Learning in brain-computer interface control evidenced by joint decomposition of brain and behavior

To cite this article before publication: Jennifer Stiso *et al* 2020 *J. Neural Eng.* in press https://doi.org/10.1088/1741-2552/ab9064

Learning in brain-computer interface control evidenced by joint decomposition of brain and behavior

Jennifer Stiso[1,2], Marie-Constance Corsi[3,4], Jean M. Vettel[5,2,6], Javier Garcia[5,2], Fabio Pasqualetti[7], Fabrizio De Vico Fallani[3,4], Timothy H. Lucas[9], and Danielle S. Bassett[2,8,9,10,11,12,13]

[1]*Neuroscience Graduate Group, Perelman School of Medicine,*
University of Pennsylvania, Philadelphia, PA 19104, USA
[2]*Department of Bioengineering, School of Engineering & Applied Science,*
University of Pennsylvania, Philadelphia, PA 19104, USA
[3]*Inria Paris, Aramis project-team, F-75013, Paris, France*
[4]*Institut du Cerveau et de la Moelle Epinire, ICM, Inserm, U 1127,*
CNRS, UMR 7225, Sorbonne Universit, F-75013, Paris, France
[5]*Human Research & Engineering Directorate, US CCDC Army Research Laboratory, Aberdeen, MD, USA*
[6]*Department of Psychological & Brain Sciences, University of California, Santa Barbara, CA, USA*
[7]*Department of Mechanical Engineering, University of California, Riverside, CA 92521*
[8]*Department of Electrical & Systems Engineering, School of Engineering & Applied Science,*
University of Pennsylvania, Philadelphia, PA 19104, USA
[9]*Department of Neurology, Perelman School of Medicine,*
University of Pennsylvania, Philadelphia, PA 19104, USA
[10]*Department of Psychiatry, Perelman School of Medicine,*
University of Pennsylvania, Philadelphia, PA 19104, USA
[11]*Department of Physics & Astronomy, College of Arts & Sciences,*
University of Pennsylvania, Philadelphia, PA 19104, USA
[12]*The Santa Fe Institute, Santa Fe, NM 87501, USA   and*
[13]*To whom correspondence should be addressed: dsb@seas.upenn.edu*

Objective: Motor imagery-based brain-computer interfaces (BCIs) use an individual's ability to volitionally modulate localized brain activity, often as a therapy for motor dysfunction or to probe causal relations between brain activity and behavior. However, many individuals cannot learn to successfully modulate their brain activity, greatly limiting the efficacy of BCI for therapy and for basic scientific inquiry. Formal experiments designed to probe the nature of BCI learning have offered initial evidence that coherent activity across spatially distributed and functionally diverse cognitive systems is a hallmark of individuals who can successfully learn to control the BCI. However, little is known about how these distributed networks interact through time to support learning.

Approach: Here, we address this gap in knowledge by constructing and applying a multimodal network approach to decipher brain-behavior relations in motor imagery-based brain-computer interface learning using magnetoencephalography. Specifically, we employ a minimally constrained matrix decomposition method – *non-negative matrix factorization* – to simultaneously identify regularized, covarying subgraphs of functional connectivity, to assess their similarity to task performance, and to detect their time-varying expression.

Main Results: We find that learning is marked by diffuse brain-behavior relations: good learners displayed many subgraphs whose temporal expression tracked performance. Individuals also displayed marked variation in the spatial properties of subgraphs such as the connectivity between the frontal lobe and the rest of the brain, and in the temporal properties of subgraphs such as the stage of learning at which they reached maximum expression. From these observations, we posit a conceptual model in which certain subgraphs support learning by modulating brain activity in sensors near regions important for sustaining attention. To test this model, we use tools that stipulate regional dynamics on a networked system (network control theory), and find that good learners display a single subgraph whose temporal expression tracked performance and whose architecture supports easy modulation of sensors located near brain regions important for attention.

Significance: The nature of our contribution to the neuroscience of BCI learning is therefore both computational and theoretical; we first use a minimally-constrained, individual specific method of identifying mesoscale structure in dynamic brain activity to show how global connectivity and interactions between distributed networks supports BCI learning, and then we use a formal network model of control to lend theoretical support to the hypothesis that these identified subgraphs are well suited to modulate attention.

INTRODUCTION

Both human and non-human animals can learn to volitionally modulate diverse aspects of their neural activity from the spiking of single neurons to the coherent activity of brain regions [36, 98, 99]. Such neural modulation is made possible by routing empirical measurements of the user's neural activity to a screen or other external display device that they can directly observe [41, 75, 98]. Referred to as a brain-computer interface (BCI), this tech-

nology can be used not only to control these external devices, but also to causally probe the nature of specific cognitive processes [6, 81, 88], and offers great promise in the treatment of neural dysfunction [71, 89, 111]. However, translating that promise into a reality has proven difficult [1, 50, 103] due to the extensive training that is required and due to the fact that some individuals who undergo extensive training will only achieve moderate control [27, 56, 75]. A better understanding of the neural processes supporting BCI learning is an important first step towards the development of BCI therapies and the identification of specific individuals who are good candidates for treatment [27, 56].

While BCIs vary widely in their nature, we focus on the common motor imagery based BCIs where subjects are instructed to imagine a particular movement to modulate activity in motor cortex. Performance on motor imagery based BCIs has been associated with a diverse array of neural features, demographic factors, and behavioral measures [3, 47, 52, 56, 62]. Neural features predicting performance are frequently identified in areas associated with either performing or imagining action; for example, better performance is associated with higher pre-task activity in supplementary motor areas [48] and larger grey matter volume in somatomotor regions [48]. Interestingly, performance has also been predicted by activity in a diverse range of other cognitive systems relevant for sustained attention, perhaps due to the high cognitive demands associated with BCI learning [56]. Specifically, better performance is associated with greater parietal power suppression in the α band, midline power suppression in the β band, and frontal and occipital activation with motor power suppression in the γ band [3, 37, 43]. The role of sustained attention in BCI control is corroborated by the fact that personality and self-report measures of attention predict successful learning [51]. The heterogeneity of predictors suggests the possibility that individual differences in the interactions between cognitive systems necessary for action, action planning, and attention might explain the idiosyncratic nature of BCI control, although these interactions are challenging to quantify [6, 29].

Assessing the interactions between cognitive systems has historically been rather daunting, in part due to the lack of a common mathematical language in which to frame relevant hypotheses and formalize appropriate computational approaches. With the recent emergence and development of network science [79], and its application to neural systems [16], many efforts have begun to link features of brain networks to BCI learning specifically and to other types of learning more generally. In this formal modeling approach [9], network nodes represent brain regions or sensors and network edges represent statistical relations or so-called *functional connections* between regional time series [30]. Recent studies have demonstrated that patterns of functional connections can provide clearer explanations of the learning process than activation alone [8], and changes in those

functional connections can track changes in behavior [5]. During BCI tasks, functional connectivity reportedly increases within supplementary and primary motor areas [50] and decreases between motor and higher-order association areas as performance becomes more automatic [24]. Data-driven methods to detect putative cognitive systems as modules in functional brain networks have been used to demonstrate that a particularly clear neural marker of learning is reconfiguration of the network's functional modules [61, 68]. Better performance is accompanied by flexible switching of brain regions between distinct modules as task demands change [7, 40, 87].

While powerful, such methods for cognitive system detection are built upon an assumption that limits their conceptual relevance for the study of BCI learning. Specifically, they enforce the constraint that a brain region may only affiliate with one module at a time [60], in spite of the fact that many regions, comprised of heterogeneous neural populations, might participate in multiple neural processes. To address this limitation, recent efforts have begun to employ so-called *soft-partitioning* methods that detect coherent patterns in mesoscale neural activity and connectivity [19, 32, 60, 67]. Common examples of such methods are independent component analysis and principal component analysis, which impose pragmatic but not biological constraints on the orthogonality or independence of partitions. An appealing alternative is non-negative matrix factorization (NMF), which achieves a soft partition by decomposing the data into the small set of sparse, overlapping, time-varying subgraphs that can best reconstruct the original data with no requirement of orthogonality or independence [66]. Previous applications of this method to neuroimaging data have demonstrated that the detected subgraphs can provide a description of time varying mesoscale activity that complements descriptions provided by more traditional approaches [60]. For example, some subgraphs identified with NMF during the resting state have similar spatial distributions to those found with typical module detection methods, while others span between modules [60]. As a minimally constrained method for obtaining a soft partition of neural activity, NMF is a promising candidate for revealing the time-varying neural networks that support BCI learning.

Here, we investigate the properties of dynamic functional connectivity supporting BCI learning. In individuals trained to control a BCI, we use a *wavelet decomposition* to calculate single trial phase-based connectivity in magnetoencephalography (MEG) data in three frequency bands with stereotyped behavior during motor imagery: α (7-14 Hz), β (15-30 Hz), and γ (31-45 Hz) (Fig. 1, step 1). We construct multimodal brain-behavior time series of dynamic functional connectivity and performance, or *configuration matrix* (Fig. 1, step 2 and 3), and apply NMF to those time series to obtain a soft partition into additive subgraphs [66] (Fig. 1, step 4). We determine the degree to which a subgraph tracks performance by defining the *performance loading* as the similarity be-

tween each subgraph's temporal expression and the time course of task accuracy (Fig. 1, step 5). We first identify subgraphs whose performance loading predicted the rate of learning and then we explore the spatial and temporal properties of subgraphs to identify common features across participants. We hypothesize that subgraphs predicting learning do so by being structured and situated in such a way as to easily modulate patterns of activity that support sustained attention, an important component of successful BCI control [56]. After demonstrating the suitability of this approach for our data (Fig. S1A-B), we test this hypothesis by capitalizing on recently developed tools in network control theory, which allowed us to operationalize the network's ability to activate sensors located near regions involved in sustained attention as the energy required for network control [45]. Collectively, our efforts provide a network-level description of neural correlates of BCI performance and learning rate, and a formal network control model that explains those descriptions.

METHODS

Participants

Written informed consent was obtained from twenty healthy, right-handed subjects (aged 27.45 ± 4.01 years; 12 male), who participated in the study conducted in Paris, France. Subjects were enrolled in a longitudinal electroencephalography (EEG) based BCI training with simultaneous MEG recording over four sessions, spanning 2 weeks. All subjects were BCI-naive and none presented with medical or psychological disorders. The study was approved by the ethical committee CPP-IDF-VI of Paris.

BCI task

Subjects were seated in a magnetically shielded room, at a distance of 90 cm from the display screen. Subjects' arms were placed on arm rests to facilitate stability. BCI control features including EEG electrode and frequency were selected in a calibration phase at the beginning of each session, by instructing the subjects to perform motor imagery without any visual feedback.

The BCI task consisted of a standard 1 dimensional, two-target box task [110] in which the subjects modulated their EEG measured α [8-12 Hz] and/or β [14-29 Hz] activity over the left motor cortex to control the vertical position of a cursor moving with constant velocity from the left side of the screen to the right side of the screen. The specific sensor and frequency selected to control the BCI were based on brain activity recorded during a calibration phase before each day of recording. Here, subjects were instructed to perform the BCI task, but received no visual feedback; specifically, the target was present on the screen, but there was no ball moving

towards the target. Each subject completed 5 consecutive runs of 32 trials each for the calibration phase. The EEG features (sensor and frequency) with the largest R-squared values for discriminating motor imagery conditions from rest conditions were used in the subsequent task.

Both cursor and target were presented using the software BCI 2000 [93]. To hit the target-up, the subjects performed a sustained motor imagery of their right-hand grasping and to hit the target-down they remained at rest. Some subjects reported that they imagined grasping objects while others reported that they simply imagined clenching their hand to make a fist. Each trial lasted 7 s and consisted of a 1 s inter-stimulus interval, followed by 2 s of target presentation, 3 s of feedback, and 1 s of result presentation (Fig 2a). If the subject successfully reached the target, the target would change from grey to yellow during the 1 s result section. Otherwise it would remain grey. The feedback portion was the only part of the trial where subjects could observe the effects of their volitional modulation of motor region activity. Specifically, the subjects saw the vertical position of the cursor change based on their neural activity, as it moved towards the screen at a fixed velocity. Brain activity was updated every 28 ms. In the present study, we therefore restricted our analysis to the feedback portion of the motor imagery task because we were interested in the neural dynamics associated with learning to volitionally regulate brain activity rather than in the neural dynamics occurring at rest.

Subjects completed 4 sessions of this BCI task, where each session took place on a different day within two weeks. Each session consisted of 6 runs of 32 trials each. Each trial had either a target in the upper quadrant of the screen, indicating increased motor imagery was needed to reach it, or a target in lower quadrant of the screen, indicating no change in activity was needed to reach it. Only signals from the motor imagery trials were analyzed. This left us with, before trial rejection due to artifacts, 16 motor imagery trials $\times$ 6 runs $\times$ 4 sessions, or 384 trials per subject. Each trial was 7 seconds in duration, leading to 3 minute long runs. Combined with the training phase, each session was 1-1.5 hours total.

