Individual traits versus invariances of cognitive functions

A model-based study of brain connectivity

Vicente Pallarés Picazo

TESI DOCTORAL UPF / year 2019

THESIS SUPERVISORS
Dr. Matthieu Gilson, Prof. Gustavo Deco and Dr. Rafael Ramirez
Department of Information and Communication Technologies

"Was sich überhaupt sagen läßt, läßt sich klar sagen; und wovon man nicht reden kann, darüber muß man schweigen."
Ludwig Wittgenstein

A Vicente y Esperanza,

Agradecimientos

Después de un camino tan largo, creo que es necesario recordar y agradecer a todos los que de una manera u otra han contribuido a que me acercara un poco más hacia esta meta. Seguro que me dejo algunos nombres, por eso pido disculpas de antemano.

En primer lugar, mi agradecimiento va a para Matthieu, Gustavo y Rafael, por darme la oportunidad, los consejos oportunos y por su infinita paciencia. He aprendido mucho durante estos años con su inestimable ayuda. Y cómo no, a mi supervisor en la sombra, colaborador y amigo. Siempre disponible casi a cualquier hora, en cualquier lugar y con cualquier tipo de café.

I want to thank Mark Bellgrove for giving me the opportunity of visiting his lab, his support and useful advice. And also thanks to Méadhbh Brosnan for her help and our scientific conversations during my time in Melbourne.

Por supuesto he de mencionar también al resto del grupo. Realmente he llegado a pensar muchas veces en que no podía existir un lugar mejor donde trabajar. En estos cuatro años la frontera entre la vida personal y la profesional se diluye, así que gracias a todos, compañeros y amigos, por el apoyo y los momentos compartidos: Víctor Saenger, Ruggero Bettinardi, Jessica De Santiago, Ane López, Manel Vila, Federico Devalle, Irene Torres, Marc Lluís Vives, Alice Del Genovese, Bea Jobst, Irene Canudas, Laura Ulysse, Adrián Ponce, Gorka Zamora, Alex Hyafil, Silvana Silva, Txema Esnaola, Iñigo Romero, Roberta Coda, Ignasi Cos, Ana Sanjuán, Mario Pannunzi, Tommaso Gagliardi, Philipp Schustek, Adrià Tauste. También a Ernest Montbrió por su esfuerzo y la orientación que me ofreció para preparar las clases. Mención especial merecen Ramón Nogueira y Josefina Cruzat, por su amistad y consejos en los momentos de incertidumbre. Y por supuesto, Ana Martín, por las conversaciones sobre trabajo y por estar siempre ahí. Quizá sin ellos no habría llegado hasta aquí.

También quiero agradecer a mi familia, a Manolo y Magda, José Luis y Elisa, Marta y Elena. Por su permanente interés y su cariño. Y por último,

me gustaría agradecer de corazón a mis padres Esperanza y Vicente, por su constante apoyo. La persona que soy hoy en día se debe en gran medida a la educación que me han proporcionado, a su implicación y el cariño que he recibido siempre.

Resumen

Es conocido en la literatura de neuroimagen que las redes cerebrales funcionales reflejan rasgos personales. Estas características individuales, podrían interferir al caracterizar la cognición entendida como la manera en que se coordinan las redes para realizar una tarea, como mantener la atención, recordar, o procesar información visual. Cómo estos aspectos individuales coexisten con mecanismos generales es, por tanto, una pregunta clave en investigación sobre conectividad cerebral. Este trabajo estudia la relación entre marcadores de conectividad específicos tanto de sujetos, como de tareas. Se centra en dos escalas temporales distintas: la variabilidad entre sesiones, y las fluctuaciones rápidas producidas durante una sesión de adquisición. Utilizamos técnicas de *machine learning* para separar cuantitativamente las contribuciones de información del sujeto y del estado cognitivo a la conectividad. La metodología presentada nos permite extraer aquellas redes representativas de ambas dimensiones, así como profundizar en su evolución, sugiriendo las escalas temporales relevantes en la cognición.

Palabras clave: neurociencia computacional, fMRI, conectividad cerebral, modelos cerebrales, aprendizaje automático, selección de features, conectividad dinámica, redes cerebrales, análisis multivariado, conectividad funcional, correlación, integración, segregación

Resum

És conegut en la literatura de neuroimatge que les xarxes cerebrals funcionals reflecteixen trets personals. Aquestes característiques individuals podrien interferir en caracteritzar la cognició entesa com la manera en què les xarxes es coordinen per realitzar una tasca, com mantenir l'atenció, recordar o processar informació visual. Cóm aquests aspectes individuals coexisteixen amb mecanismes generals, és, per tant, una pregunta clau en recerca sobre connectivitat cerebral. Aquest treball estudia la relació entre marcadors de connectivitat específics tant de subjectes, com de tasques. Se centra en dues escales temporals: la variabilitat entre sessions, i les fluctuacions ràpides produïdes durant una sessió d'adquisició. Utilitzem tècniques de *machine learning* per separar quantitativament les contribucions d'informació del subjecte i de l'estat cognitiu a la connectivitat. La metodologia presentada ens permet extreure aquelles xarxes representatives d'ambdues dimensions, així com aprofundir en la seva evolució, suggerint les escales temporals rellevants en la cognició.

Paraules clau: neurociència computacional, fMRI, connectivitat cerebral, models cerebrals, aprenentatge automàtic, selecció de trets, connectivitat dinàmica, xarxes cerebrals, anàlisi multivariat, connectivitat funcional, correlació, integració, segregació

Abstract

There is consistent evidence in the neuroimaging literature that functional brain networks reflect personal traits. Individual specificity may interfere with the characterization of cognition, in terms of coordination of brain networks to perform a task, such as sustained attention, memory retrieval or visual information processing. How individual traits coexist with invariant mechanisms is, therefore, a key question in brain connectivity research. This work aims to examine the relationship between subject- and task-specific connectivity signatures. It focuses on two different timescales: day-to-day variability and faster fluctuations exhibited within a scanning session. We adopt a machine learning approach to quantitatively disentangle the contribution of subject information and cognitive state to the connectivity patterns. The proposed methodology allows us to extract the specific brain networks that support each of the two dimensions, as well as to delve into their changes over time, suggesting the relevant timescales for cognition.

Keywords: computational neuroscience, fMRI, brain connectivity, whole-brain modelling, machine learning, feature selection, dynamic connectivity, brain networks, multivariate analysis, functional connectome, correlation, integration, segregation

Contents

List of Figures

CHAPTER 1

Introduction

1.1. Integration versus segregation

Brain can be seen as an information processing system composed of elemental units –different types of neurons– that are organized according to a hierarchical structure of specialized subsystems. Psychology and early neuroscience traditionally used to interpret the brain on a localizationist approach. Through this perspective, the brain may be studied by assigning specific functions to concrete anatomical regions. Following this basic input-output scheme, the first initial neuroimaging research attempted to identify those loci of neuronal activity that responded to a presented stimulus. Functional Magnetic Resonance Imaging (fMRI) permitted the localization of activity over the whole brain via the Blood-Oxygen-Level-Dependent signal. BOLD signal corresponds to fluctuations in the brain vascular response due to local metabolic changes. These changes may be induced –among other reasons– by an increase energy demand of neuronal ensembles. The association between the BOLD signal registered by fMRI and the activity of group of neurons, derived in its interpretation as

a proxy of neural activity. This association allowed for the localization of specific functions in the brain during some decades, whose first promising results contributed to the consolidation of the localizationist point of view in brain computation. Roughly explained, in localizationist studies the brain is viewed as a segregated system, and more or less narrowly defined groups of voxels are assigned to specific brain functions. However, cognition, especially in naturalistic settings, requires integration of information and the coordination of several basic functions, such as stimuli representations, with other higher-order processes, like attention.

The integration principle, in contrast with segregation, is at the core of an alternative understanding of brain functions that conceives the brain as a system, in which the interconnection between distinct areas matters more than their individual roles. This connectionist notion of brain hierarchical organization has become more popular during last decades (Fig. 1.1). The balance between segregation and integration conceptions of brain function represents, ultimately, the interplay between two distinct phenomena: the small-scale dynamics that locally govern the activity of neurons, and the interconnection/exchange that distant neuronal ensembles exert over each other according to the wiring scheme of anatomical connections [Varela et al., 2001, Bressler and Tognoli, 2006, Catani et al., 2013, Medaglia et al., 2015, Tononi, 2005].

Brain segregation can be quantified by comparing the activity of each voxel to a particular stimulus. All the voxels that respond similarly to the stimulus might shape one or more functional regions. The whole brain may be spatially segmented or parcellated into a group of regions that consistent with these functional areas. As an alternative to this functional parcellation, an anatomical atlas can be defined based on brain's structure and used to parcellate fMRI data. Integration studies at the whole-brain level, on the other hand, attempt to evaluate pairwise (or higher-order) interactions between distinct regions of interest (ROI). A simple and commonly used method of quantifying integration on fMRI data is via the so-called functional connectivity (FC). FC measures reflect the statistical dependence between different brain areas [Smith, 2012, van den Heuvel and Hulshoff Pol, 2010, Buckner et al., 2013]. The initial definition in fMRI

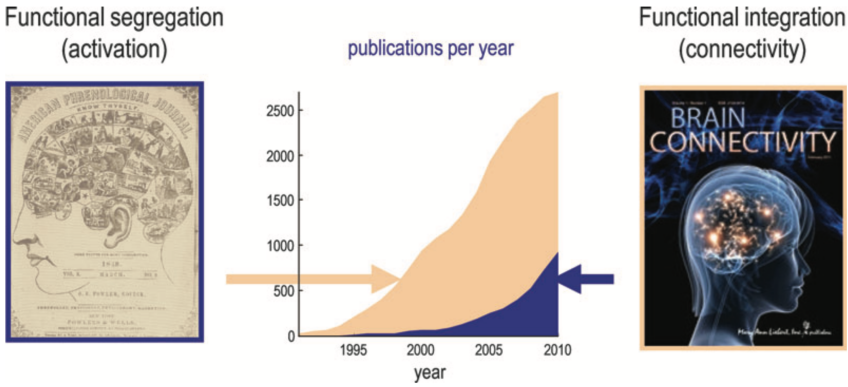

Figure 1.1: Publications per year searching for "Activation" or "Connectivity" and functional imaging. This reflects the proportion of studies looking at functional segregation (activation) and those looking at integration (connectivity). Reproduced from [Friston, 2011].

corresponds to the calculation of pairwise correlations of regional BOLD signals. Other measures of connectivity, although less popular than correlation, are also used in the literature, such as mutual information [Hlinka et al., 2011], coherence [Bowyer, 2016] or transfer entropy [Orlandi et al., 2014].

The emergence of FC patterns does not necessarily imply that two ROI are truly coupled (Fig. 1.2). The effective influence that one ROI exert over another cannot be directly captured through methods of statistical dependence, but has to be inferred through model-based approaches, such as Granger causality [Kim et al., 2011], dynamic causal modelling or DCM [Friston, 2011,Smith et al., 2011], or MOU-EC [Gilson et al., 2016]. Regardless of the method/model used, the estimated EC consists of a set of directed –and weighted– interactions between pairs of ROI. Yet, the dependency of EC on the particular generative model used makes the comparison across studies non trivial [Friston, 2011,Gilson et al., 2016]. In the model, EC determines the propagation of activity between regions. It can be seen, therefore, as another way of quantifying brain integration [Tononi et al., 1994]. FC and EC represent, then, two complementary views of the

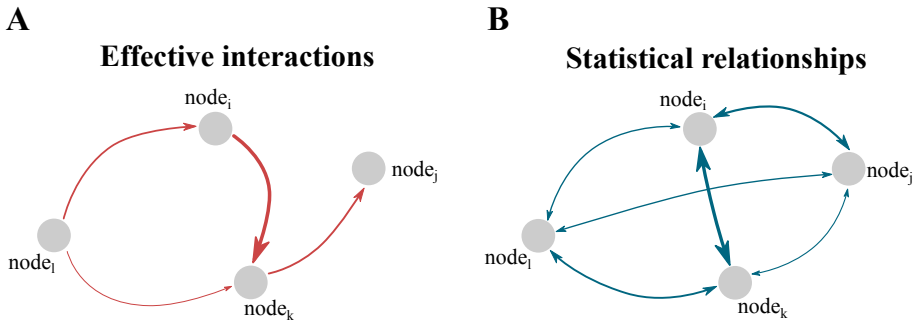

Figure 1.2: Comparison between effective interactions and statistical relationship measured in a four-node network. Toy model comparing the effective interactions that are driving the dynamics of a given network, with the undirected statistical relationships that are measured –e.g. Peason correlation– across them. **A)** Example of effective connectivity network. The arrows illustrate the influence that nodes exert over each other: $node_i$ is exerting over $node_k$, and $node_k$ over $node_j$. **B)** Example of functional connectivity network. Undirected (or bi-directional) interactions between the nodes. Two spurious interactions between $node_i$ and $node_j$, and between $node_j$ and $node_l$ emerge when computing the statistical dependence between all pairs of nodes as a consequence of indirect paths in **A)**. The thickness of the arrows indicates the strength of the interaction.

integration function: the former based on the statistical quantification of some measured observables, and the latter as a mechanistic explanation, generated by a mathematical model, of those very observables [Friston, 2011].

1.2. Sources of variability in fMRI to study cognition

In a neuroimaging study using fMRI, participants may be scanned under different cognitive conditions: they can be lying inside the scanner without receiving any specific stimulus or instruction, which is known as a resting state, or they can perform a task. Each individual subject might undergo scanning more than once, either performing the same or different tasks. These fMRI runs may be acquired during one unique session along the same day, or on separate days, perhaps with weeks or months in between. All these different circumstances entail that fMRI data will necessarily carry a complex mixing of variability coming from different sources. We will adopt as a general framework a scheme focusing on three of these sources of variability: subject, condition and time. The first main source will account for differences between subjects. This is the main dimension of variability that population studies try to minimize when comparing groups of participants that share a common trait, e.g. healthy versus clinical populations. The second source of variability will represent differences in the cognitive state imposed to the participants while they undergo scanning. This cognitive condition will depend on the cognitive state or particular task performed inside the scanner, such as awake versus asleep, rest versus task, or drug versus placebo. The third source of variability will represent variation over time. This dimension will capture differences in connectivity between several fMRI sessions of a given subject performing the same task. Importantly, this variability will manifest in the connectivity estimates extracted from fMRI runs, such as FC or EC, which are candidates to become functional biomarkers: a set of stable patterns of brain connectivity that indicate some clinical or cognitive condition.

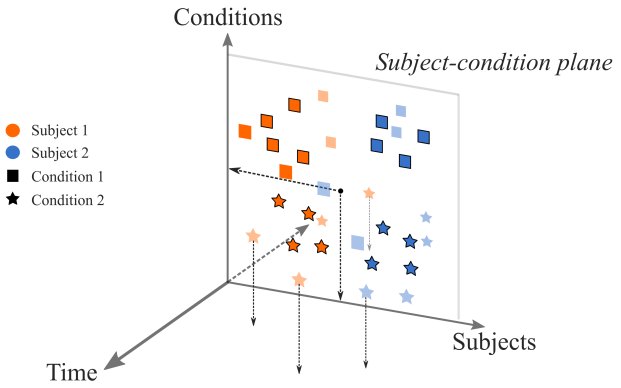

Figure 1.3: Three-dimensional coordinates space of variability. Coordinates axes representing the space comprised by the three sources of variability considered: subject, cognitive condition and time.

The three sources of variability may define a three-dimensional space (Fig. 1.3), where each axis corresponds to one variability source: subject, condition and time. According to this general structure, each individual fMRI run can be characterized as a point in the three-dimensional space. Next sections describe in detail this three dimensions of variability.

1.2.1. Individual variability

The traditional view in neuroimaging used to ignore subject differences when analyzing population studies, i.e. when comparing across groups of subjects. Under this perspective, individual differences are seen as a confound to be minimized in order to find a more significant outcome. However, last decade has witnessed a shift in the interpretation of this source of variability. In the era of the personalized medicine, individual differences appear more as a source of valuable information that can help to understand the specificity of each subject's cognitive profile [Seghier and Price, 2018]. This has a clear potential application to diagnosis and treatment of neurological and psychiatric diseases [Woo et al., 2017].
It has been shown that the information carried by BOLD time courses is

specific enough to allow for subject identification [Miranda-Dominguez et al., 2014]. This subject specificity appears to be inherited by their connectivity estimates, renaming FC as connectome fingerprint. Although FC can be used to identify subjects up to a high degree [Finn et al., 2015], many questions still remain open, such as its reproducibility, generalizability or the impact of the pre-processing pipeline [Waller et al., 2017, Anderson et al., 2011, Airan et al., 2016].
When studying the functional connectome, it has been seen that not all the brain connections have the same relevance in the characterization of subjects' individuality, but it is heterogeneously distributed over the whole brain. This suggests that some brain networks are more representative than others of subject's individual traits [Mueller et al., 2013, Chen et al., 2015].
If brain networks reflect subject's individuality, the question that naturally arises is: Can these networks be used to characterize also cognitive functions beyond individual differences? If this is possible, connectivity estimates might be used to understand the underpinnings of cognitive processes. During recent years, researchers have tried to correlate these brain networks to psychological scores such as fluid intelligence [Finn et al., 2015], sustained attention [Rosenberg et al., 2015], personality traits [Kong et al., 2018] or psychiatric disease [Kaufmann et al., 2017]. These seemingly promising results must, however, be taken with caution. Other non-cognitive confounder could be contaminating the fMRI BOLD signal and might be subject-specific as well, for example head motion [Siegel et al., 2017, Satterthwaite et al., 2012].

