Brain network dynamics fingerprints are resilient to data heterogeneity

Tommaso Menara
Department of Decoded Neurofeedback, ATR Computational Neuroscience Laboratories,
2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288, Japan
Department of Mechanical Engineering, University of California Riverside,
900 University Ave, Riverside, CA, 92521, USA

Giuseppe Lisi
Nagoya Institute of Technology,
Gokiso-cho, Showa-ku, Nagoya, Aichi, 466-8555, Japan
ATR Brain Information Communication Research Laboratory Group,
2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288, Japan

Aurelio Cortese
Department of Decoded Neurofeedback, ATR Computational Neuroscience Laboratories,
2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288, Japan
cortese.aurelio@gmail.com

January 26, 2020

Abstract

The advent of extremely large data repositories of brain activity recordings comes with no free lunch. While multi-site datasets have significantly boosted the quest to unravel the brain's mechanics, inevitable differences in physical parameters and recording protocols may lead to erroneous conclusions. In this study, we investigate

from a perspective – the viewpoint of dynamical models – which factors tend to have non-negligible effects on the recordings of hemodynamic signatures of brain activity. To this end, we make use of data-driven models to capture the dynamical wandering of the brain between large-scale networks of activity. We first confirm the emergence of robust, subject-specific dynamical patterns of brain activity in resting-state fMRI data. Next, we exploit these fingerprints to appraise the effect of an array of scanning factors in a multi-site dataset. We find that scanning sessions belonging to different sites and days tend to induce high variability in such fingerprints, while other factors affect the same metrics to a minor extent. These results concurrently indicate that each subject has its own unique trajectory of brain activity changes, but also that our ability to infer such patterns is affected by how, where and when we try to do so.

Introduction

Untagling the brain's dynamics at rest is of paramount importance in the quest to reveal the mechanisms underlying the spontaneous wandering of the mind between well-established large-scale networks of neural activity [31, 12, 2]. The characterization of the brain dynamics' spatio-temporal organization into networks of activity has greatly benefitted from the the creation of very large neuroimaging datasets [19, 47], such as the Human Connectome Project (HCP) [32, 38], the UK Biobank [36], and, in the context of neurodegenerative diseases, the Alzheimer's Disease Neuroimaging Initiative [21]. Yet, there is still a dire need to assess the effect that differences in physical parameters or scanning protocols exert on the quality of functional magnetic resonance imaging (fMRI) recordings [30, 5, 3]. In fact, it has been shown that imaging sequences are affected by site-dependent differences, such as scanner drift over time, maintenance routine, and other factors [3]. In particular, multi-site harmonization remains an outstanding problem [1, 30, 14, 3, 5, 43], and only few recent works have studied the variability in resting-state fMRI data across sites [13, 22, 29, 18]. Despite growing interest in the complications inherent to multi-site data, this line of research is still in its infancy (the first publication appeared in 2013 [44]). Interestingly, despite the fact that the brain is a dynamical system, most works rely on functional connectivity, and to the best of our knowledge no attempt exists at exploring such issues from the viewpoint of dynamical models.

Data-driven dynamical models are a promising and powerful tool for the analysis of the spatio-temporal organization of brain activity [27, 4, 40]. These models allow us to harness the vast amount of spurious information contained in large datasets [39, 23], capture the hierarchical organization of brain activity [42], enhance brain-computer interfaces [25, 24], and may even be employed in clinical settings [28, 34, 9]. However, the effect of different factors in multi-site data acquisition to the inference and identification of

dynamical models has not yet been investigated. Additionally, from the reverse point of view, dynamical models could provide notable insight into the extent of the effect of these factors on the data.

In this work, we first utilize a data-driven model to validate the findings on resting-state fMRI fingerprints reported in a previous work by Vidaurre and colleagues [42]. More in detail, we use hidden Markov modeling to infer the *hidden* brain states that describe the brain network dynamics measured by fMRI, where networks are probability distributions representing graphs. We then generalize the findings in [42] by applying the Hidden Markov Model (HMM) to a new, different dataset. Finally, we exploit our HMM to assess if, and to what extent, mixed scanning factors affect subject-specific fingerprints and, thus, resting-state fMRI recordings. To achieve our objective, we employ two datasets: the publicly available HCP dataset [32], and the Traveling-subject dataset, which consists of a novel, state-of-the-art collection of fMRI measurements of nine subjects who traveled to twelve different scanning sites [43]. We use the HCP data to build an HMM that we subsequently apply to the Traveling-subject dataset to infer subject-specific brain states and investigate how such inference process is affected by an array of scanning factors. The Traveling-subject dataset contains scanning sessions at different sites, in different days, and with varying phase encodings, number of channels per coil, scanner manufacturers, and scanner models (see Materials and Methods and SI Table 1). We illustrate our methodological approach in Fig. 1.

To anticipate, we first validate subject-specific metrics that we derive from an HMM inferred from the HCP resting-state data; next, we leverage the HMM to seek for factors that tend to have non-negligible effects in fMRI scans belonging to the Traveling-subject dataset. We depart from previous work, which relies mostly on correlation measures and smaller datasets, by exploiting brain network dynamics states. Altogether, this paper presents complementary, yet also contrasting, results with respect to resting-state data analysis: we confirm previous findings on subject-specific fingerprints, but we also shed light on the presence of factors that affect the variability of such fingerprints and, thus, the reliability of multi-site fMRI data collections and subsequent inference.

Results

Hidden Markov modeling allows us to represent the stochastic relationship between a limited number of hidden states that underlie the brain's complex dynamics, whose evolution in time is captured by the measured data. An HMM is described by a Transition Probability Matrix (TPM), which encodes the probability of transitioning from one state to another at each time step. Furthermore, the temporal characteristics of such states can be expressed as the Fractional Occupancy (FO) – the fraction of time brain dynamics spend in each state. FOs can be computed for each state, subject, or scanning session.