Neurophysiological Recordings Data

Recording

MEG and EEG data were simultaneously recorded with an Elekta Neuromag TRIUX machine (MEG) and a 74 EEG-channel system (EEG). While EEG and MEG data were recorded simultaneously, only MEG were analyzed because they are less spatially smeared than EEG signals, and therefore more appropriate for network analyses[26]. Signals were originally sampled at 1000 Hz. We also recorded electromyogram (EMG) signals from the left and right arm of subjects, electrooculograms,
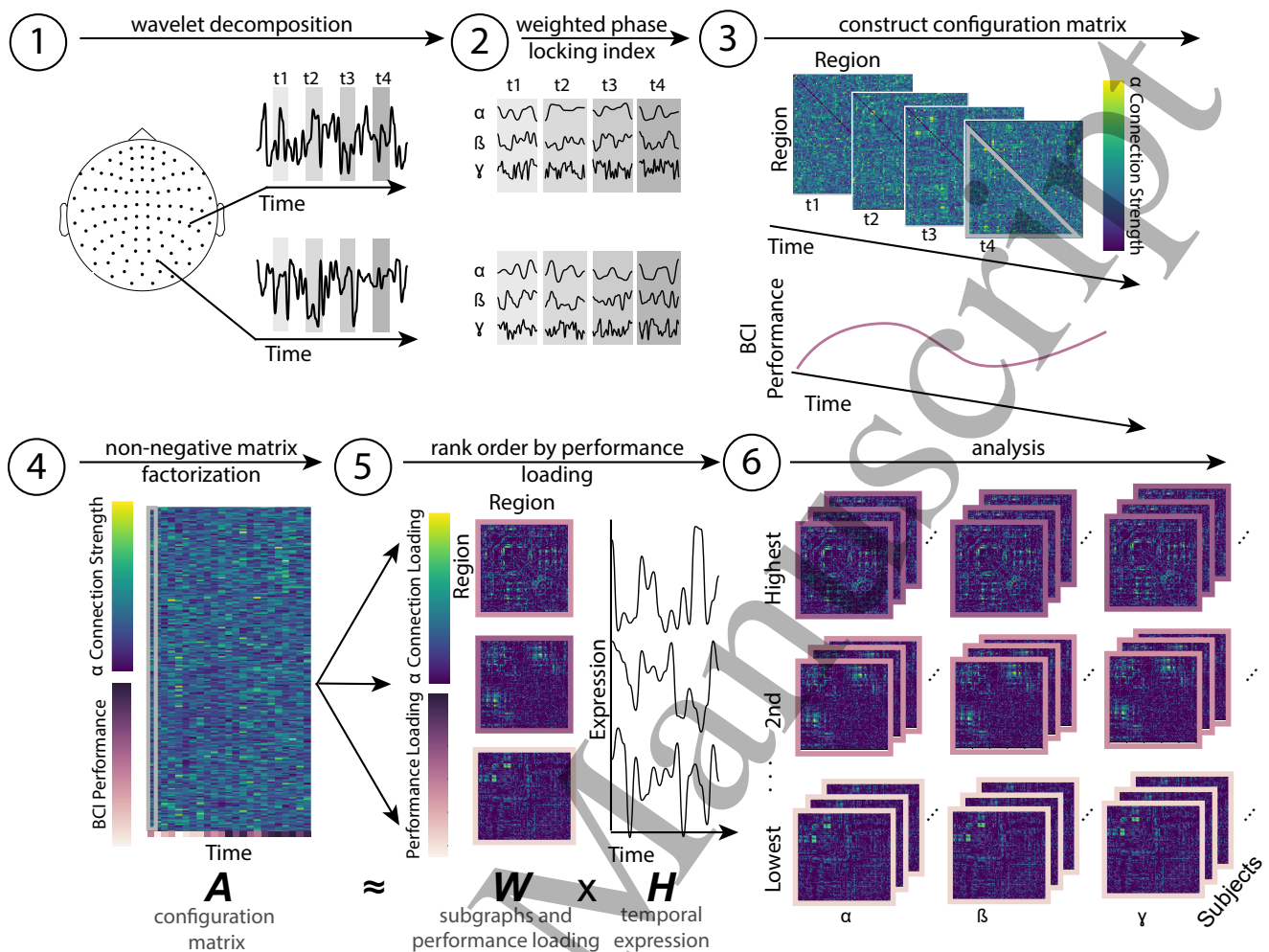
FIG. 1. **Schematic of non-negative matrix factorization.** (1) MEG data recorded from 102 gradiometers is segmented into windows (t_1, t_2, t_3, t_4, ... t_n) that each correspond to the feedback portion of a single BCI trial. (2) A Morlet *wavelet decomposition* is used to separate the signal into α (7-14 Hz), β (15-30 Hz), and γ (31-45 Hz) components. (3) In each window, and for each band, functional connectivity is estimated as the *weighted phase-locking index* between sensor time series. Only one band is shown for simplicity. The subject's performance on each trial is also recorded. (3) The lower diagonal of each trial (highlighted in grey in panel (3)) is reshaped into a vector, and vectors from all trials are concatenated to form a single *configuration matrix*. The subject's time-varying performance forms an additional row in this *configuration matrix*. This matrix corresponds to **A** in the NMF cost function. (5) The NMF algorithm decomposes the configuration matrix (composed of neural and behavioral data) into m subgraphs with a *performance loading* (where m is a free parameter), with three types of information: (i) the weight of each edge in each subgraph, also referred to as the *connection loading* (viridis color scale), (ii) the performance loading (purple color scale) and (iii) the time varying expression of each subgraph (black line graphs). The performance loading indicates how similar the time-varying performance is to each subgraph's expression. The connections and performance loadings together comprise **W** in the NMF cost function, and the temporal expression comprises **H**. (6) Across bands and subjects, we then group subgraphs by their ranked performance loading for further analysis.

and electrocardiograms. EMG activity was manually inspected to ensure that subjects were not moving their forearms during the recording sessions. If subjects did move their arms, those trials were rejected from further analyses.

Preprocessing

As a preliminary step, temporal Signal Space Separation (tSSS) was performed using MaxFilter (Elekta Neuromag) to remove environmental noise from MEG activity. All signals were downsampled to 250 Hz and segmented into trials. ICA was used to remove blink and heartbeat artifacts. An FFT of the data from each sub-

ject was inspected for line noise, although none was found in the frequency bands studied here. We note that the frequency of the line noise (50 Hz) was outside of our frequency bands of interest. In the present study, we restricted our analyses to gradiometer sensors. Gradiometers sample from a smaller area than magnetometers, which is important for ensuring a separability of nodes by network models [17]. Furthermore, gradiometers are typically less susceptible to noise than magnetometers [39]. We combined data from 204 planar gradiometers in the voltage domain using the 'sum' method from Fieldtrip's ft_combine_planar() function, resulting in 102 gradiometers (http://www.fieldtriptoolbox.org/).

Connectivity Analysis

To estimate phase-based connectivity, we calculated the weighted phase-locking index (wPLI) [107]. The wPLI is an estimate of the extent to which one signal consistently leads or lags another, weighted by the imaginary component of the cross-spectrum of the two signals. Using phase leads or lags allows us to take zero phase lag signals induced by volume conduction and to reduce their contribution to the connectivity estimate, thereby ensuring that estimates of coupling are not artificially inflated [107]. By weighting the metric by the imaginary component of the cross spectrum, we enhance robustness to noise [107]. Formally, the wPLI between two time series x and y is given by

$$\phi(x,y) = \frac{|E\{imag(\Gamma_{xy})\}|}{E\{|imag(\Gamma_{xy})|\}} , \qquad (1)$$

where $E\{\}$ denotes the expected value across estimates (here, centered at different samples), Γ_{xy} denotes the cross spectrum between signals x and y, and $imag()$ selects the imaginary component.

We first segment MEG data from gradiometers into 3 second trials, sampled at 250 Hz. The cross spectrum is then estimated using wavelet coherence [65] in each of three frequency bands of interest (α 7-14 Hz, β 15-30 Hz, and γ 31-45 Hz), with wavelets centered on each timepoint. We chose to compute the wavelet coherence – rather than Welch's method – it does not assume stationarity of the signal [65]. We implemented the procedure in the Fieldtrip package in MATLAB, with a packet width of 6 cycles and zero-padding up to the next power of two ('nextpow2'). We then calculate the wPLI as the mean of the imaginary component of the cross spectrum, divided by the imaginary component of the mean of the cross spectrum.

We then construct a network model of these statistical relationships where sensors ($N = 102$) are nodes, and the weight of the edge between node i and node j is given by the weighted phase-locking value. The graph, G, composed of these nodes and edges is a weighted, undirected graph that is encoded in an adjacency matrix $\mathbf{A}$. By constructing this network model, we can use statistics from

graph theory and computational approaches from control theory to quantify the structure of inter-sensor functional relations [6, 9].

Uniformly Phase Randomized Null Model

In order to ensure that our results are not due to choices in preprocessing, the time invariant cross-correlation of neural signals, or the autocorrelation of neural signals, we repeated all of the preprocessing and analysis steps with a uniformly phase randomized null model [53]. To enhance the simplicity and brevity of the exposition, we will also sometimes refer to this construct simply as the *null* model. Surrogate data time series from the null model were calculated using a custom function in MATLAB. Essentially, the FFT of the raw data is taken, the same random phase offset is added to every channel, and then the inverse FFT is taken to return the signal to the time domain [102]. Mathematically, this process is achieved by taking the discrete Fourier transform of a time series y_v:

$$Y(u) = \sum_{v=0}^{V-1} y_v e^{i2\pi uv/V}, \qquad (2)$$

where V is the length of the time series, v indexes time, and u indexes frequencies. We then multiply the Fourier transform by phases chosen uniformly at random before transforming back to the time domain:

$$\overline{y}_v = \frac{1}{\sqrt{V}} \sum_{v=0}^{V-1} e^{ia_u} |Y(u)| e^{-i2\pi kv/V}, \qquad (3)$$

where the phase $a_t \in [0, 2\pi)$.

Construction of a Multimodal Configuration Matrix

In this work, we wished to use a data-driven matrix decomposition technique to identify time-varying subgraphs of functional connectivity that support learning. Specifically, for each subject and each frequency band, we created a multimodal *configuration matrix* of edge weights and BCI performance over time, prior to submitting this matrix to a decomposition algorithm that we describe in more detail below (Fig. 1, step 4). We made separate matrices for each frequency band rather than concatenating them into a single matrix because it is easier for the NMF algorithm to converge if there are more time points relative to the number of edges. To construct the matrix, we first vectorize the upper triangle (not including the diagonal) of each trial's connectivity matrix, and then we concatenate all of the vectors and our one performance measure into an $E \times \tau$ matrix, where τ is the number of trials (384, if no trials were removed),

and E is the number of edges (5151) plus the number of behavioral measures (1). This concatenation process results in a 5152×384 multimodal (brain-behavior) matrix. In this task, each subject's performance is recorded as their percentage of successful trials (out of 32) on each run. This measure includes both motor imagery trials, where the target was located in the upper quadrant of the screen, and rest trials where the target was located in the lower quadrant of the screen. Because this measure was averaged over trials but the connectivity was calculated on individual trials, we interpolate the performance time series to obtain a graded estimate of the percentage of correct trials that is τ time points long. The performance vector is then normalized to have the same mean as the other rows of the configuration matrix.

Non-negative Matrix Factorization

We used a data-driven matrix decomposition method – non-negative matrix factorization (NMF) – to identify time-varying groups of neural interactions and behavior during BCI learning [66]. Intuitively, NMF decomposes a matrix into a set of additive subgraphs with time-varying expression such that a linear combination of these subgraphs weighted by temporal expression will recreate the original matrix with minimal reconstruction error [60, 66]. The NMF algorithm can also be thought of as a basis decomposition of the original matrix, where the subgraphs are a basis set and the temporal coefficients are basis weights. Unlike other graph clustering methods [80], NMF creates a soft partition of the original network, allowing single edges to be a part of multiple subgraphs. Additionally, unlike other basis decomposition methods [4, 23], NMF does not impose harsh constraints of orthogonality, or independence of the subgraphs; it simply finds the most accurate partition, given that the original matrix is non-negative. In many systems (including those whose edges reflect phase-locking), the non-negativity constraint is not difficult to satisfy; moreover, this constraint is particularly relevant to the study of physical systems, where the presence of a negative edge weight can be difficult to interpret.

