (2018).

45. Guntupalli, J. S. *et al.* A Model of Representational Spaces in Human Cortex. *Cereb. Cortex* **26**, 2919–2934 (2016).

46. Naselaris, T., Allen, E. & Kay, K. Extensive sampling for complete models of individual brains. *Current Opinion in Behavioral Sciences* **40**, 45–51 (2021).

47. Pinho, A. L. *et al.* Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping. *Sci Data* **5**, 180105 (2018).

48. Hanke, M. *et al.* A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie. *Sci Data* **1**, 140003 (2014).

49. Hanke, M. *et al.* A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation. *Sci Data* **3**, 160092 (2016).

50. Pearce, T. M. & Moran, D. W. Strategy-dependent encoding of planned arm movements in the dorsal premotor cortex. *Science* **337**, 984–988 (2012).

51. Shine, J. M. *et al.* The Dynamics of Functional Brain Networks: Integrated Network States during Cognitive Task Performance. *Neuron* **92**, 544–554 (2016).

52. Laumann, T. O. *et al.* Functional System and Areal Organization of a Highly Sampled Individual Human Brain. *Neuron* **87**, 657–670 (2015).

53. Noble, S. *et al.* Influences on the Test–Retest Reliability of Functional Connectivity MRI and its Relationship with Behavioral Utility. *Cerebral Cortex* **27**, 5415–5429 (2017).

54. Li, M. *et al.* Transitions in information processing dynamics at the whole-brain network level are driven by alterations in neural gain. *PLoS Comput Biol* **15**, e1006957 (2019).

55. Tagliazucchi, E. & Laufs, H. Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep. *Neuron* **82**, 695–708 (2014).

56. Tagliazucchi, E. & van Someren, E. J. W. The large-scale functional connectivity correlates of consciousness and arousal during the healthy and pathological human sleep cycle.

NeuroImage **160**, 55–72 (2017).

57. Seitzman, B. A. *et al.* Trait-like variants in human functional brain networks. *Proc Natl Acad Sci USA* **116**, 22851–22861 (2019).
58. Satterthwaite, T. D. *et al.* Neuroimaging of the Philadelphia neurodevelopmental cohort. *Neuroimage* **86**, 544–553 (2014).
59. Al-Aidroos, N., Said, C. P. & Turk-Browne, N. B. Top-down attention switches coupling between low-level and high-level areas of human visual cortex. *PNAS* **109**, 14675–14680 (2012).
60. Fair, D. A. *et al.* A method for using blocked and event-related fMRI data to study "resting state" functional connectivity. *NeuroImage* **35**, 396–405 (2007).
61. Barch, D. M. *et al.* Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* **80**, 169–189 (2013).
62. Dubis, J. W., Siegel, J. S., Neta, M., Visscher, K. M. & Petersen, S. E. Tasks Driven by Perceptual Information Do Not Recruit Sustained BOLD Activity in Cingulo-Opercular Regions. *Cereb Cortex* **26**, 192–201 (2016).
63. Ollinger, J. M., Shulman, G. L. & Corbetta, M. Separating processes within a trial in event-related functional MRI I. The Method. *Neuroimage* **13**, 210–217 (2001).
64. Glass, L. Moiré Effect from Random Dots. *Nature* **223**, 578–580 (1969).
65. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage* **9**, 179–194 (1999).
66. Glasser, M. F. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80**, 105–124 (2013).
67. Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* **23**, S208–S219 (2004).
68. Power, J. D. *et al.* Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* **84**, (2014).

69. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **59**, 2142–2154 (2012).

70. Fitrianto, A. & Yik, L. C. Performance Of Ridge Regression Estimator Methods On small sample size by varying correlation coefficients: a simulation study. *Journal of mathematics and Statistics* **10**, 25 (2014).

71. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **12**, 2825–2830 (2011).

72. Laumann, T. O. *et al.* Functional System and Areal Organization of a Highly Sampled Individual Human Brain. *Neuron* **87**, 657–670 (2015).

