Title: Causal emergence of task information from dynamic network interactions in the human brain

Abbreviated title: Causal emergence of task information

Authors: Ravi D. Mill[1,3], Julia L. Hamilton[1], Emily C. Winfield[1], Nicole Lalta[1], Richard H. Chen[1,2], Michael W. Cole[1]

[1]Center for Molecular and Behavioral Neuroscience, Rutgers University, Newark, NJ, 07102, USA
[2]Behavioral and Neural Sciences Graduate Program, Rutgers University, Newark, NJ, 07012, USA
[3]Lead contact

Corresponding author:
Ravi D. Mill
Center for Molecular and Behavioral Neuroscience, Rutgers University
197 University Avenue, Newark, NJ 07102
rdm146@rutgers.edu

Highlights

- Network models fit with EEG connectivity simulated the emergence of task information
- These activity flow models yield causal insights by predicting future neural activity
- The models accurately predicted empirical response information underlying behavior
- The findings clarify fundamental principles of network computation in the human brain

eTOC blurb (In Brief)

The authors comprehensively map the spatiotemporal signatures of task information in the brain, and then use a novel network modeling approach to predict future information dynamics underlying behavior. The findings link functional connectivity to a role in mediating information flow, highlighting cognitive control networks as causal drivers of this process.

Summary

How cognitive task information emerges from brain activity is a central question in neuroscience. We identified the spatiotemporal emergence of task information in the human brain using individualized source electroencephalography and dynamic multivariate pattern analysis. We then substantially extended recently developed brain activity flow models to predict the future emergence of task information dynamics. The model simulated the flow of task-evoked activity over causally interpretable resting-state functional connections (dynamic, lagged, direct and directional) to accurately predict response information dynamics underlying cognitive task behavior. Predicting event-related spatiotemporal activity patterns and fine-grained representational geometry confirmed the model's faithfulness to how the brain veridically represents response information. Simulated network "lesioning" revealed cognitive control networks (CCNs) as the dominant causal drivers of response information flow. These results demonstrate the efficacy of dynamic activity flow models in predicting the emergence of task information, thereby revealing a mechanistic role for CCNs in producing behavior.

Keywords: information, decoding, MVPA, functional connectivity, EEG, source modeling

Introduction

Clarifying the spatial and temporal signatures underlying the emergence of cognitive task information is critical to furthering our understanding of the brain. Towards this aim, multivariate pattern analysis (MVPA) approaches have enabled decoding of different types of task information (e.g. sensory information in presented stimuli or response information underlying subjects' behavior) from evoked neural activation patterns (Haynes, 2015; Kriegeskorte and Kievit, 2013; Kriegeskorte et al., 2008). Prior research has shown these multivariate approaches to be more sensitive in indexing how neural activation measures relate to cognitive and behavioral variables of functional interest than earlier univariate approaches (e.g. single-cell spike rates, event-related potentials (ERPs) and functional MRI (fMRI) general linear model activations; Grootswagers et al., 2017; Jimura and Poldrack, 2012; Kriegeskorte et al., 2006; Saxena and Cunningham, 2019). These empirical observations underpin a shift in theoretical focus from localization of function to mapping distributed functionality at the neural population level (Eichenbaum, 2018).

Decoding methods applied to animal electrophysiological data have revealed that task-related information is highly distributed throughout cortex, such that most task-related information can be decoded from most brain regions (Bernardi et al., 2020; Kauvar et al., 2020; Raposo et al., 2014; Rigotti et al., 2013; Siegel et al., 2015). However, a degree of functional specialization has been demonstrated by scrutiny of temporal decoding characteristics, such as sensory information onsetting earliest in visual regions (Hernández et al., 2010; Siegel et al., 2015). Functional neuroimaging studies have also interrogated the neural basis of task information non-invasively in humans, with fMRI highlighting a central role for large-scale spatial interactions between sensory/motor content networks and higher-order cognitive control networks (CCNs, i.e., dorsal attention, frontoparietal and cingulo-opercular networks; Cole et al., 2016a; Ito et al., 2017; Zhang et al., 2013). Recent applications of MVPA decoding in scalp electroencephalography/magnetoencephalography (EEG/MEG) data have hinted at complex information dynamics, involving short-timescale transitions between distinct whole-brain representations as sensory information is translated into a behavioral response (Gwilliams and King, 2020; Hubbard et al., 2019; King and Dehaene, 2014).

Critically, all of the aforementioned recording methods have their respective limitations: invasive electrophysiology has high spatial resolution, high temporal resolution but only partial spatial coverage; fMRI has high spatial resolution, low temporal resolution and full spatial coverage; scalp EEG/MEG has low spatial resolution, high temporal resolution and full spatial coverage. These methodological limitations have impeded a *comprehensive* spatiotemporal description of *both* spatially "where" and temporally "when" task information is decodable from brain activity.

Consequent to this lack of descriptive insight is an even greater dearth of understanding into the computational mechanisms (the "how") underpinning task information. This deeper insight critically depends on the formalization of candidate models of how representations are computed (and transferred) across the brain, which are lacking in MVPA decoding approaches per se (de-Wit et al., 2016). To this end, it is likely that connectivity between neural entities plays a formative role in the

computations that give rise to decodable information at specific spatial locations and temporal epochs (Ito et al., 2019). This follows from classical demonstrations of connectivity at the cellular level weighting the flow of action potentials via structural connectivity and Hebbian synaptic strength processes (Hebb, 1949; Levy and Steward, 1983). Such small-scale connectivity is theorized to coordinate the formation of "cell assemblies" representing task information at the population level (Caporale and Dan, 2008; Eichenbaum, 2018). A role for human neuroimaging-assessed functional connectivity (FC) as a large-scale aggregate of synaptic strengths has been hypothesized (Petersen and Sporns, 2015; Wig et al., 2011), and supported by evidence of long-term learning- and use-driven changes to FC (Lewis et al., 2009; Newbold et al., 2020a). Whilst this supports a role for repeated task coactivation in moulding connectivity, a reciprocal influence is also anticipated by Hebbian theory: the intrinsic connectivity architecture estimated via resting-state FC (restFC) should mediate the likelihood of activity propagating between brain regions. Evidence supporting this comes from recent "activity flow" models linking the emergence of task activations to communication pathways indexed by restFC. The observed high accuracy in predicting empirical task activations across a variety of cognitive tasks (Cole et al., 2016b) and in predicting dysfunctional activations and behavior in Alzheimer's disease (Mill et al., 2020) provides support for the relevance of restFC in generating cognitively relevant phenomena (Mill et al., 2017). This network modeling framework was also adapted to directly quantify the exchange of task information at the large-scale network level (Ito et al., 2017).

However, these previous network models focused exclusively on predicting the spatial signatures of task activations, and overlooked accompanying temporal dynamics. This is in part due to their sole application in fMRI which has well-known limitations in temporal resolution. We therefore applied a considerably extended version of activity flow modeling to higher temporal resolution EEG data, so as to gain greater insight into the dynamic neural computations underlying cognition in the human brain. Importantly, we could not simply apply previous activity flow models – using contemporaneous brain activity in one region to predict contemporaneous activity in an independent brain region – to EEG data. This was due to well-known instantaneous field spread artifacts in EEG (and MEG) data (Schoffelen and Gross, 2009; Stinstra and Peters, 1998), which could result in analytic circularity if a to-be-predicted EEG source region leaked some of its signal into other regions used as predictors. A substantial innovation was necessary to overcome this limitation: modifying activity flow modeling to use past activity to predict future activity. Given that it is impossible for brain activity to propagate back in time (a fundamental principle of causality and the direction of time), this completely eliminates the possibility of analytic circularity. The resultant "dynamic activity flow modeling" approach properly leverages the higher temporal resolution of EEG data to overcome a limitation of the method and, due to the use of lagged multivariate FC (Liégeois et al., 2017; Mill et al., 2017; Sanchez-Romero and Cole, 2020), open up enhanced causal inferences relative to standard models.

Further, the prior focus on non-dynamic activity flow stands in contradistinction to artificial neural network models (ANNs) applied in the field of artificial intelligence, which engineer similar activity-flow-over-connectivity architectures to output fully dynamic simulations of task information (Buonomano and Merzenich, 1995; Eliasmith et al., 2012; Yamins et al., 2014). Whilst ANNs have been heralded as charting new frontiers in simulating complex neurocognitive phenomena, they are themselves limited by their abstract and non-empirical nature. For example, connectivity weights that parameterize neural network architectures are typically determined through random

initialization followed by intensive data-driven optimization, rendering them difficult to interpret neuroscientifically (Ito et al., 2019; Sinz et al., 2019). Furthermore, whilst these engineered architectures may yield accurate performance on increasingly complex cognitive tasks (Yang et al., 2019), their embedded high degree of computational flexibility (Cybenko, 1989) means that the extent to which this high performance arises from veridical information representations (i.e comparable to those instantiated in the human brain) is open to question (Kriegeskorte and Douglas, 2018; Sinz et al., 2019; Yamins and DiCarlo, 2016).

The present report aimed to address these varied methodological and theoretical issues, substantially extending activity flow models to provide more dynamic and causally interpretable insight into the emergence of task information. We first identified information in a cross-modal sensory categorization task (Figure 1), using dynamic MVPA and anatomically individualized EEG source modeling to comprehensively describe large-scale spatial and temporal information signatures (Figure 2).

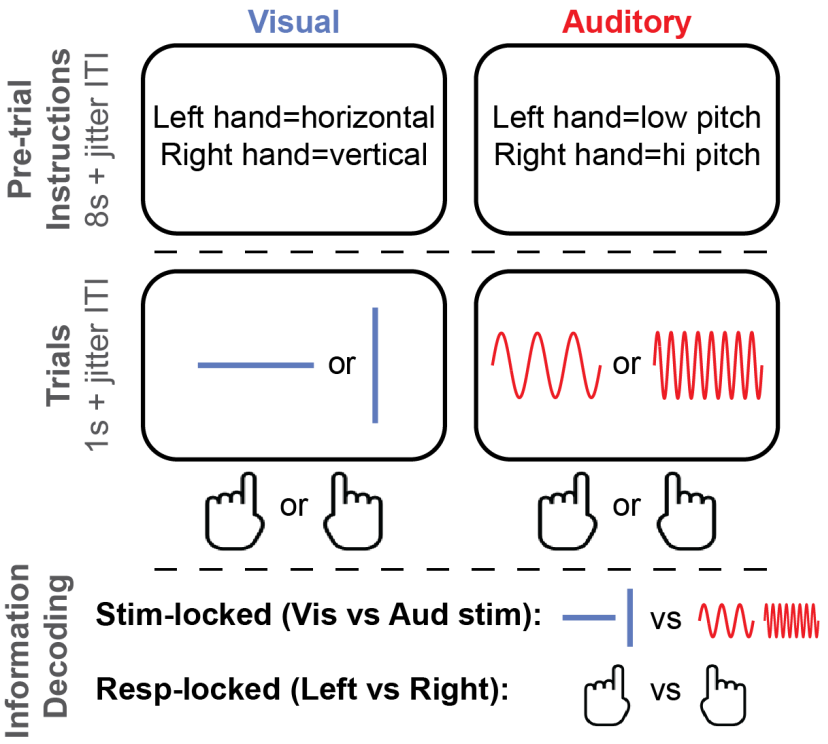

Figure 1. Task design. Depicted is one trial from each sensory block condition (10 trials per block, 12 blocks per subject session), and the types of decodable task information.

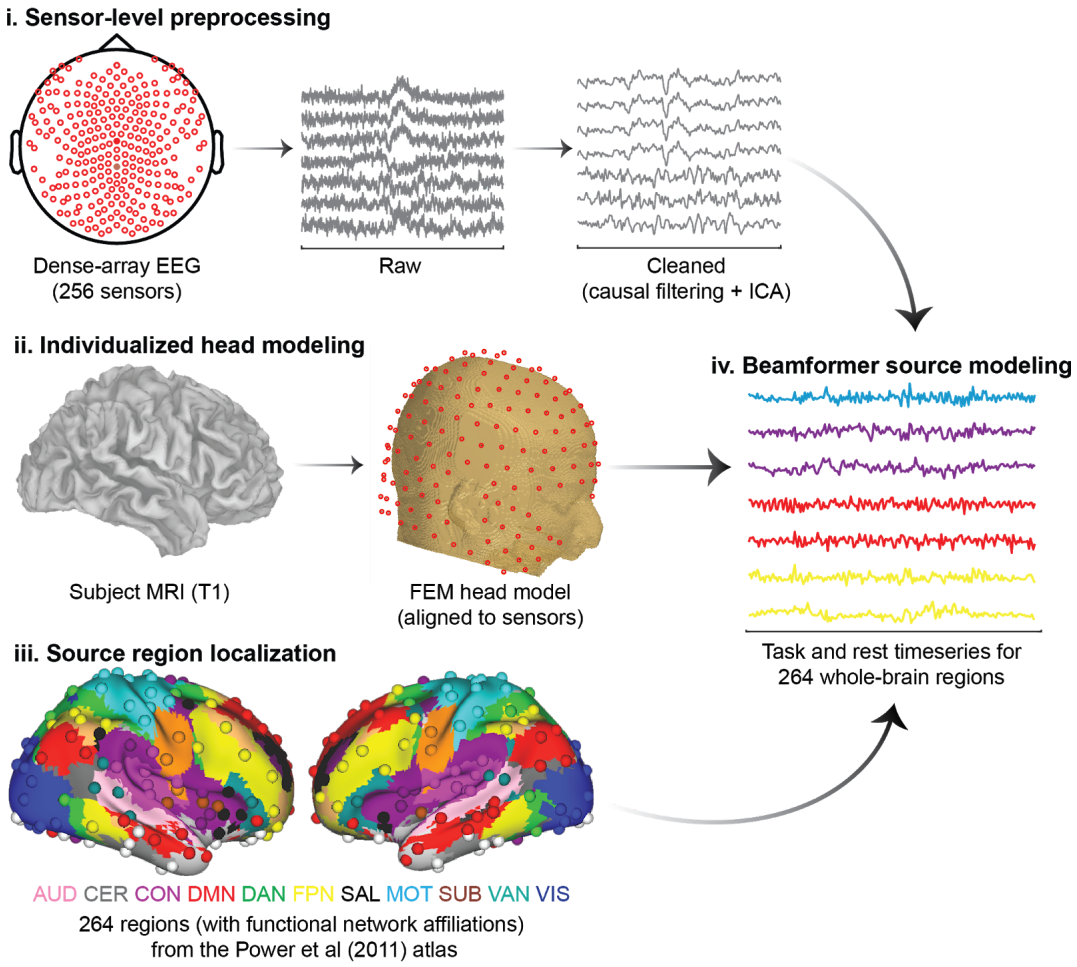

Figure 2. EEG preprocessing and source modeling pipeline. Applying this pipeline separately to task and resting-state EEG data reconstructed their respective activation timeseries. Colored text in panel iii) provides functional network affiliations for correspondingly colored regions localized from the Power functional atlas: AUD=Auditory; CER=Cerebellar; CON=Cingulo-opercular; DMN=Default mode; DAN=Dorsal attention; FPN=Frontoparietal; SAL=Salience; MOT=Motor; SUB=Subcortical; VAN=Ventral attention; VIS=Visual. See Method for full details.