Formally, the NMF algorithm will approximate an $E \times T$ configuration matrix $\hat{\mathbf{A}}$ by the multiplication of two matrices: $\mathbf{W}$, the subgraph matrix with dimensions $E \times m$, and $\mathbf{H}$, with dimensions $m \times T$. The matrices $\mathbf{A}$, $\mathbf{W}$, and $\mathbf{H}$ are shown in Fig. 1, steps 4 and 5. Here, E is the number of time varying processes (behavior and functional connections derived from MEG data), T is the number of time points, and m is the number of subgraphs. Details of how we solve for $\mathbf{W}$ and $\mathbf{H}$, as well as parameter selection can be found in the Supplemental Materials.
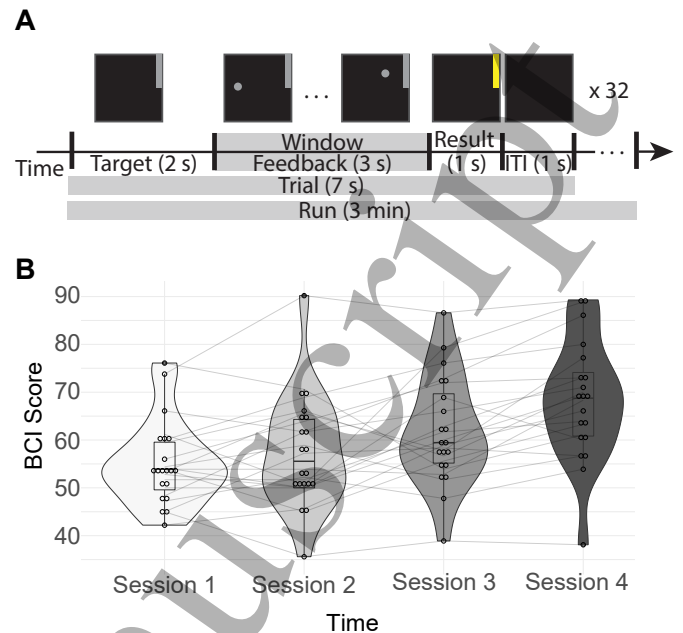

FIG. 2. **BCI task and performance.** *(A)* Schematic of the BCI task. First the target, a grey bar in the upper or lower portion of the screen, was displayed for 1 s. Next, the subjects have a 3 s feedback period, where the vertical position of the cursor is determined by their neural activity while it moves horizontally at a fixed velocity. This portion corresponds to the analysis window, indicated with a grey bar in the figure. The result is then displayed for 1 s. If the subject reached the target, it will turn yellow; otherwise it will remain grey. There is a 1 s intertrial interval (ITI) between trials where nothing is displayed on the screen. This sequence is repeated 32 times per run, with 6 runs per session. *(B)* Each subject's average performance across four days within two weeks. BCI Score is the percentage of correct trials during that session.

Subgraph Inclusion

Most subgraphs are sparse, with distributions of temporal coefficients skewed towards zero (see Fig. S4). However, for every subject and every frequency band, one subgraph showed very little regularization (no edges were equal to 0) and had a uniform, rather than skewed distribution of temporal coefficients. These subgraphs are clear outliers from the others, and appear to be capturing global phase-locking across the entire brain, rather than any unique subsystem. To answer this question about the time varying interactions between neural systems, we were particularly interested in differences between the subgraphs that were spatially localized, having edges regularized to zero. Because including these outlier subgraphs would obscure those differences, we removed these subgraphs from all further analyses.

Group Average Subgraphs

After applying NMF to the multimodal brain-behavior matrix, we next turned to a study of the nature of the detected subgraphs after ranking them by performance loading. Specifically, we were initially interested in determining which edges contributed to each ranked subgraph most consistently across the population. For this purpose, we used a consistency based approach to create a group representative subgraph for each ranked subgraph [92]. In this procedure, each subject's subgraph was first thresholded to retain only the 25% strongest connections (see Fig. S5 for evidence that results are robust to variations in this choice). We then constructed an average $N \times N$ subgraph G, where N is the number of channels and where each element G_{ij} quantifies how many subjects (out of 20) displayed an edge between region i and region j in their thresholded subgraph. In addition to visually depicting these group representative subgraphs, we also wished to summarize their content in spatial bins. It is important to note that without source reconstruction, meaningful inference about which anatomical regions correspond to which sensors is extremely difficult [82]. We therefore binned edges into 10 anatomically defined areas using montages obtained from BrainStorm [101] software (neuroimage.usc.edu/brainstorm/Tutorials/MontageEditor). For parsimony, and acknowledging the limits of anatomical inference from sensor data, we refer to each of these bins as a different lobe (frontal, motor, parietal, occipital, and temporal) in a given hemisphere (Fig. S9).

Optimal Control

Our final broad goal was to provide a theoretical explanation for why certain networks support BCI learning. We hypothesized that these regularized networks might have structures that make it easier for the brain to modulate the patterns of activity that are necessary for BCI control. This hypothesis motivated us to formulate and validate a model to explain how the sparse statistical relationships characteristic of each subgraph could support the production of brain activity patterns implicated in BCI learning [11, 44]. Additionally, this model should account for the brain's ability to reach these patterns of activity in the context of the BCI task, where there is increased volitional modulation of the left motor cortex. Here, we use tools from network control theory to satisfy these conditions [84]. Specifically, we characterize the theoretical brain activity at each sensor as a vector $x(t)$, and we use the adjacency matrix $\mathbf{A}$ of a subgraph to quantify the ease with which that activity can affect other regions. We then incorporate volitional input control as input into the brain $(u(t))$ at a specific region (given by $\mathbf{B}$). Then, by stipulating

$$\dot{x}(t) = \mathbf{A}x(t) + \mathbf{B}u(t), \qquad (4)$$

we model the linear spread of activity along the connections in $\mathbf{A}$ in the context of input to regions given in $\mathbf{B}$. We note that these dynamics are simple, and we do not expect them to fully capture the richness of observed signals; nevertheless, simple models have the notable advantages of interpretability and flexibility.

With this model of network dynamics, optimal control trajectories can be formalized and identified by developing a cost function that seeks to minimize two terms: (i) the distance of the current state from the target state and (ii) the energy required for control. Specifically, we solve the following minimization problem:

$$\min_{\mathbf{u}} \int_0^T (\mathbf{x}_T - \mathbf{x}(t))^T (\mathbf{x}_T - \mathbf{x}(t)) + \rho \mathbf{u}_\kappa(t)^T \mathbf{u}_\kappa dt,$$
$$s.t. \quad \dot{\mathbf{x}} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \quad \mathbf{x}(0) = \mathbf{x}_0, \quad and \quad \mathbf{x}(T) = \mathbf{x}_T, \qquad (5)$$

where ρ is a free parameter that weights the input constraint, $\mathbf{x}_T$ is the target state, and T is the control horizon, which is a free parameter that defines the finite amount of time given to reach the target state. During BCI control, there is specific, targeted control to a specific area of the brain (here, the left motor cortex) in addition to other ongoing control and sensory processes. We wished for our selection of the input matrix $\mathbf{B}$ to reflect this richness and also allow for computationally tractable calculations of optimal control, which is difficult for sparse control sets. Therefore, we constructed the input matrix $\mathbf{B}$ so as to allow input that was dominated by the BCI control site, while maintaining minor contributions from other areas. More specifically, rather than being characterized by binary state values, channels other than the one located over left motor cortex were given the smallest non-zeros value that assured low error calculations, approximately 5×10^{-5} at their corresponding diagonal entry in $\mathbf{B}$. See Supplement for the full derivation from [44].

It is important to note that in general the tools from linear controllability theory are not applicable to the functional networks commonly derived from neuroimaging data for two reasons. The first reason is that the model which the tools are built upon stipulates a time-dependent propagation of activity along edges; such a propagation is physically true for structural connections derived from white matter, but is not generally true for other types of connections used in network models, such as morphometric similarity or most common functional connectivity measures. The second reason is that the model assumes that interactions between nodes 'a' and 'c' are not due to node 'b', an assumption that is violated by measures of statistical similarity such as the Pearson correlation coefficient which is the measure of functional connectivity most commonly employed in neuroimaging

studies. Because we are using neither structural connectivity nor common measures of functional connectivity, it was necessary for us to first prove that the networks we are studying are consistent with our model. To address the first point regarding the propagation of activity along edges, we demonstrate that the structure of the subgraphs used have utility in predicting empirical brain state transitions, and that the relative contribution of each subgraph is related to its temporal expression (Fig. S1C-D). It is only in light of these validations that we are able to interpret our results as a potential model for driving brain activity. To address the second point regarding isolation of pairwise relations not due to third party effects, we note that the matrix $\mathbf{A}$ that we study reflects statistical similarity in phase after strict regularization that removes redundant statistical relationships (Fig. S1A-B).

Target state definition

A central hypothesis in this work is that certain regularized subgraphs are better suited to drive the brain to patterns of activity that are beneficial for BCI control than others. To test this hypothesis, we create target states that reflect these beneficial patterns, based on previous literature. Target states for motor imagery and attention are obtained for each band individually from references [3, 37, 43], and can be briefly described as follows: α contralateral motor suppression for motor imagery and parietal suppression for attention, β contralateral motor suppression and ipsilateral motor activation for motor imagery and vertex suppression for attention, and γ contralateral motor activation for motor imagery and motor cortex suppression with frontal and occipital activation for attention (Fig. S10). While acknowledging the limits of anatomical inference from sensor data, we sought to approximate these true functional systems at the sensor level by dividing channels into lobes using standard montages provided by Brainstorm [101] software (neuroimage.usc.edu/brainstorm/Tutorials/MontageEditor). The target state of channels in brain regions where we did not have specific hypotheses for their activity were set to zero; the target state of channels with activation were set to 1 and that of channels with deactivation were set to -1. Initial states were set to 0 for all channels. We then calculate the optimal energy (using the optimal control equation described above) required to reach each of these target states to test the hypothesis that subgraphs that support learning will have lower energy requirements than those that do not.

Statistical Analyses

Much of our analyses involve testing differences in distributions across subjects for different subgraphs or sessions, both for phase-randomized and empirical data. We also compare these distributions to subject learning rate defined as the slope of performance over time. For the results displayed in Fig. 2 here in the main manuscript, we used a repeated measures ANOVA to test for the presence of a main effect across conditions given that the distributions of performances were normal (see Fig. S11). In Fig. 3 here in the main manuscript, we sought to associate learning rate with ranked performance loading. After plotting quantile-quantile plots (see Fig. S12-S14) for the learning rate, and each of the performance loadings, it became clear that the lowest loadings were not normally distributed. Therefore, we used a linear model combined with non-parametric testing utilizing 5000 permutations (lmPerm package in R https://cran.r-project.org/web/packages/lmPerm). Standardized coefficients were calculated using the lm.beta package in R (https://cran.r-project.org/web/packages/lm.beta/lm.beta.pdf). We use a Bonferroni correction to control false positive errors due to multiple comparisons across all 6 predictors ($\alpha = 0.008$). To obtain an estimate of how sensitive our results are to our specific sample, we also plot summary statistics from 500 models obtained from bootstrapping a sample of equal size ($N = 60$, 3 band and 20 subjects). To examine differences in consistency (Fig. 4 here in the main manuscript), we use a linear model ($consistency \sim band + dataType + rank$) to test for a main effect of data type (null or empirical), band, and subgraph on consistency (see Fig. S15). We next sought to determine if different subgraphs had consistently different temporal expression for null and empirical data (Fig. 5 here in the main manuscript). We also used a repeated measures ANOVA to test for a main effect of subgraph across bands, and paired t-tests to test for differences amongst individual subgraphs (Fig. S16). Lastly, for the results shown in Fig. 6 here in the main manuscript, we test the relationship between learning rate and optimal control energy differences for several different models. Pearson's correlations were used, given that the data appears normally distributed and has few outliers (see Fig. S17-S20).

Data and Code

Code for analyses unique to this manuscript are available at github.com/jastiso/netBCI. Code for the NMF algorithm and the NMF parameter selection is available at github.com/akhambhati/Echobase/tree/master/Echobase/Network/Partitioning/Subgraph. Code for optimal control analyses is available at github.com/jastiso/NetworkControl. Data necessary to reproduce each figure will be made available upon request.

RESULTS

BCI Learning Performance

Broadly, our goal was to examine the properties of dynamic functional connectivity during BCI learning, and to offer a theoretical explanation for why a certain pattern of connectivity would support individual differences in learning performance. We hypothesized that decomposing dynamic functional connectivity into additive $N \times N$ subgraphs would reveal unique networks that are well suited to drive the brain to patterns of activity associated with successful BCI control. We use MEG data from 20 healthy adult individuals who learned to control a motor-imagery based BCI over four separate sessions spanning a two week period. Consistent with prior reports of this experiment [24], we find a significant improvement in performance across the four sessions (one-way ANOVA $F(3, 57) = 13.8$, $p = 6.8^{-7}$) (Fig. 2). At the conclusion of training, subjects reached a mean performance of 68%, which is above chance (approximately 55 - 60%) level for this task [77].

Dynamic patterns of functional connectivity supporting performance

To better understand the neural basis of learning performance, we detected and studied the accompanying patterns of dynamic functional connectivity. First, we calculated single trial phase-based connectivity in MEG data in three frequency bands: α (7-14 Hz), β (15-25 Hz), and γ (30-45 Hz). We then used non-negative matrix factorization (NMF) – a matrix decomposition method – to separate the time-varying functional connectivity into a soft partition of additive subgraphs. We found that the selected parameters led to an average of 7.4 subgraphs, with a range of 6 to 9, and that all frequency bands had a decomposition error lower than 0.47 (mean α error = 0.352, mean β error = 0.379, mean γ error = 0.465) (Fig. S2). The error is the Frobenius norm of the squared difference between our observed and estimated connectivity matrices (with dimensions 5152×384) and takes values between 0 and 1. For each band, the error value is low, giving us confidence that we have fairly accurately reconstructed relevant neural dynamics. To determine whether any properties of the identified subgraphs were trivially due to preprocessing choices, NMF parameters, or time-invariant autocorrelation in neural activity, we repeated the full decomposition process after permuting the phases of all time series uniformly at random. We found that the statistics of subgraph number and decomposition error were similar for the uniformly phase randomized data, indicating that any differences in subgraph and temporal expression between null and empirical data is not due to the NMF algorithm's inability to find a good decomposition, but rather due to the structure of the chosen decomposition (Fig. S2).