1.2.2. Condition-related variability

The word *condition* is used here to refer to different types of behavioral tasks perfomed inside the scanner –e.g. sustained attention, working memory, emotion–, but sufficiently stable brain states can also fall into this categorization [Haimovici et al., 2017]. It is important to note that sometimes there is no clear border between these groups: for example, different degrees of wakefulness can be defined between two clear consciousness states. Among all the variety of possible conditions, the juxtaposition of

task and resting state has been much studied in the recent neuroimaging literature. The resting state is commonly defined as a task-free condition. This makes rest a not well defined cognitive state due to the absence of a specific stimulus, in comparison with the other task-driven states. Connectivity networks extracted from both task and rest seem to be related [Finn et al., 2015, Tavor et al., 2016], perhaps, due to its shared subject dimension. Some studies propose task-based states as a more appropriate approach to study the relationship between cognition and subject's specificity [Finn et al., 2017, Greene et al., 2018, Vanderwal et al., 2017].
It is important to note that the orthogonal axes displayed in Fig. 1.3 are representing an ideal situation. The degree of independence/orthogonality between the subject and condition axes, might strongly depend on the problem at hand (see 2.3.5; Fig. 2.14). In some cases, individual subject's traits could dominate the noise mixture making the comparison across cognitive states impossible [Gratton et al., 2018]. To complicate even more the situation subject- and condition-related information could overlap in the space of brain connections, i.e. the same subnetworks might carry both subject-specific and condition-specific information. In these situations, we could enhance one or the other depending on our focus of interest [Xie et al., 2017]. Another possibility, which we propose in chapter 2, may consist in disentangling the subject- or condition-specific subnetworks. In short, understanding the two-dimensional subject-condition plane is crucial before inferring general conclusions about brain connectivity. Importantly, this plane created from subject and condition variability axes will change through a third dimension: the within- or between-sessions temporal variability.

1.2.3. Session-to-session variability

In the exposition so far we are implicitly assuming that functional connectivity is stable over time. Yet, we expect the functional connectivity to vary over repetitions of the experiment with the same subject and task. All the samples of the same subject and task extracted from separate sessions, will present, therefore, some differences. These differences should be cap-

tured in the time axis of Fig. 1.3, and may be caused by slow processes. Depending on the purpose of the study –e.g when studying learning, aging or disease evolution– this changes over time will be informative or undesired. We are going to assume in this work, that this session-to-session variability is undesired. Thus, our aim will be to minimize the impact of session-specific variability on our results. Reliability is especially crucial on resting state data because of the absence of specific stimuli and, therefore, any reference signal. This means that two fMRI resting state runs of the same subject or condition might present more differences than two stimulus-task runs. However, results showed that resting-state connectivity estimates were, in general, robust [Shehzad et al., 2009]. This test-retest reliability, which is related to the number of subjects, can explain the subject-specificity of FC described before. Remarkably, single connections present a low reliability, whereas at the multivariate level, brain connectivity shows to be robust, especially for particular subnetworks [Pannunzi et al., 2017]. These results can be extended to connectivity estimates extracted from task fMRI [Shah et al., 2016]. One way of evaluating the stability of connectivity samples is through similarity measures. Similarity may be calculated, for example, as the correlation, or the inverse of the distance –euclidean, Kolmogorov-Smirnov– between different samples. These metrics are usually applied to quantify the stability of samples that belong to the same group, either subject or condition. There are other measures that compare the similarity within a group with the dissimilarity across different groups. Some of these measures are Intra-Class Coefficient (ICC) [Shrout and Fleiss, 1979] or silhouette coefficient [Rousseeuw, 1987]. This ratio of within-group similarity over their dissimilarity –i.e. between-groups differences– is critical when applying statistical inference or machine learning techniques, as we will see in chapters 2 and 3.

Regardless of the chosen measure, within-group samples extracted from fMRI runs acquired during the same scanning session are expected to have a high degree of similarity. But, when acquired on different sessions, probably on separate days, this within-group similarity decreases [Birn et al., 2013]. Some elements that may increase the session-to-session variability

are differences in mood, awareness or time-of-day effects [Shannon et al., 2013]. The impact that other external factors have on the reliability of connectivity samples must be considered. Decisions regarding the scanning parameters, such as repetition time or length, as well as pre-processing decisions, or the parcellation scheme, are also determinant [Anderson et al., 2011, Airan et al., 2016]. Signal length has shown to be directly linked to the reproducibility of connectivity samples [Birn et al., 2013]. This is particularly relevant in resting state, for which a signal length of five minutes was established as a minimum criterion for obtaining reliable data [Van Dijk et al., 2010]. The relationship between test-retest reliability and signal length has been analyzed again in the context of individual subject identification: i.e. longer time series may capture better subject-specific characteristics [Birn et al., 2013, Finn et al., 2015].

1.2.4. Dynamic functional connectivity

One might conclude from previous section that the longer the BOLD time series, the more reliable the connectivity estimates obtained. And, as explained before, maximizing reliability is fundamental in order to be able to predict more accurately on the subject-condition plane. All the studies that adopt this perspective about maximizing reliability, propose some kind of averaging operation: concatenating or acquiring longer time series, or computing the mean of several FC estimates. These approaches are very common in the neuroimaging literature and are based on one critical assumption: the stationarity of the measured BOLD signal. This means that the computed FC is *static* in the sense that it is picturing some kind of steady internal state. Indeed, in most of studies, brain networks are extracted using an entire run of fMRI data to generate one connectivity sample. In practice, correlation and other similar statistical measures applied provide insight into overall relationships across variables. Therefore, whole-session connectivity estimates are really measuring general (or dominant) trends.

On the other hand, the brain is a dynamic organ that must constantly adapt to the rapidly changing environment. The vascular nature of BOLD sig-

nal imposes an upper limit to the measurement of this response. In addition, since BOLD signal is not directly measuring brain activity, it is unknown how much of these fluctuations in brain activity can be captured with fMRI [Liégeois et al., 2017]. Brain connectivity patterns are known to change on very different timescales that may range from years to seconds [Preti et al., 2017, Gonzalez-Castillo and Bandettini, 2018]. As exposed, in the pursuit of reproducible measures, longer time series provide more reliable samples. Yet, using whole-session measures could destroy too other shorter transient aspects of BOLD signal. These temporal features might be helpful to understand the ongoing cognitive processes.

A whole new family of methods have emerged during last years that try to bring together the extraction of faster temporal information, with the estimation of reliable functional connectomes. One of these approaches, so-called dynamic functional connectivity (dFC), is based on a wide range of techniques. These techniques go from the earliest and simplest ones, like computing FC on shorter segments of signal, to the most novel and sophisticated approaches [Preti et al., 2017]. In spite of this broad variety, the main idea that all of them share consists in computing a time-resolved FC, as opposed to its static version calculated using the entire time series. The most common way to do this involves segmenting the whole-session BOLD time courses –extracted after brain parcellation– into shorter temporal windows. Functional connectivity is computed, then, within each window, usually by means of Pearson correlation.

The dFC obtained after computing the time-resolved connectivity is a high-dimensional data structure, even if the parcellation scheme applied comprises a low number of ROI. Researchers have often transformed this structure into a format more amenable to analysis. One possibility consists in calculating some coefficients or statistics, reducing so the dFC into some summary measures. A qualitatively different solution is based on the extraction of recurrent patterns of connectivity that can define a set of discrete states [Calhoun et al., 2014, Gonzalez-Castillo et al., 2015]. Some approaches obtain this FC states by means of the factorization of dFC data, for example, via k-means [Allen et al., 2014] or singular value decomposition/PCA [Leonardi et al., 2013]. It is important to note that the

interpretation and characteristics of the obtained FC states will strongly depend on the methodology, making complicated the comparison between different techniques [Preti et al., 2017].
It is still a matter of debate whether these fluctuations and FC states are reflecting the ongoing neural activity or could be mainly caused by signal artifacts –motion or other physiological signals– or sampling variability [Siegel et al., 2017, Hindriks et al., 2016, Leonardi and Van De Ville, 2015]. Research using simultaneous recordings of both fMRI and electrophysiological signals, such as EEG, has already shown some evidence of an association between dFC patterns and neural activity [Tagliazucchi et al., 2012, Thompson et al., 2013, Chang et al., 2013]. On the contrary, some other works have also shown that it was possible to measure fluctuations even in cases where the ground-truth FC was known to be stationary [Laumann et al., 2017, Handwerker et al., 2012]. Although there are some ways to account for these artifacts and minimize their impact on the estimation of dFC [Laumann et al., 2017], establishing a proper relation to behavior and cognitive ability may come out as a valuable manner of validating the plausible neural origin of dFC [Cohen, 2018].

1.3. Objectives

Most of the literature on brain connectivity using fMRI can be categorized into one of two distinct perspectives concerning the very nature of the functional networks [Gratton et al., 2018]. One view is that *"functional networks mirror cognitive, perceptual, and motor processes, reconfiguring substantially with ongoing context, task demands, moods"*. In this sense, the measured connectivity would shape a set of possible network states dynamically defined by the underlying processing. An alternate view is that *"functional networks are fundamentally stable, [...] informative about a given person's stable traits (such as disease status or personality) regardless of that person's thoughts, mood or even behavioral response during the scan"*.
This perspective, would allow for the study of functional brain states and

trajectories, providing some foundations to reach a deeper comprehension of cognition. In this case functional networks would be, however, less useful for clinical applications, since estimates would be strongly modified by the specific context under which they were measured. According to the latter view, *"functional networks are understood as primarily determined by structural connections that maintain a functionally stable correlation structure linked to long-term histories of co-activation patterns"*. This suggests that functional networks might suitable to study personal traits and disease, with applications in personalized medicine, but would be less convenient for unveiling the underpinnings of cognitive processes.

In this work we will explore these two complementary views. To do that, we will interpret them as projections in the three-dimensional space presented in Fig. 1.3. First, we will assess the stability of personal traits using resting-state functional networks with two different test-retest datasets. The outcome will confirm the previous results of a high-degree of subject-specificity for functional networks. We will also use whole-brain modelling to extract the effective connectivity, which includes subjects' temporal information via the autocovariance, to enrich the characterization of brain networks.

The results of comparing FC and EC suggest that EC might capture stable personal traits better than FC. This could indicate that EC uses subject's temporal information contained in BOLD signal more efficiently than FC. Moreover, EC may be seen as a mechanistic explanation of FC co-activations, which provides a more intuitive understanding of brain information processing. Finally, we will use EC to differentiate between conditions: natural movie viewing versus resting state. This will show that a set of subnetworks, which are specific to a given task, can be extracted out of the whole-brain –and high-dimensional– data space. We will characterize the connectivity data on the subject-condition plane: the session-to-session variability represented in the time axis will be considered undesired.

For the second part of the thesis we will work again on the plane defined by the subject and condition axes. In this case, we will assess whether dFC can capture fast changes in functional networks by classifying differ-

ent task conditions. The results on the first data set will show an example of easy –in terms of dimensions– classification between two distinct tasks conditions. We will see how by augmenting the number of samples using dFC we can improve the classification accuracy. Different combinations of dFC parameters will be evaluated to find the more appropriate timescale.

Finally, we want to show with several examples that a machine learning is a very convenient approach to analyze this kind of data. Functional connectomes, such as FC or EC, usually comprise a large number of individual links. This, together with the noise distribution, makes complicated the application of null hypothesis testing to these data. On the contrary, machine learning tools are well-adapted to this kind of high-dimensional data. Instead of testing for multiple comparisons, the predictive power can be easily assessed and interpreted by using a proper cross-validation procedure. For this reason, most of the presented results have been validated using machine learning techniques.

CHAPTER 2

Subject- and condition-specific connectivity signatures

The study of brain communication based on fMRI data is often limited because such measurements are a mixture of session-to-session variability with subject- and condition-related information. Disentangling these contributions is crucial for real-life applications, in particular when only a few recording sessions are available. The present study[1] aims to define a reliable standard for the extraction of multiple signatures from fMRI data, while verifying that they do not mix information about the different modalities (e.g., subjects and conditions such as tasks performed by them). In particular, condition-specific signatures should not be contaminated by subject-related information, since they aim to generalize over subjects. Practically, signatures correspond to subnetworks of directed interactions

[1]This work has been performed in collaboration with Andrea Insabato and Matthieu Gilson, and has been already published: *Pallarés V., Insabato A., Sanjuán A., Kühn S., Mantini D., Deco G., Gilson M. (2018). Extracting orthogonal subject- and condition-specific signatures from fMRI data using whole-brain effective connectivity. Neuroimage 178, 238-254.*

between brain regions (typically 100 covering the whole brain) supporting the subject and condition identification for single fMRI sessions. The key for robust prediction is using effective connectivity instead of functional connectivity. Our method demonstrates excellent generalization capabilities for subject identification in two datasets, using only a few sessions per subject as reference. Using another dataset with resting state and movie viewing, we show that the two signatures related to subjects and tasks correspond to distinct subnetworks, which are thus topologically orthogonal. These results set solid foundations for applications tailored to individual subjects, such as clinical diagnostic.

2.1. Introduction

Blood-oxygen-level dependent (BOLD) signals in functional magnetic resonance imaging (fMRI) have been used for more than two decades to observe human brain activity and relate it to cognitive functions [Cordes et al., 2000, Amunts et al., 2014]. Even at rest, the brain exhibits patterns of correlated activity between distant areas [Biswal et al., 1995, Raichle et al., 2001]. Although a consensus has yet to be reached, many methods have been proposed to quantify this activity and extract functionally-relevant information, beyond the widely-used functional connectivity (FC) that measures the statistical dependencies between the BOLD activities of brain regions. For instance, interest in the temporal BOLD structure for both individual regions [He, 2011] and between areas (via the cross-covariance lags at the scale of seconds) [Mitra et al., 2015] has grown; the 'dynamic FC' was defined to quantify the BOLD correlations at the scale of minutes [Preti et al., 2017, Gonzalez-Castillo and Bandettini, 2018]. Following fundamental discoveries about brain functions, fMRI has increasingly been used to complement clinical diagnostic for neuropathologies [Matthews and Hampshire, 2016]. Resting-state fMRI has also been found to be informative about neuropsychiatric disorders [Greicius, 2008]: alterations in FC correlate with and can predict the clinical scores of several diseases [Kurth et al., 2015, Rahim et al., 2017]. This defines "sig-

natures" in the form of subnetworks of links between brain regions that consistently covary with respect to behavioral or pathological conditions.

It has recently been shown that fMRI signals are strongly biased by individual traits [Miranda-Dominguez et al., 2014, Finn et al., 2015, Finn et al., 2017, Calhoun et al., 2017], which should be taken into account when extracting task-specific signatures [Poldrack et al., 2009, Xie et al., 2017]. Generalizing, the mixture of session-to-session, subject-specific and condition-related variability in FC is a severe limitation for real-life applications when only a few sessions per subject can be recorded [Varoquaux et al., 2017, Woo et al., 2017]. In that respect, FC measures obtained from successive recording sessions have been found not to be reliable individually (i.e., at the level of single links), but only collectively [Shehzad et al., 2009, Mueller et al., 2015, Chen et al., 2015, Pannunzi et al., 2017]. Because previous studies [Finn et al., 2015, Miranda-Dominguez et al., 2014, Finn et al., 2017, Xie et al., 2017] were limited to datasets with at most 3 sessions for the same behavioral condition per subject, we aim to rigorously assess the generalization capability of prediction methods to future (unseen) data. Meanwhile, we examine how multivariate classification can disentangle signatures for distinct modalities (here subject and condition/task).

Distributed signatures in FC across the whole brain have been observed in memory tasks [Rissman and Wagner, 2012] or when the subject experiences psychological pain [Chang et al., 2015]. Moreover, the etiology of many mental disorders is unknown: They are suspected to arise from network dysfunction, as reported for large-scale FC alterations in patients with schizophrenia [Hoptman et al., 2012]. These examples strongly point in favor of whole-brain approaches to study high-level cognition [Deco et al., 2011] and brain diseases [Deco and Kringelbach, 2014]; in contrast, focusing on a few cortical areas only to test hypotheses [Goebel et al., 2003, Bastos-Leite et al., 2015] may not capture sufficient information or network effects. Such whole-brain approaches typically involve a large number of parameters to estimate, which may impair the robustness of classification based on these parameters. One aim of the present study is to provide a practical answer to this trade-off.