Supplemental information

Training Task	Same Person			Different Person			Within vs. between significance
	Accuracy	TPV	RPV	Accuracy	TPV	RPV	
All Tasks vs. Rest	.98 (.001)***	.97(.001)	.99(.0008)	.73(.01)***	.69(.01)	.93(.01)	p<.001
Individual Networks, All Tasks vs. Rest	.81(.005)***	.86(.008)	.88(.008)	.70(.01)***	.76(.02)	.71(.01)	p<.001
Multiclass	.83(.01)***			.53(.02)***			p<.001
Coherence vs. Rest	.87(.02)***	1(0)	.81(.02)	.58(.01)***	.98(.009)	.55(.01)	p<.001
Memory vs. Rest	.98(.009)***	1(0)	.97(.01)	.88(.01)***	.89(.01)	.93(.01)	p<.001
Motor vs. Rest	.98(.009)***	1(0)	.96(.01)	.82(.01)***	.99(.003)	.78(.02)	p<.001
Semantic vs. Rest	.94(.01)***	1(0)	.90(.01)	.66(.01)***	.99(.001)	.62(.01)	p<.001

Table S1. Summary table of classification performance (accuracy=M(SE)) on cross-person tests, with training on one person and testing on either the same person or a different person. Task predictive value (TPV) represents the classifier's ability to accurately label task divided by any labeling of task. Rest predictive value (RPV) represents the classifier's ability to accurately label rest divided by any labeling of rest. Model significance was determined relative to a random null created through permutation testing *p<.05, **p<.01, ***p<.001. The final column shows the difference in model performance for within-person and between-person tests (computed with permutation testing)). All Task vs. rest: binary classifier trained to distinguish all tasks from rest. Individual Network All Tasks vs Rest:similar to all task vs. rest, but used networks derived from individual parcellations (Fig. 7). Multiclass: classification models trained to distinguish between all five states (coherence, memory, motor, semantic, and rest). Coherence, memory, motor, semantic: classifiers built to distinguish a single task from rest. These analyses are described in detail in the Methods.

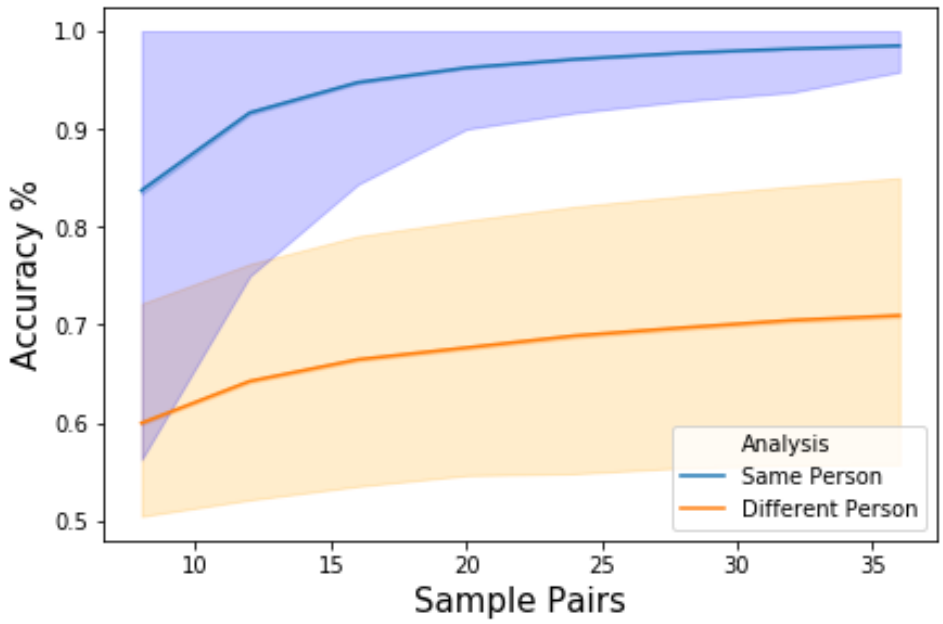

Figure S1. **Data Quantity Analysis**. Classifier performance on discriminating all tasks vs. rest for data tested on the same (blue) or a different person (orange) . Training samples were increased iteratively in pairs (rest and task) from 16 samples to 80 samples. This process was repeated with 1,000 random training sample iterations. Error bars represent the 5th and 95th percentile at each sample pair. In all cases, classifier accuracy increased with larger numbers of samples. Greater improvements were seen when testing on the same person compared to testing on a different person.