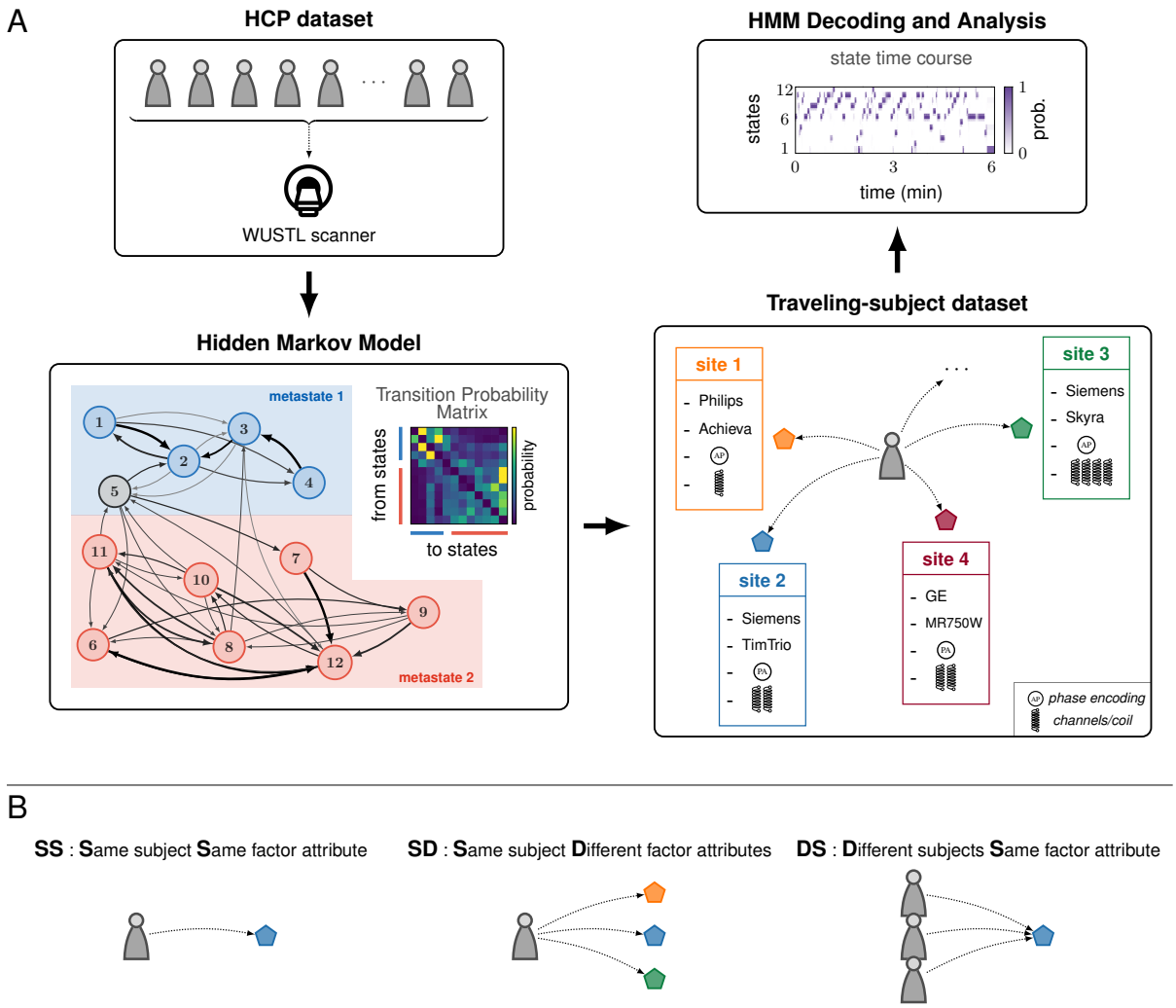

Figure 1: (A) Resting state fMRI data from the HCP dataset, which were all collected at the Washington University in St. Louis (WUSTL) Connectome-Skyra scanner, were used to infer a Hidden Markov Model (HMM). This model is described by a transition probability matrix, which encodes the probabilities of jumping from one state to another at each time step. Following [42], 12 states were identified and the graph depicted in the figure illustrates the largest transition probabilities (> 0.1) in our model (see also Fig. S2). The states are color-coded in order to distinguish which set of highly connected states (*metastate*) they belong to. HMM decoding is applied to the Traveling-subject dataset, in which subjects traveled to different sites to have their resting state fMRI data collected. The state time courses from the Traveling-subject dataset are finally used to (1) validate the subject-specific fingerprints associated to the 2-metastate structure put forth in [42], and (2) analyze the impact of different factors, such as site and scanner model, to fMRI measurements. (B) To gauge how different factors influence fMRI data collection, the state time courses obtained after decoding the HMM were compared within and across three different groups: SS, SD, and DS. In this panel, these three categories are illustrated for the factor "site", whose attributes consist of the different geographical locations.

For each state, we can stack the FO of all subjects, obtaining a vector of FOs that is as long as the number of subjects in the dataset. The pairwise correlation of these vectors yields a matrix, the FO Correlation Matrix (Materials and Methods), that displays the overall FO organization, and whose entries quantify the similarity between the FOs of each pair of states across all subjects. For each scanning session, the FO Correlation Matrix reveals the similarities and dissimilarities between brain states, and encodes the temporal characteristics of brain dynamics. In the following sections, we will make use of the FO Correlation Matrix to calculate two subject-specific metrics that will be key in answering the following questions: is HMM-estimated brain dynamics really subject-specific? Which scanning factors impact resting-state fMRI recordings? What is the magnitude of their effect?

Test-retest reliability of brain dynamics estimation

We first inferred the HMM by using the HCP resting-state fMRI data. Due to the large size of the available data, the stochastic nature of the HMM approach, and motivated by previous work [42], we inferred a battery of different models ($N = 50$ models) from the dataset and ranked them based on the free energy (i.e., an approximation of how well a model fits the data), as well as on the Euclidean distance from the ideal FO Correlation Matrix (Materials and Methods and Fig. S1). This second quantity is based on the definition of metastates, which are distinct sets of networks, or states, that the brain has a tendency to cycle within. The Euclidean distance from the ideal FO Correlation Matrix gauges how well the metastates emerge in the model's FO Correlation Matrix. We decided to also include this second measure, which assigns a significant weight in the model selection to the emergence of the metastates, because the metastates are key in the definition of the subject-specific fingerprints used in this study. We show in Fig. 2 the HMM selected and employed in this work, which is the best with respect to both the free energy and the distance from the ideal FO Correlation Matrix.

Because of the stochastic nature of the Variational Bayes approach used to infer the HMM [39], it is highly unlikely that one would obtain an exact replica of the model originally reported in [42]. However, as displayed in Fig. 2*B*, our model shows a very clear 2-metastate structure. A visual inspection of the TPM matrix alone suggests the emergence of two groups of states that tend to be more "connected". We confirm such hypothesis by employing the generalized Louvain algorithm [6] for the discovery of communities in networks. All the models trained from the datasets in this study show a 2-metastate structure, validating the claims in [42] that resting-state brain dynamics tend to be hierarchically organized in two larger sets of states, one associated with higher-order cognition, and the other one with sensorimotor and perceptual states. Importantly, we also trained the model on the Traveling-subject dataset alone while using the HCP-derived

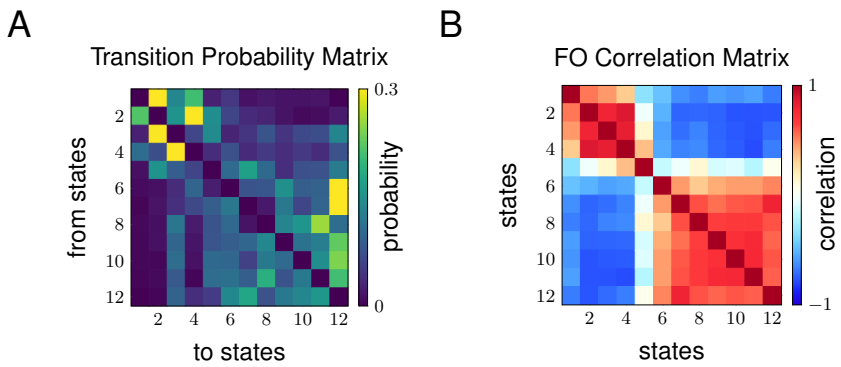

Figure 2: Matrices associated with our HMM. (A) Transition Probability Matrix. The emergence of the two metastates can be recognized by simple visual inspection, and it is confirmed by a community-detection algorithm. (B) FO Correlation Matrix. The two metastates are clearly delineated, with state 5 being mostly uncorrelated from all other states [42]. The state FOs are highly correlated (Person correlation > 0.8) within the two metastates across subjects.