The idea underlying the study of FC –in the broad sense– is that it reflects how brain areas dynamically bind to exchange and process information [Fries, 2005, Hipp et al., 2011, Betti et al., 2013]. To go beyond a phenomenological description of FC, we employ a large-scale effective connectivity (EC) relying on a model inversion [Gilson et al., 2016] to decompose FC into changes in network connectivity and local fluctuating activity. We borrow the EC terminology to describe interactions between brain regions from dynamic causal modeling [Friston, 2011] despite differences with our model, which will be discussed later. So far, DCM-based EC has been used for classification, but using 6 ROIs and two sessions only [Brodersen et al., 2011]. Here we stress the need for adequate quantitative methods to assess to which extent EC can predict the subject's identity [Miranda-Dominguez et al., 2014, Frässle et al., 2015] or behavioral condition [Gonzalez-Castillo et al., 2015], because the inference of signatures –e.g., changes in FC that "simply" correlate with behavioral conditions [Li et al., 2012, Bastos-Leite et al., 2015, Gilson et al., 2017]– does not guarantee robustness in the classification performance. In other words, parameter inference hints at putative signatures, but their predictive power must be carefully verified. As mentioned earlier, the overlap between subject- and condition-related signatures may impair the classification performance, when some links are sensitive to both modalities and should be excluded (or "conditioned out" as with probabilities) to disentangle the signatures.

The present study aims to set a new standard for extracting multivariate signatures from fMRI data, here applied to discriminate between subjects or behavioral conditions. It is organized in two parts. First, we couple whole-brain EC estimation with adequate machine learning tools to control for session-to-session variability. The focus is on the comparison between EC and FC in their generalization capabilities to unseen data for the classification of single resting-state fMRI sessions with respect to (healthy) subjects [Finn et al., 2015, Miranda-Dominguez et al., 2014]. We use datasets with large numbers of sessions per subject to provide a benchmark with an unprecedented level of control. Second, we predict both subject identity and condition (rest versus movie viewing) to verify

that EC can disentangle the two types of signatures. To do so, we examine the topological distribution of the EC links supporting the twofold classification, which allows us to quantify the overlap between the two signatures. This second part, thus, aims to assess the generalization capability of EC for multivariate classification.

2.2. Methods

Datasets A, B and C were used in this study and are described in Table 2.1 (see Appendix A for further details). The three datasets went through the workflow depicted in Fig. 2.1. After a standard pre-processing pipeline, a parcellation scheme covering the whole-brain was applied to obtain the averaged BOLD time series: 116 ROIs for the AAL parcellation and 66 ROIs for the Hagmann parcellation, with each color representing an anatomical subsystem of several ROIs (see Appendix B). Dataset A (A1+A2) was used to test the robustness of subject identification in session-to-session variability. Dataset B was used to test the generalization capability of the identification procedure for a larger number of subjects. Dataset C was used to extract both individualized and behavioral signatures.

Here we consider two versions of functional connectivity: the classical corrFC corresponding to the Pearson correlation coefficient (PCC) between pairs of time series; and the spatiotemporal FC embodied by the two covariance matrices FC0 and FC1, without and with time shift, respectively (see Eq. (2.1) and Eq. (2.2)).

In addition to these two types of functional connectivity, we also consider a model-based effective connectivity. This EC is generated by means of a whole-brain dynamic model constrained on the topology provided by the structural connectivity or SC (see section 2.2.1). In the following sections we describe in detail the procedure to obtain these different measures of connectivity.

Table 2.1: Details of the three datasets used in this chapter.

Dataset name	Subjects	Sessions per subject	Session duration
Dataset A1	6	40-50 (rest)	5 minutes
Dataset A2	50	1 (rest)	5 minutes
Dataset B	30	10 (rest)	10 minutes
Dataset C	19	3 (rest) and 2 (movie)	10 minutes

Structural connectivity (SC)

For all models, we used a generic matrix of structural connectivity (averaged over subjects) that is the same for all subjects, to determine the skeleton of the effective connectivity (i.e., existing connections). For Dataset C, structural connectivity for Dataset C was estimated from diffusion spectrum imaging (DSI) data collected in five healthy right-handed male participants [Hagmann et al., 2008]. The gray matter was first parcellated into the $N = 66$ ROIs, using the same low-resolution atlas used for the FC analysis. For each subject, we performed white matter tractography between pairs of cortical areas to estimate a neuro-anatomical connectivity matrix. In our method, the DSI values are only used to determine the skeleton: a binary matrix of structural connectivity (SC) obtained by averaging the matrices over subjects and applying a threshold for the existence of connections. The strengths of individual intracortical connections do not come from DSI values, but are optimized as explained below. For Datasets A and B, the generic SC corresponded to the AAL parcellation with $N = 116$ ROIs [Tzourio-Mazoyer et al., 2002]. A similar pipeline was used with diffusion tensor imaging.
It is known that both tractography and DTI underestimate inter-hemispheric connections [Hagmann et al., 2008]. Homotopic connections between mirrored left and right ROIs are important in order to model whole-cortex BOLD activity [Messé et al., 2014]. For both SC matrices, we added all possible homotopic connections, which are tuned during the optimization as other existing connections.

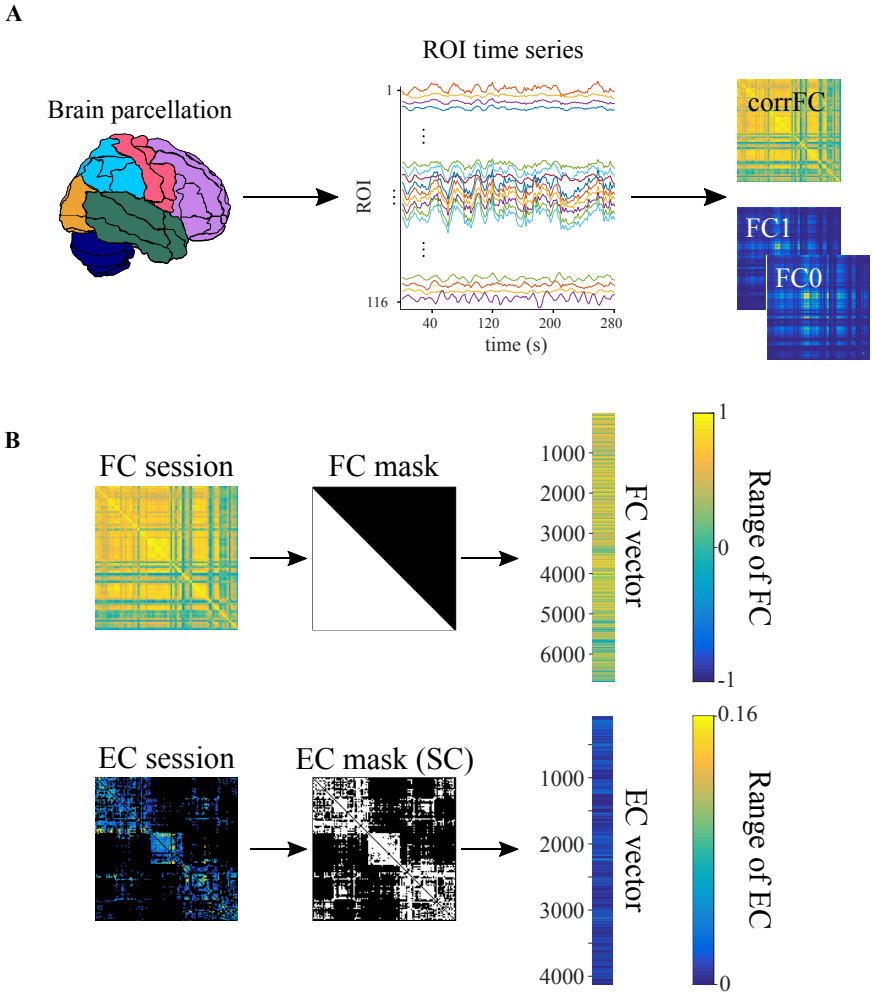

Figure 2.1: Workflow for the calculation of the connectivity measures from fMRI measurements. A) Pre-processing pipeline applied to BOLD time series. **B)** Each corrFC matrix is symmetric and has all diagonal elements equal to 1, so that only 6670 independent links are retained for identification/classification (lower triangle). Likewise, the EC matrix has 4056 non-zero elements that are used in the classification (density of 30%).

2.2.1. Connectivity measures and model estimates

Here we provide details about the calculation of the functional and effective connectivity measures introduced in Fig. 2.1.

Empirical measures of functional connectivity

For each fMRI session, the BOLD time series is denoted by s_i^t for each region $1 \leq i \leq N$ with time indexed by $1 \leq t \leq T$ (time points separated by a TR=2 seconds). The time series were first centered by removing –for each individual ROI i– the session mean $\bar{s}_i = \frac{1}{T}\sum_t s_i^t$. Following [Gilson et al., 2016], the spatiotemporal FC corresponds to covariances calculated as:

$$\begin{aligned} \widehat{Q}_{ij}^0 &= \frac{1}{T-2} \sum_{1 \leq t \leq T-1} (s_i^t - \bar{s}_i)(s_j^t - \bar{s}_j) \,, \\ \widehat{Q}_{ij}^1 &= \frac{1}{T-2} \sum_{1 \leq t \leq T-1} (s_i^t - \bar{s}_i)(s_j^{t+1} - \bar{s}_j) \end{aligned} \tag{2.1}$$

The classical BOLD correlations (corrFC in the main text) correspond to

$$\widehat{K}_{ij}^0 = \frac{\widehat{Q}_{ij}^0}{\sqrt{\widehat{Q}_{ii}^0 \widehat{Q}_{jj}^0}} \,. \tag{2.2}$$

Model of cortical dynamics

The model uses two sets of parameters to generate the spatiotemporal FC:

- The network effective connectivity (EC) between the ROIs (Fig. 2.2) is denoted by the matrix C in the following equations. Its skeleton is determined by the SC matrix, but not its weight values: Weights for absent connections are kept equal to 0 at all times, but weights for existing connections are estimated from FC matrices for each session.

- The local variability is described by the variances (one per ROI) on the diagonal of the matrix Σ.

The model FC comes from the propagation of the local variability –inputed to every ROI– that propagates via EC, generating network feedback.
Formally, the network dynamics is described by a multivariate Ornstein-Uhlenbeck process, where the activity variable x_i of node i decays exponentially with time constant τ_x –estimated using Eq. (2.7)– and evolves depending on the activity of other populations:

$$\mathrm{d}x_i = \big(\frac{-x_i}{\tau_x} + \sum_{j \neq i} C_{ij} x_j\big)\mathrm{d}t + \mathrm{d}B_i \ , \tag{2.3}$$

where $\mathrm{d}B_i$ is formally a Wiener process, a.k.a. Brownian motion, with covariance matrix Σ; note that only variances on the diagonal are non zero here.
The simplicity of the model allows for an analytic (feedforward) estimation of the covariances $Q^0_{ij} = \langle x_i(t)x_j(t)\rangle$ and $Q^1_{ij} = \langle x_i(t)x_j(t+1)\rangle$, which must reproduce the empirical $\widehat{Q}^0_{ij}$ and $\widehat{Q}^1_{ij}$, respectively. In practice, we use the two time shifts 0 and 1 TR, because this gives sufficient information to uniquely infer the network parameters (in the theory). Assuming known network parameters C and Σ, the matrix Q^0 can be calculated by solving the Lyapunov equation (for example using the Bartell-Stewart algorithm):

$$JQ^0 + Q^0 J^\dagger + \Sigma = 0 \ , \tag{2.4}$$

where the superscript $\dagger$ denotes the matrix transpose. For Q^1, it is simply given by

$$Q^1 = Q^0 e^{J^\dagger} \ . \tag{2.5}$$

Here J is the Jacobian of the dynamical system and depends on the time constant τ_x and the network effective connectivity:

$$J_{ij} = \frac{-\delta_{ij}}{\tau_x} + C_{ij} \ , \tag{2.6}$$

where δ_{ij} is the Kronecker delta; note also that $e^{J^\dagger}$ is a matrix exponential.

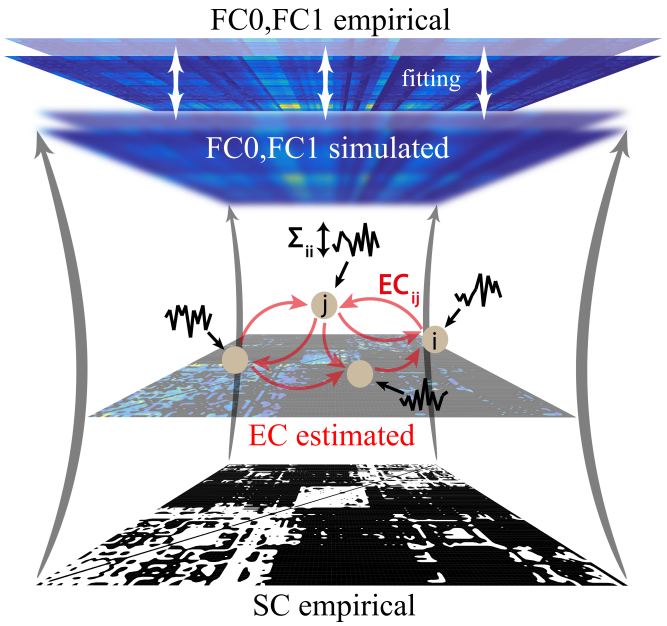

Figure 2.2: Whole-brain network model to extract effective connectivity (EC) from fMRI measurements. The local fluctuating activity (where Σ_i is the variance of the input to each region i) propagates via the recurrent EC to generate the correlation patterns at the network level. Structural connectivity (SC, bottom) obtained using DTI determines the skeleton of EC. The fitting procedure iteratively tunes EC and Σ such that the model best reproduces the empirical FC0 and FC1.

Parameter estimation procedure

In this section, we provide details about the Lyapunov optimization (or gradient descent) which is used to tune the model FC to the empirical FC of a given fMRI session. Although it differs from a maximum-likelihood estimate as classically used for multivariate autoregressive processes, it provides a single estimated value for each model parameter.
For each individual and session, we calculate the time constant τ_x associated with the exponential decay of the autocovariance averaged over all ROIs:

$$\tau_x = \frac{1}{N} \sum_{1 \leq i \leq N} \frac{1}{\log(\widehat{Q}^0_{ii}) - \log(\widehat{Q}^1_{ii})} \tag{2.7}$$

This is used to "calibrate" the model, before its optimization.
In order to invert the model (i.e., for the model FC to reproduce the experimental FC), we iteratively tune the parameters to reduce the model error defined as

$$E = \frac{1}{2} \frac{\sum_{i,j} (\Delta Q^0_{ij})^2}{\sum_{i,j} (\hat{Q}^0_{ij})^2} + \frac{1}{2} \frac{\sum_{i,j} (\Delta Q^1_{ij})^2}{\sum_{i,j} (\hat{Q}^1_{ij})^2} \, . \tag{2.8}$$

Here each term –for FC0 and FC1– is the matrix distance between the model and the data observables, normalized by the norm of the latter; for compactness, we have defined the difference matrices are $\Delta Q^0 = \widehat{Q}^0 - Q^0$ and $\Delta Q^1 = \widehat{Q}^1 - Q^1$.
The idea behind the tuning algorithm is to start from zero connectivity $C = 0$ and homogeneous Σ, then calculate the model Q^0 and Q^1 using the desired Jacobian update given by

$$\Delta J^\dagger = (Q^0)^{-1} [\Delta Q^0 + \Delta Q^1 e^{J^\dagger}] \, , \tag{2.9}$$

which decreases the model error E at each optimization step, similar to a gradient descent. The best fit corresponds to the minimum of E. Finally, the connectivity update is

$$\Delta C_{ij} = \eta_C \Delta J_{ij} \tag{2.10}$$

for existing connections only; other weights are forced at 0. We also impose non-negativity for the EC values during the optimization. To take

properly the effect of cross-correlated inputs into account, we use the Σ update as in [Gilson et al., 2017]:

$$\Delta\Sigma = -\eta_\Sigma(J\Delta Q^0 + \Delta Q^0 J^\dagger)\ . \tag{2.11}$$

As with C for non-existing connections, off-diagonal elements of Σ are kept equal to 0 at all times.
In numerical simulations, we use $\eta_C = 0.0005$ and $\eta_\Sigma = 0.05$. Further details about the derivation of the optimization updates are provided in [Gilson et al., 2016].

Comparison with state-of-the-art dynamic models and Bayesian estimation to interpret fMRI data

Compared to dynamic causal modeling [Friston et al., 2003,Friston, 2011], our model makes the simpler assumption of linearity for the local dynamics. Doing so, it ignores an explicit modeling of the mapping between the neuronal activity and the BOLD signals [Stephan et al., 2004]. Moreover, it uses a simple model of local variability (Wiener process) related to Σ to generate FC than recent development of DCM for resting state [Friston et al., 2014]. In exchange for this simplicity, we obtain a very efficient estimation procedure for networks of about 100 ROIs with 30% density, yielding ~ 3000 EC parameters. It is also worth noting that the objective function for our framework are the BOLD covariances, which are canonically related to the BOLD cross-spectrum used in the DCM for resting state [Friston et al., 2014].

2.2.2. Analysis and vectorization of EC and corrFC

Here, we provide details about the analysis of the corrFC and estimated EC matrices that are compared across sessions for subject and condition identification[2]. As illustrated in Fig. 2.1B, the connectivity measure for

[2]Note that we use a slightly different indexing in this section compared to the previous one: In the following i refers to a link in the vector $v^l = (v^l_i)$, similarly to the pair (i, j) for a matrix element C_{ij} before.

each session l was transformed into a vector v^l by extracting the lower triangle for corrFC or by applying the SC mask for EC. Following the literature in machine learning, we refer to a connectivity measure for a single session as a sample. The size of the samples v^l is $p = 6670$ for an FC session and $p = 4056$ for an EC session with Datasets A and B (corresponding to $N = 116$ ROIs). For Dataset C, we used only EC with $p = 1114$ vector elements (for $N = 66$ ROIs).