After describing these signatures, we aimed to simulate their emergence using dynamic activity flow modeling. This entailed significant refinements over previous activity flow models. Firstly, the empirical restFC weights that parameterized the model were estimated via a more causally valid approach (Figure 3A), which identified dynamic, lagged, direct and directional functional connectivity weights (see Method for details). These refinements were necessary to achieve the desired prediction outcome: moment-to-moment fluctuations in future task information. Coupled with the need to eliminate EEG field spread artifacts highlighted above, success in predicting the future information dynamics of the brain would provide compelling evidence in support of the cognitive relevance of restFC, as a large-scale analogue of synaptic strength processes governing information flow in the brain. Beyond this theoretical insight, success of the model would raise complementary benefits to the fields of empirical activity flow modeling and artificial neural network simulations, respectively by extending to fully dynamic simulations of information state and doing so on the basis of empirically-derived and causally interpretable connectivity features.

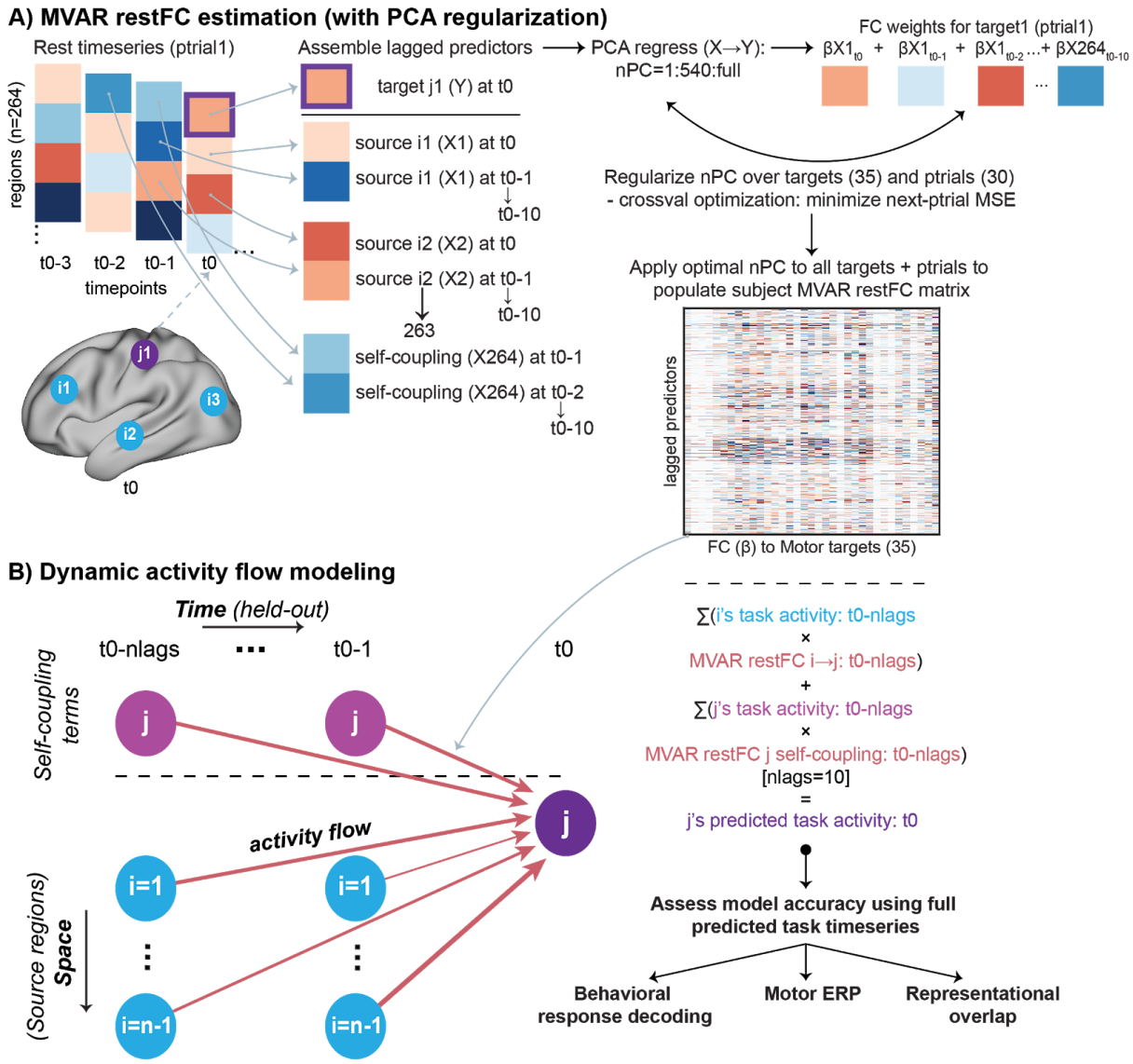

Figure 3. Approach to restFC estimation (via MVAR) and dynamic activity flow modeling. A) Estimation of MVAR restFC. Note that for demonstration purposes the schematic uses just 4 regions (1 Motor network target j1, and 3 predictor sources i1-3), whereas the full procedure iterated over 35 Motor targets and included all other 263 regions as sources. Lagged FC weights (β) from source regions and self-coupling terms to each Motor network region were calculated after regularizing the number of principal components (nPCs). This was achieved via crossvalidated minimization of the mean squared error (MSE) of the MVAR-predicted rest timeseries for held-out pseudotrials (ptrials). **B)** The lagged MVAR restFC weights were then combined with the lagged task activation timeseries to predict future Motor task activations via dynamic activity flow modeling. Iterating over all to-be-predicted Motor targets (35), trials (~30) and trial timepoints (-0.45 to 0.45s around response commission) populated the full predicted Motor task activation matrix. This predicted matrix was the basis of subsequent response information decoding (dynamic MVPA), motor ERP and representational overlap analyses that assessed model accuracy.

Results

Source modeling improves detection of cognitive task information.
We compared peak information decodability for our approach that combined individualized source modeling and dynamic MVPA, with the more commonly used sensor-level dynamic MVPA approach. Timepoint-by-timepoint behavioral response information (correct left- vs right-handed responding) was decoded separately using all source regions (SourceAll) and all scalp sensors (SensorAll) as features. Visual inspection of the group response information timecourses (Figure 4) revealed multiple significantly decodable timepoints for both feature sets, with onset prior to commission of the response. This morphology is consistent with previous response decoding of EEG (Gwilliams and King, 2020) and invasive electrophysiological (Siegel et al., 2015) data.

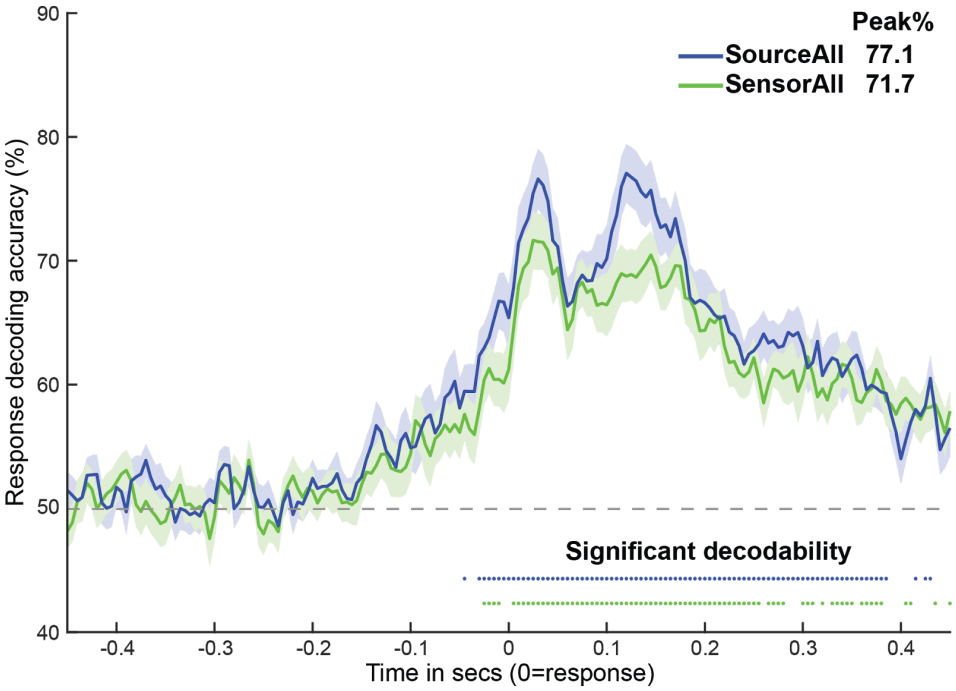

Figure 4. Detection of behavioral response information is improved for source versus sensor feature sets. Group-averaged response decoding timecourses for the SourceAll and SensorAll sets, with shaded patches reflecting the standard error of the mean across timepoints. Colored dots represent timepoints with significantly decodable information, as assessed by Wilcoxon signrank tests against 50% classification accuracy (p<.05, Bonferroni-corrected). The legend in the top right provides the peak decoding accuracy for each timecourse.

Critically, peak response information decodability was visibly higher for the source-modeled approach (Figure 4). This was formalized statistically by extracting peaks from each subject's SensorAll and SourceAll decoding timecourses, and contrasting them via paired Wilcoxon tests. The SourceAll decoding peak was significantly higher than the SensorAll peak across variation in how time-to-peaks were defined: i) from the SensorAll group time-to-peak (0.025s; 3.8% peak change; z=2.47, p=.014), ii) from the SourceAll group time-to-peak (0.120s; 8.4% peak change; z=3.57, p<.001), iii) from the peak of each individual subject's decoding timecourse (unbiased by the group results; 4.8% peak change; z=3.59, p<.001). Peak decodability of sensory information (visual vs

auditory stimulus condition) was also numerically higher for SourceAll compared to SensorAll features (see SI, Fig S1A). Hence, combining source modeling with dynamic MVPA significantly improved detection of response information. This might have arisen from improvements in signal-to-noise introduced by beamformer source modeling (Brookes et al., 2008), and opposes prior suggestions that beamforming leads to overarching cancellation of the underlying neural activations (Hui et al., 2010; Van Veen et al., 1997). These results extended our critical requirement for subsequent analyses - that spatially distinct regions were localized accurately - by demonstrating that such source localization can actually improve overall decodability.

Network decoding reveals prominent roles of Motor and Cognitive Control Networks in representing behavioral response information.

We then applied our source-modeled dynamic MVPA approach to decode spatial and temporal signatures of response information from each of the 11 major functional networks (Power et al., 2011), treating within-network regions as features (see Method). These network decoding results are depicted in Figure 5. Consistent with the morphology of the SourceAll response decoding timecourse (Figure 4), we again observed two decoding peaks across networks, which were linked via a supplementary temporal generalization analysis to distinct multivariate codes, possibly reflecting motor preparation and motor execution/feedback processes (see SI, Figure S2). Response information was broadly decodable across networks, consistent with prior findings from primate electrophysiology (Bernardi et al., 2020; Siegel et al., 2015) and rodent optical imaging (Kauvar et al., 2020).

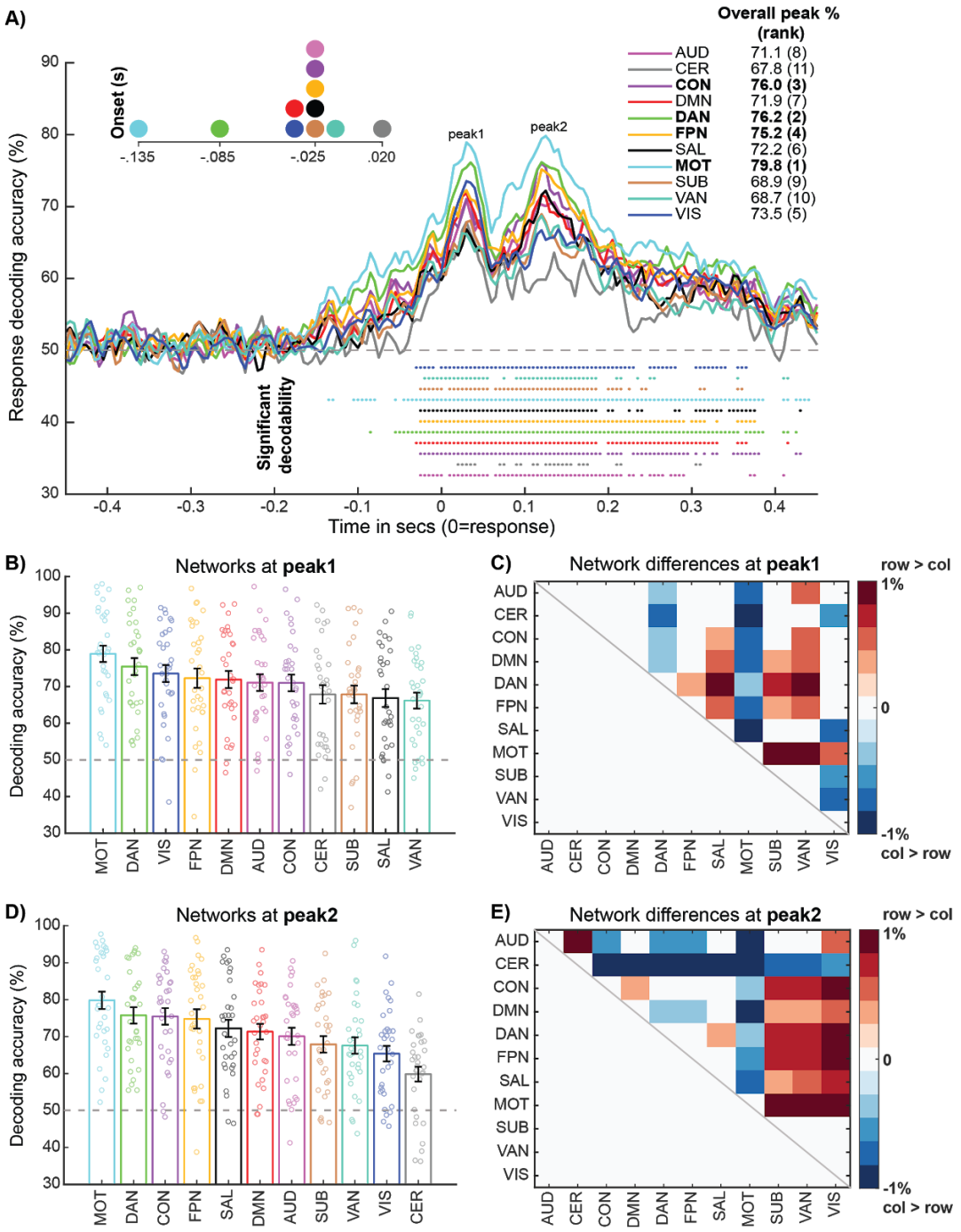

Figure 5. Network decoding of behavioral response information reveals prominent roles for Motor and Cognitive Control networks. A) Group decoding timecourses color-coded by network affiliation. Colored dots represent significantly decodable timepoints for each network (p<.05 via Wilcoxon signrank against 50% chance, Bonferroni-corrected) and the legend in the top right provides peak decoding accuracies for each network. A magnified plot of the onset of the first significant timepoint for each network is provided in the top left. **B)** Response decoding accuracy ranked across networks at peak 1 (0.03s). Each bar represents the mean and standard error for each network, with individual subject data points also overlaid. **C)** Matrix capturing the significance of cross-network differences in decoding accuracy at peak 1. Plotted is the pairwise difference in mean decoding accuracy, thresholded via paired Wilcoxon tests (p<.05, FDR-corrected). Positive values denote significantly higher decoding accuracy for the row network > the column network, and vice versa for the negative values. **D)** and **E)** follow the same conventions as C) and D) respectively, albeit focusing on peak 2 (0.125s).