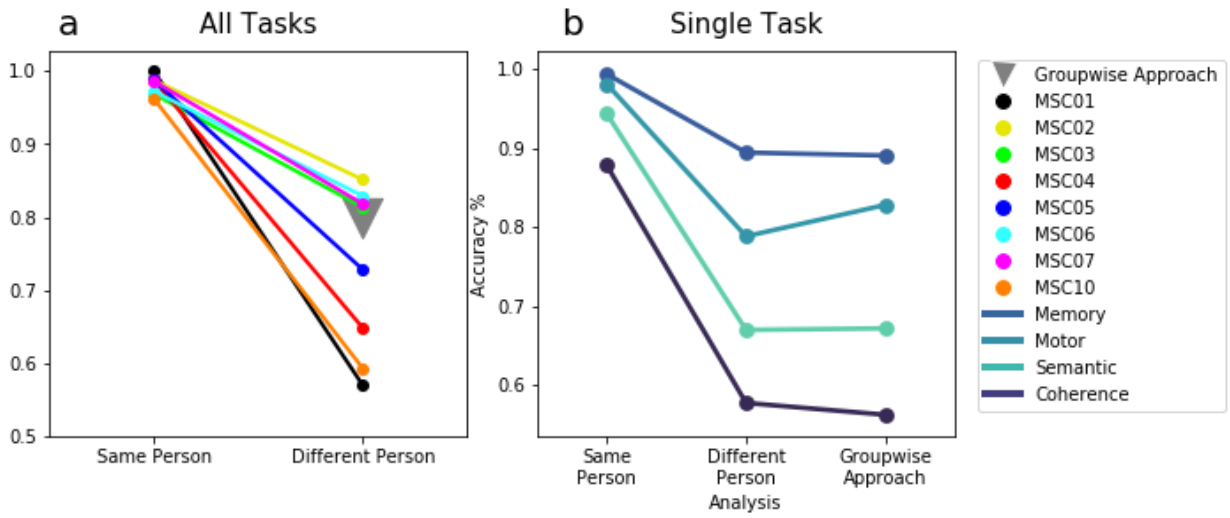

Figure S2. Comparison of individualized model to standard "groupwise" approach (a) Average accuracy of classifier performance in discriminating all tasks from rest using either individualized classifiers tested on the same or new people (as in the main text) or a more standard "groupwise" approach. The groupwise classifier was trained using a leave-one-subject-out cross validation procedure and tested on the left out subject (see Methods). **(b)** Average classifier performance for distinguishing a single task from rest using individualized classifiers tested on teh same or a different person, or classifiers built using the groupwise cross-subject approach . Performance is always higher when testing on the same person compared to testing on a different person and the groupwise approach; groupwise classification is similar to the between subject tests of individualized classifiers.