TPM as prior TPM. This choice of prior ensures that the inference starts from established initial conditions before dealing with the small size of the Traveling-subject dataset. Surprisingly, although the number of subjects in the Traveling-subject dataset n_{TS} is much smaller than the number of subjects in the HCP dataset n_{HCP}, i.e. $n_{\mathrm{TS}} \ll n_{\mathrm{HCP}}$, the 2-metastate structure still emerges in the model's matrices (Fig. S3D). This result highlights that, notwithstanding mixed scanning protocols and small sample, the metastates can be retrieved and unfold as a robust feature of resting-state data.

Metastate Profiles and Fractional Occupancies Are Robust Subject-Specific Fingerprints

We here extend previous findings purporting that brain dynamics is subject-specific and nonrandom. We first applied the HCP-trained HMM on the Traveling-subject dataset, obtaining the state time courses for each 10-minute scanning session. From the state time courses, we calculated the Metastate Profile (MP) as the FO of the second metastate (states 6-12) minus the FO of the first metastate (states 1-4). As done in [42], we excluded state 5 from our analysis as it is uncorrelated from the other states, has the highest variance, and is associated to head motion in the scanner. As a means to compare different scanning sessions for different subjects and different factors, we computed MP Differences and FO Correlations across different runs. The former is the absolute difference between the Metastate Profiles of different runs, and the latter is the pairwise correlation between the Fractional Occupancy vectors of different runs.

We find that, on average, the median MP Differences across sessions for the same subject and the same factor (SS) are significantly lower than the median MP Differences for different subjects within the same factor (DS), as shown in Fig. 3A (2-sided t-test,

$t_{10} = -3.59$, $p = 0.005$). Analogously, on average, the median FO Correlations are higher for the same subject and the same factor than the same quantity calculated for different subjects within the same factor (Fig. 3*B*) (2-sided t-test, $t_{10} = 8.15$, $p < 0.001$). Subject-wise, the median FO Correlation is greater than 0.9 (Fig. 3B, bar 'SS'), which highlights how the state time courses of the same subject within the same factor attributes tend to be exceptionally similar. This analysis corroborates the hypothesis that MP Differences and FO Correlations are robust subject-specific measures, as they are resilient to the single effect of all the factors considered in this study. Most importantly, our results indicate that differences between subjects are almost 38% larger for MP Differences than within subjects, while the FO correlations are more than 10% lower between subjects than within.

To further support our results, we use machine learning to predict individuals based on their brain dynamics fingerprints. We applied logistic regression to classify the individuals in the Traveling-subject dataset by a leave-one-attribute-out cross-validation procedure (SI Methods). In brief, for each fingerprint and for each factor, we repeated the training and validation of the classifier as many times as the number of factor attributes, using each time the samples of one left-out factor attribute as validation set and the remaining samples from all other factor attributes as training set. We find the accuracy of the classification to be consistently well above the baseline chance level (9 subjects: $1/9 \approx 0.11$), scoring on average 0.217 for the classification based on MPs, and 0.314 for the classification based on FOs (see Fig. S6).

In Resting-state fMRI Data, Some Factors are Less Equal Than Others

Given that the Traveling-subject dataset contains a considerable number of factors that could influence fMRI measurements, we inquired which, if any, of these factors influence the subject-specific fingerprints defined on the state time courses. Specifically, we asked which factors affect the MP Differences and the FO Correlations within and across subjects. We compared three different groups of MP Differences and FO Correlations, as illustrated in Fig. 1*B*, for six different factors: sites, days, phase encodings, number of channels per coil, scanner manufacturers, and scanner models. Each of these factors contains at least two attributes. For instance, the scanner manufacturers included in this study are three: GE, Philips, and Siemens. For a comprehensive list of all the attributes associated with the scanning factors in our dataset, we refer the interested reader to Table S1.

We find that some factors influence the MP Differences and the FO Correlations consistently more than others. Conversely, some factors do not seem to have any relevant effect on the aforementioned metrics. We summarize the main results of this comparison

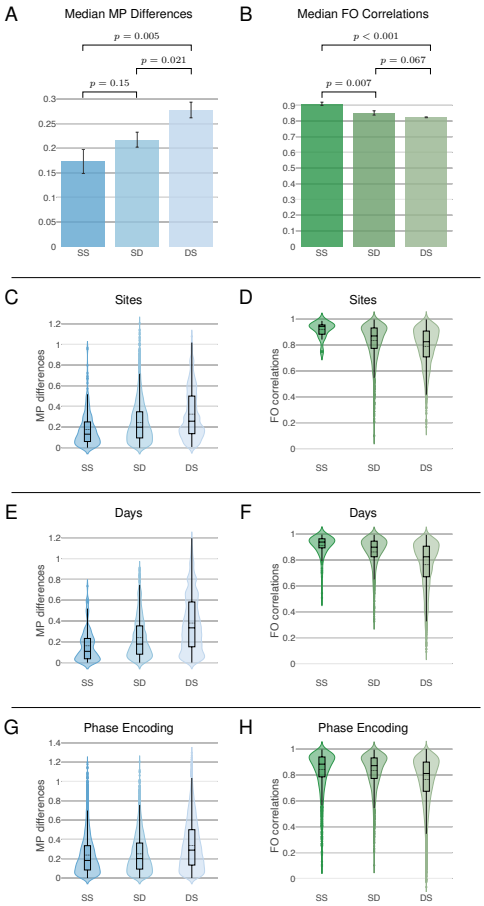

Figure 3: (A)-(B) The average median of the MP Differences and FO Correlations for the three sets SS (Same subject Same factor attribute), SD (Same subject Different factor attributes), and DS (Different subjects Same factor attribute), together with the p values from a two-sided t-test. MP Differences are the absolute difference between the Metastate Profiles of different runs, while FO Correlations are the pairwise correlation between the Fractional Occupancy vectors of different runs. We can assess that the set SS (same subjects same factor) consistently displays lower MP Differences and higher FO Correlations than the set DS (different subjects same factor), confirming the fact that such metrics are subject-specific. The fact that the set SD lies between SS and DS suggests that some scanning parameters influence the aforementioned metrics for resting-state scans of the same subject, but not as much as inter-individual differences. Panels (C) to (H) illustrate the distributions of values for both the metrics and all subjects, which have been computed as follows. The set SS comprises the MP Differences (resp., FO Correlations) computed for each subject within the same factor attribute, and the SS distribution displays these values for all subjects; the set SD consists of the MP Differences (resp., FO Correlations) computed for each subject across different attributes of the same factor, and the SD distribution displays these values for all subjects; finally, the set DS consists of the MP Differences (resp., FO Correlations) computed across all subjects within the same factor attribute, and the DS distribution displays these values for all attributes of the same factor. For all the distributions, the black dashed lines illustrate the mean. In panels (G) and (H), the difference between SS and SD distributions is not statistically significant (see Table 1).