Similarity between sessions

We used Pearson correlation coefficient (PCC) as a measure of similarity, both within and between subjects, because is the metric used in the kNN classifier; for a comparison using ICC measure see [Pannunzi et al., 2017]. For a pair of sessions l and m as reported in Fig. 2.4A and B, the similarity is:

$$S^{lm} = \mathrm{PCC}(v^l, v^m) \tag{2.12}$$

Silhouette coefficient

The silhouette coefficient [Rousseeuw, 1987] is defined for each vector $v^{l,s}$ with indices l for the session and s for the subject. Here, each subject s is a cluster and the similarity in Eq. (2.12) is taken as the metric, but the indices l and m are replaced by doublets of the type (l, s) here. For a given sample l, s, we have the average similarity within his own cluster s defined as

$$a^{l,s} = \mathrm{mean}_{m \neq l}(S^{(l,s)(m,s)}) \ , \tag{2.13}$$

and the maximum –over all other clusters– of the same average similarity, but with elements from another cluster:

$$b^{l,s} = \max_{s' \neq s}[\mathrm{mean}_{m'}(S^{(l,s)(m',s')})] \ . \tag{2.14}$$

The silhouette is then given by the following contrast between the cohesion of the element within its cluster ($a^{l,s}$) and the separation from other

clusters ($b^{l,s}$):

$$\sigma^{l,s} = \frac{b^{l,s} - a^{l,s}}{\max(a^{l,s}, b^{l,s})} \ . \tag{2.15}$$

Values of silhouette range from 1 for fully separated clusters to -1 for fully overlapping clusters.

Within-session z-scoring

The values of the EC and corrFC links were z-scored within each session, using the mean and standard deviation of the corresponding vectorized connectivity measure v^l in Fig. 2.1B:

$$\hat{v}_i^l = \frac{v_i^l - \mathrm{mean}_i(v_i^l)}{\mathrm{std}_i(v_i^l)} \ , \tag{2.16}$$

where the vector elements are indexed by i, corresponding to a link in the EC or corrFC matrix. The z-scored vectors $\hat{v}^l$ are the inputs of the classifiers, which means that the classification relies on the ranking of the vectorized elements v^l rather than the absolute values of their elements.

Classification of sessions to attribute them to subjects

Fig. 2.3 shows the classification procedure applied to identify subjects using connectivity measures and estimates. First, a fixed number of sessions per subject were selected from each data set, corrFC and EC matrices. These matrices were vectorized (recall Fig. 2.1B) and individually z-scored using Eq. (2.16), using the mean and standard deviation of each v^l. Then, the corresponding classifier –1-nearest-neighbor (1NN) or multinomial logistic regression (MLR)– was trained using two different approaches. In all cases, the accuracy of a classifier is evaluated by its prediction of samples in the test set.
We used Dataset A1 to study the effect of increasing the samples in the train set, and Dataset B for the effect of increasing the number of subjects. The same method was also applied to Dataset C by training two MLR classifiers, one for subjects and one for conditions. Both classifiers are available in the scikit-learn package [Pedregosa et al., 2011].

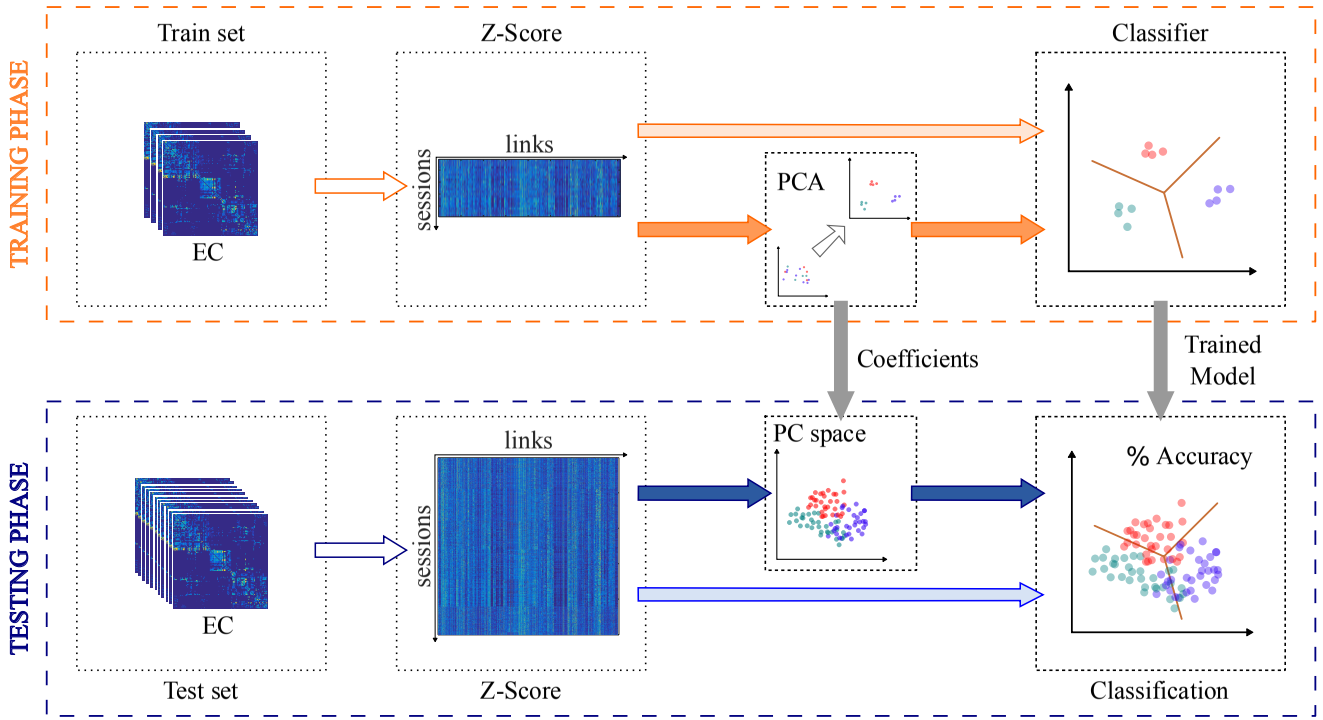

Figure 2.3: Classification pipeline. The full set of connectivity measures (here EC) over all fMRI sessions was split into two groups: a train set and a test set. We use z-scores calculated over the elements of each session matrix; see Eq. (2.16) in Methods for details. We trained the classifier –with or without previously applying PCA– and evaluated the classification accuracy on the test set.

1NN classifier

A kNN classifier is a technique that assigns to a new sample the class to which belong the majority of its k closest neighbors. In our case, we use $k = 1$ with a single nearest neighbor. Moreover, we use the PCC-based similarity measure in Eq. (2.12) as the metric to evaluate the inverse distance between two samples (here sessions). Like with clustering algorithm in general, closest samples (i.e., most similar sessions) are grouped together. Because the PCC similarity is not linear, it can be considered as a non-linear classifier, in comparison to a linear classifier such as a perceptron or a MLR. In practice, the train set has an equal number of sessions per subject –either vectorized and z-scored corrFC or EC– ranging from 1 to 40 for Dataset A1. The identity of the each target session l from the test set is predicted by the identity of the most similar session from the train set $(S^{l,1}, S^{l,2}, \cdots, S^{l,D})$ with D the size of the train set.

MLR classifier

MLR is the canonical tool for high-dimensional linear classification. It tunes a weight for each feature (one per dimension), thus selecting the important ones. The parameters (or regressors) of the model are adjusted in order to predict the probabilities of new samples of belonging to each class (or category). It relies on the following logistic function that relates to the probability for the vectorized connectivity measure v^l (with elements indexed by i) for session l to be in the subject class s:

$$\Pr(v^l \in s) = \phi\Big(\sum_i w_i^s v_i^l\Big) \,, \tag{2.17}$$

where ϕ is the sigmoid function (ranging from 0 to 1). The training is performed by a regression to find the classification weight w_i^s such that $\Pr(v^l \in s)$ discriminates the class s against the last subject s'. A refinement that does not appear in Eq. (2.17) is that there is an extra weight to correct for the possibly non-zero mean of the samples v^l. Note also that there are $M - 1$ regressors for M subjects, such that the weights are well constrained. In practice, we used train sets with equal numbers of sessions per subject.

Extraction of discriminative support networks with RFE

We examined which links strongly contributed to the classification. To do that, we evaluated the contribution of individual links that supported the classification, forming networks of most discriminative elements in EC. In order to extract these support networks, we employed a commonly used method in machine learning, recursive feature elimination (RFE), to rank the links –taken as features– according to their relevance for the classification [Guyon et al., 2002].

For each application of the RFE algorithm, the train set was composed of 90% randomly chosen samples (to capture the full variability of the data) and test set of the 10% remaining samples. After fitting the MLR classifier, RFE removes the link with the smallest classification weight w_i^s in the MLR formula in Eq. (2.17), which measures the contribution of the link to the classification. The removing procedure is repeated recursively on the shrinking subset of links until only one is left. This gives a ranking for the links according to their relevance for the classification. We then evaluated the accuracy of MLR on the test set when increasing the number of features following the order given by the RFE ranking. This training and testing procedure was repeated 100 times with different train and test sets each time. We selected the number of features for which the mean test set accuracy was maximum. In order to find the maximum we chose the number of features for which the numerical derivative of the mean was less than 10^{-6}. In order to reduce the impact of fluctuations due to the random selection of samples, we smoothed the curve of means with a rolling average of width 2 features. Since the accuracy is expected to increase initially as a function of the number of features and then either saturate or decrease, this method allows for finding the number of features for which the classifier performance is maximum or adding more features has no practical benefit.

In contrast, kNN cannot be easily used for RFE since it does not estimate weights associated to links. Therefore, for kNN to be used with RFE, one needs to put the model in a wrapper to compare the effect of removing each combination of links on the performance. However, given the high amount

of features of our setting, wrappers cannot be evaluated on all subsets of features ($\sim 10^{300}$ tests would be required for 1000 features). Wrappers may rely on approximate greedy algorithms, for example, eliminating the feature that scores worst. It is well known that greedy algorithms might produce inappropriate solutions if the problem does not have optimal substructure. In addition the computation time almost scales as p^2 with p the number of links, while for RFE it is linear in p.

Software tools

The computer code for the model optimization and classification is available online with Dataset C at `github.com/MatthieuGilson/WBLEC_toolbox`. It is written in the open-source language python and uses the numpy and scipy[3] libraries, as well as scikit-learn[4] library [Pedregosa et al., 2011] for machine-learning routines.
Connectivity measures and estimates, as well as similarity analyses were performed in MATLAB 2016 (TM). Network plots in Fig. 2.17 were done in Gephi 0.9.1[5].

2.3. Results

2.3.1. Functional and effective connectivity as measures of brain network dynamics

In this study we used fMRI data from the three datasets described in Table 2.1. After applying a standard preprocessing pipeline to the BOLD signals for Datasets A and B, we parcellated the brain into 116 regions of interest (ROIs) according to an anatomical space (see Fig. 2.1A). Likewise, Dataset C was parcellated into 66 ROIs covering the cortex.
Classical functional connectivity (corrFC) was calculated using the pairwise Pearson correlation coefficient (PCC) between the time courses of

[3] `http://www.scipy.org`
[4] `http://scikit-learn.org`
[5] `http://gephi.org`

the N regions of interest (ROIs), obtaining an $N \times N$ symmetric matrix for each recorded session ($N = 116$ for Datasets A and B, $N = 66$ for Dataset C); see Eq. (2.2) in Methods. In parallel, we used the whole-brain dynamic model [Gilson et al., 2016, Gilson et al., 2017] in Fig. 2.2: Each ROI is a node in a noise-diffusion network whose topology (skeleton) is determined by the structural connectivity (SC) obtained from diffusion tensor imaging (DTI) or similar techniques. In the model, the global pattern of FC arises from the local variability Σ_i (for node i, $1 \leq i \leq N$) that propagates via the network connections EC_{ij} (from node j to node i, $1 \leq i \leq N$, $1 \leq j \leq N$, $i \neq j$). To fit each fMRI session, all relevant EC_{ij} and Σ_i parameters are iteratively tuned such that the model spatio-temporal FC –as measured by FC0 (no temporal lag) and FC1 (covariance with temporal lag of 1 TR), see Eq. (2.1) in Methods– best reproduces the empirical counterpart. A detailed description of the model and the estimation procedure is provided in Methods. In essence, the model inversion decomposes the empirical matrices (FC0, FC1) into two estimates EC and Σ, which can be seen as multivariate biomarkers for the brain dynamics in each fMRI session.

2.3.2. Structure of individual session-to-session variability for EC and FC

Before considering classification, we examined the distributions of connectivity measures extracted from repeated scans for individual subjects. Using Datasets A and B, we compared the capability of the four connectivity measures (corrFC, FC0, FC1 and EC), as well as Σ, in terms of within- and between-subject similarity (WSS and BSS, respectively), as a first step toward subject identification. For each pair of sessions, the similarity S_X was calculated using the PCC between two vectorized connectivity measures X in Fig. 2.1B (non-zero elements for EC, low-triangle elements for corrFC).

In the matrix of S_{EC} values for Dataset A1 (Fig. 2.4A), six diagonal blocks with larger values corresponding to the WSS can be noticed; the remaining matrix elements correspond to BSS. Fig. 2.4B compares the distributions

of S_{EC} and S_{corrFC}. The distribution of WSS was obtained by using all pairs of vectors v^l and v^m with $l \neq m$ from the same subject in Dataset A1. To compute the distribution of BSS, all possible combinations of vector pairs v^l and v^m from distinct subjects were used separately in Dataset A1 and A2.

WSS and BSS distributions are better separated for EC than for corrFC. In other words, sessions from the same subject are more similar to one another, and more different from those of other subjects, viewed from the EC than the corrFC perspective. This suggests a better capability for EC to discriminate between subjects. Note that the BSS from Datasets A1 (6 subjects) and A2 (50 subjects) remarkably overlap for both corrFC and EC, showing that BSS for six subjects generalizes well to larger numbers. These qualitative observations are confirmed by Table 2.2 that summarizes the Kolmogorov-Smirnov (KS) distance between the similarity distributions (blue versus red and blue versus green in Fig. 2.4B). We used the KS distance between the distributions because the number of WSS and BSS values were different. EC gave larger KS distance than corrFC and all other FC-related measures. Note that we also calculated KS distance using only the links in corrFC and FC0 corresponding to the 4056 existing connections in EC (determined by SC), in order to compensate for the relative sparsity of EC links as compared to corrFC and FC0; however, this did not change the results (Fig. 2.5). Note that our version of vectorized FC0 does not involve variances (diagonal elements). In contrast, the calculation of each element corrFC values incorporates the variances, describing the ratio of covariance with respect to the variances of the two nodes; see Eq. (2.2). This appears to enhance the information content of corrFC compared to FC0. Lastly, the diagonal elements of Σ showed the smallest distances. In the following, we focus on EC and corrFC. Fig. 2.5) shows the similarity distributions for FC0, FC1, corrFC/SC and Σ using Datasets A1 and A2.

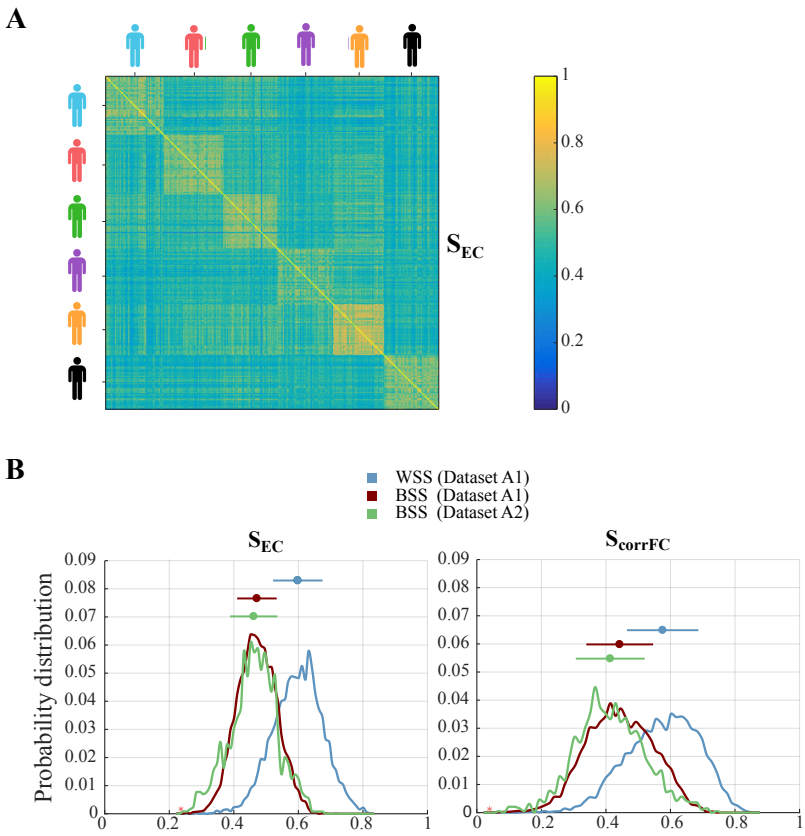

Figure 2.4: Within- and between-subject similarity (WSS and BSS, respectively) for EC and corrFC. A) Matrix of similarity values S_{EC} between all pairs of sessions from Dataset A1. EC matrices from all sessions are transformed into vectors (see Fig. 2.1B), from which the PCC is calculated to obtain S_{EC} for all pairs of sessions (see Eq. (2.12) in Methods). Here the sessions are grouped by subjects, as indicated by the colored symbols. **B)** Distributions of WSS (blue) and BSS (red) of S_{EC} values for Datasets A1 –corresponding to diagonal and off-diagonal blocks in panel **A)**, respectively– and of BSS (green) for Dataset A2. The right panel shows the corresponding distributions for corrFC. The above error bars indicate the means and standard deviations, indicating a smaller overlap between WSS and BSS for EC.