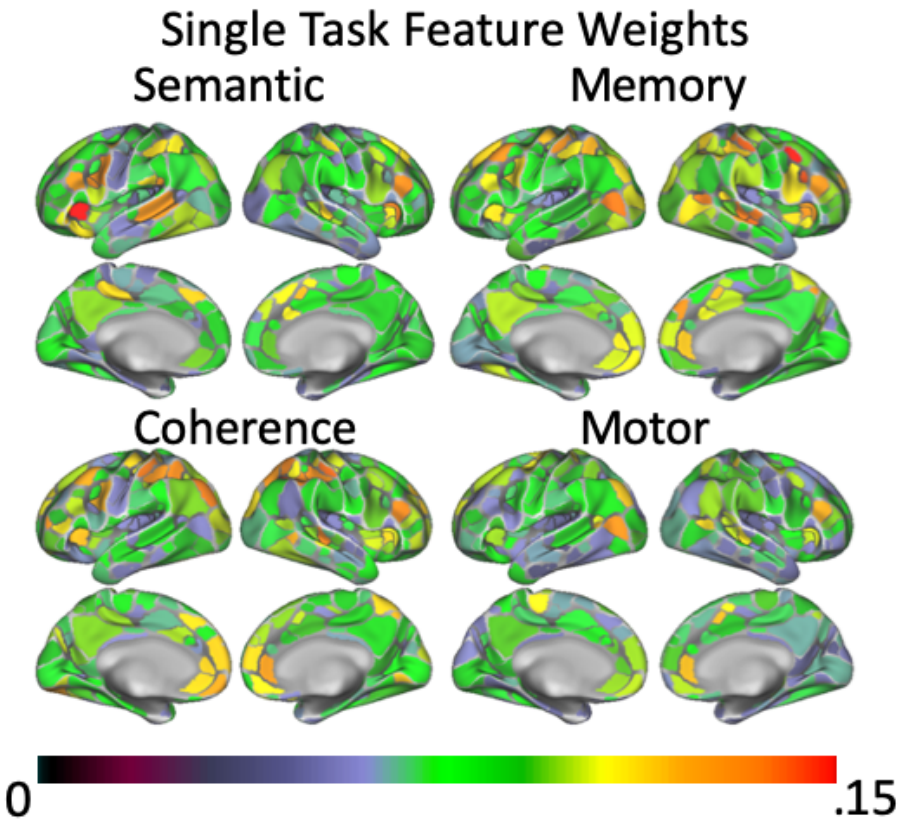

Figure S4 Feature Weights Across Task Average feature weights on the single task analysis for a single person (MSC05). Feature weights vary by task.

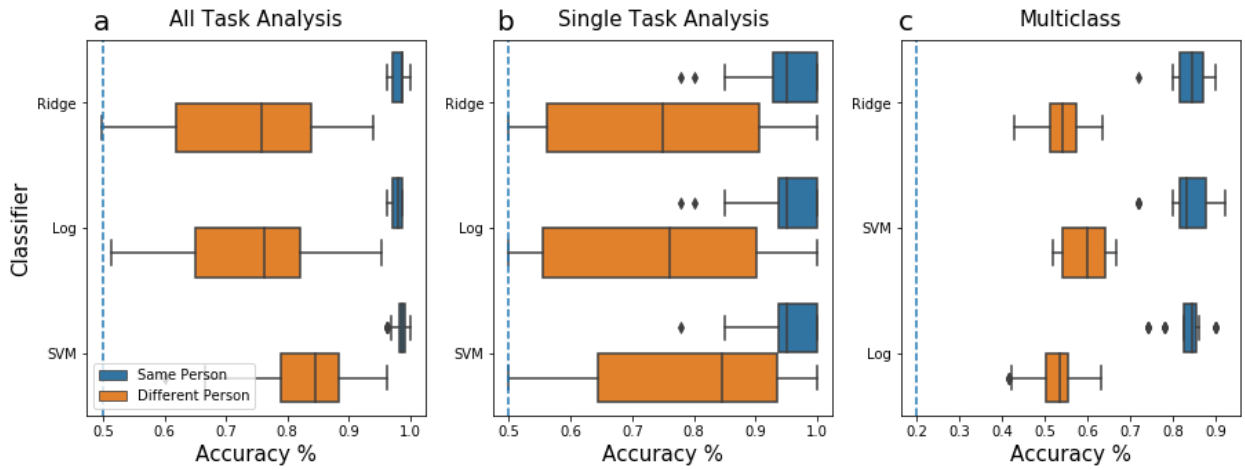

Figure S4. Model performance across different algorithms Task state prediction performance across different machine learning algorithms when tested on independent data from the same (blue) or a different person (orange), shown as box and whisker plots. The whiskers represent the upper and lower quartile to the highest/lowest value that is within the 25th and 75th percentile. Points outside the range (solid diamonds) reflect outliers. Three different classifiers were tested: ridge regression ("ridge"; the primary method in the main text), logistic regression ("log"; with L2 regularization and lbfgs solvers) and support vector classification ("SVM"; using LinearSVC with default parameter settings). **(a)** Classifier performance for data tested on the same or different person when trained to discriminate between all tasks and rest data for a person. **(b)** Classifier performance for classifiers trained to distinguish a single task from rest (averaged across tasks). **(c)** Multiclass performance for classifiers trained to distinguish among all four tasks and rest. Performance was similar across algorithms, with a consistent benefit for within person tests relative to between person tests.