However, examination of network timecourse features revealed content-appropriate specialization: response information onset earliest and peaked strongest in the Motor network. Prominent response decodability was also observed in the CCNs, all three of which yielded the highest response information peaks after the Motor network. We statistically formalized these between-network effects by contrasting decoding timecourse peaks extracted from individual subjects (see Method). We extracted subject-level peaks separately for peak 1 (motor preparation, Figure 5B) and peak 2 (motor execution/feedback, Figure 5D), from the respective cross-network group peak timepoints within these windows (0.03s and 0.125s). Motor network activity was decoded significantly higher relative to other networks across both peak 1 (Figure 5C) and peak 2 (Figure 5E). Interestingly, whereas Visual network activity was prominently decoded during peak 1 (see network ranking in Figure 5B), it was much less involved in peak 2 (Figure 5D). This strengthens the notion that this earlier period reflects motor preparation operations that translate sensory into response information, whereas the later peak 2 period elicits purer motor representations underlying response execution and feedback (Elsayed et al., 2016; Raposo et al., 2014; Wascher and Wauschkuhn, 1996).

The transition from peak 1 to peak 2 also charted an interesting differentiation in the engagement of the CCNs. The DAN was more prominently involved in the motor preparation peak 1, and differed significantly from both the CON and FPN (Figure 5B). This profile of preferential DAN engagement was also observed when sensory information was more directly decoded from the stimulus-locked data (see SI, Figure S1B), suggesting that similar sensory-related computations might be occurring at peak 1 in the response timecourse. However, the later peak 2 elicited a different network profile: information was higher for all three CCNs relative to the other networks but was not reliably differentiated between them (even at an uncorrected threshold of $p<.05$, Figure 5E).

Overall, the network decoding results highlight the utility of decoding information from neuroimaging data that is well-resolved in both the spatial and temporal domains. Beyond recovering the content-appropriate specialization of the Motor network for response information, this approach led to a segregation of dynamic CCN profiles: from preferential engagement of the DAN during translation of sensory to response information at peak 1, to more collaborative engagement of all three CCNs to generate purer response representations at peak 2.

Dynamic activity flow modeling predicts future response information dynamics.

The preceding network decoding results described in detail which spatial networks prominently represent response information ("where") and their accompanying temporal profiles ("when"). We next sought more mechanistic insight into "how" these representations emerge computationally in the brain. To interrogate this, we applied a novel network modeling approach (dynamic activity flow) to predict empirical future response information dynamics from past task activations flowing over communication pathways parameterized by restFC. Based on the content-appropriate specialisation of the Motor network demonstrated above (Figure 5), we considered regions within this network as to-be-predicted "targets" for the emergence of neural representations driving overt behavior.

Functional connectivity to these regions from the rest of the brain was estimated via multivariate autoregression (MVAR) in a separate resting-state EEG session. This MVAR approach was causally

constrained to output dynamic, lagged, direct and directional estimates of restFC (see Figure 3A and Method for details). An embedded PCA optimization step ensured that restFC regression weights for each subject were appropriately regularized for each individual[1]. These lagged restFC weights were then combined with lagged task activations to generate future model-predicted activation timeseries for all Motor network regions across the entirety of the task (Figure 3B).

We first assessed the accuracy of the dynamic activity flow model in capturing the emergence of behavioral response information. The predicted Motor region activation timeseries underwent the same dynamic MVPA procedure applied previously to the actual data. This yielded a predicted Motor network response information timecourse, which is depicted in Figure 6A along with the actual Motor timecourse for comparison. The plot demonstrates that overt behavioral information was decodable at multiple timepoints from the predicted data which, to reiterate, was generated exclusively from FC weights derived from a held-out rest session, and held-out (past) temporal task information. The predicted decoding peak was in fact significantly higher in magnitude (82.3%) than the actual peak (79.8%), as formalized by a paired Wilcoxon contrast of the subject-level peaks (extracted from the actual group time-to-peak at 0.125s): $z=2.36$, $p=.018$. Furthermore, comparison of the overlap between the predicted and actual decoding timecourses confirmed that the model faithfully captured the temporal morphology of the information timecourse. This held at the group-level ($r=.97$, $p<.00001$) and, critically, also at the subject-level ($r=.78$, $p<.00001$); further emphasizing the accurate individualized predictions obtained. Similarly high model prediction accuracy was achieved when using more common non-causal filtering during preprocessing (see SI, Figure S3). In the Supplement we also highlight our rigorous regression approach to controlling for field spread artifacts (Figure S4).

[1] Results of the PCA regression regularization (achieved via cross-validated optimization): cross-subject mean squared error (MSE) in predicting future resting-state activation timeseries=0.006 (std=0.003); optimal nPCs=540 (59.4% of subjects) or 1080 (40.6%).

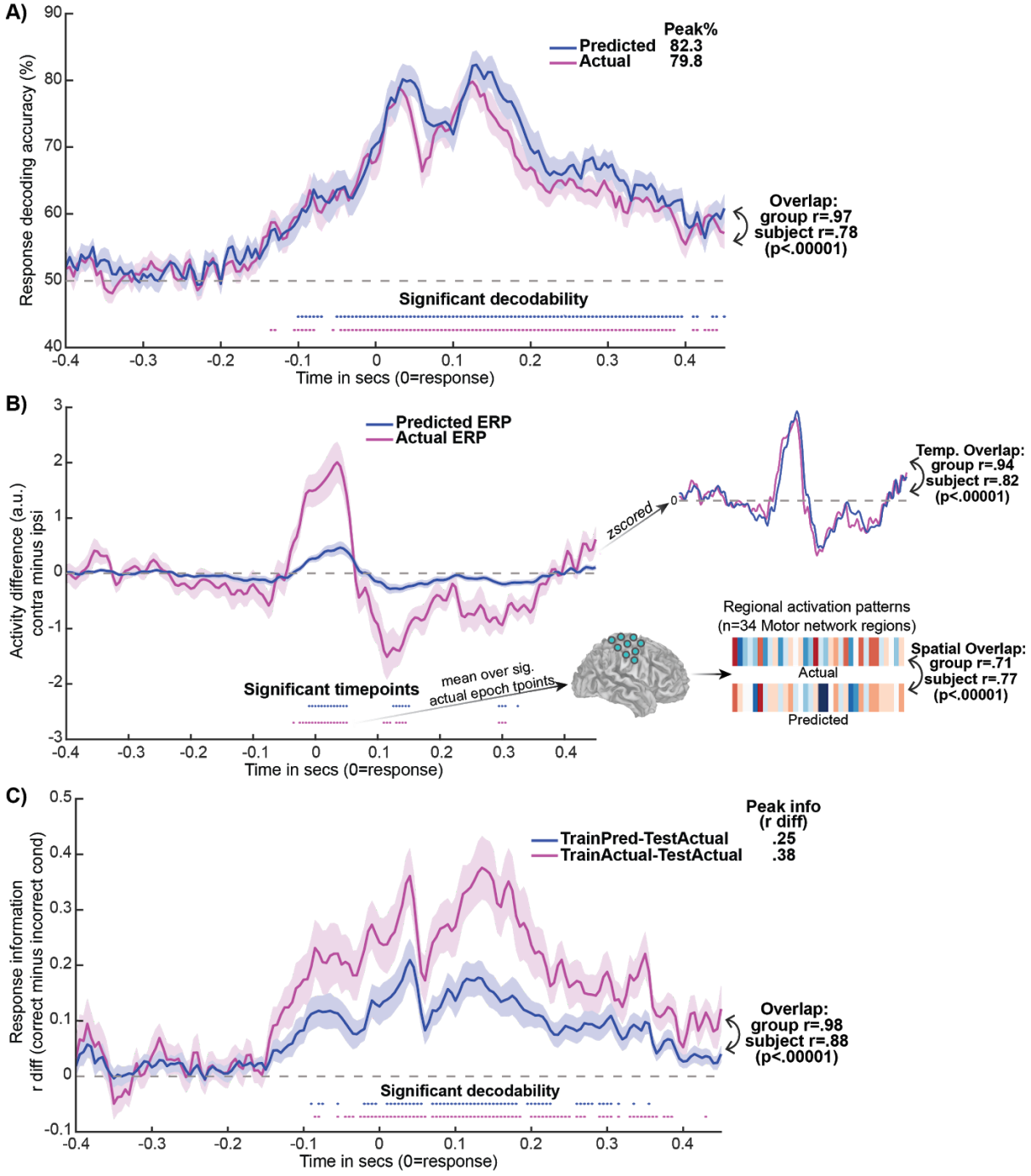

Figure 6. Dynamic activity flow modeling accurately and veridically predicts future behavioral response information. A) The model accurately predicted future response information dynamics in the Motor network. Predicted and actual group decoding timecourses are plotted, with overall decoding peaks (top right), and predicted-to-actual timecourse overlap (Pearson r) at group and subject levels. **B)** The motor ERP waveform (and spatial pattern) was also successfully predicted, demonstrating fidelity to the underlying activations. Group averaged ERP difference waves are plotted (contralateral minus ipsilateral; a.u.=arbitrary units). To aid visualization, the top right depicts the zscored group waveforms, along with the temporal predicted-to-actual overlap (at group and subject levels). The bottom right provides the spatial overlap between the predicted and actual Motor region activation vectors, extracted over the -0.035 to 0.050s epoch of significance in the actual data. **C)** Representational overlap analyses highlight veridicality of the model-predicted representations, given their ability to decode the actual data. The group information timecourse resulting from training on predicted representations and testing on the actual data (TrainPred-TestActual) is depicted, as well as the result of training and testing on the actual data (TrainActual-TestActual) for comparison. Response information was

quantified as the difference in Pearson r values computed for the test activation pattern at each timepoint with correct minus incorrect response condition templates (see Method). Overall information peaks (top right) and temporal overlap between the two timecourses (bottom right) are provided. For all panels, significance at each timepoint was assessed via Wilcoxon signrank tests (versus 50% chance for panel A, and versus 0 for B/C) with Bonferroni correction.

The predicted dynamic MVPA results hence attest to the fidelity of the dynamic activity flow model in capturing the network-weighted flow of response information in a highly temporally resolved and individualized fashion. This provides clear mechanistic evidence for the cognitive relevance of restFC in predicting behaviorally relevant task information dynamics.

Dynamic activity flow modeling also predicts the raw activation timeseries (motor ERP).
We then probed the model's accuracy in predicting the activations in Motor network regions that were the basis of the dynamic MVPA analysis above. This would demonstrate fidelity to the underlying neural activation timeseries, and thereby challenge the possibility that the strong response information decoding observed in the predicted data emerged from factors somehow artificially introduced by the model.

We targeted accurate prediction of an established motor event-related potential (ERP) termed the lateralized readiness potential (Cheyne et al., 2006; Deecke et al., 1976; Kutas and Donchin, 1980): greater activation at the time of response commission for Motor network regions that are contralateral to the response hand. The predicted and actual group motor ERP waveforms are provided in Figure 6B (as contralateral minus ipsilateral difference waves, see Methods). Multiple significant activation timepoints were observed in the actual and predicted timecourses, with high overlap in their temporal morphology at both the group and subject levels ($r=.94$ and $r=.82$ respectively, both $p<.00001$). We also probed the predicted-to-actual overlap in the spatial domain, by correlating the respective activation vectors for Motor network regions averaged across the epoch of significance in the actual data (-0.035 to 0.050s, Figure 6B). This revealed a high degree of spatial overlap between the predicted and actual spatial activation patterns, at the group ($r=.71$) and subject levels ($r=.77$, both $p<.00001$). This high spatial overlap held when the predicted and actual vectors concatenated timepoints across the entire trial, at both the group ($r=.70$) and subject levels ($r=.61$, both $p<.00001$). Overall, the motor ERP results illustrate that the dynamic activity flow model maintained high fidelity to the actual task activations; across both temporal and spatial domains, and even at the individual subject level.

The model is directly faithful to the veridical representational geometry.
Whilst the previous section reported highly accurate predictions of the underlying Motor network activations, we sought to more directly demonstrate that the strong response information decoding obtained via dynamic activity flow modeling was driven by multivariate representations that overlapped with how the brain veridically represents response information. This aligns with mounting emphasis on the need to optimize representational overlap between neural network models and the brain (Kriegeskorte and Douglas, 2018; Yamins and DiCarlo, 2016; Yang and Wang, 2020).

To address this, we investigated whether actual response information in the Motor network could be dynamically decoded from the model-predicted representations. The group information timecourses are presented in Figure 6C: 1) for trained representations in the predicted data applied to decode the actual data ("TrainPred-TestActual"), and 2) for trained representations in the actual data applied to decode the actual data ("TrainActual-TestActual"). The figure highlights multiple significantly decodable timepoints in the TrainPred-TestActual timecourse, providing direct evidence that the dynamic activity flow model captured neurally valid multivariate representations. The morphology of the resultant response information timecourse also temporally overlapped with that obtained for the TrainActual-TestActual model, at the group and subject level ($r=.98$ and $r=.88$ respectively, both $p<.00001$). This suggests that the dynamics in representational geometry over the entire trial were well captured. Overall, this representational overlap analysis provides compelling evidence that the dynamic activity flow model preserves high fidelity to the veridical neural codes underlying response information.

Model-simulated network lesioning suggests prominent causal influence of the CCNs.