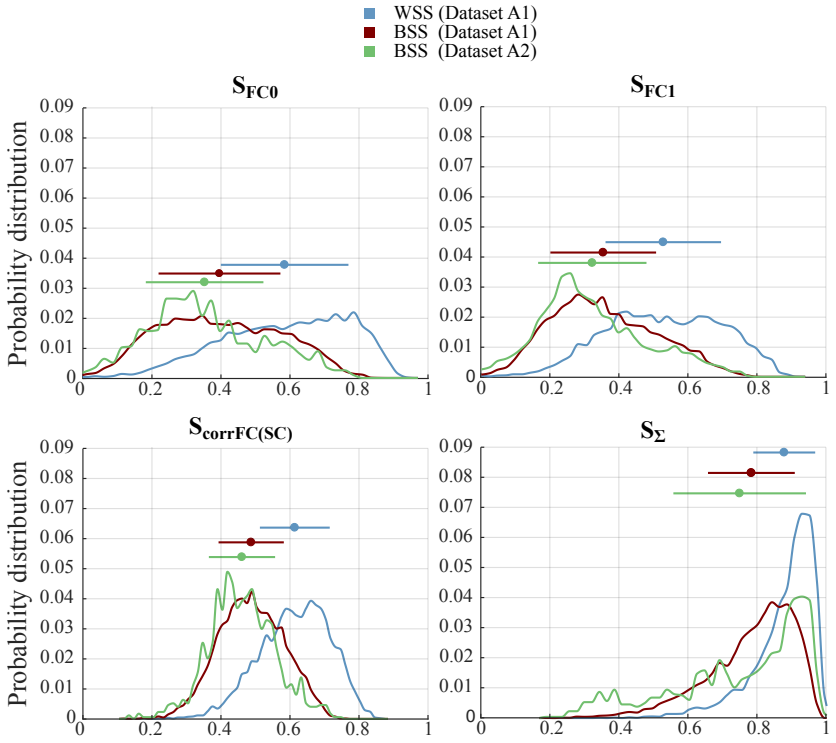

Figure 2.5: Within- and between-subject similarity (WSS and BSS, respectively) for FC0, FC1, Σ and corrFC. Upper panels show the similarity distribution for FC0 (left) and FC1 (right). Bottom left panel shows the similarity distribution of corrFC after extracting those links determined by SC. The obtained result is very similar to the right plot in Fig. 2.4B indicating that the effect noticed in the distributions of S_{EC} is not caused by the SC mask. Bottom right panel shows the similarity distributions of Σ.

Table 2.2: Kolmogorov-Smirnov (KS) distance between WSS and BSS distributions. Comparison of the connectivity measures (corrFC, FC0, FC1) and model estimates (EC, Σ). The third column corresponds to the distance between the blue and red distributions in Fig. 2.4B and Fig. 2.5, and the fourth column to the blue and green distributions.

Measure	Description	Dataset A1	Dataset A1 & A2
EC	EC weights estimated with the model	0.6440	0.6581
corrFC	FC computed as Pearson correlation	0.4517	0.5477
FC0	FC computed as 0-lag covariance matrix	0.3685	0.4888
FC1	FC computed as 1-lag covariance matrix	0.4139	0.5118
corrFC/SC	FC as Pearson correlation and masked with SC	0.4769	0.5729
FC0/SC	FC as 0-lag covariance matrix and masked with SC	0.3770	0.4644
Σ	Local variability for each ROI estimated with the model	0.3778	0.3287

Dimensionality analysis

The high dimensionality of our connectivity measures may reduce their predictive power, an effect known as the Hughes phenomenon [Hughes, 1968]. This is especially important in our case where the number p of dimensions for the multivariate measures (Dataset A1: $p = 4056$ for EC, $p = 6670$ for corrFC) exceeds the number n of samples ($n = 6$ subjects $\times$ 50 sessions = 300 samples). To study visually how the variability of the data is spread over in the space with high dimension p, we applied principal component analysis (PCA) to extract the main directions (principal components, or PCs) that capture the largest portion of the data variance.

PCA was applied to the whole Dataset A1 (6 subjects, 40-50 sessions per subject) and the first 6 PCs were retained. Fig. 2.6 compares corrFC and EC: each panel corresponds to PC1 to PC3 on the one hand, and PC4 to PC6 on the other hand. To the naked eye, the colored clouds representing all sessions for each subject exhibit smaller overlap for EC than corrFC.

We quantified the clustering degree of these clouds using a silhouette coefficient for each session, ranging from -1 for poor clustering to 1 for perfect clustering [Rousseeuw, 1987]; see Section 2.2.2 in Methods. In Fig. 2.7A, the silhouette coefficient was computed for each point in these clouds in a six-dimensional PC space for Dataset A1.

As shown in Fig. 2.7A, EC produced larger (almost all positive) silhouette values than corrFC, confirming the visual impression of Fig. 2.6. Silhouette coefficients were calculated on the data projected onto the first six PCs, i.e. the number of PCs that maximized the silhouette coefficient (Fig. 2.8). As can be seen in Fig. 2.7B, the silhouette coefficients for the data in the original link space (left violin plots) are smaller than those for data in the PCs space (right); for Dataset B (in blue), the first 30 PCs are used. Consequently, PCA may facilitate the identification of subjects by reducing the dimensionality of the data, but this has to be properly tested.

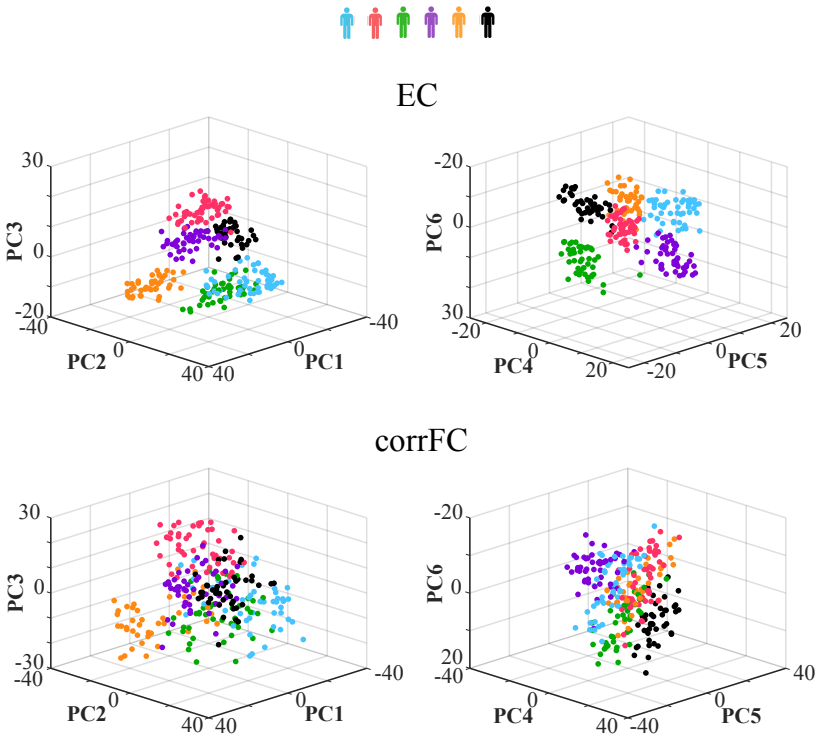

Figure 2.6: Projection of Dataset A1 onto the first 6 PCs. Visualization of the sessions of Dataset A1 in the space of the first six principal components, or PCs (split into the left and right panels), obtained from PCA applied to EC (top row) and corrFC (bottom row) measures. Each point corresponds to a session and each color to one of the 6 subjects.

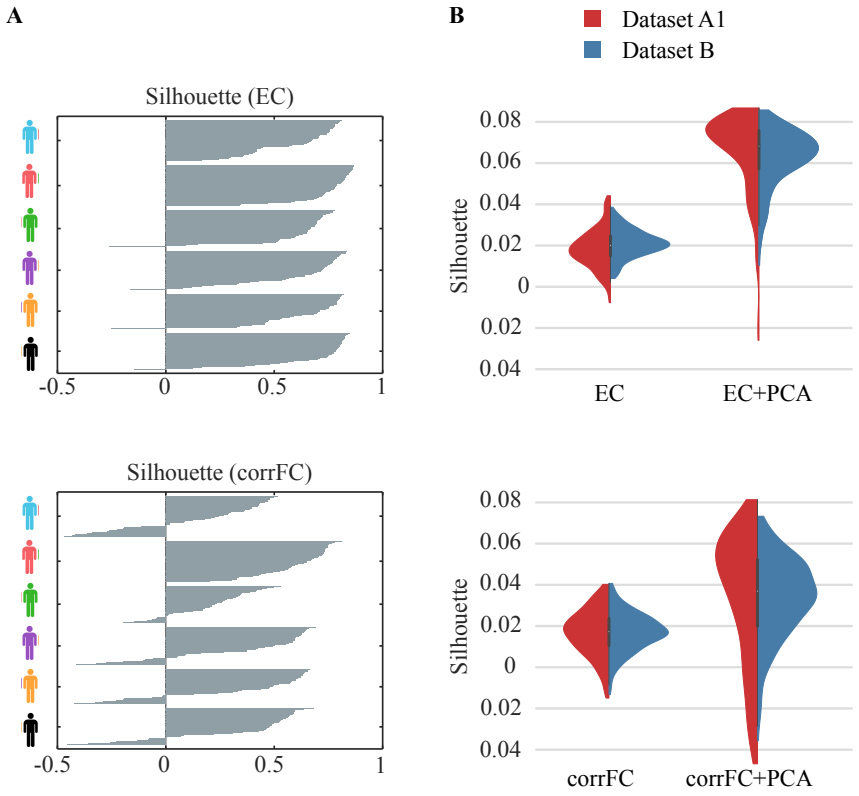

Figure 2.7: Silhouette coefficients obtained from Datasets A1 and B. **A)** Silhouette coefficients of each session in Fig. 2.6. Sessions with value close to 1 are well clustered and those close to -1 are poorly clustered; see Eq. (2.15) in Methods for further details. **B)** Distribution of the silhouette coefficients for EC (top panel) and corrFC (bottom panel): comparison between the original link space (left) and the PCA space. Both Datasets A1 (6 subjects with 6 PCs, in red) and B (30 subjects with 30 PCs, in blue) are represented by the violin plots; see also Fig. 2.8 about the choice for the number of PCs. Note the larger silhouette coefficients for EC than for corrFC.

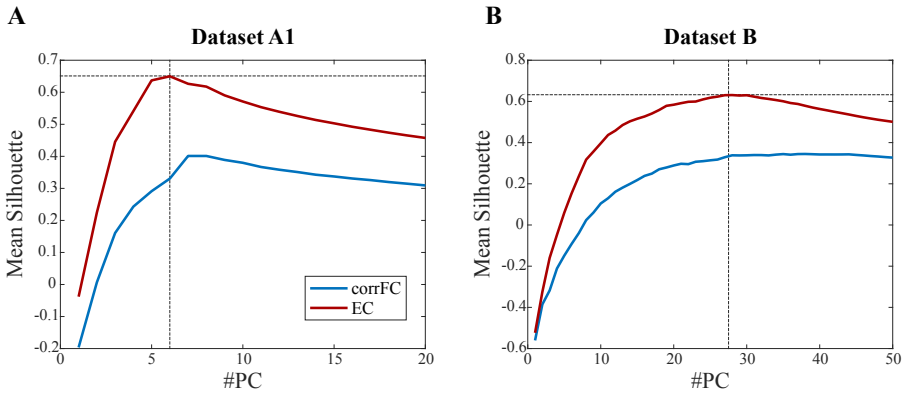

Figure 2.8: Average silhouette over number of principal components used. A) The silhouette reaches a maximum for EC and 6 PCs on Dataset A1. **B)** Mean silhouette coefficient for EC (in red) and corrFC (in blue) measures. The maximum is between 28-30 PC. Note that both numbers are practically equal to the number of subjects in each dataset.

2.3.3. Subject identification using EC is more robust than using FC

Now we turn to the main goal of our study: the classification of single sessions based on EC or corrFC. In this section, we start with attributing sessions to subjects. Subject identification for a large cohort (~ 100 subjects) was recently pioneered [Finn et al., 2015], relying on a k-nearest-neighbor (kNN) classifier with $k = 1$ and PCC as metric. In order to classify a target session, the PCC between the target and one known reference session for each subject (called database[6]) is calculated; the predicted identity for the target is that of the subject corresponding to the closest (most similar) session (see Methods for details). In contrast with previous studies using 1NN [Finn et al., 2015, Finn et al., 2017, Kaufmann et al., 2017], our method relies on a multinomial logistic regression (MLR) classifier, a classical tool in machine learning. MLR uses a linear model to predict

[6]The term *database* is equivalent here to train set

the probability that an input sample belongs to a class (subject here). A technical comparison of both approaches is further detailed in Methods (Section 2.2.2).
In classification algorithms the problem of overfitting describes the situation where the algorithm performs very well with the data it is trained with, but fails to generalize to new samples. Due to the high dimensionality of the connectivity measures [Hughes, 1968], it is essential to control for overfitting with an appropriate training and test procedure. Our train-test procedure and the use of large test-retest datasets –unlike previous studies [Finn et al., 2015, Finn et al., 2017, Vanderwal et al., 2017]– aims to improve the understanding of the classification performance by considering other classifiers and larger train-test sets. Fig. 2.3 describes the train-test general procedure for the identification of subjects:

- fMRI sessions (EC in the figure) are randomly split in train and test datasets;
- after preprocessing (orange arrows) involving within-session z-score (to obtain a normalized ranking of the EC/corrFC vector values, see Section 2.2.2 in Methods for details);
- when PCA is used, the high-dimensional z-scored vectors are projected into a space of lower dimension determined by a number of PCs (others are discarded);
- the classifier is then optimized, as illustrated for the MLR with boundaries that best predict the training dataset;
- the test set is used to verify the generalization capability of the classifier (blue arrows), by measuring to which extent the classifier boundaries, estimated with the train set, correctly classify single sessions from the test set.

We first used Dataset A1 and increased the number of training sessions per subject from 1 to 40 to evaluate how many training sessions are necessary for satisfactory accuracy. As shown in Fig. 2.9A, EC (in red) outperformed corrFC (in blue) by more than one standard deviation (shaded

area around the curve), for both MLR and 1NN. For both classifiers, the Mann-Whitney test over 100 repetitions gives a p-value $< 10^{-11}$ until 10 sessions for the superiority of EC. Moreover, almost perfect classification was reached with MLR for only 5 training sessions, whereas 10-15 were necessary for 1NN. This is important when only a few training sessions per subject are available, as expected with clinical applications. Fig. 2.9B displays the classification accuracy for Dataset B, used to verify the robustness with respect to the number of subjects to be classified. We trained the classifiers with 1 session per subject and evaluated the performance varying the number of subjects from 2 to 30 (test set comprised the remaining 9 sessions per subject). Again, EC is more robust than corrFC: While performance with corrFC rapidly deteriorates as the number of subjects is increased, classification using EC is barely affected by the number of subjects. Again, the Mann-Whitney test over 100 repetitions gives a p-value $< 10^{-5}$ in all cases (more than two subjects). This is our core technical result: EC and MLR largely outperform corrFC and 1NN, respectively. Other connectivity measures such as FC0 showed similar performance to corrFC (not shown).

PCA is a preprocessing step commonly used in machine learning to remove noise while keeping the dimensions that capture most of the variability of the data. Using it assumes that the largest part of the data variances captures the relevant information for the classification. Unlike with the analysis of the data structure (Fig. 2.7B), PCA only marginally increased the classification performance here (not shown). The difference with Fig. 2.7B may be explained because here PCA can only be applied to the train set. Another possible reason is that the classification accuracy is already very high for MLR and EC with 1 training session. As shown in Fig. 2.8, subject-specific information is conveyed by many PCs, corresponding to a rather large subspace of the EC links. This supports the use of proper machine learning tools to extract this distributed information.