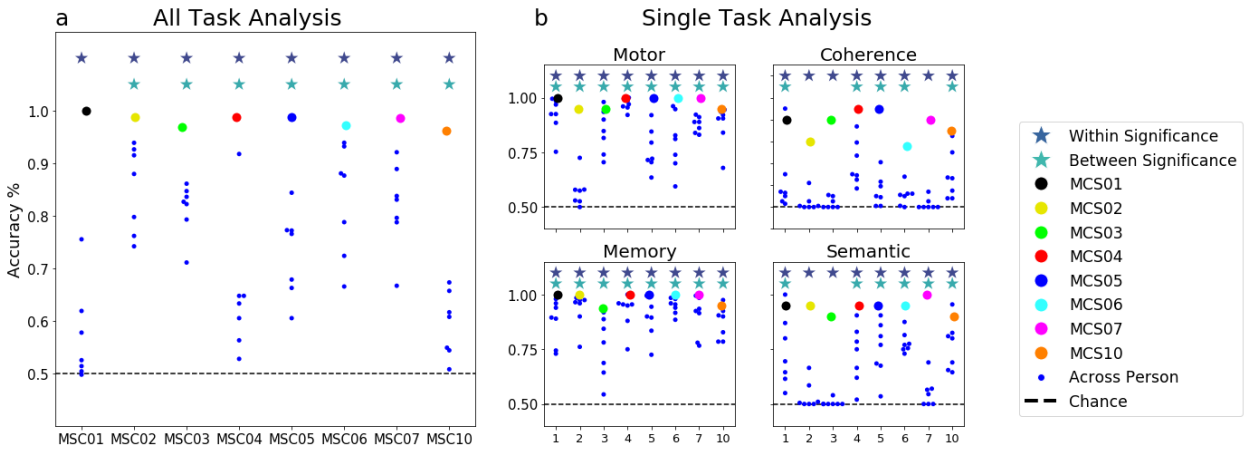

Figure S5. Model performance per person. Comparison of model performance for models trained on a particular individual (x-axis) and tested on either the same individual (large colored dots) or other individuals (blue dots). Figures depict the average accuracy across folds, with chance (50%) represented as a dashed black line. Significance per individualized classifier plotted as stars (dark blue within, light blue between; based on comparing true scores to a permuted null; *p<0.05). (**a**) Task state classification performance for classifiers trained to distinguish between all tasks and rest **(b)** Classification performance for classifiers trained to distinguish single tasks from rest. In all cases, performance is significantly higher when testing on the same person compared to between people.