We quantified the similarity between each subgraph's temporal expression and the time course of performance, and we refer to this quantity as the subgraph's performance loading (Fig. 1). Here, performance is calculated as the percentage of accurate trials over a run of 32 trials. We hypothesized that the ranked performance loading would be associated with task learning, as operationalized by the slope of performance over time. It is important to note the distinction between performance and learning: performance is defined as task accuracy and therefore varies over time, while learning is defined as the linear rate of change in that performance over the course of the experiment (384 trials over 4 days). We tested whether learning was correlated with the performance loading of subgraphs. Because the minimum number of subgraphs in a given subject was 6, we decided to investigate the top four highest performance loading subgraphs, and the smallest and second smallest nonzero loading subgraphs. We found a general trend that the performance loading from high loading subgraphs was negatively associated with learning rate, and the performance loading from low loading subgraphs was positively associated with learning rate (Fig. 3AB). We assessed the statistical significance of these trends and found that only the third highest loading subgraph displayed a performance loading that was significantly correlated with learning rate after Bonferroni correction for multiple comparisons (linear model with permutation tests $slope \sim loading3 + band : p = 0.005$). Performance loading from uniformly phase randomized surrogate data for this subgraph was not associated with learning rate ($p = 0.292$). The direction of the observed effect in the empirical data is notable; subjects with *lower* loading onto *high* loading subgraphs learned the task better, suggesting that learning is facilitated by a dynamic interplay between several subnetworks. It is also notable that the highest loading subgraphs do not have the strongest associations with learning, indicating that the subgraphs that most closely track performance are not the same as the subgraphs that track changes in performance.

Spatial properties of dynamic patterns of functional connectivity

Next we sought to better understand why the third highest loading subgraph was most robustly associated with learning. We hypothesized that because of this subgraph's strong association across subjects, it might recruit sensors near consistent brain regions and reflect the involvement of specific cognitive systems across subjects. To evaluate this hypothesis, we began by investigating the shared spatial properties of this subgraph in comparison to the others. To identify shared spatial features we grouped subgraphs together by their ranked performance loading, and then quantified how consistent edges were across participants [92] (see Methods). We found that the average consistency varied by frequency
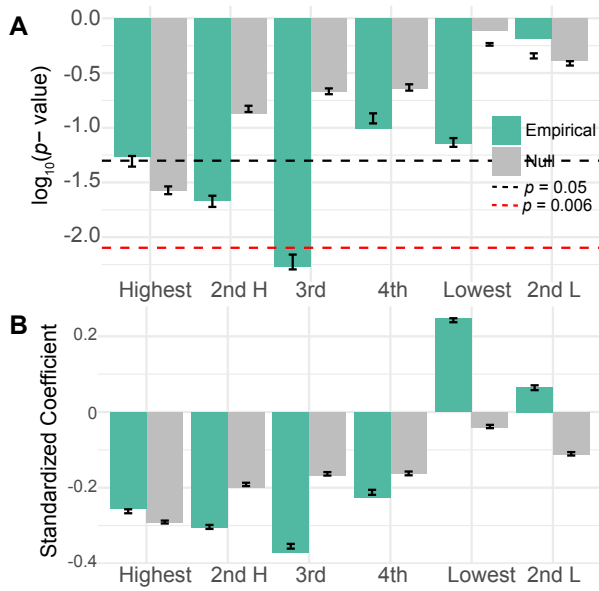
FIG. 3. **Performance loading is associated with** learning.
(A) Here we show the *p*-values for empirical (green) and uniformly phase randomized (grey) data for linear models relating the slope of performance with ranked performance loading from each frequency band. The black line corresponds to $p = 0.05$, while the red dashed line corresponds to the Bonferroni corrected $\alpha = 0.008$. Error bars show the standard error and median of *p*-values from 500 models with bootstrapped samples. *(B)* The standardized regression coefficients for the same models. Error bars show the standard error and mean of coefficients from 500 models with bootstrapped samples.

band, and differed between the empirical and surrogate data, but not across ranked subgraphs (linear model $consistency \sim band + rank + data : F_{band}(2, 17) = 90.36$, $p_{band} = 9.00 \times 10^{-10}$, $F_{data}(1, 17) = 41.8$, $p_{data} = 5.78 \times 10^{-6}$). The α band had the most consistent edges, followed by the γ band, and then the β band ($t_{\alpha\beta} = -12.68$, $p_{\alpha\beta} = 4.3 \times 10^{-10}$, $t_{\alpha\gamma} = -10.41$, $p_{\alpha\gamma} = 1.2 \times 10^{-8}$). In the uniformly phase randomized surrogate data, we observed less consistent subgraphs than those observed in the empirical data ($t = -6.47$, $p = 5.78 \times 10^{-6}$). These observations support the conclusion that across the population, despite heterogeneous performance, similar regions interact to support performance and learning to varying degrees.

In order to approximate system-level activation with sensor level data, we used lobe montages provided by Brainstorm (see Methods). Spatially, subgraphs were dominated by connectivity in the frontal lobe sensors, with subtle differences in the pattern of connections from the frontal lobe sensors to sensors located in other areas of the brain (Fig. 4). To determine which functional edges were most consistent in each subgraph and frequency band, we calculated the average consistency over each lobe and motor cortex in both hemispheres (for the

same analysis in surrogate data, see Fig. S6). In the α band, the most consistent edges on average were located in the left frontal lobe in the highest performance loading subgraph, in the left occipital lobe in the second highest performance loading subgraph, between right frontal and right motor in the third highest performance loading subgraph, and between left frontal lobe and right parietal lobe in the lowest performance loading subgraph. In the β band, the most consistent edges were located between right and left frontal lobe for the highest and second highest performance loading subgraph, between left frontal lobe and right motor for the third highest performance loading subgraph, and between left and right frontal lobe for the lowest performance loading subgraph. In the γ band, the most consistent edges were located in the left frontal and right frontal lobes for the highest performance loading subgraph, in the left frontal lobe and right motor for the second highest performance loading subgraph, and in left frontal and right frontal lobe for the third highest and lowest performance loading subgraphs. We wished to demonstrate that the consistent involvement of more frontal sensors across subgraphs was not due to the presence of electro-oculogram (EOG) artifacts that persisted after removal of eye blinks with ICA. We therefore calculated the weighted phase-locking index between both vertical and horizontal EOG sensors and all neural sensors. Qualitatively, we did not observe any consistently strong connectivity between EOG channels and more frontal sensors, indicating that the frontal connectivity identified in our analysis is likely not due to residual artifacts from eye movements (Fig. S7). We also note that the most consistent individual edges for each subgraph are still only present in 10-12 individuals, indicating a high amount of individual variability. Collectively, these observations suggest widespread individual variability in the spatial composition of ranked subgraphs, with the most consistent connectivity being located in the frontal lobe during BCI learning.

Temporal properties of dynamic patterns of functional connectivity

Importantly, subgraphs can be characterized not only by their spatial properties, but also by their temporal expression. We therefore next examined the temporal properties of each subgraph to better understand why the third highest performance loading subgraph was most robustly associated with learning. As a summary marker of temporal expression, we calculated the total energy of the time series operationalized as the sum of squared values, as well as the time of the peak value of the time series. Across frequency bands, we found no significant dependence between energy and subgraph ranking. We did find a significant effect of rank for the peak time of temporal expression obtained from the empirical data (repeated measures ANOVA $peak \sim rank + band : F_{rank}(3, 215) = 6.67$, $p_{rank} = 2.53 \times 10^{-4}$ but not from the uniformly
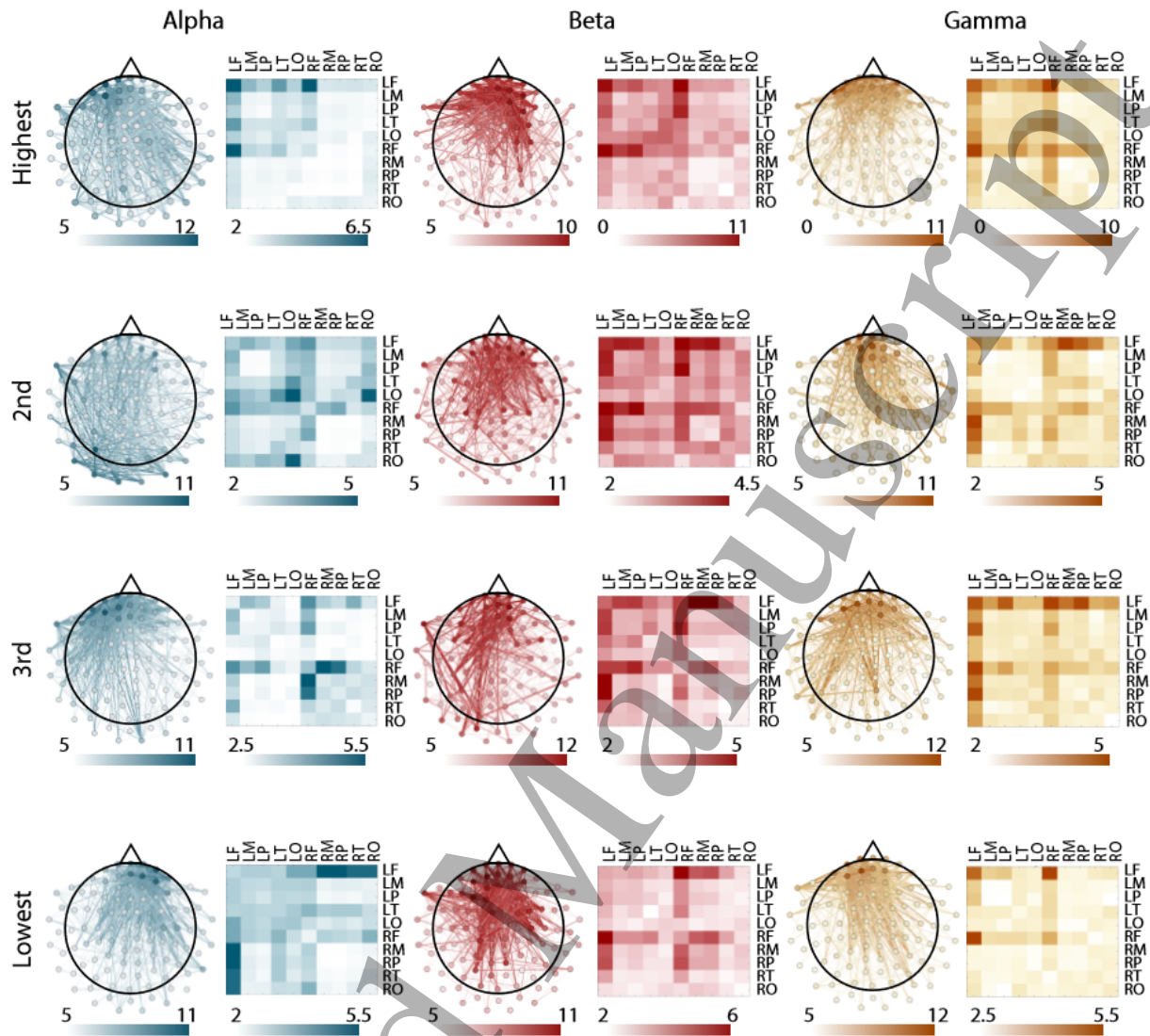
FIG. 4. **Spatial distribution of subgraph edges that are consistent across participants.** Consistent edges for each frequency band and for each ranked subgraph. Left images show individual edges plotted on a topographical map of the brain. Right images show the mean edge weight over sensors for a given region. We studied 10 regions, including the frontal lobe, temporal lobe, parietal lobe, occipital lobe, and motor cortex in both hemispheres. The weight of the edge corresponds to the number of individual participants for whom the edge was among the 25% strongest for that subgraph.

phase randomized surrogate data ($F_{rank}(3, 215) = 1.28$, $p = 0.282$). Overall, peak times are widely distributed across individuals. However we find that across bands, the highest performance loading subgraph has a later peak, which is intuitive since performance is generally increasing over time and these subgraphs most strongly track performance.