Table 1: Kolmogorov-Smirnov test results for MP Differences and FO Correlations. All p-values have been FDR-adjusted [45] and all $p < 10^{-3}$ when the null hypothesis is rejected. SS: Same subject Same factor attribute. SD: Same subject Different factor attributes. DS: Different subjects Same factor attribute.

	MP Differences			FO Correlations		
Parameter	SS vs SD	SS vs DS	SD vs DS	SS vs SD	SS vs DS	SD vs DS
1. Site	1	1	1	1	1	1
2. Day	1	1	1	1	1	1
3. Phase	0	1	1	0	1	1
4. Channels/Coil	1	0	1	1	1	1
5. Manufacturer	1	1	1	1	1	1
6. Scanner	1	1	1	1	1	0

in Fig. 3. We also report in Table 1 the results of the Kolmogorov-Smirnov nonparametric test between all the distributions of values for the groups of MP Differences and FO Correlations. More in detail, by comparing the distributions of values for both metrics between the sets SS (same subject and same factor attribute) and SD (same subject and different factor attributes), we find them to be statistically different ($p < 0.001$, see Table 1) for all factors except for the phase encoding direction. Importantly, the factors that influence both the MP Differences and FO Correlations the most are sites and days, in agreement with previous findings on the variability of functional connectivity across sites [29]. In fact, the subject-specific fingerprints that we employ have, on average, smaller difference and variance for the values in the set SS than the values in the set DS. Instead, the distributions of MP Differences and FO Correlations values with respect to different phase encoding directions (posterior to anterior or anterior to posterior) are not significantly different (Table 1), as visible also in Fig. 3*G*-*H*. We report in Fig. S4 the distributions of values for the parameters that lie between the aforementioned extremum cases – all significantly different ($p < 0.001$). It is worth noting that, albeit smaller than between subjects, the difference in MP Differences for the same subject within the same factor attribute is, according to our data, 26% smaller than between different attributes; compatibly, the FO Correlations are 7% higher.

Finally, machine learning classification from brain dynamics fingerprints (SI Methods) reveals that, for both fingerprints, the accuracy in predicting individual subjects is the lowest when the training and validation sets are based on different days or sites (Table S3).

Discussion

Subject-specific fingerprints of brain dynamics are robust to different physical and temporal factors affecting the data that populates multi-site collections of neurophysiological and hemodynamic signals. This study corroborates and complements previous work that found that the emergence of temporal patterns of brain activity tends to repeat more similarly within the same subject and over time [42, 40, 11]. In this work, we show that such dynamical patterns are robust to a battery of different scanning factors in resting-state fMRI recordings. Furthermore, quantities that are specific to the same person, such as Mestastate Profile Differences and Fractional Occupancy Correlations, can be used to measure the effect of distinct physical and nuisance factors to the quality and reproducibility of fMRI data. We summarize in Fig. 4 our findings, where we compare the impact of sites, days, phase encoding directions, number of channels per coil, scanner manufacturers, and machine models, to the distributions of MP Differences and FO Correlations. Panel (A) illustrates the effect size (Cohen's d) of the factors on the differences between distributions, highlighting how the dissimilarity between the distributions of values of brain dynamics fingerprints is the greatest when comparing, for the same scanning factor, values from the same subject and values from different subjects. This complements, from a dynamical point of view, both seminal and more recent work reporting more dissimilar resting-state networks inter-subject than intra-subject [44, 3].

In this study, we also investigate which factors tend to influence resting-state fMRI data recordings the most. We find that different sites and days have the biggest effect on both MP Differences and FO Correlations, with a particular emphasis on the former, as we show in Fig. 4. More specifically, in panel (A) it is easy to see that in all three color-coded groups, site and day are consistently amongst the factors with the biggest effect size. Moreover, panel (B) highlights that the MP Differences seem to be impacted the most by the factors site and day. On the other hand, other scanning factors seem to have a rather small effect on the fingerprints distributions. These are the number of channels per coil and the phase encoding direction, which have a limited impact on the metrics considered in this work. Our findings relate to recent work reporting that different sites and vendors influenced fMRI functional connectivity maps [3, 29].

Warnings and caveats in multi-site studies

Recent years have witnessed a growing interest in the identification and characterization of the factors that tend to make multi-site fMRI recordings spurious [30, 18, 5, 3, 29], endangering the reproducibility and the overall quality of the results that may be inferred from these data. The first work raising a warning flag on multi-site studies was [44], where the authors investigated the undesirable relationships with nuisance variables on inter-individual variation in the functional connectome. A number of works has fol-