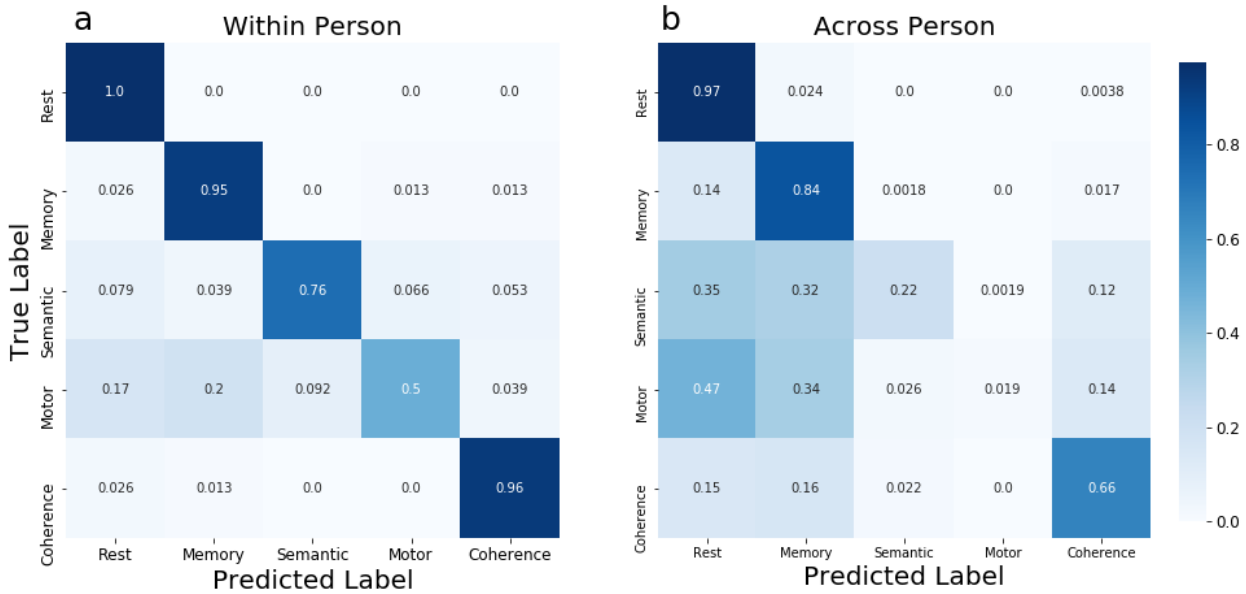

Figure S6. Average Multiclass Confusion Matrices (**a**) within person and (**b**) across person. Classifiers performed relatively accurately when classifying new data in the same person. When classifying across people there were more errors, especially for motor which is often misclassified as rest. Classifier performance is averaged across participants (see Figure S7 for single participant results).