We then performed *post-hoc* paired t-tests corrected for multiple comparisons (Bonferroni correction $\alpha = 0.006$) between the highest performance loading subgraph and all other ranked subgraphs in each band. In the α band, the highest performance loading subgraph only peaked significantly later than the lowest (paired t-test $N =$

20, $t_{low} = 8.06$, $p_{low} = 1.49 \times 10^{-7}$) after Bonferroni correction ($\alpha = 0.006$). In the β band, the highest performance loading subgraph peaked significantly later than all others (paired t-test $N = 20$, $t_{2H} = 10.9$, $p_{2H} = 1.39 \times 10^{-9}$; $t_{3H} = 7.56$, $p_{3H} = 3.57 \times 10^{-7}$; $t_{low} = 8.07$, $p_{low} = 1.49^{-7}$). In the γ band, the highest performance loading subgraph peaked significantly later than the second highest, and lowest loading subgraphs (paired t-test $N = 20$, $t_{2H} = 4.50$, $p_{2H} = 2.46 \times 10^{-4}$; $t_{low} = 8.06$, $p_{low} = 1.49 \times 10^{-7}$). (Fig. 5). Finally, we asked whether the time of the peak in the third highest performance loading subgraph was associated with learning. We did not find a relationship between peak time
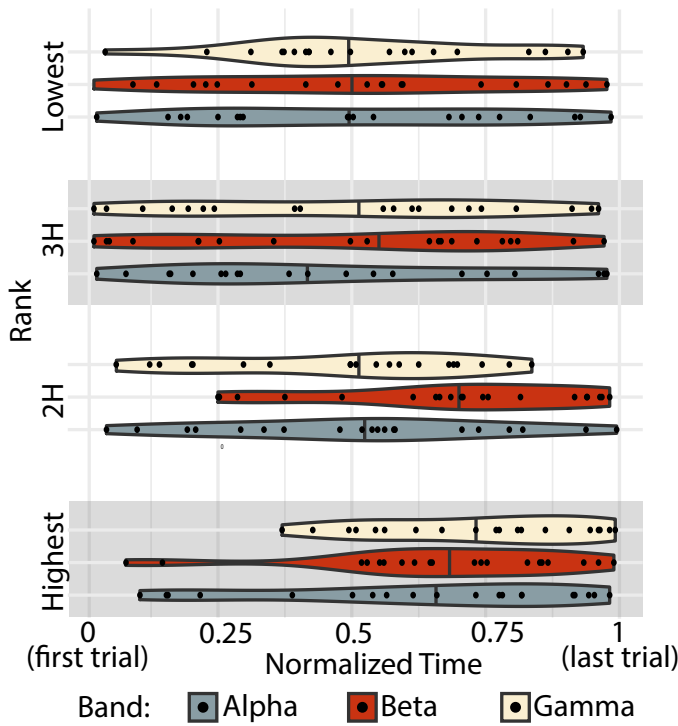
FIG. 5. **Temporal expression of ranked subgraphs.** The peak temporal expression for every subject (black data point), for each frequency band (indicated by color) and for each subgraph (ordered vertically). Violin plots show the density distribution of all subjects' peaks. The median is marked with a solid line through the violin plot.

and learning in any frequency band (Pearson's correlation: $\alpha : r = 0.005$, $p = 0.98$, $\beta : r = 0.047$, $p = 0.84$, $\gamma : r = -0.21$, $p = 0.037$). To summarize these findings, we note that across participants and especially in the β band, subgraphs that support performance are highly expressed late in learning, when performance tends to be highest. However, subgraphs that support learning do not have consistent peaks across subjects, and each individual's peak does not relate to their learning rate, indicating that some other feature of these subgraphs must explain their role in learning.

Explaining dynamic patterns of functional connectivity supporting BCI learning via network control theory

Lastly we asked how the third highest loading subgraph could facilitate successful BCI performance, as shown in Fig. 3. Here, we considered an edge – extracted under penalties of spatial and temporal sparsity – as a potential path for a brain region to affect a change in the activity of another brain region [35, 109]. Assuming the true connectivity structure is sparse, the regularization

applied in the NMF algorithm can remove large statistical relationships between regions that are not directly connected, but might receive common input from a third region [28] (see Methods for addition discussion, and see Fig. S1A-B for the effect of regularization on the prevalence of triangles). We hypothesized that the pattern of edges in this subgraph would facilitate brain states, or patterns of activity, that were predictive of BCI literacy. Specifically, we expected that when the brain mirrored the connectivity of the third subgraph, the brain could more easily reach states of sustained motor imagery or sustained attention than when the brain mirrored the connectivity of the lowest performance loading subgraph. To operationalize these hypotheses from sensor level data, we identified sensors near motor and attention areas with montages from Brainstorm and set those as targets (see Methods). We also hypothesized that the magnitude of this difference would be associated with each subject's learning rate. To test these hypotheses, we used mathematical models from network control theory to quantitatively estimate the ease with which the brain can reach a desired pattern of activity given a pattern of connectivity (see Methods and Fig. S1C-D for analyses demonstrating the efficacy of the regularized subgraphs in linearly predicting changes in activity). Specifically we calculated the optimal control energy required to reach a target state (either sustained motor imagery or sustained attention) from an initial state when input is applied primarily to the left motor cortex, which was the site of BCI control (Fig. 6A-B).

We tested whether the third highest performance loading subgraph supported the transition to states of sustained motor imagery or sustained attention with smaller energy requirements than other subgraphs that did not support learning in the same way. We chose the lowest performance loading subgraph for comparison because it was the only subgraph with a large positive standardized regression coefficient for fitting learning, which contrasts sharply with the large negative coefficient for the third subgraph. For both states (motor imagery and attention), we found no population level differences in energy requirements by the two subgraphs (paired t-test $N = 20$, motor imagery: $t_\alpha = -0.005$, $p_\alpha = 0.565$, $t_\beta = 1.38$, $p_\beta = 0.184$, $t_\gamma = -1.00$, $p_\gamma = 0.329$. attention: $t_\alpha = -1.35$, $p_\alpha = 0.193$, $t_\beta = -0.344$, $p_\beta = 0.735$, $t_\gamma = -0.937$, $p_\gamma = 0.360$). We next tested whether the magnitude of the difference in energy required by the two subgraphs to reach a given state tracked with learning rate. In the β band, we observed a significant correlation between the magnitude of the energy difference to reach attentional states and learning rate over subjects (Pearson's correlation coefficient $r = 0.560$, $p = 0.0103$, Bonferroni corrected for multiple comparisons across frequency bands; Fig. 6). Notably, the relationship remained significant when controlling for subgraph density (linear model $slope \sim energy\_difference + density\_difference$: $t_{energy} = 2.68$, $p_{energy} = 0.0158$, $t_{density} = -0.266$, $p_{density} = 0.794$). When using sub-

graphs derived from the uniformly phase randomized surrogate data, the relationship was not observed (Pearson's correlation $r = -0.0568$, $p = 0.819$). We next asked which subgraph contributed most to this effect. We found no significant relationship between learning rate and the energy required to reach the attentional state by the third highest performance loading subgraph (Pearson's correlation $r = -0.389$, $p = 0.702$) or by the lowest performance loading subgraph (Pearson's correlation $r = 0.227$, $p = 0.335$). This finding suggests that learning rate depends on the relative differences between subgraphs, rather than the energy conserving architecture of one alone. As a final test of specificity, we assessed whether this difference was selective to the third highest and lowest performance loading subgraph. We found no significant relationship when testing the difference of the highest with the third highest performance loading subgraph (Pearson's correlation $r = -0.554$, $p = 0.586$), the highest with the lowest performance loading subgraph (Pearson's correlation $r = 0.40$, $p = 0.077$), the second highest with the third highest performance loading subgraph (Pearson's correlation $r = 0.266$, $p = 0.257$), or the second highest with the lowest performance loading subgraph (Pearson's correlation $r = -0.072$, $p = 0.764$). This pattern of null results underscores the specificity of our finding.

Reliability and specificity of inferences from network control theory

Collectively, our findings are consistent with the hypothesis that during BCI learning, one subnetwork of neural activity arises, separates from other ongoing processes, and facilitates sustained attention. An alternative hypothesis is that our results are due to trivial factors related to the magnitude of the attentional state, or could have just as easily been found if we had placed input to a randomly chosen region of the brain, rather than to the left motor cortex which was the actual site of the BCI control. To determine whether these less interesting factors could explain our results, we performed the same network control calculation but with a spatially non-overlapping target state, and then – in a separate simulation – with a mirrored input region (right motor cortex rather than left motor cortex). We performed the spatial shifting by ordering the nodes anatomically (to preserve spatial contiguity), and then circular shifting the attention target state by a random number between 1 and $N - 1$. For 500 circularly shifted states, only 3 (0.6%) had a correlation value equal to or stronger than the one observed (Fig. S8). Furthermore, we found no significant relationship between learning rate and the difference in energy required by the two subgraphs to reach the true attention state when input was applied to the right motor cortex instead of the left motor cortex (Pearson's correlation $t = 0.711$, $p = 0.313$). Together, these two findings suggest that the relationship identified is specific to BCI control.

Finally, we assessed the robustness of our results to choices in modeling parameters. First we performed the computational modeling with two different sets of control parameter values (see Supplement). In both cases, the significant relationship remained between learning rate and the difference in energy required by the two subgraphs to reach the attentional state (set one Pearson's correlation coefficient $r = 0.476$, $p = 0.0338$; set two Pearson's correlation coefficient $r = 0.514$, $p = 0.0204$). Second, since our target states were defined from prior literature, there was some flexibility in stipulating features of those states. To ensure that our results were not unduly influenced by these choices, we tested whether ideologically similar states would provide similar results. Namely, we assessed (i) the impact of varying the magnitude of (de)activation by changing (-)1 to (-)2, (ii) the impact of the neutral state by changing 0 to 1, and (iii) the impact of negative states by changing -1, 0 and 1 to 1, 2, and 3. We found a consistent relationship between learning rate and the difference in energy required by the two subgraphs to reach the attentional state when we changed the magnitude of activation/deactivation (Pearson's correlation coefficient $r = 0.560$, $p = 0.0103$), as well as when we changed the neutral state (Pearson's correlation coefficient $r = 0.520$, $p = 0.0188$). However, we found no significant relationship when removing negative states (Pearson's correlation coefficient $r = 0.350$, $p = 0.130$), indicating that this result is dependent on our choice to operationalize deactivation as a negative state value. After performing these robustness checks, we conclude that a selective separation of the third highest and lowest performance loading subgraphs impacts their ability to drive the brain to patterns of sustained attention in the β band in the context of BCI control. This result is robust to most of our parameter choices, is selective for biologically observed states, and is not observed in surrogate data.

DISCUSSION

In this work, we use a minimally constrained decomposition of dynamic functional connectivity during BCI learning to investigate which groups of phase locked brain regions (subgraphs) support BCI control. The performance loading onto these subgraphs favors the theory that dynamic involvement of several subgraphs during learning supports successful control, rather than extremely strong expression of a single subgraph. Additionally, we find a unique association for the third highest loading subgraph with learning at the population level. This result shows that learning is not simply explained by the subset of edges that has the most similar temporal expression to behavior, but rather that a subnetwork with a middling range of similarity has the strongest relationship with performance improvement. While the spatiotemporal distribution of this subgraph was vari-
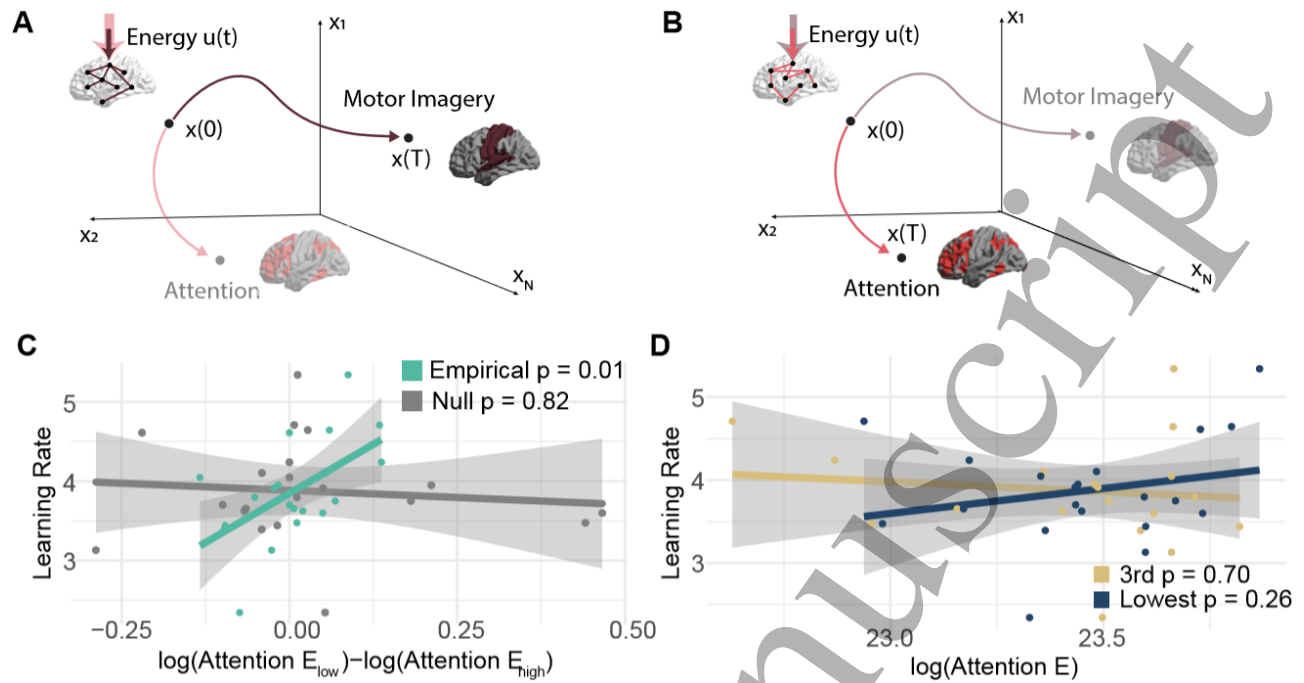
FIG. 6. **Separation of the ability to modulate attention is associated with learning.** Different patterns of connections will facilitate transitions to different patterns of brain activity. We hypothesize that the ease with which connections in certain regularized subgraphs facilitate transitions to patterns of activity that support either motor imagery *(A)* or attention *(B)* will be associated with learning rate. We use network control theory to test this hypothesis. We model how much energy ($u(t)$) is required to navigate through state space from some initial pattern of activity $x(0)$ to a final pattern of activity $x(T)$. Some networks (e.g., the brown network in panel *A*) will require very little energy (schematized here with a smaller, solid colored arrow) to reach patterns that support motor imagery, while other networks (e.g., the pink network in panel *B*) will have small energy requirement to reach patterns of activity that support attention. *(C)* The relationship between learning rate and the difference in energy required to reach the attention state when the underlying network takes the form of the lowest versus third highest performance loading subgraphs for empirical data (green) and uniformly phase randomized surrogate data (grey). *(D)* The relationship between the learning rate and the energy required to reach the attention state when the underlying network takes the form of the lowest performance loading subgraph, or when the underlying network takes the form of the third highest performance loading subgraph.

able across individuals, we did observe some consistencies at the group level. Spatially, the third highest loading subgraph showed strong edges between left frontal and right motor cortices for low frequencies, and left frontal and left motor cortices for the γ band. Lower frequencies showed stronger connectivity to the ipsilateral (to imagined movement) motor cortex, suggesting a possible role in suppression for selective control. This subgraph also showed the highest expression earlier than the other ranked subgraphs we investigated, perhaps linking it to the transition from volitional to automatic control.