We next modified dynamic activity flow models to provide additional causal insight into the functional network processes underlying the generation of overt behavioral responses. Separate models predicting response information in the Motor network were run that systematically lesioned all but one of the remaining functional networks (see Methods). Beyond this iterative restriction of source predictor terms (Figure 3B), the procedure was identical to that detailed in the dynamic MVPA analysis of the model-predicted data. This yielded 10 response decoding timecourses for each functional network model, capturing their likely causal influence on information flow to the Motor network.

The network-lesioned group decoding timecourses are presented in Figure 7A. Visual inspection revealed significant decodability for all network models, however this peaked highest for the CCN models (top right Figure 7A), with onset also preceding the other networks (top left Figure 7A). The timecourses once again followed a 2-peak morphology, with CCN-derived decodability visibly varying across the earlier motor preparation epoch and the later motor execution/feedback epoch (as in Figure 5).

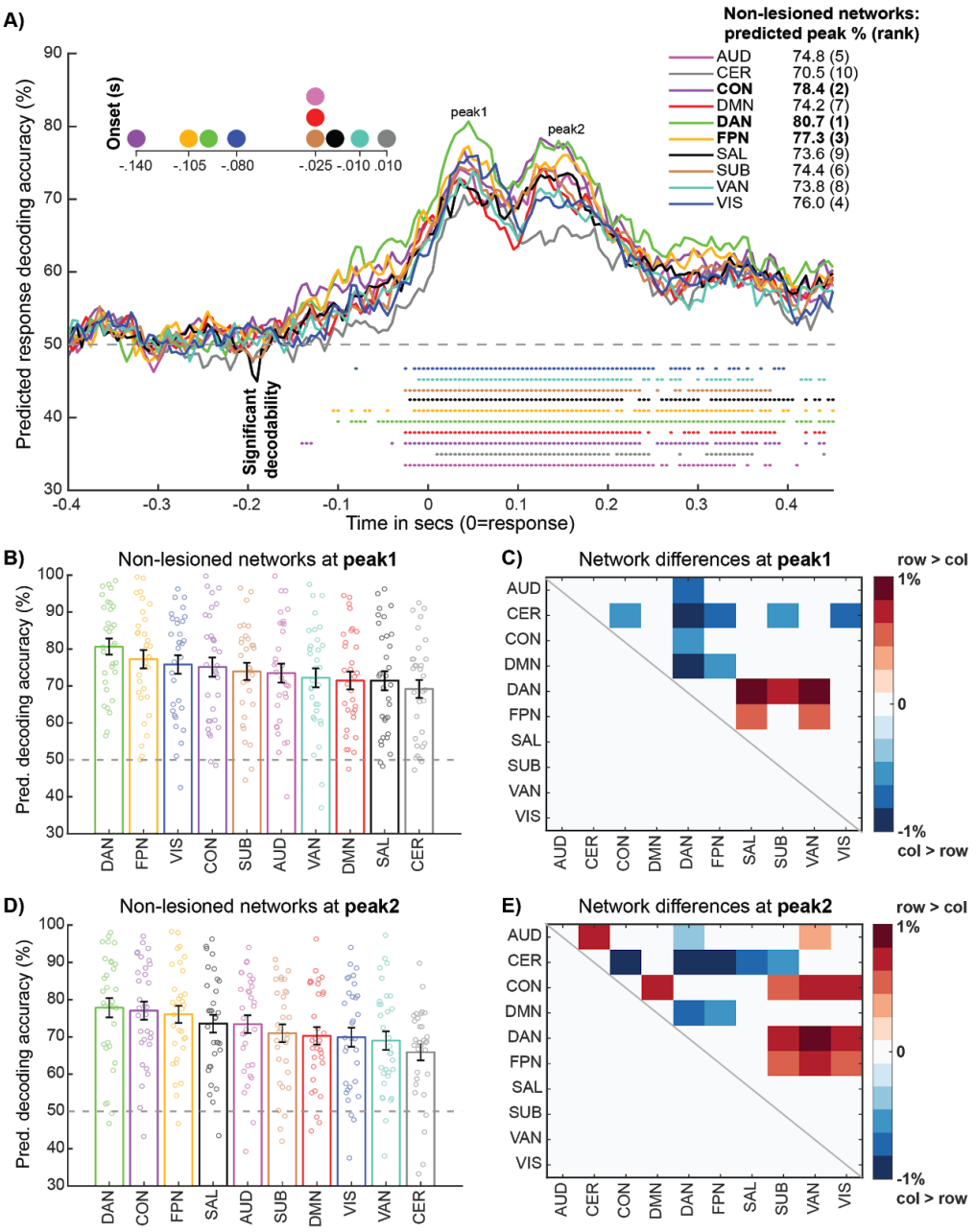

Figure 7. Network lesioning reveals the likely causal contribution of individual functional networks in driving future behavior. Unlike the visually similar Figure 5 that provided descriptive insight, these analyses provided explanatory insight into which spatial networks likely causally drive response information representations in the Motor network, and their accompanying temporal signatures. For these analyses all networks were lesioned except for the indicated (non-lesioned) network. **A)** Group predicted decoding timecourses for each of the network-lesioned models, color-coded by affiliation as before. Magnified decoding onsets (top left), overall decoding peaks (top right) and significant decodability at each timepoint (assessed via bonferroni-corrected Wilcoxon tests, as before) are provided for each network model. **B)** Predicted response decoding accuracy ranked across network models at peak 1 (0.045s). Each bar represents the mean and standard error for each network, with overlaid subject data points. **C)** Matrix capturing cross-network differences in predicted decoding accuracy at peak 1. Plotted is the pairwise difference in mean decoding accuracy, thresholded via paired Wilcoxon ($p<.05$, FDR-corrected). Positive values denote significantly higher predicted decoding accuracy for the row network > the column network, and vice versa for the negative values. **D)** and **E)** follow the same conventions as C) and D) respectively, except now focusing on peak 2 (0.155s).

However, these network lesioning analyses critically extended the descriptive results in Figure 5, highlighting that whilst the CCNs exerted a broadly strong causal influence across the trial, this pattern of influence differed across the two peaks. These differing network profiles were statistically formalized by between-network contrasts of subject-level predicted decoding peaks, which were extracted separately from identifiable group peak timepoints for peak 1 (0.045s) and peak 2 (0.155s). At the motor preparation peak 1 (Figure 7B), the DAN network model yielded response decodability that was significantly higher than the CON ($p<.05$ FDR-corrected, Figure 7C) and marginally higher than the FPN ($p<.05$ uncorrected). Considering the strong influence of the Visual network also observed at peak 1 (Figure 7B), the results again support more selective engagement of the DAN during this likely period of sensory-related processing (e.g. translating sensory into response information). At the later motor execution/feedback peak (Figure 7D), all three CCNs exerted the strongest casual influence on response decodability, and were not reliably differentiated amongst each other (Figure 7E), even at lenient $p<.05$ uncorrected thresholds. The pattern of network differences is again indicative of more collaborative engagement across the CCNs in generating purer response representations during this later period.

The network lesioning results therefore extended previous description of spatiotemporal signatures underlying response information to elucidate the likely causal influence of the CCNs on network information flow driving overt behavior. This prominent influence of the CCNs was also recovered when using subject-level prediction accuracy for each network model as an alternative metric to the predicted decoding peaks used above (see SI, Figure S5).

Discussion

The present report aimed to first provide a detailed description of the spatiotemporal signatures underlying the emergence of task information in the human brain, and then to provide deeper insight into the computational mechanisms that generated them. In targeting these aims, we benefited from recent advances in anatomically individualized EEG source modeling (Klamer et al., 2015; Vorwerk et al., 2018), which we combined with dynamic MVPA methods to decompose spatially "where" and temporally "when" task information was present across the entire brain. This extended prior applications of dynamic MVPA in scalp-level EEG, which is inherently incapable of elucidating the spatial loci of neural coding schemes and is contaminated to a greater degree (relative to source EEG) by field spread artifacts (Haufe et al., 2013; Van de Steen et al., 2019). Beyond fulfilling our specific need for valid spatial reconstruction, we observed generally improved information detection when sources were treated as MVPA features rather than raw sensors. This finding should motivate increased application of source modeling in dynamic EEG/MEG decoding studies moving forwards.

These methodological refinements enabled us to probe the large-scale human analogues of population coding schemes reported extensively in animal electrophysiology research. Whilst the spatial resolution recovered by these invasive methods remains considerably higher than our non-invasive approach, we nevertheless were able to decode with equivalent temporal resolution and superior whole-brain coverage in humans. Maximizing whole-brain coverage is a non-trivial benefit, given that regions not implanted with recording electrodes are effectively missing from descriptive maps of information content, and (more critically) can act as unobserved confounders to causal inferences of how task information emerges (Pearl and Mackenzie, 2018). Our network decoding approach revealed broad representation of stimulus and response information across all major functional networks, paralleling the spatially distributed information profiles reported in animal studies (Bernardi et al., 2020; Kauvar et al., 2020; Raposo et al., 2014; Rigotti et al., 2013; Siegel et al., 2015). Despite this broad decodability, analysis of the information peak and temporal onset revealed a degree of network specialization: response information emerged earliest and maximally in the Motor network, and sensory information emerged earliest and maximally in the Visual network. Our findings align with recent electrophysiological research challenging a strictly compartmentalized or modular view of functional organization (which would not anticipate spatially distributed information), in favor of a more graded organization (Kauvar et al., 2020; Siegel et al., 2015).

In accounting for this graded organization, previous reports have highlighted the need to examine contributions from regions beyond sensory/motor content networks to neural coding schemes that translate sensory inputs to a behavioral response (Gwilliams and King, 2020; Kauvar et al., 2020). To this end, we identified strong involvement of the CCNs (the DAN, FPN and CON) in representing response information underlying overt behavior, in keeping with previous fMRI reports (Cole et al., 2013; Ito et al., 2017). Critically, the enhanced temporal resolution of source EEG enabled identification of distinct CCN network profiles emerging dynamically over two epochs concomitant to the generation of behavior. These two epochs were shown to elicit distinct response

representations, with the pattern of CCN engagement supporting a link between these epochs and distinct motor preparation and motor execution/feedback processes respectively (Elsayed et al., 2016; Hanks et al., 2015; Raposo et al., 2014; Wascher and Wauschkuhn, 1996). Specifically, we observed prominent information in the DAN and Visual networks during earlier motor preparation, whereas later motor execution/feedback was characterized by lower Visual network engagement and stronger, less differentiated CCN engagement. The pattern of results suggest a spatiotemporally resolved network coding scheme, with the DAN preferentially associated with the translation of sensory to response information during motor preparation, and more collaborative engagement of the CCNs underlying the generation of later purer response representations. This extended observation of similar response coding schemes via invasive recordings confined to the motor cortex (Ames et al., 2019; Elsayed et al., 2016) to demonstrate that similar sequential response codes operate amongst large-scale sensory/motor content networks and CCNs.

Our subsequent network modeling analyses sought insight into the computational principles that determined these spatiotemporal signatures. This represents a significant advance over merely describing task information, through formalizing a model of how that information is computed and transferred across the brain – a lacuna in the information decoding literature highlighted previously (Ito et al., 2019; Kriegeskorte and Douglas, 2018; de-Wit et al., 2016). The resultant "dynamic activity flow model" used connectivity estimated empirically during task-free rest to parameterize a neural network architecture, wherein these connectivity estimates weighted the likelihood of task information propagating to Motor network regions from the rest of the brain.

This activity flow model implemented key output refinements compared to previous versions (Cole et al., 2016b; Ito et al., 2017; Mill et al., 2020). Firstly, the model produced fully dynamic outputs of behavioral response information (with millisecond-scale resolution), unlike previous versions that predicted trial-averaged fMRI beta activations. Extending to fully dynamic predictions was facilitated by application in high temporal resolution source EEG data, and indeed the success of the model in this imaging modality (the first application in a non-fMRI modality) testifies to the general validity of its mechanistic tenets. A second output refinement was that predictions of future information dynamics were accomplished through an autoregressive modeling framework (past lagged task activations weighted by lagged restFC). Predicting the future information state of the brain represents a significant advance over more standard approaches of predicting held-out data at the same trial timepoint, given that this integrates a fundamental principle of causality and the direction of time(i.e. causes temporally precede effects; Granger, 1969; Pearl and Mackenzie, 2018). Applying this future-predicting property to simulate the emergence of response information specifically ensured functional significance, given that this information drove subjects' overt behavior (Williams et al., 2007; de-Wit et al., 2016).

The above output features (fully dynamic and prospective predictions of task information) bring the activity flow framework into the realm of artificial neural network models (ANNs) that have emerged as the dominant method of simulating complex cognitive information in the field of artificial intelligence (Ito et al., 2020; Yamins and DiCarlo, 2016; Yang and Wang, 2020). A critical benefit of our approach over these artificial models is its foundation in empirical data, which enabled 1) empirical estimation of network connectivity weights and 2) direct comparison of model-derived representational geometry with actual empirical data.

Regarding the first point, our ability to simulate future information dynamics in a causally interpretable way (i.e. via past activations propagating over lagged restFC) was critically dependent on causally interpretable estimation of restFC (Mill et al., 2017; Reid et al., 2019). We used an autoregressive form of multiple linear regression to estimate parameters capturing dynamic, lagged, direct and directional connectivity between brain regions (see Methods). This yielded more interpretable and neuroscientifically valid connectivity weights than those estimated in ANNs via random initialization followed by intensive optimization. These connectivity weights can also be conceptualized as empirically derived "inductive biases", previously identified as necessary additions to ANN architectures to improve learning and generalization performance towards human levels (Ito et al., 2019; Sinz et al., 2019). The fact that these connectivity weights generalized from predicting future activations in the rest data (embedded in how FC was estimated, see Figure 3A and Methods) to making similar predictions during task performance attests both to the intrinsic (task-general) functional network architecture present at rest (Cole et al., 2014; Krienen et al., 2014), as well as the utility of autoregressive modeling methods in capturing it (Gates et al., 2010; Liégeois et al., 2017; Mill et al., 2016). From a more functional perspective, the success of dynamic activity flow modeling provides powerful evidence of the cognitive relevance of restFC (Mill et al., 2017), which yielded accurate predictions even in this difficult scenario of simulating future information dynamics in high temporal resolution data. The results therefore support the hypothesized role for restFC as a large-scale analogue of synaptic weight processes that coordinate information processing across the brain (Eichenbaum, 2018; Hebb, 1949; Wig et al., 2011).