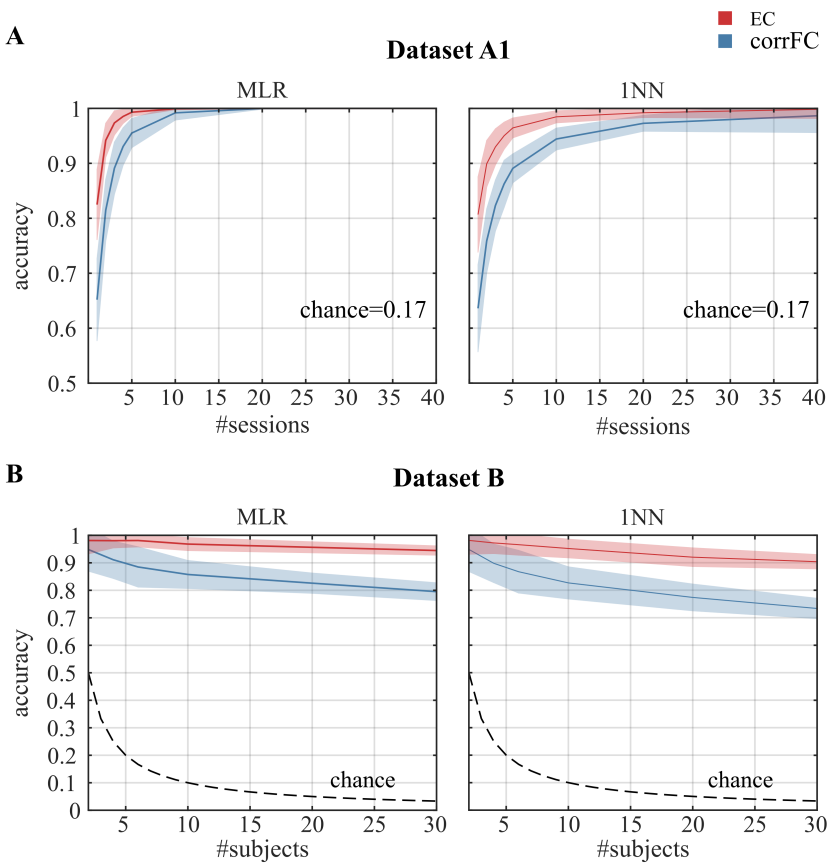

Figure 2.9: Comparison between EC and FC in subject identification. A) Performance of multinomial logistic regression (MLR, left panel) and 1-nearest-neighbor (1NN, right panel) classifiers when increasing the number of sessions per subject used as training set with Dataset A1. The mean (solid curve) and standard deviation (colored area) were calculated for 100 folds of randomly-split train-test sets (the number of sessions per subject for the training set is indicated by the value of the x-axis; the test set comprised all remaining sessions). **B)** Same as A when varying the number of subjects using Dataset B, using a single training session per subject (leaving 9 sessions per subject as test).

2.3.4. Signature network of links supporting the classification

Another important advantage of MLR over kNN is its efficiency in characterizing the links that contribute to the classification. This can be used to extract signatures related to a given modality (subject or condition), as we show now for subject identification. In contrast to the previous section were all links are used as features in the classification, recursive feature elimination (RFE) can be used to rank the links according to their weight in the classification, giving the lowest number of links for which MLR achieved the maximum classification performance; see Methods for details (Section 2.2.2). As mentioned before, RFE is particularly adapted to the MLR classifier: the same procedure with kNN would require many more computations, recalculating the closest neighbor for all combinations of links (here the number of links is $p > 1000$; see Methods for further arguments). A last technical point here is that RFE is used without cross-validation, using the full dataset as train set. The support network obtained using RFE on Dataset A1 had 18 links, compared to 44 links for Dataset B. In both cases, subject identification using only those links achieved perfect accuracy (90% of all available sessions were used for training and 10% for testing). The two support networks are shown in Fig. 2.10 in the same matrix: remarkably, the networks are very sparse and non-uniformly distributed across the whole brain. This is the signature of the most subject-discriminative ROIs: Frontal and cingulate cortices, as well as the temporal and occipital regions, seem to play a major role here. It is worth noting that the EC adjacency matrix is not symmetric, which implies different roles for nodes as receivers (especially frontal ROIs) or senders (cingulate).

The sparsity of the signature in Fig. 2.10 hides the fact that the rankings for Datasets A1 and B are close (PCC=0.59, p-value $\ll 10^{-50}$), indicating that similar links convey information about the identity of subjects. Fig. 2.11 shows the remarkable correspondence of the mean RFE ranking of EC connections for Datasets A1 and B, where the source and target ROIs are grouped anatomically (in group labels in color in Fig. 2.10;

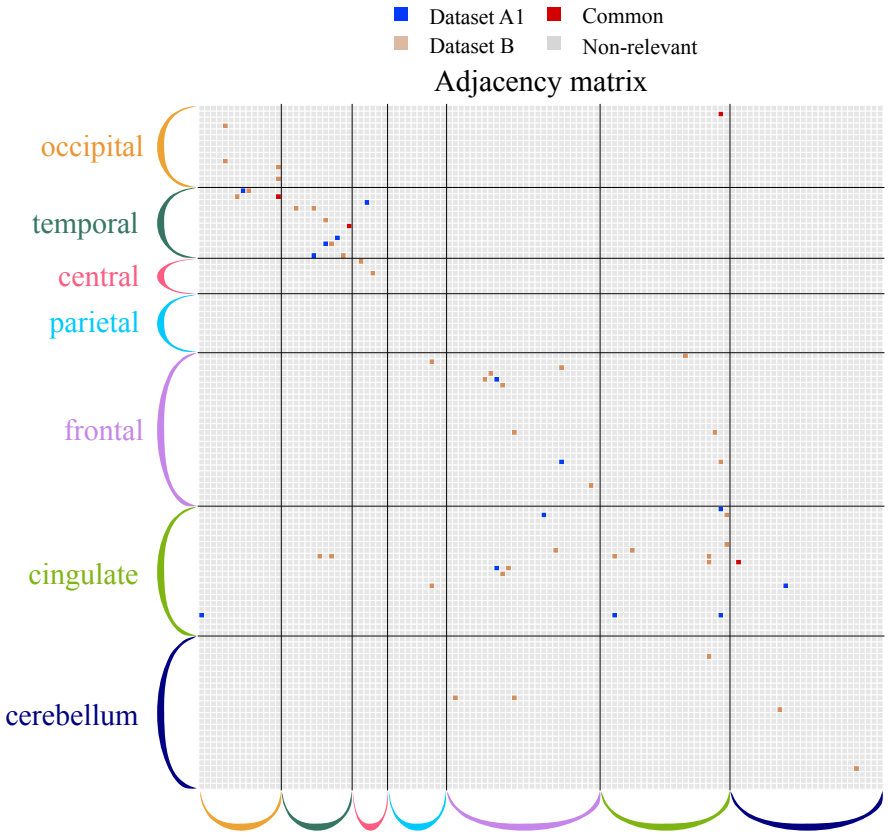

Figure 2.10: Signature of EC links to identify subjects. Extracted links that contribute to the classification for Datasets A1 and B, obtained using recursive feature elimination (RFE). The ROIs are grouped in anatomical pools, as detailed in Appendix B.

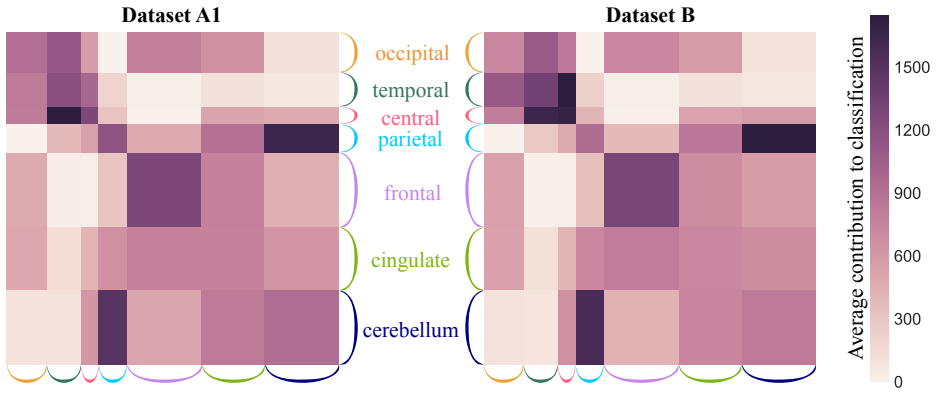

Figure 2.11: Average RFE ranking over the ROI anatomical pools. Darker color indicates more important links for the classification. Subject-specific extracted networks are similar for Datasets A1 and B.

see Appendix B for details). This confirms that the two corresponding signatures coincide at the subnetwork level. Compared to previous studies that based classification on parietal and frontal regions [Finn et al., 2015], we found that temporal and occipital regions are discriminative. An explanation for this difference may lie in that EC consists of directional links and may capture more refined information than FC.

To further measure the overlap between these networks, we selected the subset of links with the highest ranking for each dataset and computed the number of common links. Fig. 2.12A shows that the proportion of common links exceeds by far its expectation under the hypothesis of random rankings (shaded gray area). This indicates a good agreement between the support networks from the two datasets even at the single-link level.

Expecting the size of the support network in Fig. 2.10 to grow with the number of subjects, we quantified it using Dataset B and found a sublinear relationship, as illustrated in the log-log plot of Fig. 2.13A. The extrapolation to 100 and 1000 subjects gave network sizes of 53 and 200 links, the latter corresponding to 5% of the total number of EC links (i.e., the signature remained sparse). Likewise, we fitted a curve to predict the EC-based

accuracy when increasing the number of subjects (see Fig. 2.13B), which gave 92% and 88% for 100 and 1000 subjects. These results implied that the complexity of EC biomarkers related to individual variability should not "explode", which is good news for the generalization capabilities of EC.

2.3.5. Twofold classification of subject identity and behavioral condition

Using another dataset, we applied the knowledge gained from the subject identification to reach our final objective, the extraction of several signatures corresponding to distinct modalities and the evaluation of the overlap between them. From sessions in Dataset C, we extracted a signature for the subject identity and another for the behavioral/cognitive condition. The twofold classification was schematically depicted in Fig. 1.3 with three fictive dimensions: The information about subject identity corresponds to the horizontal axis and information about the condition to the vertical axis; The session-to-session variability, which should be ignored, spreads along the depth axis. In this idealized scenario, it is possible to classify a session with respect to both subjects and conditions using different dimensions of the data (Fig. 2.14A). In the high dimensional case, this occurs when different sets of links support the two classifications. In contrast, the two signatures may involve overlapping subsets of links, corresponding to non-orthogonal axes (Fig. 2.14B) for which the condition "score" for classification (projection on axis) also involves individual information.

MLR with EC achieved very high performance (accuracy $> 90\%$) for subject identification and perfect classification for the condition. For condition classification the samples are grouped, i.e. some fMRI sessions for each condition come from the same subject, so predicting the class for an out-of-sample subject is in principle more difficult than for subjects present in the training set [Poldrack et al., 2009]. In our case we found no difference between standard random permutation cross-validation and leave-one-group-out cross-validation: both provide perfect classification

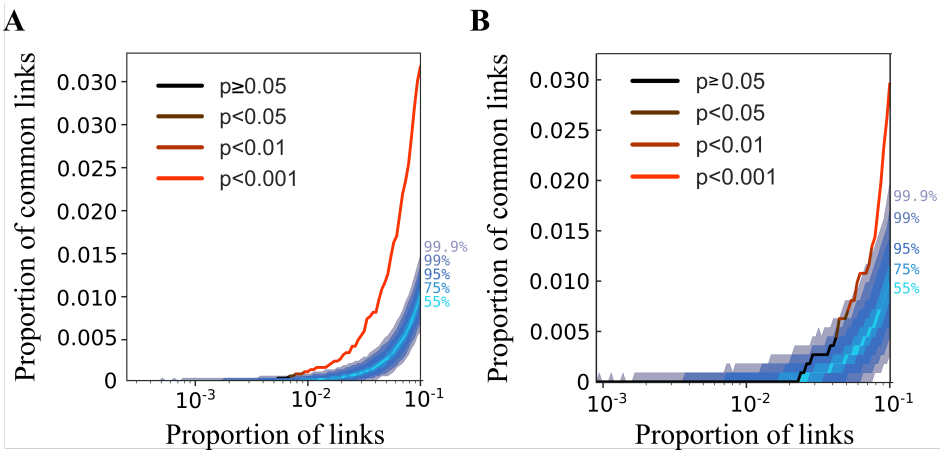

Figure 2.12: Proportion of common links between signatures. A) Overlap between the two signatures for Datasets A1 and B as a function of selected links. The curve represents the amount of common links in the data. Shaded areas represent different quantiles of the surrogate distribution of common links under the null-hypothesis of random rankings. The color of the curve indicates the probability of the corresponding amount of common links under the null-hypothesis (here p-value < 0.001 when considering more than 1% of the total links, namely 40 links). **B)** Proportion of common links between the subject and condition signatures as a function of selected links for Dataset C (in the order of the RFE ranking). The two signatures are significantly different, i.e., with a number of common links corresponding to the null hypothesis with p-value ≥ 0.05 (cf. legend) with up to 4% of the total links.

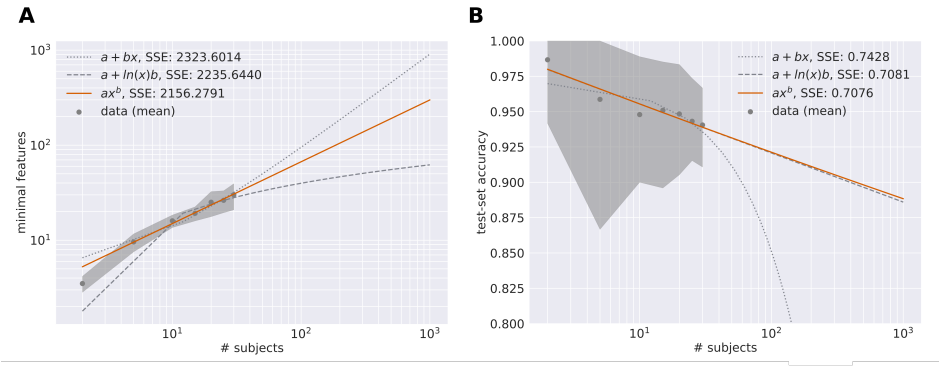

Figure 2.13: Extrapolation of the size of the support network and accuracy. A) Extrapolation of the size of the support network in Fig. 2.10 for Dataset B when increasing the number of subjects up to 1000 (x-axis), for 10 repetitions. The shaded area corresponds to 10 repetitions. The curves compare three approximations, the best one displayed in orange (as indicated by the fitting error SSE) corresponds to a sublinear power law (exponent equal 0.6). **B)** Extrapolation of the accuracy based on EC. The plotted data points correspond to EC in Fig. 2.9B and the variability to 100 repetitions. The best approximation also corresponds to a power law (in orange), which is very close to a lognormal relationship.

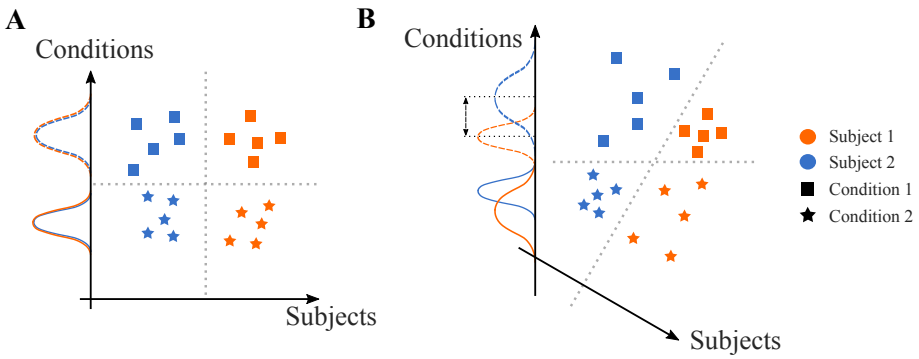

Figure 2.14: Projection of samples onto orthogonal or non-orthogonal subspaces. Different schemes for the the twofold classification where the biomarkers for all sessions (colored symbols) are projected onto two subspaces, one for subjects (orange versus blue) and one for conditions (squares versus stars). The two panels compare the situations where the subspaces (here axes) are orthogonal **(A)** or not **(B)**, respectively. In both cases, the dashed separatrices correspond to perfect classification. In the right panel, however, the condition "score" (projected value) mixes information about subjects, as indicated by the non-overlapping solid and dashed distributions.

of the conditions. We then sought the smallest subsets of links that achieved the maximum performance of each classification using RFE (Fig. 2.16), as done before (see Methods for details). Both support networks were again very sparse and distributed across the brain, as can be seen in their adjacency matrix (Fig. 2.15). More links are necessary to identify the subjects (57) than the behavioral conditions (13), indicating a higher complexity for the former.

To quantify the overlap between the signatures, we computed the number of common links supporting both subject and condition identifications, when increasing the two subsets according to the order given by RFE as before in Fig. 2.12A. The overlap fell within the expected values for the null hypothesis up to 30% of all links, as illustrated in Fig. 2.12B. This indicated that distinct subsets of links were relevant for the subject identities and behavioral conditions, ensuring the "orthogonality" of the two signatures for this dataset. In addition, the few links in common could be removed and replaced by further links in the RFE ranking to further disentangle the signatures. In any case, this suggested a good generalization property of the multivariate signature extraction for several modalities, in the sense that disjoints subsets of links can support the classifications. This application further demonstrated the above mentioned advantage of the MLR classifier equipped with RFE: Working in the original space of links allows us to identify the most discriminative links individually, as well as combinations of them in a subsequent step. In contrast, PCA-like approaches gives PCs corresponding to linear combinations of many links, making it difficult to interpret results at the level of single links.