We next wished to posit a theory of how these subgraphs fit with previously identified neural processes important for learning, despite their heterogeneity across subjects. After quantifying the extent to which NMF regularization removed potentially redundant relationships between regions (Fig. S1A-B), we suggested that the regularized pattern of statistical relationships identified in this subgraph could comprise an avenue through which

brain activity could be modulated via cognitive control or external input. We then hypothesized that these networks would be better suited to modulate activity in either regions implicated in attention or in motor imagery than other subgraphs, and further that individuals whose networks better modulated activity in these regions would display greater task learning [56]. We chose to operationalize the "ease of modulation" with a metric from network control theory called *optimal control energy*. Optimal control energy quantifies the minimum input needed to drive the brain from an initial pattern of activity to a final pattern of activity, while also assuring that the pattern of activity stays close to the target state at every point in time. This last constraint ensures that we are unlikely to pass through biologically unfeasible patterns of activity. The notion of optimal control energy that we use here assumes a particular linear model of how neural dynamics change given potential avenues of communication between regions. Importantly, in the

supplement (Fig. S1C-D) we show that our subgraphs predict empirical brain state changes according to this model, and that the contribution of each subgraph to empirical changes in brain state is related to its temporal expression. Using this model, we did not find any population differences in optimal control energy when the simulation was enacted on the third highest performance loading subgraph compared to the lowest performance loading subgraph. However, we did find that the magnitude of this difference was associated with learning in individual subjects. This result was specific to the β band and to brain regions implicated in attention. Critically, the relation to learning could not be explained by the energy of either subgraph alone, was not present in surrogate data derived from a uniformly phase randomize null model, and was robust to parameter choices. Overall, the observations support our hypothesis that in the β band the subgraphs we identified that support learning are well suited to modulate activity in brain regions associated with attention.

A delicate balance of interactions is required for BCI learning

Our initial analysis explored the relationship between performance loading and learning. It is important to note the behavioral difference between performance and learning: we use the term performance to refer to task accuracy over time, whereas we use the term learning to refer to how well a subject is able to increase that accuracy. With that distinction in mind, we aimed to better understand how subgraphs that vary similarly to performance (those with high performance loading) relate to learning. We found that the subgraph with the third highest performance loading was most strongly associated with learning and that a narrow distribution of performance loading across all subgraphs was associated with better learning. Together, these two observations are in line with previous research in motor and spatial learning, which shows that some brain structures display differential activity during learning that is independent of performance [86, 95]. Our work adds to this literature by demonstrating that in addition to targeted differences in individual brain regions or networks, a minimally constrained decomposition of dynamic functional connectivity across the whole brain reveals that separable processes are most associated with performance and with learning.

Additionally, we find that BCI learning is not explained simply by the processes most strongly associated with performance and learning individually, but by a distributed loading across many different subgraphs. This notion is supported by the sign of beta value for ranked subgraphs. Generally, subgraphs with higher ranked loading were negative betas, while subgraphs with lower ranked loading were positive betas. A wealth of whole brain connectivity analyses have similarly shown that the interaction between systems is an important component of skill learning specifically, and other domains

of learning more generally [2, 8]. While we observed marked interactions between many regions, the majority were located in the frontal lobe for all frequency bands. Even for α and β frequencies in the highest loading subgraph, we see involvment of frontal regions and heterogeneity across individuals. This suggests that the NMF method did not extract a network that was trivially related to the deterministic mapping between brain activity and cursor location determined by the BCI2000 software. Previous work has also demonstrated changes in frontal-motor [59] and fronto-parietal [69] connectivity during motor skill learning. In BCI learning specifically, the strength of white matter connectivity between frontal and occipital regions predicts control of motor imagery based BCIs [97]. Additionally, analyses of this same experiment have shown task related changes in functional connectivity were spatially diffuse, and found in frontal, temporal, and occipital regions in the α band [24], and were strongest in frontal, motor, central, and parietal regions in the β band. Our results add to these findings by demonstrating that the most consistent regions that covary in their functional connectivity are interactions between the frontal lobe and other regions. Our work shows that broad motifs like the dynamic integration of multiple systems (including cognitive systems involving the frontal lobe) found in other types of learning are also important for BCI learning. Additionally, we add to previous work on BCI learning specifically by quantifying the structure of covarying subgraphs of connectivity.

BCI learning is heterogenous across individuals

We find population level consistencies in spatial and temporal properties of ranked subgraphs despite having no constraint to assure consistency across individuals. However, we also note that there is a high degree of variability in both of these measures. The variability is mirrored in the subjects' performance, with final performances varying from 38.1 % to 89.3 %. Our observations are in line with previous literature demonstrating variability in subjects' performance and learning for psychological, cognitive, and neurological predictors [49, 56]. Such pervasive and marked individual differences present a challenge for the use of BCIs clinically [14]. To address this challenge, researchers have explored ways to optimize BCI features and algorithms for neurofeedback itself [64, 106] and to identify selection criteria for BCI based therapies [49, 57]. The results of our study support the idea that different individuals will have slightly different neural correlates of both performance and learning based on a variety of features such as demographics [94], spatial manipulation skills [108], relationship with the technology [13], and attention span [42, 43]. Our findings also highlight the importance of studying models fit to each individual when searching for selection criteria for BCI therapies. Here, despite temporal and edge level heterogeneity, our minimally constrained, individual

specific method of brain connectivity decomposition revealed a robust association with learning with a theoretical role that aligns well with previous literature. Further development and expansion of this model to incorporate resting state neuroimaging data and other physiological predictors could be a promising direction for selection of candidates for BCI therapies before training.

Role of beta oscillations in BCI learning

Prominent theories describing the neural processes that give rise to cognition and shape our behavior often involve integration of complex multimodal information using a combination of top-down predictions (built from prior experience) and bottom-up, sensory-driven representations of the dynamic world around us [63, 100?]. These generalized frameworks, in turn, require the precise coordination of ensemble neural activity both within and between brain regions. Several theoretical approaches have examined how these two scales of functional activity may harmonize to produce the desired behavior [91], and empirical research has shown that there is consistent cross-talk between these scales [90]. Within human neuroimaging work, synchronous oscillations have been critical to understanding this complex coordination, where cortico-cortical propagation delays and membrane potentials give rise to observed oscillatory activity in the brain [10, 96]. Here, we study the time varying connectivity within α, β, and γ bands. Much like how specialized functions arise from different brain regions, different narrowband oscillations have been implicated in diverse but specialized processes, where some generalizable theories suggest a role for α in disengagement of task irrelevant areas or a lack of sensory processing [83], β in sustaining the current cognitive state [34] and γ in task active local cortical computation [38]. Specifically in the context of motor imagery based BCIs, α and β bands have prominent signatures in motor imagery [74]. Our results show that only the β band's functional connectivity is well suited to modulate patterns of activity that support sustained attention (not motor imagery), which is a critical process for BCI control. While our results are in line with generalized theories on the role of oscillations in cognition, the specificity of the β band in our results extends classic studies that discuss the role of this oscillation in attention [85] and in maintaining the current cognitive state [34]. Our results suggest that this maintenance, a consistent control (or attention to) internally generated activity, may play a crucial role in longterm BCI use.

Methodological Considerations

NMF Non-negative matrix factorization is a machine learning technique for separating, in our case, a multimodal configuration matrix into a soft-partition of subgraphs with time-varying expression. This process has several advantages, such as being able to link behavioral and neural data, and creating a quantification of mesoscale structure where brain regions can participate in multiple functional groups. Nevertheless, the method also faces several limitations that are common to other large-scale machine learning techniques. NMF yields a low rank approximation of a large configuration matrix, and can sometimes be rank deficient for large number of subgraphs, for very large datasets, or for datasets with high covariance. Because of this sensitivity, we were not able to test our data against independently phase randomized null models.

MEG Functional Connectivity We chose to complete our analyses in sensor, rather than source space. Ultimately, this choice was motivated by the fact that if any of our findings could be applicable to clinicians monitoring learning during real-time BCI learning they would need to be obtained in the sensor space. However, this choice has two major methodological consequences: (1) it limits the anatomical resolution of our data, and therefore the specificity of the claims that we can make about the spatial distribution of the regions involved and (2) it does not protect as well against false positive connectivity estimates[82, 114]. We were not interested in the finer anatomical resolution of the identified subgraphs, but more in the process of identifying them, in validating the hypothesis that features of these subgraphs are associated with learning, and in their theoretical functions. We used montages provided by Brainstorm to approximate lobes and systems at the sensor level; however, we acknowledge that even claims made about specific systems (motor, and attention) at the source level are best interpreted in light of controls. Our use of spatial permutation tests is thus particularly important, because they demonstrate that similar contiguous states do not show the same relationship between energy and learning. Additionally, we cannot fully eliminate the possibility that parts of our data are due to false positive interaction from signal spread, and our conclusions should be interpreted in light of this fact. That being said, we have taken several steps to reduce the influence of false positives in our connectivity estimates. First, we use a connectivity estimate that does not include zero-phase lag contributions that could arise from signal spread [107]. However removing zero-phase lag contributions on its own is not enough to prevent against false positive from source spread from true connections[82]. While source reconstruction partially addresses this problem, it does not eliminate it entirely[82], and it additionally requires many parameter choices and has potentially confounding effects on estimates of functional connectivity [12, 22, 54]. Secondly, all results of interest are compared to a phase-randomized null model with the same static covariance structure as the original data, which should lessen the effect of spurious connectivity estimates.

Optimal Control We chose to use tools from network control theory to quantify the ease with which each network can modulate brain activity. Network control the-

ory relies on several assumptions that should be considered when interpreting these results [105]. First, the model of dynamics that we employ is linear and noise free, unlike the brain [46], but has proven useful in gaining intuitions about the behavior of nonlinear systems [55, 76]. However, we still sought to quantify the ability of this linear model to explain empirical changes in brain state. Specifically, we asked two questions: (1) do the regularized subgraphs used in our analyses have the ability to predict state transitions, and do they do so better than randomly rewired networks, and (2) is the contribution of each subgraph to explaining a given state transition proportional to its temporal expression, and is it more proportional than a different subgraph's temporal expression? To evaluate these questions, we generated brain states for every trial (band specific power at each channel) and simulated Eq. 5 (see Supplement). Regarding the similarity of predicted and empirical state transitions, we find modest correlation values (mean Pearson's $r = 0.25$) that are significantly greater than the correlations observed from randomized networks. Similarly for our second question, we found small but positive correlations between the contribution of each subgraph to a given transition and its temporal expression (mean Pearson's $r = 0.03$), which was also significantly greater than correlations to temporal expression from mismatched subgraphs. While it is unsurprising that our linear model did not fully capture neural dynamics across a three second trial, it is worth considering extensions that can maximize this similarity for future analyses investigating how connections between regions facilitate changes to activity. One option is to use effective connectivity [70, 78] – that solve for a network of connections that best predicts the evolution of brain states in time. However, effective connectivity matrices are often sparse, and therefore not well suited to the NMF matrix decomposition used in the present work. Alternatively, one could use non-linear models of dynamics [58] and non-linear control theory [112] to capture a wider range of dynamic behaviors, although non-linear control does not currently support the same scope of tools available for linear control theory. Lastly, future work could use functional approximation [15] in order to identify a set of simple basis functions that well approximate the data. If a sparse approximation can be found, it supports the idea that the underlying non-linear dynamics can be captured with linear combinations of these basis functions, and therefore are suitable to be modeled with simplified linear models.

Additionally, network control is typically applied to time invariant, structural connections that have a clear role as an avenue along which brain activity can propagate. Here we used functional connectivity (weighted phase locking) which is a statistical relationship that (1) does not imply the presence of a physical connection and (2) is not time invariant. Due to (1), our original functional connectivity matrix can have large values between two regions that are not directly connected, but might

both connect to the same region. This situation would lead to a triangle composed of three connections in a functional connectivity matrix where in reality there are only two connections. However, the regularization applied by the NMF algorithm mitigates this concern in a manner that is similar to the regularization applied in effective connectivity metrics [28, 70]. We also explicitly quantify the effect of regularization on triangles in our subgraphs and find a dramatic reduction from the original functional connectivity (Fig. S1A-B). This quantification, along with the two validations discussed above, show that our model is a suitable way to evaluate the role of regularized subgraphs in modulating different patterns of activity. In relation to (2), we note that functional connectivity in not time-invariant, unlike the state matrix more commonly employed in linear control models. However, it is important to note that NMF identifies subgraphs that are separable from their temporal expression, and that we expect that the hypothesized role in control would only be prominent when the subgraph was highly expressed.