A second benefit of our empirical network modeling approach over ANNs is that it enabled direct assessment of the veridical overlap of model-derived representations with those instantiated in the human brain. Given that the degree of computational flexibility is high even in dramatically simplified neural networks (Cybenko, 1989), this raises the possibility that accurate simulation of task information might arise from non-veridical representational codes (Kriegeskorte and Douglas, 2018; Sinz et al., 2019; de-Wit et al., 2016; Yamins and DiCarlo, 2016). To address this, we first demonstrated the model's accuracy in predicting established ERP effects in the "raw" task activations that were the basis of the MVPA decoding analyses. The veridicality of the model's representations was even more directly evidenced by the representational overlap analyses, which demonstrated reliable decoding of the actual data from the model-predicted representations. Predicting outputs in the same neural units as the empirical data (i.e. task activation representations in the Motor network) simplified the process of comparing the predicted and actual representational geometry, in comparison to the abstract units typically implemented in ANNs. Although representational similarity approaches exist that can align the task representation space across abstract model and empirical neural units (Kriegeskorte and Douglas, 2018; Kriegeskorte et al., 2008; see also Yang et al., 2019), these typically depend on the presence of many task conditions and result in alignment in an abstract representational space (as opposed to the actual neural space used here). Critically, with continued advancement in computing hardware and access to large training datasets, it seems likely that ANNs will continue to yield accurate simulations of increasingly complex neurocognitive phenomena. It is therefore vital that future neural network simulations aspiring towards neuroscientific insight are interrogated for their neural veridicality, rather than judging them solely on their ability to perform cognitive tasks.

Our final extension of dynamic activity flow modeling to simulate lesioning of different networks provided additional causal insight into which networks were dominant drivers of behavioral

response information. This confirmed the strong collaborative influence of the CCNs that was alluded to by the descriptive network decoding results, with activity flow models built selectively from these three networks yielding the highest predicted response information and with the earliest onsets. Whilst the implied roles for the CCNs are partially consistent with previous fMRI findings linking the DAN preferentially to sensory control (Corbetta et al., 2008; Ito et al., 2017), the CON to motor control (Newbold et al., 2020b) and the FPN to flexible task control more generally (Cocuzza et al., 2020; Cole et al., 2013), separating these networks based on the precise timing and geometry of response representations was uniquely accomplished by our superior temporal resolution. It will hence require future work with similarly spatiotemporally resolved approaches to assess whether the CCN functional profiles are reconcilable with previous fMRI-centric views, or whether these need to be updated.

A sound basis for causal inference within our modeling framework was derived from the strong causal constraints imposed on its core computations (causally interpretable estimation of restFC) and output properties (fully dynamic and prospective predictions of behaviorally relevant response information). This was furthered in the lesioning analysis specifically through simulating perturbations to the system, as an analogue to classical lesioning approaches in neuropsychology (Sadaghiani et al., 2018) and disruptions to neural processing induced by brain stimulation techniques (e.g. transcranial magnetic stimulation, TMS; transcranial alternating current stimulation, TACS). The latter class of methods will be necessary to verify the accuracy of our simulated lesioning results e.g. whether inhibiting (via TMS, Huang et al., 2005) or desynchronizing (via TACS, Polanía et al., 2012) CCN regions during task performance disproportionately disrupts the emergence of response information and consequent behavior, relative to disrupting other networks.

This potential for dynamic activity flow modeling to generate predictions for stimulation interventions raises more general utility in empirically deriving individualized network models of a particular cognitive function. The efficacy in individualizing the models is demonstrated by our consistently high subject-level prediction accuracy, and previous reports highlighting the highly individualized “trait” information contained in restFC (Gordon et al., 2017; Gratton et al., 2018). The simulated lesioning approach gives a particularly clear example of how individualized network models can be manipulated/perturbed to generate insights into individual subjects’ neural information coding profiles. Beyond increasing theoretical understanding into computational principles of task information, such insights could be useful in modeling dysfunction in clinical contexts, and the successful implementation of the network models in empirical human imaging data (with spatiotemporal neural features preserved) makes this clinical translation easier than for more abstract ANNs (Mill et al., 2020; Woo et al., 2017). In achieving this long-term clinical aim, future work will undoubtedly be necessary to test the efficacy of dynamic activity flow modeling in simulating information in more complex and naturalistic tasks, and in clinical disorders. Nevertheless, our preliminary findings attest to the potential of harnessing the causal computational tenets of the activity flow framework to build accurate, individualized models of neurocognitive function in health and disease.

Methods

CONTACT FOR RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Ravi Mill (ravi.mill@rutgers.edu).

MATERIALS AVAILABILITY

This study did not generate new unique reagents.

DATA AND CODE AVAILABILITY

Raw and preprocessed EEG data used in the present report will be uploaded to www.osf.io upon acceptance. The critical analysis code used to 1) calculate restFC via the MVAR approach, and 2) generate task activation predictions via dynamic activity flow modeling are available here https://github.com/ColeLab/DynamicSensoryMotorEGI_release.

KEY RESOURCES TABLE

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw and preprocessed EEG data		(Will be uploaded to www.osf.io upon acceptance)
Software and algorithms		
MATLAB	Mathworks	RRID:SCR_001622; https://www.mathworks.com/
Eprime	Psychology Software Tools (PST)	RRID:SCR_009567; https://pstnet.com/products/e-prime/
Fieldtrip	Oostenveld et al., 2011; Vorwerk et al., 2018	RRID:SCR_004849; http://www.fieldtriptoolbox.org
EEGLAB	Delorme and Makeig, 2004	RRID:SCR_007292; http://sccn.ucsd.edu/eeglab/index.html
CosmoMVPA	Oosterhof et al., 2016	RRID:SCR_014519; http://cosmomvpa.org

LibSVM	Chang and Lin, 2011	RRID:SCR_010243; http://www.csie.ntu.edu.tw/~cjlin/libsvm
RestFC estimation via MVAR and dynamic activity flow modeling	In-house MATLAB scripts	https://github.com/ColeLab/DynamicSensoryMotorEGI_release

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants and study design

The sample consisted of 32 healthy young adults (age mean=21.50 years, range=18-30; 14 female) out of a total of 33 recruited (one subject was excluded due to a computer malfunction aborting their session). All subjects provided informed consent prior to participating, as per the ethical guidelines from the Rutgers University Institutional Review Board (IRB).

Each subject participated in two sessions: an MRI session (in which T1 structural MRI images were collected for EEG source modeling, in addition to T2 structural and 10 minute resting-state fMRI sequences) and an EEG session (consisting of 10 minute pre-task resting-state, ~35 minute task and 10 minute post-task resting-state sequences). The fMRI resting-state and EEG post-task resting-state sequences were not analyzed for the present report.

METHOD DETAILS

MRI AND EEG data acquisition

MRI data was acquired using a 3T Siemens Trio scanner housed at the Rutgers University Brain Imaging Center (RUBIC), with a 32-channel head coil. From the full MRI session (see Participants and study design section for details), only the T1 structural images (MPRAGE, 0.8mm isotropic voxels) were used in the present report to construct realistic head models for EEG source modeling.

EEG data was acquired using the HydroCel Geodesic Sensory Net (HC-GSN) dense-array system manufactured by EGI (1000Hz sampling rate, 256 sensors, Cz reference electrode). Saline solution was used as electrolyte, and impedances across all channels were lowered to <50 kΩ at the start of the session. All resting-state and task components of the experiment were presented to subjects on a laptop running Eprime 2.

Task design

For the EEG pre-task resting-state session (used to estimate FC weights for the dynamic activity flow model, see later Method section), subjects were instructed to rest whilst fixating centrally and to not fall asleep. The ensuing EEG task session consisted of a cued sensory categorization paradigm with a 2-alternative forced choice response format (2AFC; Figure 1 for design schematic). Subjects were cued at the start of each block as to which sensory modality would be presented on the ensuing block of 10 trials, and the associated stimulus-response mappings: visual (horizontal vs vertical lines) or auditory (low vs high pitch tones). Subjects responded with the left or right index fingers across

both sensory modalities, with stimulus-response mappings held constant (e.g. horizontal lines were always categorized with the left index finger). The interval between the cue and first block was randomly jittered (5.5, 6 or 6.5s in duration). On each trial, subjects had 1s to respond after stimulus onset, followed by a randomly jittered inter-trial-interval (1, 1.5 or 2s). The order of visual and auditory sensory conditions was randomized across blocks. All subjects were able to perform the task, with accuracy significantly above chance for both conditions (via one-sample ttest; Visual mean=92.5%, $t(31)=35.2$; Auditory mean=91.5%, $t(31)=23.6$; both $p<.00001$) and no significant difference in accuracy between conditions (via paired ttest, $t(31)=0.69$, $p=.496$).

QUANTIFICATION AND STATISTICAL ANALYSES

Causal temporal filtering prevents circularity in dynamic activity flow modeling.
Importantly, all temporal filters applied in the EEG preprocessing pipeline used a "causal" filter design (de Cheveigné and Nelken, 2019; Rousselet, 2012; Widmann and Schröger, 2012; Widmann et al., 2015). This prevented circularity from entering into our dynamic activity flow modeling analysis (see later Method section) due to leakage of task information from the present timepoint (t0) to past lagged timepoints (t0-n lags) after filtering. Such leakage could arise from more commonly applied "non-causal" filters, wherein through both forward and backward application of the filter its output at t0 is influenced by both past (t0-n lags) and future (t0+n lags) timepoints. The backward step involving future timepoints could introduce circularity into our autoregressive dynamic activity flow modeling approach, from to-be-predicted target region activity at timepoint t0 leaking into t0-n lagged timepoints. This could lead to the target t0 timepoints being predicted to some degree by themselves. To eliminate this possibility, we used a causal filter design (one-pass, hamming-windowed, minimum phase finite impulse response (FIR); Widmann and Schröger, 2012), which only used past timepoints to calculate the output timepoints at t0. This allowed us to benefit from the critical improvements in artifact removal and signal-to-noise ratio resulting from temporal filtering, whilst not introducing circularity into the modeling analyses.

A limitation of causal filtering is the introduction of a delay in the filtered output signal relative to the input (Luck, 2005; Widmann and Schröger, 2012). Hence, we advocate caution in inferring too much into the *precise absolute* timing of the reported activation and information timecourses; rather, we focus on comparisons of *relative timing* (e.g. comparing information decoding onset across spatially distinct functional networks, which have all undergone identical causal filtering and hence are subject to equivalent output delays). Nevertheless for completeness, in the Supplementary Information (Figure S3) we provide a full set of results after application of more conventional non-causal filters during preprocessing (design: two-pass, hamming-windowed, zero-phase Butterworth (infinite impulse response, IIR)). Given that non-causal filters are designed to correct for output latency delays by backwards application of the filter, and as these filters are more commonly applied in EEG/MEG research, these supplemental results should provide some means of comparing the absolute timing of our decoding results with previous reports. Albeit it is again worth highlighting the supplementary nature of these results, given their potential for introducing leakage/circularity in the dynamic activity flow modeling results. We also encourage caution in inferring too much into absolute timing even for the non-causal filter results, as previous studies have established that these filters in practice overcorrect for the latency delay by shifting the onset of signals prior to when they actually occur (de Cheveigné and Nelken, 2019).

EEG preprocessing.
All EEG preprocessing was conducted in Matlab using the Fieldtrip toolbox (Oostenveld et al., 2011), with an identical pipeline run on the pre-task rest and task data (Figure 2 for an overview). Notch filtering was firstly applied to the continuous data to remove line noise (cutoffs at 60Hz, and higher harmonics 120Hz and 180Hz), followed by high-pass filtering (1Hz cutoff). For all filters, a causal filter design was used (as detailed in the preceding section) and filter order was optimized in a lower-level Fieldtrip function to ensure the highest slope for the cutoff frequencies. Noisy sensors were then identified via an automated procedure, wherein all sensors with a zscored mean absolute amplitude (across all timepoints) greater than 3 were removed. The continuous data was segmented into trials: -0.5 to 1.5s for the task data around trial stimulus onset, and into 20s non-overlapping "pseudotrials" for the rest data. Noisy trials/pseudotrials were identified via a similar automated procedure: all trials with a zscored mean absolute amplitude (averaged across all trial timepoints and sensors) greater than 3 were removed.

Temporal independent components analysis (ICA; Jung et al., 2001) was used to identify eye blink and eye movement artifacts. Extended ICA was run on each subject's task and rest data using the "binica" method, accessible in Fieldtrip via an EEGLAB plugin (Delorme and Makeig, 2004). After estimating the component scores, a semi-automated procedure was used to identify components capturing ocular artifacts. Two vertical electro-oculogram channels (EOG) located above and below the right eye were subtracted from each other to provide a timecourse capturing blink artifacts with a high signal-to-noise ratio (Chaumon et al., 2015). Similarly, two horizontal EOG channels located adjacent to the right and left eyes were subtracted from each other for the eye movement/saccade timecourse. These vertical and horizontal EOG signals were used as "artifact templates" that were separately correlated with each ICA component score's timecourse (following a similar approach in Pontifex et al., 2017), with likely artifact-related components identified as those with a zscored absolute Pearson r value > 3. These candidate artifact components were finally accepted/rejected after visual inspection. The ICA-cleaned data were then low-pass filtered (50Hz cutoff), baseline corrected (treating -0.5 to 0 as the baseline for the task trial data, and the mean over each entire pseudotrial as the rest baseline) and re-referenced to the common sensor average.

EEG source modeling.
The preprocessed task and rest EEG data then underwent source modeling to recover the underlying neural source activations from the sensor signals (Figure 2). Source modeling was critical for analyses requiring valid recovery of spatial signatures of task information representation, versus relying on sensor-level analyses that are inherently spatially ambiguous (Haufe et al., 2013; Pernet et al., 2020; Reid et al., 2019; Van de Steen et al., 2019). This endeavour was helped by our use of a dense-array EEG system (256 sensors), given that such a high number of sensors has been previously shown to improve EEG source localization accuracy to match the millimeter precision observed with MEG (Klamer et al., 2015; Michel and Brunet, 2019; Mill et al., 2016).

Realistic, individualized head model construction began by segmenting the anatomical T1 MRI images for each subject into five skull tissue types (grey matter, white matter, cerebrospinal fluid, skull and scalp). This enabled finite element modeling (FEM; Vorwerk et al., 2018) of the distortion of electric dipole fields generated by neural activity as they propagate through skull tissues towards the scalp sensors. The FEM head model was then aligned to the dense-array EEG sensors based on

common fiducial landmarks. Source region locations were specified in MNI space based on the whole-brain functional atlas described by Power et al (Power et al., 2011), which provides independent functional network affiliations for 264 regions. These network affiliations were identified via clustering analyses applied to task and resting-state fMRI data in the original paper, and have subsequently been verified via alternative (Gordon et al., 2016) and multi-modal (Glasser et al., 2016; Ji et al., 2019) human parcellation approaches. Convergent network profiles have also been observed in animals (Stafford et al., 2014; Wang et al., 2013), highlighting the general validity of this functional atlas. The source regions were aligned linearly with the head model and sensors. The full forward model was then estimated via normalized leadfields describing electric field propagation from the neural sources through skull tissues to the recorded EEG electrodes.