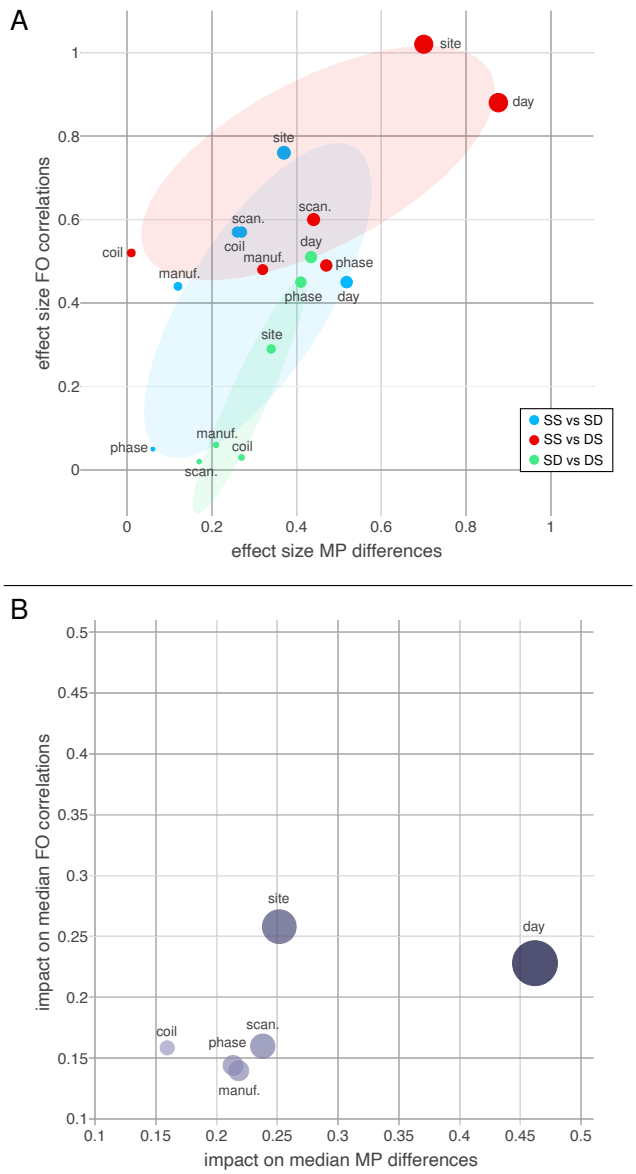

Figure 4: (A) The x and y axes represent the Cohen's d values obtained by comparing the log-transformed distributions of the MP differences and FO correlations for the three groups SS (Same subject Same factor attribute), SD (Same subject Different factor attributes), and DS (Different subjects Same factor attribute), within and across different scanning session factors, respectively (e.g., factor = channels per coil, attribute = 8, 12, 24, 32 channels). The ellipses represent the least squares minimization of the distance from the cloud of points for each of the three sets [17], and are depicted to ease the visual clustering of the data points. It is easy to see that the largest effect on both the metrics is consistently caused by the factors site and day, for all the comparisons between groups of distributions. (B) The impact due to different parameters on the medians of the SS, SD, and DS distributions of values, for all different scanning parameters. The impact is computed as the sum of the median differences. Since the possible values lie within the same range, the two axes can be directly compared, showing that the MP Differences are impacted the most by all scanning factors.

lowed, reporting mostly consistent results. For example, [5] provides a method to assess multisite reproducibility in resting-state functional connectivity fingerprints, [18] reports that hierarchical clustering may be able to identify structural and functional scans from different participants images on different scanners across time, [3] finds that consistency (spatial Pearson's correlation) of seven canonical resting-state networks is affected by sites and vendors, and [29] reports that voxel-wise connectivity is sensitive to site and scanner manufacturer effects.

With respect to previous work, this study contains a larger and more detailed dataset. Specifically, this dataset contains the largest number of traveling subjects out of all the aforementioned studies. To date, only [3] has more sites than the Traveling-subject dataset used in this study, but it has the drawback of scanning only a single subject. Furthermore, the Traveling-subject dataset allows for the analysis of some scanning factors – such as the numbers of channels per coil or different scanner models within the same vendor – that have not been taken into consideration in previous work, giving more breadth and depth to our findings.

Differently from [18, 3], where time seems to play a negligible effect, we find that different scanning days affect brain network dynamics. This is likely due to the fact that the HMM allowed us to analyze the state time courses associated to the subjects' brain dynamics, whereas previous study assessed *static* measures, such as parcellations and functional connectivity. Yet, our finding does not go against the claim that functional connectivity networks remain a reliable subject-specific fingerprint over long period of times, but rather we suggest that brain state trajectories can differ extensively between days. To note, due the limited number of scanning sessions taken on different days in the Traveling-subject dataset (see Table S1), this fact will need further validation by future studies.

Altogether, all the studies on multi-site datasets come to the common conclusion that the neuroscience community needs to be aware of all the pitfalls and drawbacks that are inherent to such collections of data, and emphasize the importance of harmonization methods [15, 46, 43].

Going Beyond the Classical Sliding Window Approach

Hidden Markov modeling and the sliding window approach seek for the same goal: the identification of functional networks that repeat over time [26]. If the goal is a robust and detailed description of the system's dynamics, the HMM approach requires large amounts of data for training purposes, thus appearing not suitable to analyze small cohorts of subjects. However, in this study we give proof that, although HMM inference requires a very large sample size, one can use a very large dataset (i.e. HCP) to infer an HMM and then apply it to a smaller dataset. We show with our results that the procedure is

robust. On the other hand, if only a relatively small number of subjects is available for the inference, it is still possible to recover a coarser – and nonrandom – representation of the brain dynamics by using the TPM inferred from a large dataset as a prior, as we report in Fig. S3*D*.

In light of the aforementioned observations, an HMM has advantages that make it superior to a sliding window approach in several scenarios. In fact, a HMM is as fast as the data modality allows, since it provides instantaneous likelihood of high correlation between brain signals. Instead, the sliding window approach, albeit being intrinsically easier to set up, has two crucial limitations. First, the choice of the window is constrained by a tradeoff between time resolution and quality of the results. Second, since the sliding time window approach relies on a user-defined window, some changes observed in resting-state fMRI dynamic functional connectivity may be explained by sampling variability [20]. Because we can estimate functional connectivity for each state by using the entire pool of subjects, this latter problem does not affect the HMM.

In conclusion, given the considerable recent advances in inference techniques [39, 7, 8], and the ever-increasing availability of computational power, our work further suggests that the HMM is, and, most importantly, will be, a powerful technique to explain and interpret the dynamic aspects of the brain. Furthermore, the possibility of inferring an HMM on a very large dataset to apply it to a much smaller one has important implications for clinical applications. In the future, perhaps with even more data, these general models could be built and then utilized to infer subject-specific fingerprints in other smaller cohorts and be used for a more personalized approach to treatments. In other words, a one-size-fits-all approach could be employed to build the model, consequently allowing us to move to a personalized course of action by evaluating the model at the individual level. For instance, closed-loop fMRI neurofeedback [10] could significantly benefit from these models, which will allow for a more holistic approach to the dynamical properties of mental and cognitive processes, particularly from a clinical perspective [37, 35].

Methodological Considerations

Despite its capabilities, hidden Markov modeling is based on some assumptions and comes at a cost. A discussion on the assumptions underlying the HMM is in order. Firstly, the states are assumed to be mutually exclusive. However, such assumption may be relaxed by considering different time scales. For instance, the time spent in each state or certain repeating sequences of states may capture states overlapping functional connectivity patterns or a continuous variation in mean BOLD activation. The HMM is also based on the Markovian assumption. This theorizes that we can predict, based on the state we are at time t, which state is more likely to follow at time $t + 1$. Yet, it is established that the brain features a number of long-range dependencies [19]. While the

HMM does not explicitly model such interactions, it does not exclude their existence. In fact, FO Correlations and MP Differences inherently display information that appears only at longer time scales. Further, obtaining a good HMM from group inference requires a considerably large number of subjects and scanning sessions. Notwithstanding, as we show in Fig. S1D, the metastate structure can still be recovered even when only nine subjects are used to infer the HMM. Thus, detailed analyses and claims based on hidden Markov modeling should be gauged on the size of the available data.