Conclusion and Future Directions

Future research that builds on this work could explore ways to increase sensitivity to an individual's learning rate. Given that EEG and MEG sensors capture some unique information [72] and provide increased discriminability in clinical applications including BCIs [21, 25], it would be interesting to investigate whether the concurrently collected EEG data in this study better captures relevant neural dynamics for performance and learning, respectively. Such an effort, combined with source reconstruction, would be a useful next step in basic scientific inquiries directed towards characterizing these separable networks involved in learning. However, combining EEG and MEG sources would greatly increase the number of variables relative to the number of observations in the connection matrix to be decomposed, and would make the NMF algorithm less likely to converge. It may thus be necessary to use connectivity estimates from smaller time windows. Clinical utility could potentially be achieved if similar methods could be applied to resting state data to identify network properties that separate individuals by their learning rate, thereby eliminating the need for any BCI training. Finally, confirmatory studies with a larger sample of individuals would both validate the current results, and provide a better assessment of potential clinical utility.

In conclusion, we use a minimally constrained method of matrix decomposition that is specific to each human participant to investigate the dynamic neural networks that support BCI learning. We find that the subgraphs that most tightly mirror performance are not the same subgraphs that most strongly support learning. Additionally, we find that the interaction between many different neural processes is important for BCI learning. While

the subgraphs identified are heterogeneous (as is subject performance), we find consistent involvement of frontal and motor cortices in subgraphs that support learning. We also observe differential temporal expression amongst subgraphs, and perhaps most notably that the subgraphs that vary more similarly with performance reach their highest expression later in learning. Lastly, we test the hypothesis that subgraphs that support learning are better suited to modulate activity in brain regions important for attention than other subgraphs. We find evidence to support this hypothesis in the β band specifically, ultimately suggesting that the separation of processes for maintaining attention is important for successful BCI learning. Our results align with prior work from dynamic functional connectivity in other types of skill learning, and also highlight a method for identifying individual predictors of successful BCI control with theoretical support.

Citation Diversity Statement

Recent work in neuroscience and other fields has identified a bias in citation practices such that papers from women and other minorities are under-cited relative to the number of such papers in the field [18, 20, 31, 33, 73, 104]. Here we sought to proactively consider choosing references that reflect the diversity of the field in thought, form of contribution, gender, and other factors. We used automatic classification of gender based on the first names of the first and last authors [33, 113], with possible combinations including male/male, male/female, female/male, and female/female. Excluding self-citations to the senior authors of our current paper, the references contain 55.2% male/male, 11.5% male/female, 21.9% female/male, 11.5% female/female, and 1.0% unknown categorization. We look forward to future work that could help us to better understand how to support equitable practices in science.

ACKNOWLEDGEMENTS

[1] M. Ahn and S. C. Jun. Performance variation in motor imagery brain-computer interface: A brief review. *Journal of Neuroscience Methods*, 243:103–110, 2015.

[2] N. Altman and M. Krzywinski. Points of Significance: Clustering. *Nature Methods*, 14(6):545–546, 2017.

[3] A. Bamdadian, C. Guan, K. K. Ang, and J. Xu. The predictive role of pre-cue EEG rhythms on MI-based BCI classification performance. *Journal of Neuroscience Methods*, 235:138–144, 2014.

[4] D. J. Bartholomew. Principal components analysis. *International Encyclopedia of Education*, 2:374–377, 2010.

[5] D. S. Bassett and M. G. Mattar. A Network Neuroscience of Human Learning: Potential to Inform Quantitative Theories of Brain and Behavior. *Trends in Cognitive Sciences*, 21(4):250–264, 2017.

[6] D. S. Bassett and O. Sporns. Network neuroscience. *Nature Neuroscience*, 20(3):353–364, 2017.

[7] D. S. Bassett, N. F. Wymbs, M. A. Porter, P. J. Mucha, J. M. Carlson, and S. T. Grafton. Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences*, 108(18):7641,

2010.

[8] D. S. Bassett, M. Yang, N. F. Wymbs, and S. T. Grafton. Learning-induced autonomy of sensorimotor systems. *Nature Neuroscience*, 18(5):744–751, 2015.

[9] D. S. Bassett, P. Zurn, and J. I. Gold. On the nature and use of models in network neuroscience. *Nature Reviews Neuroscience*, pages 1–13, jul 2018.

[10] A. M. Bastos, W. M. Usrey, R. A. Adams, G. R. Mangun, P. Fries, and K. J. Friston. Canonical Microcircuits for Predictive Coding. *Neuron*, 76(4):695–711, nov 2012.

[11] R. F. Betzel, S. Gu, J. D. Medaglia, F. Pasqualetti, and D. S. Bassett. Optimally controlling the human connectome: The role of network topology. *Scientific Reports*, 6:1–14, 2016.

[12] M. J. Brookes, J. R. Hale, J. M. Zumer, C. M. Stevenson, S. T. Francis, G. R. Barnes, J. P. Owen, P. G. Morris, and S. S. Nagarajan. Measuring functional connectivity using MEG: Methodology and comparison with fcMRI. *NeuroImage*, 56:1082–1104, 2011.

[13] M. J. Brosnan. The impact of computer anxiety and self-efficacy upon performance. *Journal of Computer*

Assisted Learning, 14(3):223–234, 1998.

[14] C. Brunner, V. Kaiser, B. Z. Allison, G. Pfurtscheller, G. R. Müller-Putz, and C. Neuper. Toward a hybrid braincomputer interface based on imagined movement and visual attention. *Journal of Neural Engineering*, 7(2):026007, 2010.

[15] S. L. Brunton, J. L. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

[16] E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Neuroscience*, 2009.

[17] C. T. Butts. Revisting the foundations of network analysis. *Science*, 325(5939):414, 2009.

[18] N. Caplar, S. Tacchella, and S. Birrer. Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nature Astronomy*, 1(6):0141, 2017.

[19] L. R. Chai, A. N. Khambhati, R. Ciric, T. M. Moore, R. C. Gur, R. E. Gur, T. D. Satterthwaite, and D. S. Bassett. Evolution of brain network dynamics in neurodevelopment.

[20] P. Chakravartty, R. Kuo, V. Grubbs, and C. McIlwain. # communicationsowhite. *Journal of Communication*, 68(2):254–266, 2018.

[21] R. A. Chowdhury, Y. Zerouali, T. Hedrich, M. Heers, E. Kobayashi, J. M. Lina, and C. Grova. MEGEEG Information Fusion and Electromagnetic Source Imaging: From Theory to Clinical Application in Epilepsy. *Brain Topography*, 28(6):785–812, 2015.

[22] G. L. Colclough, M. W. Woolrich, P. K. Tewarie, M. J. Brookes, A. J. Quinn, and S. M. Smith. How reliable are MEG resting-state connectivity metrics? *NeuroImage*, 138:284–293, 2016.

[23] P. Comon, P. Comon, I. Component, and P. Comon. Independent Component Analysis , a new concept? *Signal Processing*, 1684(94):287–314, 2015.

[24] M.-C. Corsi, M. Chavez, D. Schwartz, N. George, L. Hugueville, A. E. Khan, S. Dupont, D. S. Bassett, and F. De Vico Fallani. Spatiotemporal neural correlates of brain-computer interface learning. (2), 2018.

[25] M. C. Corsi, M. Chavez, D. Schwartz, L. Hugueville, A. N. Khambhati, D. S. Bassett, and F. De Vico Fallani. Integrating EEG and MEG Signals to Improve Motor Imagery Classification in Brain-Computer Interface. *International Journal of Neural Systems*, 29(1):1–12, 2019.

[26] B. N. Cuffin and D. Cohen. COMPARISON OF THE MAGNETOENCEPHALOGRAM AND ELECTROENCEPHALOGRAM. Technical report, 1979.

[27] E. A. Curran and M. J. Stokes. Learning to control brain activity: A review of the production and control of EEG components for driving brain-computer interface (BCI) systems. *Brain and Cognition*, 51(3):326–336, 2003.

[28] A. Das, A. L. Sampson, C. Lainscsek, L. Muller, W. Lin, J. C. Doyle, S. S. Cash, E. Halgren, and T. J. Sejnowski. Interpretation of the precision matrix and its application in estimating sparse brain connectivity during sleep spindles from human electrocorticography recordings. Technical Report 3, 2017.

[29] F. De Vico Fallani and D. S. Bassett. Network neuroscience for optimizing brain-computer interfaces.

Physics of Life Reviews, 1:6–11, 2018.

[30] F. De Vico Fallani, J. Richiardi, M. Chavez, and S. Achard. Graph analysis of functional brain networks: Practical issues in translational neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1653), 2014.

[31] M. L. Dion, J. L. Sumner, and S. M. Mitchell. Gendered citation patterns across political science and social science methodology fields. *Political Analysis*, 26(3):312–327, 2018.

[32] O. Dipasquale, L. Griffanti, M. Clerici, R. Nemni, G. Baselli, and F. Baglio. High-dimensional ica analysis detects within-network functional connectivity damage of default-mode and sensory-motor networks in alzheimer's disease. *Frontiers in Human Neuroscience*, 9(FEB):1–7, 2015.

[33] J. D. Dworkin, K. A. Linn, E. G. Teich, P. Zurn, R. T. Shinohara, and D. S. Bassett. The extent and drivers of gender imbalance in neuroscience reference lists. *bioRxiv*, 2020.

[34] A. K. Engel and P. Fries. Beta-band oscillationsásignalling the status quo? *Current Opinion in Neurobiology*, 20:156–165, 2010.

[35] F. Ferreri, F. Vecchio, D. Ponzo, P. Pasqualetti, and P. M. Rossini. Time-varying coupling of EEG oscillations predicts excitability fluctuations in the primary motor cortex as reflected by motor evoked potentials amplitude: An EEG-TMS study. *Human Brain Mapping*, 35(5):1969–1980, 2014.

[36] E. E. Fetz. Operant conditioning of cortical unit activity. *Science*, 163(3870):955–958, 1969.

[37] J. Frey, C. Mühl, F. Lotte, and M. Hachet. Review of the Use of Electroencephalography as an Evaluation Method for Human-Computer Interaction. *Proceedings of the International Conference on Physiological Computing Systems*, page 2019, 2013.

[38] P. Fries. Neuronal Gamma-Band Synchronization as a Fundamental Process in Cortical Computation. *Annual Review of Neuroscience*, 32(1):209–224, 2009.

[39] P. Garcés, D. López-Sanz, F. Maestú, and E. Pereda. Choice of Magnetometers and Gradiometers after Signal Space Separation.

[40] R. T. Gerraty, J. Y. Davidow, K. Foerde, A. Galvan, D. S. Bassett, and D. Shohamy. Dynamic flexibility in striatal-cortical circuits supports reinforcement learning. *J Neurosci*, 2084:17, 2018.

[41] B. Graimann, B. Allison, and G. Pfurtscheller. Brain-Computer Interfaces. pages 1–28, 2010.

[42] M. Grosse-Wentrup and B. Schölkopf. High gamma-power predicts performance in sensorimotor-rhythm brain-computer interfaces. *Journal of Neural Engineering*, 9(4), 2012.

[43] M. Grosse-Wentrup, B. Schölkopf, and J. Hill. Causal influence of gamma oscillations on the sensorimotor rhythm. 2011.

[44] S. Gu, R. F. Betzel, M. G. Mattar, M. Cieslak, P. R. Delio, S. T. Grafton, F. Pasqualetti, and D. S. Bassett. Supplementary Information for Optimal Trajectories of Brain State Transitions List of Figures. pages 1–28, 2016.

[45] S. Gu, R. F. Betzel, M. G. Mattar, M. Cieslak, P. R. Delio, S. T. Grafton, F. Pasqualetti, and D. S. Bassett. Optimal trajectories of brain state transitions. *NeuroImage*, 148(January):305–317, 2017.

[46] S. Gu, F. Pasqualetti, M. Cieslak, Q. K. Telesford, A. B. Yu, A. E. Kahn, J. D. Medaglia, J. M. Vettel, M. B. Miller, S. T. Grafton, and D. S. Bassett. Controllability of structural brain networks. *Nature Communications*, 6:1–10, 2015.

[47] A. Guillot, C. Collet, V. A. Nguyen, F. Malouin, C. Richards, and J. Doyon. Functional neuroanatomical networks associated with expertise in motor imagery. *NeuroImage*, 41(4):1471–1483, 2008.

[48] S. Halder, D. Agorastos, R. Veit, E. M. Hammer, S. Lee, B. Varkuti, M. Bogdan, W. Rosenstiel, N. Birbaumer, and A. Kübler. Neural mechanisms of brain-computer interface control. *NeuroImage*, 55(4):1779–1790, 2011.

[49] S. Halder, E. M. Hammer, A. Kübler, B. Blankertz, C. Sannelli, T. Dickhaus, G. Curio, and K.-R. Müller. Neurophysiological predictor of SMR-based BCI performance. *NeuroImage*, 51(4):1303–1309, 2010.

[50] M. Hamedi, S. H. Salleh, and A. M. Noor. Electroencephalographic motor imagery brain connectivity analysis for BCI: A review. *Neural Computation*, 28(6):999–1041, 2016.

[51] E. M. Hammer, S. Halder, B. Blankertz, C. Sannelli, T. Dickhaus, S. Kleih, K. R. Müller, and A. Kübler. Psychological predictors of SMR-BCI performance. *Biological Psychology*, 89(1):80–86, 2012.