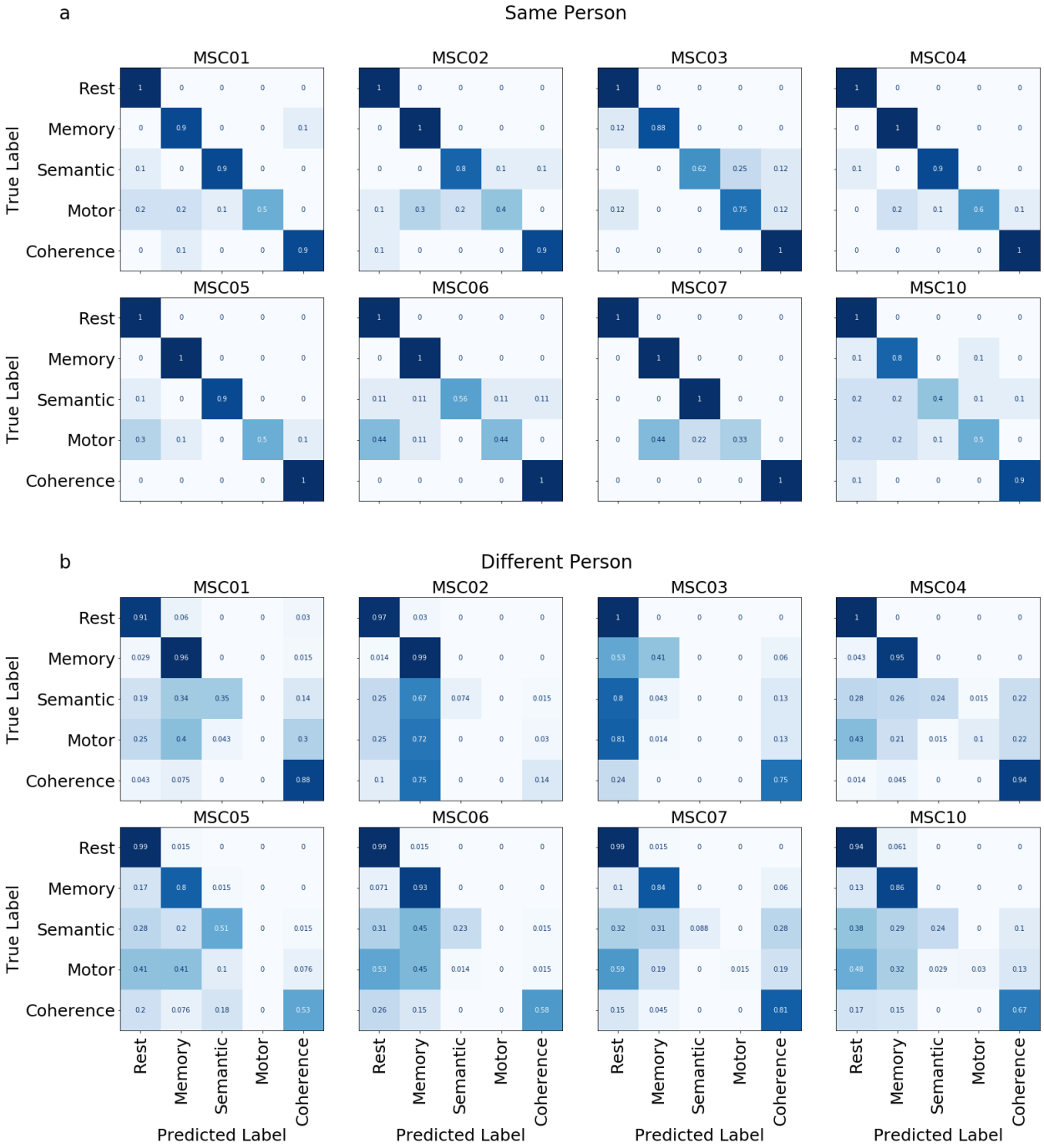

Figure S7. Multiclass confusion matrices of individualized classifiers. Multiclass output of confusion matrices for each individualized classifier when testing on left out sessions from either (a) the same person or (b) a different person. Performance was relatively similar across individualized classifiers. In all cases, within subject tests showed a strong diagonal (correct performance) with relatively lower errors. Between subject tests were more varied, with biased errors in classification (e.g., for motor and semantic especially).

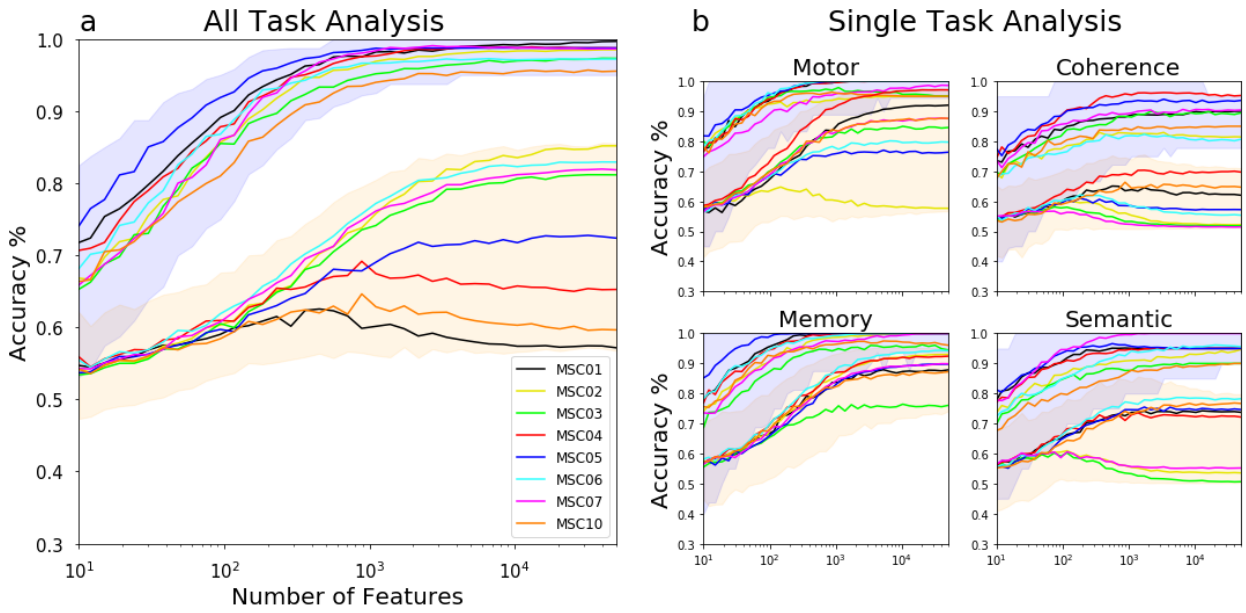

Figure S8 Random feature selection split up by individuals. Here we present random feature selection (10 - 50,000 number of features) when training the classifiers to (a) discriminate all tasks from rest or (b) single tasks from rest. We've plotted how individual test sets perform when tested on the same person as the training set (95% range in blue) compared to testing along a different person (orange). The mean for each individual is shown in a colored line within each set.