Beamforming in the time domain via the linearly constrained minimum variance approach (LCMV; Van Veen et al., 1997) was used to invert the forward model and reconstruct the activation timeseries for each source location. We opted for beamformer-based source modeling as this has been shown to mitigate artifactual field spread influences in source connectivity analyses, compared to alternative minimum norm estimate (MNE) approaches (Schoffelen and Gross, 2009). The orientation of the reconstructed source timeseries was constrained to the direction yielding maximum power (based on singular value decomposition of the 3-dimensional xyz dipole moments). This entire source modeling procedure was applied separately to the task and rest EEG data, reconstructing the relevant activation timeseries for each session from the same 264 spatial locations.

Dynamic multivariate pattern analysis (MVPA).

The source-modeled task EEG data were then submitted to dynamic multivariate pattern analysis (MVPA) to decode task information with millisecond-scale temporal resolution. The analyses focused primarily on decoding behavioral response information (left- vs right-handed responding, Figure 1 for task design) from response-locked data, but decoding of sensory information (visual vs auditory stimulation) from stimulus-locked data was also conducted as a supplementary analysis (see SI, Figure S1). The features used for dynamic MVPA were separately “SensorAll” (timeseries from all 256 EEG channels), “SourceAll” (timeseries from all 264 source regions localized from the Power atlas) or “network sources”. For the latter, dynamic MVPA was performed for each of the 11 Power atlas networks[2] separately (treating within-network source regions as features), yielding 11 information decoding timecourses.

The steps adopted were the same across these feature sets. For response information decoding, the trial data were first rearranged from stimulus- to response-locked (segmented -.45 to .45 seconds around response commission), using the trial reaction times. The data were then downsampled from 1000Hz to 200Hz (by taking every 5th sample sequentially) to boost signal-to-noise for the classifications (as recommended in Grootswagers et al., 2017). Incorrect trials were removed and the remaining trials arranged into the two conditions of interest: correct left and correct right response trials, collapsed across visual and auditory stimulation conditions. Hence, the resulting multivariate

[2]For all analyses targeting inferences at the network level (i.e. the network dynamic MVPA and all dynamic activity flow modeling analyses), we excluded all regions from the "Uncertain" and "Retrieval" Power atlas networks, given that these networks have not been reliably recovered in more recent parcellation schemes (e.g. Ji et al., 2019). For similar reasons, we collapsed the original “Somatomotor-hand” and “Somatomotor-face” (depicted in orange in Fig 2, panel iii) networks into a single “Motor” network definition.

response representations are likely cross-modal due to collapsing across balanced numbers of auditory and visual condition trials (Figure 1). For each response condition, trials were averaged into smaller "subtrial" sets to boost signal-to-noise for the classifications (Grootswagers et al., 2017). Specifically, the ~120 trials in each condition were averaged over ~14 trials to yield ~8 independent subtrials per condition, per subject ('~' reflects variable removal of noisy and incorrect trials across subjects during preprocessing). The trial indices going into these subtrial averages were randomized over 10 iterations, meaning that the entire decoding analysis was run 10 times per subject to ensure robustness.

For each of the subtrial averaging iterations, linear support vector machine (SVM) classifiers were trained separately at each trial timepoint to distinguish between left- and right-handed responses, as implemented by libSVM (Chang and Lin, 2011) and the cosmoMVPA toolbox (Oosterhof et al., 2016). Ten-fold crossvalidation was used, in which trials were randomly split into 80% training and 20% testing sets over 10 folds. Note that the number of train/test trials selected from each response condition was equated on each fold to prevent imbalanced trial counts from influencing the classifiers (Poldrack et al., 2019). The accuracy (%) in classifying between left- and right-handed responses was averaged over folds, with this classification process repeated for each individual timepoint to yield a decoding timecourse capturing response information dynamics for that subject. The final decoding timecourse for each subject was taken as the average across the 10 subtrial averaging iterations.

Statistical significance was assessed via a non-parametric approach treating subjects as a random effect, wherein the classification accuracies at each timepoint were contrasted against chance classification (50%) via the Wilcoxon signrank test (Grootswagers et al., 2017). The resulting p values were Bonferroni-corrected for multiple comparisons across timepoints. For analyses contrasting decoding peaks via paired Wilcoxon signrank tests, peak classification accuracies were extracted from each subject's decoding timecourse from the relevant group time-to-peak (see Results for details). For the SourceAll feature set, we also performed a supplementary temporal generalization analysis (King and Dehaene, 2014) to probe whether the multivariate representations underlying the response decoding timecourse varied or remained constant over time (see SI, S2).

Multivariate autoregressive (MVAR) estimation of resting-state functional connectivity: rationale. The source-modeled rest EEG data were used to estimate the lagged functional connectivity weights (Figure 3A) that parameterized the dynamic activity flow model (Figure 3B). Note that we only used the pre-task rest phase here (see Participants and study design section), given the potential for influences of the task on FC in the post-task rest phase. These could encompass second-order connectivity changes induced by the task (Muraskin et al., 2016) or first-order coactivation effects from cognitive replay of the task (Cole et al., 2019), which could have introduced circularity in predicting activation/information in the earlier task phase.

Our motivations for using a multivariate autoregressive (MVAR) FC approach stemmed from the overarching aim of imposing strong causal constraints on the subsequent dynamic activity flow modeling. Firstly, MVAR permits estimation of dynamic aspects of FC through the inclusion of predictors over temporally lagged timepoints (Liégeois et al., 2017; Mitra and Raichle, 2016), in contradistinction from static FC methods (e.g. Pearson correlation or multiple linear regression estimated over an entire rest session). Secondly, MVAR approaches allow for separation of

contemporaneous (connectivity of region A at t0 with region B at t0) and lagged (connectivity of region A at t0-1 lag with region B at t0) contributions to FC (Gates et al., 2010; Kim et al., 2007). This is unlike typical sliding window dynamic FC approaches, which assess fluctuations in contemporaneous influences over discrete time windows (Hutchison et al., 2013), and have been criticised for potentially capturing sampling variability (i.e. noisy estimates of static FC) rather than true dynamic interactions (Laumann et al., 2016; Liégeois et al., 2017). Estimation of dynamic FC with clear separation of lagged influences was essential to our goal of predicting *future* activation/information states via dynamic activity flow modeling.

Thirdly, our MVAR approach fitted all spatial ("source" predictor regions) and temporal (t0-n lags) terms simultaneously via multiple linear regression models (Figure 3A), meaning that the approach is fully multivariate in both the temporal and spatial domains. This prevents indirect spatial or temporal influences from serving as unobserved causal confounders on the FC weights, which would arise if alternative bivariate FC estimation methods were used (Pearl and Mackenzie, 2018; Reid et al., 2019; Sanchez-Romero and Cole, 2020). Fourthly, our exclusive focus on the lagged MVAR estimates for dynamic activity flow modeling (see later Method section for details) imparted directionality to the approach, given that past t0-n lagged terms predicted future target t0 terms (Figure 3B). This autoregressive form of directionality is the basis of Granger Causality methods (Geweke, 1982; Granger, 1969), which have been fruitfully applied to both fMRI and source modeled MEG/EEG data (Cheung et al., 2010; Mill et al., 2016; Roebroeck et al., 2011).

To summarize, MVAR yielded FC estimates that were dynamic, lagged, direct and directional; thereby increasing the overall causal validity (Mill et al., 2017; Reid et al., 2019) of this version of activity flow modeling beyond previous versions.

MVAR estimation of resting-state functional connectivity: specific approach.
The rest data (already segmented into 20s non-overlapping pseudotrials[3], see EEG preprocessing section) was firstly downsampled from 1000Hz to 200Hz. This matched the downsampling applied to the task data and reduced computation time for the intensive MVAR procedure. Consistent with our focus on predicting activation/information states in the Motor network via dynamic activity flow modeling (see next section), the multiple linear regression models comprising the MVAR step treated the 35 Motor network regions as to-be-predicted (target, j in Figure 3A) regions, and the remaining network regions as predictor (source, i) regions. Timeseries for all regions were firstly zscored within each trial. A model order of 10 lags was selected (i.e. t0-10 timepoints extending 50ms into the past of each target t0 timepoint), based on evidence from invasive electrophysiology suggesting 50ms as a reasonable task information timescale (Crowe et al., 2013; Murray et al., 2014; Siegel et al., 2015) and our previous work directly involving autoregressive modeling in EEG/MEG data (Cole et al., 2010; Mill et al., 2016).

When selecting MVAR predictor terms, our general approach was to estimate all possible spatial and temporal terms to eliminate indirect influences arising from unobserved confounders, with more selective use of predictors in the later dynamic activity flow modeling step. Hence, all non-target Motor network regions were included in the MVAR source set, despite these regions being excluded

[3] We selected 20s as the length of each pseudotrial to capture the lower bound of the low frequency fluctuations considered to contribute to resting-state FC (i.e. 0.1-1Hz, with 20s allowing 2 cycles of 10s; Kucyi et al., 2018; Vincent et al., 2007), as well as contributions from higher frequencies.

from dynamic activity flow (to focus the model on long-distance connections, see next section). Similarly, both contemporaneous (t0) and lagged (t0-10 lags) terms were included for source predictor regions during the MVAR step, despite basing the dynamic activity flow modeling solely on the lagged source terms (to preserve our aim of predicting *future* activation/information states). This follows previous simulation work demonstrating that inclusion of contemporaneous terms improves the precision of lagged estimates, versus fitting the lagged terms alone (Gates et al., 2010). The MVAR models for each target region also included a set of autoregressive lagged terms (over t0-10, excluding t0 to prevent circularity) that estimated serial autocorrelation or self-coupling as per common guidelines (Gates et al., 2010; Geweke, 1982; Liégeois et al., 2017).

The MVAR FC weights for the input spatial and temporal terms were estimated via a regularized form of multiple linear regression, involving principal component analysis (PCA; Figure 3A). We have applied a non-autoregressive variant of this FC approach recently (Mill et al., 2020), and the regularization here was similarly focused on identifying the number of predictor PCs (nPCs) included in each MVAR regression that would reduce overfitting to noise. The approach employs cross-validation to optimize regression beta coefficients to better fit held-out data. To ensure tractability during this computationally intensive procedure, nPCs were varied from 1 to the max number of PCs (i.e. the rank of the predictors, ~2700) in increments of 540. Hence, for a given Motor target region on a given rest pseudotrial, PCA was run on the accompanying predictor matrix (contemporaneous source + lagged source + lagged self-coupling terms). The nPCs retained were varied as per the regularization (1:540:max), with the target region's timeseries then regressed onto those selected nPC scores. The resulting beta weights were transformed back to the original predictor space, and then applied to predict the target timeseries for the next pseudotrial. The mean squared error (MSE) between the actual and predicted next-pseudotrial timeseries was stored, with this process repeated for all trials in sequence (i.e. train on pseudotrial1, test on pseudotrial2; train on pseudotrial2, test on pseudotrial3 etc). Note that this corresponds to a "sliding window" cross-validation scheme that preserves the serial ordering of the input timeseries during model training (unlike more general k-fold cross-validation schemes) to theoretically improve generalization (Cerqueira et al., 2019). The full regularization scheme hence generated an MSE output variable with dimensions nPCs x target regions x test pseudotrials, with the optimal nPCs finally selected as that yielding the minimal MSE (averaged across target regions and test pseudotrials) for that subject. This optimal nPC value was used to estimate the final MVAR restFC weights for all Motor network targets across all pseudotrials for that subject, with the cross-pseudotrial average restFC matrix used in the next dynamic activity flow modeling step.

Dynamic activity flow modeling.

We applied dynamic activity flow modeling to capture the emergence of future response information representations in the Motor network (Figure 3B). The present model built on previous versions (Cole et al., 2016b; Ito et al., 2017; Mill et al., 2020) by including stronger causal constraints. As introduced in the previous MVAR FC estimation sections, these constraints ensured that predictions of future response information states were made from activity flowing over functional network connections in a dynamic, lagged, direct and directional fashion. This causal grounding was strengthened by focusing the modeling on response information representations, given that these underlie generation of overt behavior and hence likely mark a causal end-point of cognitive task processing (Williams et al., 2007; de-Wit et al., 2016). This justified imposing directionality from t0-n lagged non-Motor sources→t0 Motor targets in the model (Figure 3B). Such temporal directionality

would be harder to ascribe in dynamic activity flow models of sensory information in the Visual network (following up on the sensory decoding results, see SI Figure S1), given greater alternation between feedforward (activity flow from lagged non-Visual sources→Visual targets) and feedback cycles (lagged Visual sources→non-Visual targets; Gwilliams and King, 2020).

To reiterate, certain predictor terms that were modeled in the MVAR restFC step were excluded from the ensuing dynamic activity flow modeling: non-target Motor regions and contemporaneous (t0) source terms. These exclusions focused the predictions on spatially long-range and temporally in-the-future activity flow processes respectively. Lagged autoregressive terms were included to account for local self-coupling processes that likely contribute to dynamic activity flow. The source-modeled task EEG data were firstly downsampled to 200Hz. Predicted activations for each target Motor network region (j in Figure 3B), for each task trial, for each timepoint (t0), were then generated from the dot product of lagged task activations (for sources and self-coupling terms, over the same model order of t0-10 samples) and lagged rest MVAR restFC weights (capturing causally valid FC *from* the sources and self-coupling terms *to* the targets). This generated a matrix of predicted Motor region task activations with the same dimensions (Motor regions x task trials x timepoints) as the actual task activations used for the previous decoding analyses. This predicted activation matrix was the basis of subsequent assessments of model accuracy: response information decoded via dynamic MVPA, motor ERP analyses and representational overlap (see ensuing sections).

Some further features of the model that aimed to rigorously rule out circularity are worth highlighting. To reiterate, causal filtering was used during preprocessing to prevent introduction of circularity due to temporal leakage (i.e. task activation from targets at t0 leaking to lagged predictors after filtering, see earlier Method section). Note also that the MVAR FC weights were derived from a resting-state session that was entirely separate from the task activation data, which prevented circularity in estimating FC from the same task session as the to-be-predicted data (due to first-order task coactivation effects contaminating the FC weights; Cole et al., 2019). We also mitigated circularity introduced by EEG field spread artifacts, which could lead to target t0 activations spreading to spatially proximal predictor regions. Firstly, we employed beamformer source modeling which has been shown to minimize such artifacts compared to alternative methods (see EEG source modeling section; Schoffelen and Gross, 2009). Our use of the Power functional atlas (Power et al., 2011) to identify source regions also helped this goal, given that each region is spaced at least 10mm apart from each other and hence avoids the most severe field spread arising between highly proximal regions (Schoffelen and Gross, 2009). Our exclusion of the contemporaneous (t0) source terms from the activity flow step also mitigates field spread, as these artifacts are theoretically instantaneous and hence less likely to influence lagged terms (Nolte et al., 2004; Stinstra and Peters, 1998). As a final rigorous control, we also regressed out the task timeseries for a given target (t0) from all predictor source regions (at the same t0), prior to rearranging them into lagged predictors for dynamic activity flow modeling. The influence of this step in numerically improving model prediction accuracy (versus omission of this step) is detailed in the Supplementary Information (see SI, Figure S4).