One common limitation of studies involving traveling subjects is the size of the dataset. Although this work contains one of the biggest datasets of this type with respect to number of states, scanning factors, and subjects, a larger number of traveling subjects would certainly add to the robustness of the claims made in this study. Further, the claims associated to the data in panels (A) and (B) of Fig. 3 will require further validation by consequent studies, due to the small ($n = 6$) number of values for each set (SS, SD, and DS) from which the medians are computed and compared.

Finally, it is also worthwhile to touch upon the differences between the HCP dataset, which was used for the HMM inference, and Traveling-subject dataset employed in our study. These differences relate to the preprocessing pipeline, the scanning protocols, and even the countries in which the data were collected. On the one hand, these differences may weaken our claims. On the other hand, while such differences are necessary for the analyses performed in this work, they also strongly corroborate the result that brain dynamics fingerprints are subject-specific. Indeed, the metrics used in this study are shown to be resilient to all the aforementioned differences between the two datasets.

Conclusion

In this work, we address the important issues of reproducibility and variability of fMRI data. We show, as a proof of concept, that dynamical states can be estimated reliably. We leveraged the large HCP collection of resting-state data to infer a hidden Markov model capable of describing the brain state time courses at the subject level. By applying such a model to a dataset of traveling subjects, we find that brain network dynamics displays a signature fingerprint that is robust to different scanning factors and distinctive for each subject. This result enables and promotes further investigations on the dynamical characteristics of brain states. Once a good model is inferred, it can be applied to a battery of different goals, such as the analysis of task-based datasets, the examination of data collections from subjects with neurological disorders, and the promising use in clinical or rehabilitation settings, for instance by using brain state inference in clinical populations to estimate the best time for providing a given treatment.

Materials and Methods

Datasets

The two dataset used in this study are (1) the HCP 1200-distribution and (2) the Traveling-subject dataset. The former consists of resting-state fMRI data from $N = 1206$ healthy subjects (age 22-35) that were scanned twice (two 15-minute runs) on two different days, one week apart, on a Siemens 3T Connectome-Skyra scanner. For each subject, four 15-minute runs of resting state fMRI time series data with a temporal resolution of 0.73 s and a spatial resolution of 2-mm isotropic were available. For our analysis, we used time series from the 1003 subjects with 4 complete scanning sessions. The HCP dataset provides the required ethics and consent needed for study and dissemination, such that no further institutional review board (IRB) approval is required.

The Traveling-subject dataset consists of 9 healthy subjects (all men; age range 24–32; mean age 27 ± 2.6y), who were all scanned at each of the 12 sites, producing a total of 411 10-minute scanning sessions [43]. There were two phase-encoding directions (posterior to anterior [$P \rightarrow$ A] and anterior to posterior [A $\rightarrow$ P]), three MRI manufacturers (Siemens, GE, and Philips), four numbers of channels per coil (8, 12, 24, and 32), and seven scanner types (TimTrio, Verio, Skyra, Spectra, MR750W, SignaHDxt, and Achieva). Furthermore, subjects were scanned for 5 cycles at one of the sites. All participants in all datasets provided written informed consent. All recruitment procedures and experimental protocols were approved by the institutional review boards of the principal investigators' respective institutions (Advanced Telecommunications Research Institute International [ATR] [approval numbers: 13–133, 14–133, 15–133, 16–133, 17–133, and 18–133], Hiroshima University [E-38], Kyoto Prefectural University of Medicine [KPM] [RBMR-C-1098], SWA [B-2014-019 and UMIN000016134], the University of Tokyo [UTO] Faculty of Medicine [3150], Kyoto University [C809 and R0027], and Yamaguchi University [H23-153 and H25-85]) and conducted in accordance with the Declaration of Helsinki.

FO Correlation Matrix and Fingerprints Computation

By applying (i.e., decoding) an HMM to a dataset with multiple subjects, we obtain the state time courses for each subject, from which it is possible to compute the vector of the FO of every state for each subject. Stacking such vectors in a matrix yields the FO Matrix $\mathbf{R}$, which is a *(no. of subjects)* $\times$ *(no. of states)* matrix. Each element $\mathbf{R}_{ij}$ of this matrix denotes the fraction of time spent by subject i in state j. To compare different scanning sessions for different subjects and different factors, we compute the Metastate Profile (MP) Differences and the Fractional Occupancy (FO) Correlations across different runs. The MP for subject i and run k is the FO of the second metastate (states 6-12)

minus the FO of the first metastate (states 1-4), and is computed as follows:

$$\mathrm{MP}_{i,k} = \mathbf{R}_{i,6:12} - \mathbf{R}_{i,1:4}.$$

Then, the MP Difference between run k_1 for subject i_1 and run k_2 for subject i_2 reads as

$$\text{MP difference} = |\mathrm{MP}_{i_1,k_1} - \mathrm{MP}_{i_2,k_2}|.$$

Further, by taking the pairwise correlation of the columns of the FO Matrix $\mathbf{R}$, we obtain the *(no. of states)* $\times$ *(no. of states)* FO Correlation Matrix $\mathbf{C} = \mathrm{corr}(\mathbf{R}_{:,k}, \mathbf{R}_{:,\ell})$, where $\mathbf{R}_{:,k}$ denotes the column vector of the FO of all subjects for the k-th state.

Finally, it is worth noting that exploiting and comparing the two metrics defined above gives us a remarkable advantage with respect to utilizing only the model's TPM. Namely, because of the stochastic nature of the model inference, we are able to avoid the non-uniqueness issue of the TPM and, at the same time, to reliably capture the temporal characteristics of the state time courses.

Model Training and Selection

To train the HMM, we used the publicly available toolbox `HMM-MAR` (`https://github.com/OHBA-analysis/HMM-MAR`) [41]. Following [42], extensively preprocessed HCP time series [32, 33] were used for the training, combined with resampled time series from the Traveling-subject dataset. The latter were resampled in order to match the same repetition time (TR) as the HCP data. Both dataset were reduced to a 50-dimensional ICA space by means of group-spatial ICA. Before the training of the HMM, all time series were standardized so that each scanning session has 0 mean and unitary standard deviation, and finally concatenated along the time direction. Full details are in SI Methods.

To select the model that best fits the data, we computed two values for each of the fifty different models that we have inferred: the free energy and the Euclidean distance from the ideal FO Correlation Matrix (Fig. S1). The former provides a bound on the log-evidence for any model, and is composed of a tradeoff between the model complexity and the model accuracy [16]. Because the data sets have different sizes (HCP only, HCP plus Traveling-subject, and Traveling-subject only), we corrected the free energy according to the size of the dataset used for the model inference in order to compare different models in a fair way. The second value that we computed for each model is

$$d_i = \left\| \mathbf{C}_i - \begin{bmatrix} \mathbf{1}_{4\times4} & \mathbf{0}_{4\times1} & -\mathbf{1}_{4\times7} \\ \mathbf{0}_{1\times4} & 1 & \mathbf{0}_{1\times7} \\ -\mathbf{1}_{7\times4} & 0_{7\times1} & \mathbf{1}_{7\times7} \end{bmatrix} \right\|$$

for $i = 1, \ldots, 50$, where $\mathbf{C}_i$ if the FO Correlation Matrix of model i, $\mathbf{1}$ is a matrix of all

ones, $\mathbf{0}$ is a zero matrix, and $\|\cdot\|$ denotes the Euclidean norm. The model depicted in Fig. 2 is the one with the lowest free energy and the smallest d_i.