Last, we compared the topological properties of the support networks in Fig. 2.15, interpreting the most discriminative EC links extracted using the MLR with RFE as graph. These directed networks are represented in the top panels of Fig. 2.17 (A for subject and B for condition). The 66 ROIs are colored to indicate the brain subsystem they belong to and the directed nature of links can be appreciated by zooming. Similar to Datasets A1 and B, subject identification of Dataset C largely concerned the frontal and cingulate systems. Condition identification was also supported by occipital and temporal cortices, which were found to have the

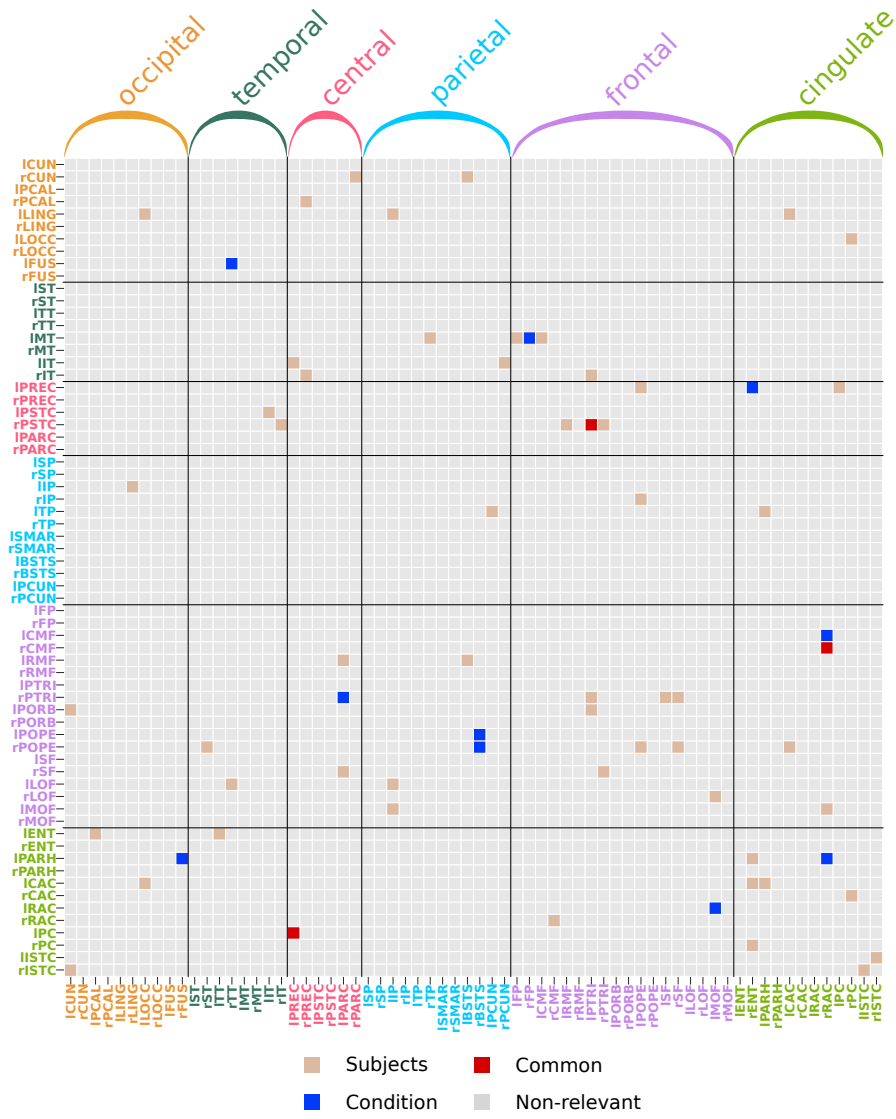

Figure 2.15: Subject and condition signatures extracted using EC. Signatures of the most discriminative EC links (estimated with RFE, see Methods for details) for the twofold classification of Dataset C: 54 links for subject classification in brown, 10 for condition classification in blue, 3 common links in red. The ROIs are grouped in anatomical pools, as detailed in Appendix B.

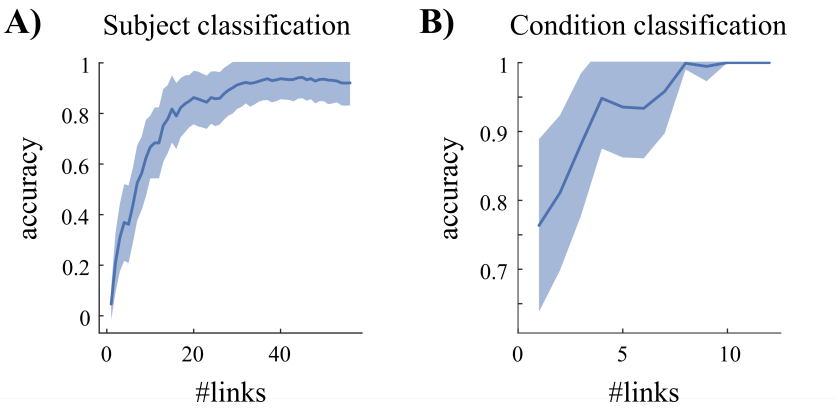

Figure 2.16: Classification subjects and conditions using Dataset C. Performance of the classification for 19 subjects **(A)** and 2 conditions **(B)** using Dataset C as a function of number of links. Note the distinct scales for the y-axis, because the subject identification is a harder problem.

strongest activity modulations during movie viewing [Gilson et al., 2017]. Because we only considered two tasks, it was somehow expected that the task network would be much simpler than the support network for the 19 subjects. However, these two support networks also corresponded to distinct types of graphs: The subject network had a large connected component with several central nodes (hubs, indicated by their large size), located in frontal and cingulate regions, in addition to two small components. In comparison, the condition network was segregated into 8 small isolated components. This suggested that elaborate collective dynamics involving high-level ROIs differed across subjects, which might be used as a biomarker to examine its relationship with individual cognitive traits, as was done with FC [Finn et al., 2015].

The bottom plots in Fig. 2.17 show the lateralization of the support links, stressing the asymmetries between the two hemispheres: Most of the important links are ipsilateral (i.e., within the same hemisphere) and many belong to the left hemisphere for the subject network, whereas they are mainly contralateral for the condition network. This is also in line with strong inter-hemispheric interactions observed during movie viewing for

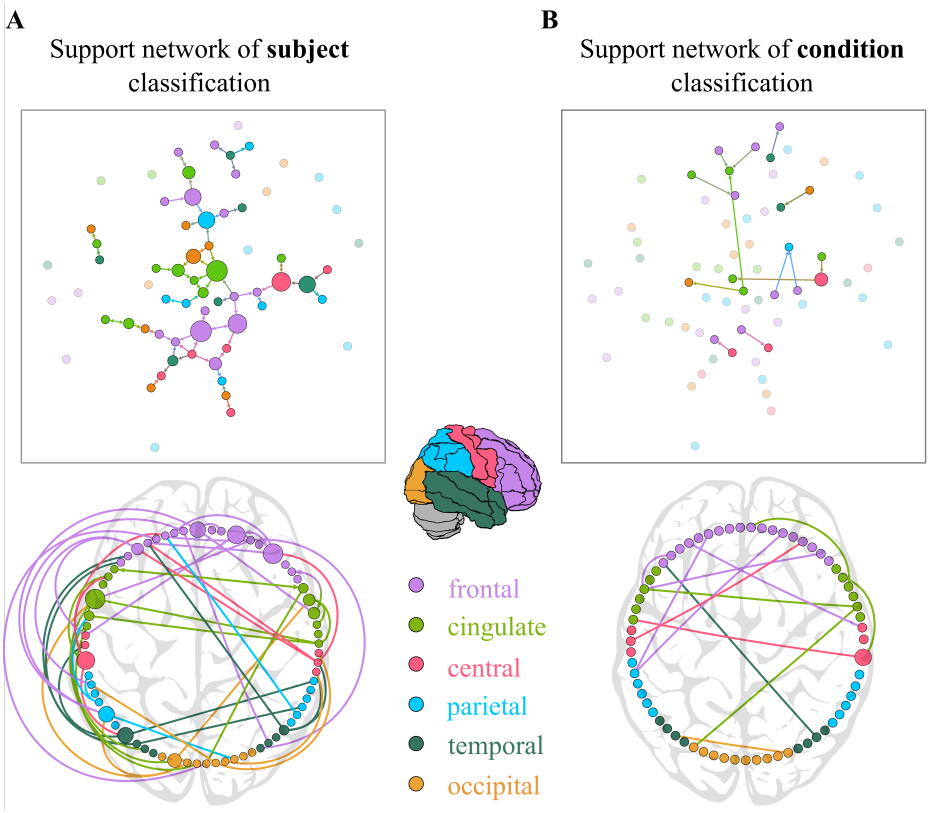

Figure 2.17: Support networks of subject and condition. A) The top graph plot represents the 66 ROIs or nodes and the 57 most discriminative EC links supporting the classification of subjects (same as in Fig. 2.15). The color of the circles indicates the corresponding anatomical brain; it is dimmed for nodes with no connection in the extracted networks. The size of each node represents its betweenness centrality in the extracted network, with most central regions being located in the frontal and cingulate cortices. The bottom circular plot shows the asymmetry and lateralization of the network. Links that are inside the circle correspond to contralateral connections, while links outside the circle correspond to ipsilateral connections. The link direction is indicated by the color of the source node (i.e. a blue line indicates an interaction from a blue node to another node). **B)** Similar graph and circular plots as A for the 13 links supporting the classification between the two conditions (rest versus movie viewing).

the same dataset, but relying on community analysis [Gilson et al., 2017]. This shows that the non-overlap between the signatures is not only quantitative, but also qualitative in the sense that these signatures involve distinct types of links.

2.4. Discussion

In this study, we have proposed a framework to extract signatures related to various modalities from fMRI data, allowing for a quantification of the interference between signatures. The twofold classification of subjects and behavioral conditions goes beyond previous studies that used FC as a "fingerprint" [Miranda-Dominguez et al., 2014, Finn et al., 2015, Calhoun et al., 2017, Finn et al., 2017]. It is precisely because fMRI signals are dominated to some extent by individual traits [Finn et al., 2015] that a proper multivariate classification is necessary. In particular, we have shown how machine learning tools such as the MLR enable a quantification of the topological orthogonality between signatures, in addition to efficiently extracting them. Our study also demonstrates that signatures based on EC are very robust to session-to-session variability (much more than with FC) and can be obtained relying on a limited number of sessions, namely 4-5 recording sessions (of 5 minutes each) to classify 40+ other sessions. This corresponds to a train-test ratio of 10-90, which is unusual low; in contrast, most studies use ratio above 50-50, which raises concerns about generalization capabilities [Varoquaux et al., 2017]. Our method yields consistent results for fMRI datasets collected using different scanners and preprocessed with different pipelines, which further supports its general scope. Taken together, our results define a solid ground to scale up the multivariate classification to larger datasets with more subjects and tasks. We now discuss specific points.

We found that very sparse signatures (a few % of all EC links) are sufficient to obtain perfect classification because of the datasets we used, which involved 30 subjects maximum and 2 conditions. The size of the corresponding support networks (Figs. 2.10) is expected to increase with

the complexity of the "environment" to be represented, with more subjects and tasks. Open-access resources are becoming available to quantitatively test the framework on larger scales [Zuo et al., 2014, Gordon et al., 2017]. An encouraging result for the generalization capability of EC is the sublinear dependency of the signature size with the subject number (Fig. 2.13A). The sparsity of signatures means that the complexity of the variability of EC biomarkers remains "controlled" in the sense that it can be described using a number of dimensions much smaller than the numbers of elements in the corresponding modality (here subjects or conditions).

The support networks for the twofold classification (subject and condition) show several noticeable differences (Fig. 2.15). The subject network is large, almost fully connected, distributed over the two hemispheres (with more links within the left one) and concentrated in the cingulate and frontal areas. This suggests subject-specific dynamics between areas involved in high-level functions and overlapping with the default mode network [Raichle et al., 2001]. These discriminative EC patterns may reflect heterogeneities in the interactions between the different neural subsystems (e.g., cingulate to frontal in Fig. 2.15) and the propagation of information between them [Ekstrom, 2010, Engel et al., 2013]. As expected with the movie viewing condition studied here, links in the visual and temporal areas are discriminative. In addition, we also found a much higher percentage of contralateral links for condition than subject. This suggests that EC-based biomarkers may also be interpreted in terms of brain communication as was shown recently [Gilson et al., 2017, Senden et al., 2017], beyond simply supporting the classification.

A fundamental advancement of our study is the development of a reliable and well-benchmarked method, extending previous proofs of concept for subject or condition identification [Poldrack et al., 2009, Miranda-Dominguez et al., 2014, Finn et al., 2015, Gonzalez-Castillo et al., 2015]. Our core technical result is that (whole-brain) EC discriminates subjects better than corrFC (Fig. 2.7 and Fig. 2.9), which is used in the vast majority of previous studies [Finn et al., 2015, Finn et al., 2017, Kaufmann et al., 2017, Calhoun et al., 2017, Varoquaux et al., 2017, Woo et al., 2017, Xie et al., 2017, Drysdale et al., 2017]. The generalization capability of EC is

much more robust than FC when the classification becomes harder (many subjects to identify with few sessions per subject, Fig. 2.9). This confirms that the BOLD temporal structure –captured by the EC after the bandpass filtering of the BOLD signals– reflects the identity of the subject [Miranda-Dominguez et al., 2014], as previously shown for a task involving attention (or not) [He, 2011] or for wake versus sleep [Mitra et al., 2015]. The use of z-scores in the classification shows that the EC ranking (i.e., which brain connections have large weights among all) conveys the relevant information. This is also in line with our previous studies [Gilson et al., 2017, Senden et al., 2017] that showed how changes in task-evoked fMRI activity is captured by whole-brain EC. Moreover, the directed nature of EC reflects the propagation of BOLD signals and can be interpreted in terms of brain communication. Our results thus support a change in biomarkers used for multivariate classification, where EC unfolds the relevant information of BOLD signals in a suitable space. Here the focus was on EC because it performed better than Σ estimates for the resting-state fMRI, but it has been shown that Σ may be strongly affected by tasks [Gilson et al., 2017], so Σ might further improve the classification for conditions, reflecting sensory stimuli.

On the machine learning side, our results show that the MLR classifier performs better than the 1NN (Fig. 2.9), even though the accuracy is good for both. Note that this also holds for kNN with $k \geq 2$ (results not shown). This suggests that the EC/FC pools related to subjects/conditions are linearly separable in their high dimensional spaces, in a even easier manner for EC (related to the better performance). Unlike the kNN classifier [Finn et al., 2015], the MLR discards uninformative links with little reliability to discriminate the desired modalities, so taking whole-brain EC with many dimensions (links) without a-priori selection is not an issue. Even though the MLR achieves a very good performance without regularization, the latter is an interesting direction to explore in the future to obtain various types of signatures (e.g., more or less sparse). Therefore, the multivariate classification can be implemented using distinct subsets of links, discarding those that mix several modalities. Note that RFE is used without cross-validation here (because of the computation speed), separating the

evaluation of the ranking of the EC links from the generalization capability of the MLR. Together, our results underline the importance of adequate machine learning tools to obtain a powerful and flexible framework that can scale up with the complexity of modalities.

Formally, our whole-brain dynamic model is a continuous-time network with linear feedback that incorporates topological constraints from SC. EC is estimated using a gradient descent (or Lyapunov optimization) that takes into account the network feedback and can be very efficiently calculated for the whole brain with 100 ROIs and each session with 300 time points per ROI [Gilson et al., 2016, Gilson et al., 2017]. Each session gives a parameter estimate (for each EC link), whose distributions across subjects and conditions are used for classification. Our dynamic model and estimation procedure are simpler than the dynamic causal model with hemodynamics and Bayesian machinery [Friston et al., 2003,Stephan et al., 2004], which has been used for classification relying on a few ROIs only [Brodersen et al., 2011]. Nonetheless, they provide powerful signatures in a much richer (high-dimensional) space that can be used for modality discrimination. Our study focused on two coarse parcellations covering the whole brain [Tzourio-Mazoyer et al., 2002] or cortex [Hagmann et al., 2008]. Although the two parcellations were applied to different datasets, we did not observe significant difference in the performance of the classifiers. Much work has been done recently to correct the bias due to the use of specific parcellations [Da Mota et al., 2014]; for our purpose, more refined parcellations may entail better discriminability in higher-dimensional spaces, but raise issues for the EC estimation robustness. This motivated us to choose rather coarse parcellations with large ROIs as a first step. We have used a generic SC with 30% density, ensuring a sufficiently rich potential biomarker with thousands of estimated EC links in total. Inaccuracies about SC may affect some % of all links, but this is unlikely to hardly affect the collective predictive power of EC. Moreover, preprocessing using PCA was surprisingly found to have very little influence on the performance. Nonetheless, PCA may be useful for datasets with larger number of subjects and conditions [Preti et al., 2017]. We also tried the classification procedure with additional global signal regression of the BOLD sig-

nals and results were similar in terms of performance. Although refining the preprocessing pipeline may improve the (already excellent) classification performance, we expect EC to perform better in general.