Assessment of dynamic activity flow model accuracy: dynamic MVPA.

The accuracy of dynamic activity flow modeling was assessed in a number of ways (Figure 3B). Our primary approach was to apply the identical dynamic MVPA procedure to the model-predicted

Motor region activation timeseries as was applied to the actual data (see earlier Method section). This yielded a predicted response information timecourse, for which significant decodability at each timepoint was assessed via Wilcoxon signrank against 50% chance (Bonferroni-corrected across multiple timepoint comparisons, as before). The recovery of significantly decodable timepoints in the predicted timecourse would provide evidence that our dynamic activity flow modeling approach accurately captured the emergence of future response information. The success of the model in capturing the temporal morphology of the response information timecourse was also quantified via Pearson correlation of the predicted and actual timecourses. This was computed both after averaging the timecourses across subjects (group-level overlap), as well as with a random effects approach that correlated the predicted and actual timecourses for each subject and contrasted this r value (after Fisher-z transform) against 0 via one-sample ttest (subject-level overlap).

Assessment of dynamic activity flow model accuracy: motor event-related potential (ERP).
We also examined the model's accuracy in capturing the activations in Motor network regions that underpinned the dynamic MVPA analyses. This targeted recovery of the well-established motor ERP termed the lateralized readiness potential (Cheyne et al., 2006; Deecke et al., 1976; Kutas and Donchin, 1980). This reflects greater activation around response commission in the Motor network hemisphere contralateral to the response hand. For each subject, the model-predicted trial activations were arranged into "contralateral" (left-response for right hemisphere regions and right-response for left hemisphere regions) and "ipsilateral" (left-response for left hemisphere and right-response for right hemisphere) conditions, and averaged across regions and trials. The difference wave for the resulting contralateral minus ipsilateral waveforms hence indexed recovery of the motor ERP for each subject, with this entire process applied separately to the model-predicted and actual data. In both predicted/actual cases, significance was assessed by contrasting the subject motor ERP fluctuations at each timepoint against 0 via Wilcoxon signrank test, with Bonferroni correction for multiple timepoint comparisons.

The accuracy of the model in predicting the raw task activation effects was assessed firstly as the ability to recover significant activations in the predicted motor ERP waveform. We also quantified the model's ability to recover the morphology of the motor ERP by correlating the predicted and actual difference waves, at both the group- and subject-levels. Whilst this constituted an assessment of "temporal overlap", we were also able to probe predicted-to-actual overlap in the spatial domain as the ERP analyses did not aggregate information across spatially distinct regions (i.e were spatially univariate unlike the dynamic MVPA analyses). The subject motor ERPs were averaged over significant epochs identified in the group actual waveform, separately for each Motor region, and separately for the predicted and actual data. The predicted and actual activation vectors were then correlated at the group- and subject-levels to interrogate how well the dynamic activity flow model captured the spatial pattern of Motor network activations. To ensure that spatial overlap was not critically dependent on the selected significant epoch, we also assessed group- and subject-level overlap between predicted and actual activation vectors that concatenated all Motor regions and all trial timepoints (i.e. capturing the degree of "spatiotemporal overlap" across the entire trial timecourse).

Assessment of dynamic activity flow model accuracy: representational overlap.
We also probed the extent to which the multivariate representations underlying the accurate decoding of response information in the activity flow-predicted data overlapped with the

representations in the actual data. Such direct "representational overlap" between the predicted and actual data would increase confidence that dynamic activity flow modeling is capturing how the brain veridically represents response information. To address this, we tested whether response information in the actual data could be decoded from multivariate representations trained in the predicted data. We employed an approach inspired by representational similarity analysis (Diedrichsen and Kriegeskorte, 2017; Ito et al., 2017; Mur et al., 2009), which was capable of more fine-grained assessment of the similarity in representational geometry than the SVM classification approach used for the main dynamic MVPA analyses.

Predicted task activations for Motor regions were generated via dynamic activity flow modeling as described above, which were then averaged into subtrials separately for the two response conditions (correct left and correct right) and assigned into training and test sets via 10-fold cross-validation (as with the dynamic MVPA analyses). Representational templates were created at each timepoint for each response condition by averaging Motor region activations over relevant trial indices in the *predicted* data. These timepoint-by-timepoint predicted representations were then applied to decode the *actual* Motor activations in the held-out test trials. For each timepoint in each test trial, Pearson correlations were computed between the actual activation vector and both the "correct" predicted condition template (e.g. left response template correlated with actual left response test trial) and the "incorrect" predicted condition template (e.g. right response template correlated with actual left response test trial). Hence, the difference value for correct r minus incorrect r provided an estimate of response information decodability at each timepoint in the actual test data (with values > 0 denoting presence of information). Iterating this process over all timepoints, testing folds and subtrial iterations for each subject generated a timecourse capturing the degree of overlap between the predicted and actual response information representations.

Statistical significance of the representational timecourse was again examined via Wilcoxon signrank tests against 0, with Bonferroni correction across multiple timepoint comparisons. Recovery of significantly decodable timepoints in this "TrainPred-TestActual" timecourse provided evidence that the dynamic activity flow model accurately captured representations underlying future information decoding in the actual data. For comparison, we repeated the same representational similarity approach with trained condition templates computed in the actual data, thereby generating a "TrainActual-TestActual" timecourse. Correlating the TrainPred-TestActual and TrainActual-TestActual timecourses (at both the group- and subject-levels) hence clarified how well the activity flow model captured dynamics in representational geometry over the entire trial.

Network lesioning extension of dynamic activity flow modeling.

To extend the dynamic activity flow model towards insight into principles of functional brain organization, we applied a modified "network lesioning" variant of the model. This followed a similar approach to generating activity flow-predicted response information timecourses for the Motor network via dynamic MVPA, as described above. The critical difference here was that information timecourses were predicted via separate activity flow models that selectively "lesioned" all except one of the 11 functional networks from the Power atlas. Hence, only regions within one specific network were included in the predictor source set (Figure 3B) for each dynamic activity flow model, with lagged self-coupling terms also excluded. This yielded 10 timecourses capturing the unique contributions to response information decoding made by each individual network. Comparing peak decodability across network models hence provided mechanistic insight into which networks were

dominant drivers of response information. Such comparisons were made via visual inspection of the 10 lesioned information timecourses, as well as by contrasting decoding peaks (extracted for each subject from the group time-to-peak across networks) via paired Wilcoxon signrank tests (with FDR correction for multiple pairwise network comparisons). As an alternative to contrasting peak decoding accuracies, we also contrasted the subject-level prediction accuracy for each network model (i.e. Pearson r for each network-predicted timecourse with the actual Motor network timecourse) via paired Wilcoxon signrank tests (with FDR correction for multiple pairwise network comparisons). This captured how well each lesioned network model predicted the morphology of the actual response information timecourse (see SI, Figure S5).

Acknowledgements
The authors acknowledge support by the US National Institutes of Health under awards R01 AG055556 and R01 MH109520 to MWC. The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies. The authors acknowledge the Office of Advanced Research Computing (OARC) at Rutgers, The State University of New Jersey for providing access to the Amarel cluster and associated research computing resources that have contributed to the results reported here.

Author contributions
RDM and MWC designed the task paradigm and analysis approach. RDM, JLH, ECW, NL and RHC were responsible for data collection. RDM performed the analysis under the supervision of MWC, with help from JLH, ECW and NL. RDM and MWC wrote the manuscript, with feedback received from all other authors.

Conflict of interest
The authors declare no competing financial interests.

References

Ames, K.C., Ryu, S.I., and Shenoy, K.V. (2019). Simultaneous motor preparation and execution in a last-moment reach correction task. Nat. Commun. *10*.

Bernardi, S., Benna, M.K., Rigotti, M., Munuera, J., Fusi, S., and Salzman, C.D. (2020). The Geometry of Abstraction in the Hippocampus and Prefrontal Cortex. Cell *183*, 954-967.e21.

Brookes, M.J., Mullinger, K.J., Stevenson, C.M., Morris, P.G., and Bowtell, R. (2008). Simultaneous EEG source localisation and artifact rejection during concurrent fMRI by means of spatial filtering. NeuroImage *40*, 1090–1104.

Buonomano, D., and Merzenich, M. (1995). Temporal information transformed into a spatial code by a neural network with realistic properties. Science *267*, 1028–1030.

Caporale, N., and Dan, Y. (2008). Spike Timing–Dependent Plasticity: A Hebbian Learning Rule. Annu. Rev. Neurosci. *31*, 25–46.

Cerqueira, V., Torgo, L., and Mozetic, I. (2019). Evaluating time series forecasting models: An empirical study on performance estimation methods. ArXiv190511744 Cs Stat.

Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol. *2*, 1–27.

Chaumon, M., Bishop, D.V.M., and Busch, N.A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. J. Neurosci. Methods *250*, 47–63.

Cheung, B.L.P., Riedner, B.A., Tononi, G., and Van Veen, B. (2010). Estimation of Cortical Connectivity From EEG Using State-Space Models. IEEE Trans. Biomed. Eng. *57*, 2122–2134.

de Cheveigné, A., and Nelken, I. (2019). Filters: When, Why, and How (Not) to Use Them. Neuron *102*, 280–293.

Cheyne, D., Bakhtazad, L., and Gaetz, W. (2006). Spatiotemporal mapping of cortical activity accompanying voluntary movements using an event-related beamforming approach. Hum. Brain Mapp. *27*, 213–229.

Cocuzza, C.V., Ito, T., Schultz, D., Bassett, D.S., and Cole, M.W. (2020). Flexible Coordinator and Switcher Hubs for Adaptive Task Control. J. Neurosci. *40*, 6949–6968.

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

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

Cole, M.W., Bassett, D.S., Power, J.D., Braver, T.S., and Petersen, S.E. (2014). Intrinsic and Task-Evoked Network Architectures of the Human Brain. Neuron *83*, 238–251.

Cole, M.W., Ito, T., and Braver, T.S. (2016a). The Behavioral Relevance of Task Information in Human Prefrontal Cortex. Cereb. Cortex *26*, 2497–2505.

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

Cole, M.W., Ito, T., Schultz, D., Mill, R., Chen, R., and Cocuzza, C. (2019). Task activations produce spurious but systematic inflation of task functional connectivity estimates. NeuroImage *189*, 1–18.

Corbetta, M., Patel, G., and Shulman, G.L. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. Neuron *58*, 306–324.

Crowe, D.A., Goodwin, S.J., Blackman, R.K., Sakellaridi, S., Sponheim, S.R., MacDonald, A.W., and Chafee, M.V. (2013). Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. Nat. Neurosci. *16*, 1484–1491.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. Math. Control Signals Syst. *2*, 303–314.

Deecke, L., Grozinger, B., and Kornhuber, H.H. (1976). Voluntary finger movement in man: Cerebral potentials and theory. Biol. Cybern. *23*, 99–119.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods *134*, 9–21.
Diedrichsen, J., and Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. PLOS Comput. Biol. *13*, e1005508.
Eichenbaum, H. (2018). Barlow versus Hebb: When is it time to abandon the notion of feature detectors and adopt the cell assembly as the unit of cognition? Neurosci. Lett. *680*, 88–93.
Eliasmith, C., Stewart, T.C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., and Rasmussen, D. (2012). A Large-Scale Model of the Functioning Brain. Science *338*, 1202–1205.
Elsayed, G.F., Lara, A.H., Kaufman, M.T., Churchland, M.M., and Cunningham, J.P. (2016). Reorganization between preparatory and movement population responses in motor cortex. Nat. Commun. *7*.
Gates, K.M., Molenaar, P.C.M., Hillary, F.G., Ram, N., and Rovine, M.J. (2010). Automatic search for fMRI connectivity mapping: An alternative to Granger causality testing using formal equivalences among SEM path modeling, VAR, and unified SEM. NeuroImage *50*, 1118–1125.
Geweke, J. (1982). Measurement of Linear Dependence and Feedback between Multiple Time Series. J. Am. Stat. Assoc. *77*, 304–313.
Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., et al. (2016). A multi-modal parcellation of human cerebral cortex. Nature *536*, 171–178.
Gordon, E.M., Laumann, T.O., Adeyemo, B., Huckins, J.F., Kelley, W.M., and Petersen, S.E. (2016). Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. Cereb. Cortex *26*, 288–303.
Gordon, E.M., Laumann, T.O., Gilmore, A.W., Newbold, D.J., Greene, D.J., Berg, J.J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H., et al. (2017). Precision Functional Mapping of Individual Human Brains. Neuron *95*, 791-807.e7.
Granger, C.W.J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica *37*, 424.
Gratton, C., Laumann, T.O., Nielsen, A.N., Greene, D.J., Gordon, E.M., Gilmore, A.W., Nelson, S.M., Coalson, R.S., Snyder, A.Z., Schlaggar, B.L., et al. (2018). Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. Neuron *98*, 439-452.e5.
Grootswagers, T., Wardle, S.G., and Carlson, T.A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. J. Cogn. Neurosci. *29*, 677–697.
Gwilliams, L., and King, J.-R. (2020). Recurrent processes support a cascade of hierarchical decisions. ELife *9*.
Hanks, T.D., Kopec, C.D., Brunton, B.W., Duan, C.A., Erlich, J.C., and Brody, C.D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. Nature *520*, 220–223.
Haufe, S., Nikulin, V.V., Müller, K.-R., and Nolte, G. (2013). A critical assessment of connectivity measures for EEG data: A simulation study. NeuroImage *64*, 120–133.
Haynes, J.-D. (2015). A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. Neuron *87*, 257–270.
Hebb, D.O. (1949). The organization of behavior: a neuropsychological theory (Mahwah, N.J: L. Erlbaum Associates).
Hernández, A., Nácher, V., Luna, R., Zainos, A., Lemus, L., Alvarez, M., Vázquez, Y., Camarillo, L., and Romo, R. (2010). Decoding a Perceptual Decision Process across Cortex. Neuron *66*, 300–314.
Huang, Y.-Z., Edwards, M.J., Rounis, E., Bhatia, K.P., and Rothwell, J.C. (2005). Theta Burst Stimulation of the Human Motor Cortex. Neuron *45*, 201–206.
Hubbard, J., Kikumoto, A., and Mayr, U. (2019). EEG Decoding Reveals the Strength and Temporal