[52] E. M. Hammer, T. Kaufmann, S. C. Kleih, B. Blankertz, and A. Kübler. Visuo-motor coordination ability predicts performance with brain-computer interfaces controlled by modulation of sensorimotor rhythms (SMR). *Frontiers in Human Neuroscience*, 8(AUG):1–9, 2014.

[53] S. Heitmann and M. Breakspear. Putting the dynamic back into dynamic functional connectivity. *Network Neuroscience*, pages 1–61, 2018.

[54] A. Hillebrand, G. R. Barnes, J. L. Bosboom, H. W. Berendse, and C. J. Stam. Frequency-dependent functional connectivity within resting-state networks: An atlas-based MEG beamformer solution. *NeuroImage*, 59(4):3909–3921, 2012.

[55] C. J. Honey, R. Kötter, M. Breakspear, and O. Sporns. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Sciences*, 104(24):10240–10245, 2007.

[56] C. Jeunet, B. N'kaoua, and F. Lotte. Advances in user-training for mental-imagery-based BCI control: Psychological and cognitive factors and their neural correlates 1. *Progress in Brain Research*, 228, 2016.

[57] C. Jeunet, B. Nkaoua, S. Subramanian, M. Hachet, and F. Lotte. Predicting Mental Imagery-Based BCI Performance from Personality, Cognitive Profile and Neurophysiological Patterns. *PLoS ONE*, 10(12):1–21, 2015.

[58] V. Jirsa and H. Haken. Field Theory of Electromagnetic Brain Activity. Technical Report 5, 1996.

[59] H. T. Karim, T. J. Huppert, K. I. Erickson, M. E. Wollam, P. J. Sparto, E. Sejdi, and J. M. VanSwearingen. Motor sequence learning-induced neural efficiency in functional brain connectivity. *Behavioural Brain Research*, 319:87 – 95, 2017.

[60] A. N. Khambhati, M. G. Mattar, N. F. Wymbs, S. T. Grafton, and D. S. Bassett. Beyond modularity: Fine-scale mechanisms and rules for brain network reconfiguration. *NeuroImage*, 166:385–399, 2018.

[61] A. N. Khambhati, A. E. Sizemore, R. F. Betzel, and D. S. Bassett. Modeling and interpreting mesoscale network dynamics, 2018.

[62] S. C. Kleih and A. Kbler. Psychological factors influencing brain-computer interface (bci) performance. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 3192–3196, Oct 2015.

[63] P. Kok, J. F. M. Jehee, and F. P. de Lange. Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, 75(2):265–270, 2012.

[64] A. Kubler, M. Tangermann, K.-R. Muller, J. Hohne, F. Cincotti, R. Rupp, J. d. R. Millan, D. Mattia, R. Leeb, and G. Muller-Putz. Towards Noninvasive Hybrid BrainComputer Interfaces: Framework, Practice, Clinical Application, and Beyond. *Proceedings of the IEEE*, 103(6):926–943, 2015.

[65] J.-P. Lachaux, A. Lutz, D. Rudrauf, D. Cosmelli, M. Le Van Quyen, J. Martinerie, and F. Varela. Estimating the time-course of coherence between single-trial brain signals: an introduction to wavelet coherence. Technical report, 2002.

[66] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[67] N. Leonardi, J. Richiardi, M. Gschwind, S. Simioni, J. M. Annoni, M. Schluep, P. Vuilleumier, and D. Van De Ville. Principal components of functional connectivity: A new approach to study dynamic brain connectivity during rest. *NeuroImage*, 83:937–950, 2013.

[68] Q. Li, X. Wang, S. Wang, Y. Xie, X. Li, Y. Xie, and S. Li. Dynamic reconfiguration of the functional brain network after musical training in young adults. *Brain Structure and Function*, 224(5):1781–1795, 2019.

[69] C. h. J. Lin, M. C. Chiang, A. D. Wu, M. Iacoboni, P. Udompholkul, O. Yazdanshenas, and B. J. Knowlton. Enhanced Motor Learning in Older Adults Is Accompanied by Increased Bilateral Frontal and Fronto-Parietal Connectivity. *Brain Connectivity*, 2(2):56–68, 2012.

[70] Y. Liu and S. Aviyente. Quantification of effective connectivity in the brain using a measure of directed information. *Computational and Mathematical Methods in Medicine*, 2012, 2012.

[71] N. Lofthouse, L. E. Arnold, S. Hersch, E. Hurt, and R. DeBeus. A Review of Neurofeedback Treatment for Pediatric ADHD. *Journal of Attention Disorders*, 2012.

[72] F. H. Lopes Da Silva, H. J. Wieringa, and M. J. Peters. Source Localization of EEG Versus MEG: Empirical Comparison Using Visually Evoked Responses and Theoretical Considerations. Technical Report 2, 1991.

[73] D. Maliniak, R. Powers, and B. F. Walter. The gender citation gap in international relations. *International Organization*, 67(4):889–922, 2013.

[74] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw. Mu and beta rhythm topographies during motor imagery and actual movements. Technical Report 3, 2000.

[75] K. A. Moxon and G. Foffani. Brain-machine interfaces beyond neuroprosthetics. *Neuron*, 86(1):55–67, 2015.

[76] S. F. Muldoon, F. Pasqualetti, S. Gu, M. Cieslak, S. T. Grafton, J. M. Vettel, and D. S. Bassett. Stimulation-Based Control of Dynamic Brain Networks. *PLoS Computational Biology*, 12(9), 2016.

[77] G. R. Müller-Putz, R. Scherer, C. Brunner, R. Leeb, and G. Pfurtscheller. Better than random? A closer look on BCI results. *Ijbem.Org*, 10(1):52–55, 2008.

[78] A. Neumaier and T. Schneider. Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Trans. Math. Softw.*, 27:27–57, 2001.

[79] M. E. J. Newman. *Networks: An Introduction.* 2010.

[80] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, (10), 2003.

[81] Y. O. Okazaki, J. M. Horschig, L. Luther, R. Oostenveld, I. Murakami, and O. Jensen. Real-time MEG neurofeedback training of posterior alpha activity modulates subsequent visual detection performance. *NeuroImage*, 107:323–332, 2015.

[82] J. M. Palva, S. H. Wang, S. Palva, A. Zhigalov, S. Monto, M. J. Brookes, J. M. Schoffelen, and K. Jerbi. Ghost interactions in MEG/EEG source space: A note of caution on inter-areal coupling measures. *NeuroImage*, 173(March 2017):632–643, 2018.

[83] S. Palva and J. M. Palva. New vistas for α-frequency band oscillations. *Trends in Neurosciences*, 30(4):150–158, 2007.

[84] F. Pasqualetti, S. Zampieri, and F. Bullo. Controllability metrics, limitations and algorithms for complex networks. *IEEE Transactions on Control of Network Systems*, 1(1):40–52, 2014.

[85] G. Pfurtscheller A'b', C. Neuper, D. Flotzinger, and M. Pregenzer. EEG-based discrimination between imagination of right and left hand movement. Technical report, 1997.

[86] A. Purushotham, R. D. Seidler, K. Uğurbil, J. Ashe, S.-G. Kim, and D. Willingham. Cerebellum Activation Associated with Performance Change but Not Motor Learning. Technical Report 5575, 2002.

[87] A. I. Ramos-Nuñez, S. Fischer-Baum, R. C. Martin, Q. Yue, F. Ye, and M. W. Deem. Static and dynamic measures of human brain connectivity predict complementary aspects of human cognitive performance. *Frontiers in Human Neuroscience*, 11(August):1–13, 2017.

[88] M. Reiner, R. Rozengurt, and A. Barnea. Better than sleep: Theta neurofeedback training accelerates memory consolidation. *Biological Psychology*, 95:45–53, 2014.

[89] T. Renton, A. Tibbles, and J. Topolovec-Vranic. Neurofeedback as a Form of Cognitive Rehabilitation Therapy Following Stroke: A Systematic Review. *Archives of Physical Medicine and Rehabilitation*, 96(12):e27, 2015.

[90] C. G. Richter, W. H. Thompson, C. A. Bosman, and P. Fries. Top-down beta enhances bottom-up gamma. *Journal of Neuroscience*, 37(28):6698–6711, 2017.

[91] J. Riddle, K. Hwang, D. Cellier, S. Dhanani, and M. D'esposito. Causal evidence for the role of neuronal oscillations in topdown and bottomup attention. *Journal of Cognitive Neuroscience*, 31(5):768–779, 2019.

[92] J. A. Roberts, A. Perry, G. Roberts, P. B. Mitchell, and M. Breakspear. Consistency-based thresholding of the human connectome. *NeuroImage*, 145:118–129, 2017.

[93] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw. BCI2000: A general-purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering*, 51(6):1034–1043, jun 2004.

[94] J. Schumacher, C. Jeunet, and F. Lotte. Towards explanatory feedback for user training in brain-computer interfaces. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 3169–3174, Oct 2015.

[95] A. L. Shelton and J. D. E. Gabrieli. Neural Correlates of Individual Differences in Spatial Learning Strategies. *American Psychological Association*, 18(3):442–449, 2004.

[96] W. Singer, A. K. Engel, and P. Fries. Dynamic predictions: oscillations and synchrony in top-down processing. Technical Report 10, 2001.

[97] R. Sitaram, S. Halder, N. Birbaumer, M. Bogdan, W. Rosenstiel, B. Varkuti, and A. Kübler. Prediction of brain-computer interface aptitude from individual brain structure. *Frontiers in Human Neuroscience*, 7(April):1–9, 2013.

[98] R. Sitaram, T. Ros, L. Stoeckel, S. Haller, F. Scharnowski, J. Lewis-Peacock, N. Weiskopf, M. L. Blefari, M. Rana, E. Oblak, N. Birbaumer, and J. Sulzer. Closed-loop brain training: The science of neurofeedback. *Nature Reviews Neuroscience*, 18(2):86–100, 2017.

[99] N. J. Steiner, E. C. Frenette, K. M. Rene, R. T. Brennan, and E. C. Perrin. Neurofeedback and cognitive attention training for children with attention-deficit hyperactivity disorder in schools. *Journal of Developmental and Behavioral Pediatrics*, 35(1):18–27, 2014.

[100] T. J. Sussman, J. Jin, and A. Mohanty. Top-down and bottom-up factors in threat-related perception and attention in anxiety. *Biological Psychology*, 121:160–172, 2016.

[101] F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, and R. M. Leahy. Brainstorm: A user-friendly application for meg/eeg analysis. *Intell. Neuroscience*, 2011:8:1–8:13, Jan. 2011.

[102] J. Theiler. Testing for nonlinearity in timeseries: the method of surrogate data. Technical report.

[103] R. T. Thibault, M. Lifshitz, and A. Raz. The self-regulating brain and neurofeedback: Experimental science and clinical promise. *Cortex*, 74:247–261, 2016.

[104] Y. Thiem, K. F. Sealey, A. E. Ferrer, A. M. Trott, and R. Kennison. Just Ideas? The Status and Future of Publication Ethics in Philosophy: A White Paper. Technical report, 2018.

[105] C. Tu, R. P. Rocha, M. Corbetta, S. Zampieri, M. Zorzi, and S. Suweis. Warnings and Caveats in Brain Controllability.

[106] C. Vidaurre, C. Sannelli, K. R. Müller, and B. Blankertz. Co-adaptive calibration to improve BCI efficiency. *Journal of Neural Engineering*, 8(2), 2011.

[107] M. Vinck, R. Oostenveld, M. Van Wingerden, F. Battaglia, and C. M. Pennartz. An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *NeuroImage*, 55(4):1548–1565, 2011.

[108] A. Vuckovic and B. A. Osuagwu. Using a motor imagery questionnaire to estimate the performance of a Brain-Computer Interface based on object oriented motor imagery. *Clinical Neurophysiology*, 124(8):1586–1595, 2013.

[109] A. Weigand, A. Horn, R. Caballero, D. Cooke, A. P. Stern, S. F. Taylor, D. Press, A. Pascual-Leone, and M. D. Fox. Prospective validation that subgenual connectivity predicts antidepressant efficacy of transcranial magnetic stimulation sites. *Biological Psychiatry*, 84(1):28 – 37, 2018. Mechanisms of Depression and Antidepressant Treatment.

[110] J. Wolpaw, D. McFarland, T. Vaughan, and G. Schalk. The wadsworth center brain-computer interface (bci) research and development program. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):204–207, jun 2003.

[111] B. M. Young, Z. Nigogosyan, A. Remsik, L. M. Walton, J. Song, V. A. Nair, S. W. Grogan, M. E. Tyler, D. F. Edwards, K. Caldera, J. A. Sattin, J. C. Williams, and V. Prabhakaran. Changes in functional connectivity correlate with behavioral gains in stroke patients after therapy using a brain-computer interface device. *Frontiers in Neuroengineering*, 7(JUL):1–12, 2014.

[112] J. G. T. Zañudo, G. Yang, and R. Albert. Structure-based control of complex networks with nonlinear dynamics. 2016.

[113] D. Zhou, E. J. Cornblath, J. Stiso, E. G. Teich, J. D. Dworkin, A. S. Blevins, and D. S. Bassett. Gender diversity statement and code notebook v1.0, Feb. 2020.

[114] J. M. Zumer, H. T. Attias, K. Sekihara, and S. S. Nagarajan. Probabilistic algorithms for MEG/EEG source reconstruction using temporal basis functions learned from data. *NeuroImage*, 41(3):924–940, 2008.