References

[1] A. Abraham, M. P. Milham, A. D. Martino, R. C. Craddock, D. Samaras, B. Thirion, and G. Varoquaux. Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example. *NeuroImage*, 147:736 – 745, 2017.

[2] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun. Tracking Whole-Brain Connectivity Dynamics in the Resting State. *Cerebral Cortex*, 24(3):663–676, 11 2012.

[3] A. Badhwar, Y. Collin-Verreault, P. Orban, S. Urchs, I. Chouinard, J. Vogel, O. Potvin, S. Duchesne, and P. Bellec. Multivariate consistency of resting-state fmri connectivity maps acquired on a single individual over 2.5 years, 13 sites and 3 vendors. *NeuroImage*, 205:116210, 2020.

[4] K. Bansal, J. D. Medaglia, D. S. Bassett, J. M. Vettel, and S. F. Muldoon. Data-driven brain network models differentiate variability across language tasks. *PLOS Computational Biology*, 14(10):1–25, 10 2018.

[5] S. Bari, E. Amico, N. Vike, T. M. Talavage, and J. Goñi. Uncovering multi-site identifiability based on resting-state functional connectomes. *NeuroImage*, 202:115967, 2019.

[6] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

[7] M. Bonomi, C. Camilloni, A. Cavalli, and M. Vendruscolo. Metainference: A bayesian inference method for heterogeneous systems. *Science Advances*, 2(1), 2016.

[8] D. Bzdok and B. T. Yeo. Inference in the age of big data: Future perspectives on neuroscience. *NeuroImage*, 155:549 – 564, 2017.

[9] M. Charquero-Ballester, B. Kleim, D. Vidaurre, C. Ruff, E. Stark, J. J. Tuulari, H. McManners, Y. Bar-Haim, L. Bouquillon, A. Moseley, S. C. R. Williams, M. Woolrich, M. L. Kringelbach, and A. Ehlers. Effective psychological treatment for ptsd changes the dynamics of specific large-scale brain networks. *bioRxiv*, 2020.

[10] A. Cortese, K. Amano, A. Koizumi, M. Kawato, and H. Lau. Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. *Nature Communications*, 7(1):1–18, 2016.

[11] C. E. Curtis and M. D'Esposito. Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, 7(9):415 – 423, 2003.

[12] J. S. Damoiseaux, S. A. R. B. Rombouts, F. Barkhof, P. Scheltens, C. J. Stam, S. M. Smith, and C. F. Beckmann. Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences*, 103(37):13848–13853, 2006.

[13] R. A. Feis, S. M. Smith, N. Filippini, G. Douaud, E. G. P. Dopper, V. Heise, A. J. Trachtenberg, J. C. van Swieten, M. A. van Buchem, S. A. R. B. Rombouts, and C. E. Mackay. ICA-based artifact removal diminishes scan site differences in multi-center resting-state fMRI. *Frontiers in Neuroscience*, 9:395, 2015.

[14] J.-P. Fortin, N. Cullen, Y. I. Sheline, W. D. Taylor, I. Aselcioglu, P. A. Cook, P. Adams, C. Cooper, M. Fava, P. J. McGrath, M. McInnis, M. L. Phillips, M. H. Trivedi, M. M. Weissman, and R. T. Shinohara. Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*, 167:104 – 120, 2018.

[15] J.-P. Fortin, D. Parker, B. Tunç, T. Watanabe, M. A. Elliott, K. Ruparel, D. R. Roalf, T. D. Satterthwaite, R. C. Gur, R. E. Gur, R. T. Schultz, R. Verma, and R. T. Shinohara. Harmonization of multi-site diffusion tensor imaging data. *NeuroImage*, 161:149 – 170, 2017.

[16] K. Friston, J. Mattout, N. Trujillo-Barreto, J. Ashburner, and W. Penny. Variational free energy and the Laplace approximation. *NeuroImage*, 34(1):220 – 234, 2007.

[17] R. Halır and J. Flusser. Numerically stable direct least squares fitting of ellipses. In *Proc. 6th International Conference in Central Europe on Computer Graphics and Visualization. WSCG*, volume 98, pages 125–132. Citeseer, 1998.

[18] C. Hawco, J. D. Viviano, S. Chavez, E. W. Dickie, N. Calarco, P. Kochunov, M. Argyelan, J. A. Turner, A. K. Malhotra, R. W. Buchanan, and A. N. Voineskos. A longitudinal human phantom reliability study of multi-center t1-weighted, dti, and resting state fmri data. *Psychiatry Research: Neuroimaging*, 282:134 – 142, 2018.

[19] B. J. He. Scale-free properties of the functional magnetic resonance imaging signal during rest and task. *Journal of Neuroscience*, 31(39):13786–13795, 2011.

[20] R. Hindriks, M. H. Adhikari, Y. Murayama, M. Ganzetti, D. Mantini, N. K. Logothetis, and G. Deco. Can sliding-window correlations reveal dynamic functional connectivity in resting-state fmri? *NeuroImage*, 127:242 – 256, 2016.

[21] C. R. Jack Jr., M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L.

Gunter, D. L. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, G. Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, and M. W. Weiner. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging*, 27(4):685–691, 2008.

[22] K. Jann, D. G. Gee, E. Kilroy, S. Schwab, R. X. Smith, T. D. Cannon, and D. J. Wang. Functional connectivity in BOLD and CBF data: Similarity and reliability of resting brain networks. *NeuroImage*, 106:111 – 122, 2015.

[23] T. Kaufmann and et al. Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nature Neuroscience*, 2019.

[24] G. Lisi, D. Rivela, A. Takai, and J. Morimoto. Markov switching model for quick detection of event related desynchronization in EEG. *Frontiers in Neuroscience*, 12:24, 2018.

[25] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of Neural Engineering*, 15(3):031005, apr 2018.

[26] D. J. Lurie, D. Kessler, D. S. Bassett, R. F. Betzel, M. Breakspear, S. Keilholz, A. Kucyi, R. Liégeois, M. A. Lindquist, A. R. McIntosh, R. A. Poldrack, J. M. Shine, W. H. Thompson, N. Z. Bielczyk, L. Douw, D. Kraft, R. L. Miller, M. Muthuraman, L. Pasquini, A. Razi, D. Vidaurre, H. Xie, and V. D. Calhoun. Questions and controversies in the study of time-varying functional connectivity in resting fMRI. *Network Neuroscience*, 2019. In Press.

[27] H. Markram. The Blue Brain Project. *Nature Reviews Neuroscience*, 7(2):153, 2006.