Dynamics of Goal-Relevant Representations. Sci. Rep. *9*.
Hui, H.B., Pantazis, D., Bressler, S.L., and Leahy, R.M. (2010). Identifying true cortical interactions in MEG using the nulling beamformer. NeuroImage *49*, 3161–3174.
Hutchison, R.M., Womelsdorf, T., Allen, E.A., Bandettini, P.A., Calhoun, V.D., Corbetta, M., Della Penna, S., Duyn, J.H., Glover, G.H., Gonzalez-Castillo, J., et al. (2013). Dynamic functional connectivity: Promise, issues, and interpretations. NeuroImage *80*, 360–378.
Ito, T., Kulkarni, K.R., Schultz, D.H., Mill, R.D., Chen, R.H., Solomyak, L.I., and Cole, M.W. (2017). Cognitive task information is transferred between brain regions via resting-state network topology. Nat. Commun. *8*.
Ito, T., Hearne, L., Mill, R., Cocuzza, C., and Cole, M.W. (2019). Discovering the Computational Relevance of Brain Network Organization. Trends Cogn. Sci. S1364661319302402.
Ito, T., Yang, G.R., Laurent, P., Schultz, D.H., and Cole, M.W. (2020). Constructing neural network models from brain data reveals representational transformations underlying adaptive behavior (Neuroscience).
Ji, J.L., Spronk, M., Kulkarni, K., Repovš, G., Anticevic, A., and Cole, M.W. (2019). Mapping the human brain's cortical-subcortical functional network organization. NeuroImage *185*, 35–57.
Jimura, K., and Poldrack, R.A. (2012). Analyses of regional-average activation and multivoxel pattern information tell complementary stories. Neuropsychologia *50*, 544–552.
Jung, T.-P., Makeig, S., McKeown, M.J., Bell, A.J., Lee, T.-W., and Sejnowski, T.J. (2001). Imaging brain dynamics using independent component analysis. Proc. IEEE *89*, 1107–1122.
Kauvar, I.V., Machado, T.A., Yuen, E., Kochalka, J., Choi, M., Allen, W.E., Wetzstein, G., and Deisseroth, K. (2020). Cortical Observation by Synchronous Multifocal Optical Sampling Reveals Widespread Population Encoding of Actions. Neuron *107*, 351-367.e19.
Kim, J., Zhu, W., Chang, L., Bentler, P.M., and Ernst, T. (2007). Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data. Hum. Brain Mapp. *28*, 85–93.
King, J.-R., and Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn. Sci. *18*, 203–210.
Klamer, S., Elshahabi, A., Lerche, H., Braun, C., Erb, M., Scheffler, K., and Focke, N.K. (2015). Differences Between MEG and High-Density EEG Source Localizations Using a Distributed Source Model in Comparison to fMRI. Brain Topogr. *28*, 87–94.
Kriegeskorte, N., and Douglas, P.K. (2018). Cognitive computational neuroscience. Nat. Neurosci. *21*, 1148–1160.
Kriegeskorte, N., and Kievit, R.A. (2013). Representational geometry: integrating cognition, computation, and the brain. Trends Cogn. Sci. *17*, 401–412.
Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based functional brain mapping. Proc. Natl. Acad. Sci. *103*, 3863–3868.
Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini, P.A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. Neuron *60*, 1126–1141.
Krienen, F.M., Yeo, B.T.T., and Buckner, R.L. (2014). Reconfigurable task-dependent functional coupling modes cluster around a core functional architecture. Philos. Trans. R. Soc. B Biol. Sci. *369*, 20130526–20130526.
Kucyi, A., Schrouff, J., Bickel, S., Foster, B.L., Shine, J.M., and Parvizi, J. (2018). Intracranial Electrophysiology Reveals Reproducible Intrinsic Functional Connectivity within Human Brain Networks. J. Neurosci. *38*, 4230–4242.
Kutas, M., and Donchin, E. (1980). Preparation to respond as manifested by movement-related brain potentials. Brain Res. *202*, 95–115.
Laumann, T.O., Snyder, A.Z., Mitra, A., Gordon, E.M., Gratton, C., Adeyemo, B., Gilmore, A.W., Nelson, S.M., Berg, J.J., Greene, D.J., et al. (2016). On the Stability of BOLD fMRI Correlations. Cereb. Cortex.

Levy, W.B., and Steward, O. (1983). Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. Neuroscience *8*, 791–797.
Lewis, C.M., Baldassarre, A., Committeri, G., Romani, G.L., and Corbetta, M. (2009). Learning sculpts the spontaneous activity of the resting human brain. Proc. Natl. Acad. Sci. *106*, 17558–17563.
Liégeois, R., Laumann, T.O., Snyder, A.Z., Zhou, J., and Yeo, B.T.T. (2017). Interpreting temporal fluctuations in resting-state functional connectivity MRI. NeuroImage *163*, 437–455.
Luck, S.J. (2005). An introduction to the event-related potential technique (Cambridge, Mass: MIT Press).
Michel, C.M., and Brunet, D. (2019). EEG Source Imaging: A Practical Review of the Analysis Steps. Front. Neurol. *10*.
Mill, R.D., Bagic, A., Bostan, A., Schneider, W., and Cole, M.W. (2016). Empirical validation of directed functional connectivity. NeuroImage *146*, 275–287.
Mill, R.D., Ito, T., and Cole, M.W. (2017). From connectome to cognition: The search for mechanism in human functional brain networks. NeuroImage *160*, 124–139.
Mill, R.D., Gordon, B.A., Balota, D.A., and Cole, M.W. (2020). Predicting dysfunctional age-related task activations from resting-state network alterations. NeuroImage *221*, 117167.
Mitra, A., and Raichle, M.E. (2016). How networks communicate: propagation patterns in spontaneous brain activity. Philos. Trans. R. Soc. B Biol. Sci. *371*, 20150546.
Mur, M., Bandettini, P.A., and Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI--an introductory guide. Soc. Cogn. Affect. Neurosci. *4*, 101–109.
Muraskin, J., Dodhia, S., Lieberman, G., Garcia, J.O., Verstynen, T., Vettel, J.M., Sherwin, J., and Sajda, P. (2016). Brain dynamics of post-task resting state are influenced by expertise: Insights from baseball players. Hum. Brain Mapp. *37*, 4454–4471.
Murray, J.D., Bernacchia, A., Freedman, D.J., Romo, R., Wallis, J.D., Cai, X., Padoa-Schioppa, C., Pasternak, T., Seo, H., Lee, D., et al. (2014). A hierarchy of intrinsic timescales across primate cortex. Nat. Neurosci. *17*, 1661–1663.
Newbold, D.J., Laumann, T.O., Hoyt, C.R., Hampton, J.M., Montez, D.F., Raut, R.V., Ortega, M., Mitra, A., Nielsen, A.N., Miller, D.B., et al. (2020a). Plasticity and Spontaneous Activity Pulses in Disused Human Brain Circuits. Neuron *107*, 580-589.e6.
Newbold, D.J., Gordon, E.M., Laumann, T.O., Seider, N.A., Montez, D.F., Gross, S.J., Zheng, A., Nielsen, A.N., Hoyt, C.R., Hampton, J.M., et al. (2020b). Cingulo-Opercular Control Network Supports Disused Motor Circuits in Standby Mode (Neuroscience).
Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., and Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. Clin. Neurophysiol. *115*, 2292–2307.
Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Comput. Intell. Neurosci. *2011*, 1–9.
Oosterhof, N.N., Connolly, A.C., and Haxby, J.V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. Front. Neuroinformatics *10*.
Pearl, J., and Mackenzie, D. (2018). The book of why: the new science of cause and effect (New York: Basic Books).
Pernet, C., Garrido, M.I., Gramfort, A., Maurits, N., Michel, C.M., Pang, E., Salmelin, R., Schoffelen, J.M., Valdes-Sosa, P.A., and Puce, A. (2020). Issues and recommendations from the OHBM COBIDAS MEEG committee for reproducible EEG and MEG research. Nat. Neurosci.
Petersen, S.E., and Sporns, O. (2015). Brain Networks and Cognitive Architectures. Neuron *88*, 207–219.
Polanía, R., Nitsche, M.A., Korman, C., Batsikadze, G., and Paulus, W. (2012). The Importance of Timing in Segregated Theta Phase-Coupling for Cognitive Performance. Curr. Biol. *22*, 1314–1318.
Poldrack, R.A., Huckins, G., and Varoquaux, G. (2019). Establishment of Best Practices for Evidence

for Prediction: A Review. JAMA Psychiatry.
Pontifex, M.B., Miskovic, V., and Laszlo, S. (2017). Evaluating the efficacy of fully automated approaches for the selection of eyeblink ICA components: Eyeblink component identification. Psychophysiology *54*, 780–791.
Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., et al. (2011). Functional Network Organization of the Human Brain. Neuron *72*, 665–678.
Raposo, D., Kaufman, M.T., and Churchland, A.K. (2014). A category-free neural population supports evolving demands during decision-making. Nat. Neurosci. *17*, 1784–1792.
Reid, A.T., Headley, D.B., Mill, R.D., Sanchez-Romero, R., Uddin, L.Q., Marinazzo, D., Lurie, D.J., Valdés-Sosa, P.A., Hanson, S.J., Biswal, B.B., et al. (2019). Advancing functional connectivity research from association to causation. Nat. Neurosci. *22*, 1751–1760.
Rigotti, M., Barak, O., Warden, M.R., Wang, X.-J., Daw, N.D., Miller, E.K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature *497*, 585–590.
Roebroeck, A., Formisano, E., and Goebel, R. (2011). The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution. NeuroImage *58*, 296–302.
Rousselet, G.A. (2012). Does Filtering Preclude Us from Studying ERP Time-Courses? Front. Psychol. *3*.
Sadaghiani, S., Dombert, P.L., Løvstad, M., Funderud, I., Meling, T.R., Endestad, T., Knight, R.T., Solbakk, A.-K., and D'Esposito, M. (2018). Lesions to the Fronto-Parietal Network Impact Alpha-Band Phase Synchrony and Cognitive Control. Cereb. Cortex.
Sanchez-Romero, R., and Cole, M.W. (2020). Combining Multiple Functional Connectivity Methods to Improve Causal Inferences. J. Cogn. Neurosci. 1–15.
Saxena, S., and Cunningham, J.P. (2019). Towards the neural population doctrine. Curr. Opin. Neurobiol. *55*, 103–111.
Schoffelen, J.-M., and Gross, J. (2009). Source connectivity analysis with MEG and EEG. Hum. Brain Mapp. *30*, 1857–1865.
Siegel, M., Buschman, T.J., and Miller, E.K. (2015). Cortical information flow during flexible sensorimotor decisions. Science *348*, 1352–1355.
Sinz, F.H., Pitkow, X., Reimer, J., Bethge, M., and Tolias, A.S. (2019). Engineering a Less Artificial Intelligence. Neuron *103*, 967–979.
Stafford, J.M., Jarrett, B.R., Miranda-Dominguez, O., Mills, B.D., Cain, N., Mihalas, S., Lahvis, G.P., Lattal, K.M., Mitchell, S.H., David, S.V., et al. (2014). Large-scale topology and the default mode network in the mouse connectome. Proc. Natl. Acad. Sci. *111*, 18745–18750.
Stinstra, J.G., and Peters, M.J. (1998). The volume conductor may act as a temporal filter on the ECG and EEG. Med. Biol. Eng. Comput. *36*, 711–716.
Van de Steen, F., Faes, L., Karahan, E., Songsiri, J., Valdes-Sosa, P.A., and Marinazzo, D. (2019). Critical Comments on EEG Sensor Space Dynamical Connectivity Analysis. Brain Topogr. *32*, 643–654.
Van Veen, B.D., Van Drongelen, W., Yuchtman, M., and Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans. Biomed. Eng. *44*, 867–880.
Vincent, J.L., Patel, G.H., Fox, M.D., Snyder, A.Z., Baker, J.T., Van Essen, D.C., Zempel, J.M., Snyder, L.H., Corbetta, M., and Raichle, M.E. (2007). Intrinsic functional architecture in the anaesthetized monkey brain. Nature *447*, 83–86.
Vorwerk, J., Oostenveld, R., Piastra, M.C., Magyari, L., and Wolters, C.H. (2018). The FieldTrip-SimBio pipeline for EEG forward solutions. Biomed. Eng. OnLine *17*.
Wang, Z., Chen, L.M., Négyessy, L., Friedman, R.M., Mishra, A., Gore, J.C., and Roe, A.W. (2013). The Relationship of Anatomical and Functional Connectivity to Resting-State Connectivity in Primate Somatosensory Cortex. Neuron *78*, 1116–1126.
Wascher, E., and Wauschkuhn, B. (1996). The interaction of stimulus- and response-related

processes measured by event-related lateralizations of the EEG. Electroencephalogr. Clin. Neurophysiol. *99*, 149–162.
Widmann, A., and Schröger, E. (2012). Filter Effects and Filter Artifacts in the Analysis of Electrophysiological Data. Front. Psychol. *3*.
Widmann, A., Schröger, E., and Maess, B. (2015). Digital filter design for electrophysiological data – a practical approach. J. Neurosci. Methods *250*, 34–46.
Wig, G.S., Schlaggar, B.L., and Petersen, S.E. (2011). Concepts and principles in the analysis of brain networks: Brain networks. Ann. N. Y. Acad. Sci. *1224*, 126–146.
Williams, M.A., Dang, S., and Kanwisher, N.G. (2007). Only some spatial patterns of fMRI response are read out in task performance. Nat. Neurosci. *10*, 685–686.
de-Wit, L., Alexander, D., Ekroll, V., and Wagemans, J. (2016). Is neuroimaging measuring information in the brain? Psychon. Bull. Rev. *23*, 1415–1428.
Woo, C.-W., Chang, L.J., Lindquist, M.A., and Wager, T.D. (2017). Building better biomarkers: brain models in translational neuroimaging. Nat. Neurosci. *20*, 365–377.
Yamins, D.L.K., and DiCarlo, J.J. (2016). Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. *19*, 356–365.
Yamins, D.L.K., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., and DiCarlo, J.J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc. Natl. Acad. Sci. *111*, 8619–8624.
Yang, G.R., and Wang, X.-J. (2020). Artificial Neural Networks for Neuroscientists: A Primer. Neuron *107*, 1048–1070.
Yang, G.R., Joglekar, M.R., Song, H.F., Newsome, W.T., and Wang, X.-J. (2019). Task representations in neural networks trained to perform many cognitive tasks. Nat. Neurosci. *22*, 297–306.
Zhang, J., Kriegeskorte, N., Carlin, J.D., and Rowe, J.B. (2013). Choosing the Rules: Distinct and Overlapping Frontoparietal Representations of Task Rules for Perceptual Decisions. J. Neurosci. *33*, 11852–11862.