[28] T. J. Mitchell, C. D. Hacker, J. D. Breshears, N. P. Szrama, M. Sharma, D. T. Bundy, M. Pahwa, M. Corbetta, A. Z. Snyder, J. S. Shimony, and E. C. Leuthardt. A Novel Data-Driven Approach to Preoperative Mapping of Functional Cortex Using Resting-State Functional Magnetic Resonance Imaging. *Neurosurgery*, 73(6):969–983, 09 2013.

[29] S. Noble, D. Scheinost, E. S. Finn, X. Shen, X. Papademetris, S. C. McEwen, C. E. Bearden, J. Addington, B. Goodyear, K. S. Cadenhead, H. Mirzakhanian, B. A. Cornblatt, D. M. Olvet, D. H. Mathalon, T. H. McGlashan, D. O. Perkins, A. Belger, L. J. Seidman, H. Thermenos, M. T. Tsuang, T. G. van Erp, E. F. Walker, S. Hamann, S. W. Woods, T. D. Cannon, and R. T. Constable. Multisite reliability of MR-based functional connectivity. *NeuroImage*, 146:959 – 970, 2017.

[30] R. A. Poldrack, K. Whitaker, and D. Kennedy. Introduction to the special issue on reproducibility in neuroimaging. *NeuroImage*, page 116357, 2019.

[31] M. E. Raichle, A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, and G. L. Shulman. A default mode of brain function. *Proceedings of the National Academy of Sciences*, 98(2):676–682, 2001.

[32] S. M. Smith, C. F. Beckmann, J. Andersson, E. J. Auerbach, J. Bijsterbosch, G. Douaud, E. Duff, D. A. Feinberg, L. Griffanti, M. P. Harms, M. Kelly, T. Laumann, K. L. Miller, S. Moeller, S. Petersen, J. Power, G. Salimi-Khorshidi, A. Z. Snyder, A. T. Vu, M. W. Woolrich, J. Xu, E. Yacoub, K. Uğurbil, D. C. V. Essen, and M. F. Glasser. Resting-state fMRI in the Human Connectome Project. *NeuroImage*, 80:144 – 168, 2013. Mapping the Connectome.

[33] S. M. Smith, D. Vidaurre, C. F. Beckmann, M. F. Glasser, M. Jenkinson, K. L. Miller, T. E. Nichols, E. C. Robinson, G. Salimi-Khorshidi, M. W. Woolrich, D. M. Barch, K. Uğurbil, and D. C. V. Essen. Functional connectomics from resting-state fmri. *Trends in Cognitive Sciences*, 17(12):666 – 682, 2013. Special Issue: The Connectome.

[34] J. Stiso, A. N. Khambhati, T. Menara, A. E. Kahn, J. M. Stein, S. R. Das, R. Gorniak, J. Tracy, B. Litt, K. A. Davis, F. Pasqualetti, T. H. Lucas, and D. S. Bassett. White matter network architecture guides direct electrical stimulation through optimal state transitions. *Cell Reports*, 28(10):2554 – 2566.e7, 2019.

[35] L. Stoeckel, K. Garrison, S. Ghosh, P. Wighton, C. Hanlon, J. Gilman, S. Greer, N. Turk-Browne, M. deBettencourt, D. Scheinost, C. Craddock, T. Thompson, V. Calderon, C. Bauer, M. George, H. Breiter, S. Whitfield-Gabrieli, J. Gabrieli, S. LaConte, L. Hirshberg, J. Brewer, M. Hampson, A. V. D. Kouwe, S. Mackey, and A. Evins. Optimizing real time fMRI neurofeedback for therapeutic discovery and development. *NeuroImage: Clinical*, 5:245 – 255, 2014.

[36] C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, B. Liu, P. Matthews, G. Ong, J. Pell, A. Silman, A. Young, T. Sprosen, T. Peakman, and R. Collins. UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine*, 12(3):1–10, 03 2015.

[37] V. Taschereau-Dumouchel, A. Cortese, T. Chiba, J. D. Knotts, M. Kawato, and H. Lau. Towards an unconscious neural reinforcement intervention for common fears. *Proceedings of the National Academy of Sciences*, 115(13):3470–3475, 2018.

[38] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, and K. Ugurbil. The WU-Minn Human Connectome Project: An overview. *NeuroImage*, 80:62 – 79, 2013. Mapping the Connectome.

[39] D. Vidaurre, R. Abeysuriya, R. Becker, A. J. Quinn, F. Alfaro-Almagro, S. M. Smith, and M. W. Woolrich. Discovering dynamic brain networks from big data in rest and task. *Neuroimage*, 180:646–656, 2018.

[40] D. Vidaurre, L. T. Hunt, A. J. Quinn, B. A. E. Hunt, M. J. Brookes, A. C. Nobre, and M. W. Woolrich. Spontaneous cortical activity transiently organises into frequency specific phase-coupling networks. *Nature communications*, 9(1):2987, 2018.

[41] D. Vidaurre, A. J. Quinn, A. P. Baker, D. Dupret, A. Tejero-Cantero, and M. W. Woolrich. Spectrally resolved fast transient brain states in electrophysiological data. *Neuroimage*, 126:81–95, 2016.

[42] D. Vidaurre, S. M. Smith, and M. W. Woolrich. Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences*, 114(48):12827–12832, 2017.

[43] A. Yamashita, N. Yahata, T. Itahashi, G. Lisi, T. Yamada, N. Ichikawa, M. Takamura, Y. Yoshihara, A. Kunimatsu, N. Okada, H. Yamagata, K. Matsuo, R. Hashimoto, G. Okada, Y. Sakai, J. Morimoto, J. Narumoto, Y. Shimada, K. Kasai, N. Kato, H. Takahashi, Y. Okamoto, S. C. Tanaka, M. Kawato, O. Yamashita, and H. Imamizu. Harmonization of resting-state functional mri data across multiple imaging sites via the separation of site differences into sampling bias and measurement bias. *PLOS Biology*, 17(4):1–34, 04 2019.

[44] C.-G. Yan, R. C. Craddock, X.-N. Zuo, Y.-F. Zang, and M. P. Milham. Standardizing the intrinsic brain: Towards robust measurement of inter-individual variation in 1000 functional connectomes. *NeuroImage*, 80:246 – 262, 2013. Mapping the Connectome.

[45] D. Yekutieli and Y. Benjamini. Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference*, 82(1):171 – 196, 1999.

[46] M. Yu, K. A. Linn, P. A. Cook, M. L. Phillips, M. McInnis, M. Fava, M. H. Trivedi, M. M. Weissman, R. T. Shinohara, and Y. I. Sheline. Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fmri data. *Human Brain Mapping*, 39(11):4213–4227, 2018.

[47] A. Zalesky, A. Fornito, L. Cocchi, L. L. Gollo, and M. Breakspear. Time-resolved resting-state brain networks. *Proceedings of the National Academy of Sciences*, 111(28):10341–10346, 2014.