CHAPTER 3

Classification of condition based on dynamic functional connectivity

3.1. Introduction

A large part of the research being currently done on neuroimaging uses connectivity measures –quantifying how different neural populations interact between each other– as an approach to better understand the fundamentals of cognition. Proposed originally at the end of the 1990s, the so-called functional connectivity (FC) is typically computed as the similarity across time between the activity of two given regions. It is a measure of statistical dependence, such as correlation or mutual information. FC patterns can be obtained, for example from fMRI, by computing the pairwise correlation between brain regions BOLD signals.

The connectivity patterns from a single subject have shown to be reproducible –in spite of showing some degree of session-to-session variability– for periods of time ranging from days to months. The reliability of these connectivity measures has been assessed either when the subject is per-

forming a task, or lying inside the scanner without receiving any external stimulus, which is known as resting state, [Pannunzi et al., 2017, Chen et al., 2015, Birn et al., 2013]. A subset of networks that is heterogeneously distributed across the brain appears to be consistently stable over time, independently of the subject's emotional or cognitive state [Mueller et al., 2013, Gratton et al., 2018]. This characterization of individual traits in FC patterns seems to be also related to practical aspects such as session length, and other pre-processing steps, like the parcellation scheme, filtering or the removal of the global signal (GSR) [Finn et al., 2017, Finn et al., 2015, Birn et al., 2013, Laumann et al., 2017, Airan et al., 2016].

Different aspects should be considered when extracting connectivity biomarkers, such as distinct sources of noise and how they interact with the measure, as seen in the previous chapter. Obtaining more reliable samples can be accomplished by averaging groups of FC networks of the same category but from repeated scanning sessions [Gratton et al., 2018]. Another option is recording fMRI signals as long as possible and then extracting the overall response through correlation of the entire time series, concatenating sessions if possible [Finn et al., 2015, Rosenberg et al., 2015]. However, an important and usually overlooked consequence of averaging across sessions is the reduction in the total number of samples. The size of the training set, and especially the ratio of samples to number of features, is critical when applying machine learning to high-dimensional data [Hughes, 1968]. It is a fact that the explosion during last decades in the number of machine learning applications is mostly due to an increase in the amount of available data, more than to any substantial improvement in the algorithms.

Computer vision has been one of the most benefited fields by this increase since the number of parameters to fit frequently exceeds the number of image samples. In this type of problem, having a large number of images in the training set provides the model with more information about the sample, which helps for a better generalization. When having more samples is not possible, the training set can be artificially increased by creating new ones through the application of some geometrical transformations to the original images –such as translation, rotations, or the addition of noise–, a

technique known as data boosting or data augmentation. These transformations add more variability to the training set and act as a form of regularization against overfitting, a crucial issue when dealing with spaces that have many more parameters than samples.

For fMRI, classification of connectivity estimates usually presents this dimensionality problem, where the dataset typically consists of less than a hundred samples, and has more than several thousands of dimensions (one per each link in the connectivity measure). For example, applying a low resolution parcellation of 50 regions of interest (ROI) to a hypothetical dataset would yield an FC data space of 1225 links or dimensions[1]. Most likely, the number of acquired samples –i.e fMRI sessions– would be one or two orders of magnitude less than the number of links. In this context, averaging FC patterns to remove variability would reduce even more the samples/dimensions ratio.

Here we present a method for the application of data augmentation to the classification of connectivity estimates extracted from BOLD fMRI signals. In contrast with classification of static images, connectivity patterns are extracted from data that change over time, i.e. a set of time series of a certain length. Duration of fMRI acquisitions commonly varies from a few minutes to hours, both for resting state and task conditions [Birn et al., 2013], but there is no practical reason to consider each fMRI run or session as the minimal unit possible to compute FC patterns. In practice, most techniques applied to estimate FC are based on the assumption of stationarity but there is no guarantee that this conditions is fulfilled along an entire fMRI session. Even in resting state, subject's vigilance state, attention, or mood may vary. Thus, in the time span of a session, BOLD signal can contain shorter non-stationary information that may be removed with the calculation of its covariance matrix.

Different time-resolved methods have emerged as an attempt to capture the fluctuations of BOLD signal in what has been referred to as dynamic functional connectivity (dFC) [Hutchison et al., 2013]. The most common approach to compute dFC is based on a sliding-window technique that

[1]The equation $\frac{N \times N}{2} - N$ yields the number non-duplicate connectivity links, being N the number of ROI)

estimates the instantaneous correlation inside a time window of a given length W that is shifted over time.

Although there are results that have shown some evidence about the association between dFC and behavior [Gonzalez-Castillo and Bandettini, 2018], it is unclear to what extend these temporal fluctuations are dominated by other non-neural aspects, or be representative of the underlying cognitive processes under study [Xie et al., 2017, Gratton et al., 2018]. Thus, using dFC to predict some cognitive-related scores or states can be a manner of assessing this association [Cohen, 2018].

We propose to use dFC to predict a set of cognitive-related states. From this perspective, time series could be split into a handful of shorter temporal chunks, providing the classifier with a more detailed description of the data space. In this sense, we would *artificially* increase the number of samples by adding more variability –akin to the data augmentation approach applied to image classification– and in the same way, regularizing the model to control for overfitting. The operation of splitting the time series of a whole session into shorter signals can be seen as a generalized dFC technique, in which the parameters W and T can be freely set.

In this work, we explore the parameter space of dFC (varying both W and T) in order to assess whether data augmentation can in general enhance the classification accuracy. Our hypothesis is that we could improve the classification of conditions or subjects when the amount of data is limited by creating new connectivity samples from shorter BOLD time series. It has been proven that correlations computed by using very short signals may elicit an estimation noise that can contaminate the obtained patterns [Hindriks et al., 2016, Leonardi and Van De Ville, 2015]. Despite of being noisier, these shorter signals might be intrinsically more meaningful –describing more precisely fluctuating states– for the classification than using the whole-session time series. In addition, from a data analysis perspective, having more samples could also aid to the generalization of the predictive model. The classifier could, therefore, take advantage of the created variability, which might be informative about the data space and classes the algorithm is trying to learn from. The trade-off across these three elements will determine the outcome of the method.

Firstly, we apply data augmentation to the classification of FC patterns and compare its performance with a whole-session approach. We use Dataset C, that can be categorized into two task conditions: natural movie viewing and resting-state. We show that two clusters corresponding to the two conditions dominate the data space –even over the generally stronger subject classes [Xie et al., 2017, Gratton et al., 2018]–, which makes the classification easily separable in terms of parameters/dimensions. This dataset is also use to visualize how dFC is transforming the original whole-session data space.
Then, we evaluate this technique on a second dataset with less separable classes. To this end, we use Dataset D that can be classified into four categories. This dataset comprises multi-session fMRI data from subjects that are performing a Go-NoGo task under the effect of four different attention-related drugs. In this case, we assign a large part of the dataset to the train set, and assess the impact of data augmentation on the accuracy of the prediction of unseen sessions.
Results show that in both cases data augmentation yields better performance than using whole-session FC. The choice of the classifier turns out to be crucial as well, being logistic regression the model that exploits this technique more effectively.

3.2. Methods

3.2.1. Whole-session FC and data augmentation

First of all, the fMRI runs from each dataset were split into train and test sets. Whole-session FC (wFC) samples were extracted using the test data, while data augmentation was only applied to the runs that comprised the train set.
Time series from each individual fMRI run in the train set were segmented by applying the sliding-window technique (see Fig. 3.1A). This approach is based on the displacement of a temporal window of a certain length W over time. For each time shift (of a distance T), a new intermediate FC_i –a Pearson's correlation matrix– is computed using that part of the

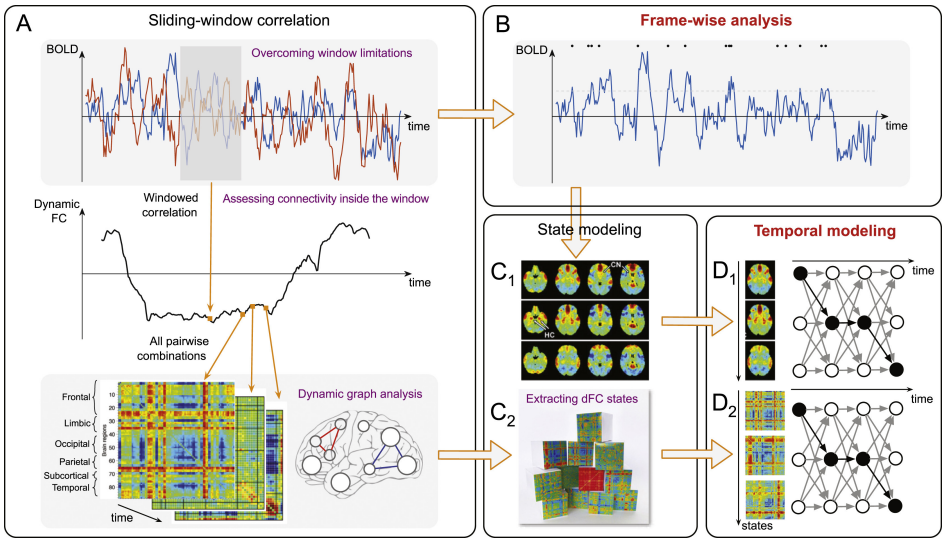

Figure 3.1: Summary figure of existing dFC analytical strategies. (A) Sliding-window technique for the extraction of the dynamic FC. The functional connectivity (FC) from BOLD time series is computed pairwise within each time window (upper panel). This window is shifted over time every step T, generating a connectivity timecourse (middle panel). One connectivity matrix (FC_i) is obtained per shift when all the time series are included (lower left panel). **(B)** A recent conceptual alternative to the sliding window framework is a frame-wise description of timecourses, where only moments when the BOLD signal exceeds a threshold are retained for the analysis . These frames can be used for the generation of voxel-wise brain states (C1), the co-activation patterns (CAPs). Alternatively, the connectivity matrices obtained from (A) can be used to retrieve dFC states (C2). Through temporal modeling, parameters describing CAPs (D1) or connectivity states (D2) and their relationship can be inferred, so that amongst all possible state trajectory options (denoted by the set of white dots linked by light grey arrows), the observed path (black dots and arrows) is the most likely. Compared to sliding window analysis, frame-wise analysis and temporal modeling are two suggested, conceptually innovative directions for future dynamic functional connectivity work. Reproduced from [Preti et al., 2017].

time series that fall within the window. Both parameters, W and T, are measured in time points or TR units. For the analyses performed in this work, the window length W was increased from 2 TR to the duration of the entire time series. For each particular length W, the minimum value of T=1 provided the largest number of FC_i for that run. Thus, increasing the parameter T was equivalent to sub-sampling the dFC.
Given W and T, the total number N of FC_i per run obtained is:

$$N = \frac{L - W + 1}{T} \,, \tag{3.1}$$

being L the length of the original time series (L = 300 time points in Dataset C, and L = 174 time points in Dataset D).
For the fMRI sessions comprising the test set, a single wFC was computed per run –also by means of the Pearson's correlation– between all possible pairs of ROI signals.

3.2.2. Pipeline

The resulting matrices after computing dFC and wFC were converted into vectors by removing diagonal and duplicates –by keeping the lower triangle–, as shown in chapter 2. Each vectorized FC (wFC or FC_i) constitutes one sample composed of p elements or features; the number p depends on the number of ROI (p = 1114 for Dataset C and p = 6670 for Dataset D).
All the training vectors coming from dFCs were stacked together vertically, creating a matrix with p columns and m rows ($m = N \times$ the number of sessions/runs). For example, for $W = 2$, $T = 1$, and $m = 10$ fMRI runs from Dataset C ($L = 300$), the size of the resulting matrix will be using Eq. (3.1): $[m \times p] = [(10 \times \frac{300-2+1}{1}) \times p] = [2990 \times 1114]$. This would be the design matrix used to train the classifier.
A similar classification pipeline to the one presented in chapter 2 was applied to calculate the accuracy of the model in the discrimination of conditions. Two different classifiers were again used: multinomial logistic regression (MLR) and k nearest neighbors (kNN) with k = 1. The proce-

dure was cross-validated following an iterative train-test routine. A certain number of fMRI runs was used to train the model in each iteration, leaving the remaining runs for test. This operation was repeated a certain number of times, selecting randomly the runs for train and test. In each one of these iterations, data in the training set was augmented computing the dFC as described before, and the classifier was evaluated on the wFC samples comprising the test set.
Following this procedure there were chances of having fMRI runs of the same subject in both training and test sets. Due to the within-subject similarity observed in FC estimates (see chapter 2), we decided to control for this within-subject information and the possibility that distinct samples of the same subject would be used in train and test. This situation might increase the performance by distorting the classification of conditions due to the classification of subjects. For this reason, we repeated the whole process but controlling for each subject data to be either in training or in test sets, but not in both.

3.2.3. Classes defined

Dataset C was used to discriminate between two task conditions: natural movie viewing (MV) and resting state (RS). The accuracy was calculated using all the runs from the same subject in training (3 movie and 2 rest samples), and the remaining subjects for test. The cross-validation procedure comprised a total of 22 iterations, the same as the number of subjects. This approach implicitly implied controlling for subject, since samples from the same participant could not be both in train and test sets.
Dataset D consisted of four different drug conditions: placebo (PLAC), (MPH), (ATM), (CIT). In this case, 70% of the dataset was used for training, and the 30% remaining for test. The accuracy was calculated after repeating the process 10 times, selecting randomly with replacement the fMRI runs for train and test. When the selection of runs was controlled for subject, 70% of the number of subjects were used for training, leaving the remaining subjects/runs for test. Further details of both datasets can be found in appendix A.

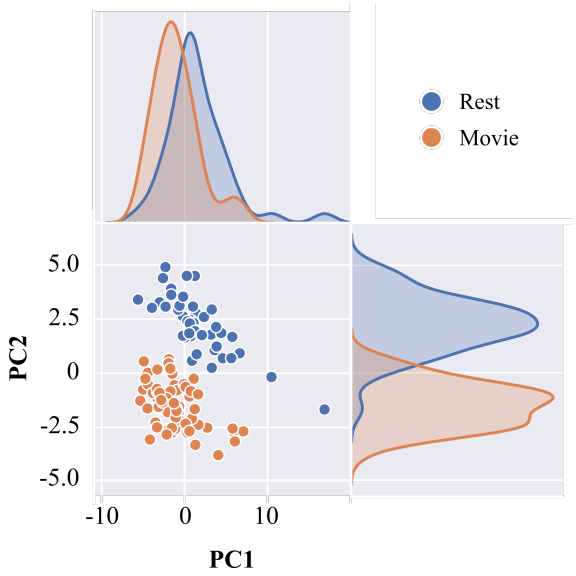

Figure 3.2: Projection of Dataset C onto the PC1-PC2 plane after applying PCA. Each dot corresponds to one wFC computed using the entire fMRI session. Rest (in blue) and movie (in orange) samples can be perfectly –and linearly– separated on the two-dimensional plane defined by the two first principal components (PCs). This indicates that the two classes are easily separable in terms of dimensions.

3.3. Results

3.3.1. Classification of condition: movie viewing vs. resting state

As depicted in Fig. 3.2, Dataset C comprises two classes (MV and RS) that are linearly separable along one or two dimensions. Due to its high degree of separability, we only used the fMRI runs from only one subject (2 RS and 3 MV) to train the classifier after applying data augmentation. The accuracy was measured on the remaining sessions (42 RS and 63 MV) from the remaining 21 subjects. This operation was repeated 22 times:

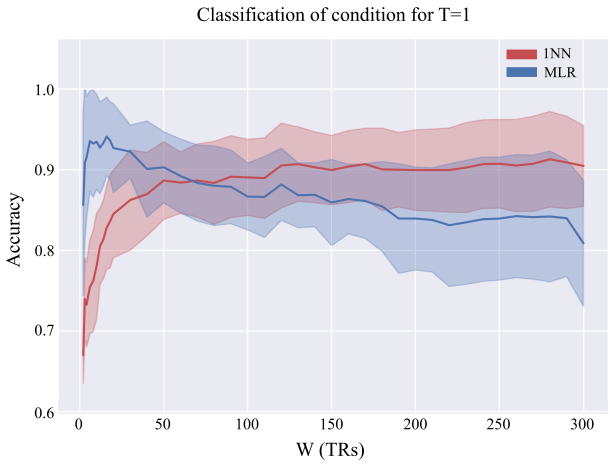

Figure 3.3: Classification accuracy applying data augmentation on Dataset C Accuracy depending on W when T=1 for MLR (blue) and 1NN (red) classifiers. MLR shows higher accuracy for short window lengths. These curves are equivalent to the first column of the plane W-T shown in Fig.3.4.

each subject was only once in the train set. The data in the train set were augmented in each iteration as explained before, using different values of W and T. The accuracy was re-calculated for each pair W-T, using the same sessions in train and test.

Mean accuracy for all the combination of W and T values as well as for both classifiers is shown in Fig. 3.4. Average performance of MLR increases from 81% using wFC to 94% after augmenting the data sample. Remarkably, even the shortest window length ($W = 2$) yields better accuracy than using wFC. The maximum mean accuracy is reached with MLR and for the ranges $6 < W < 20$ and $1 < T < 10$.

On the other hand, the performance using a kNN with k=1 for the prediction of test subjects has a different profile: lower accuracy for short window lengths before it stabilizes around W=50 approximately, and, then, a soft slope until reaching 90.5%. In this case, the increase in the number of training samples does not improve the accuracy. The main reason for

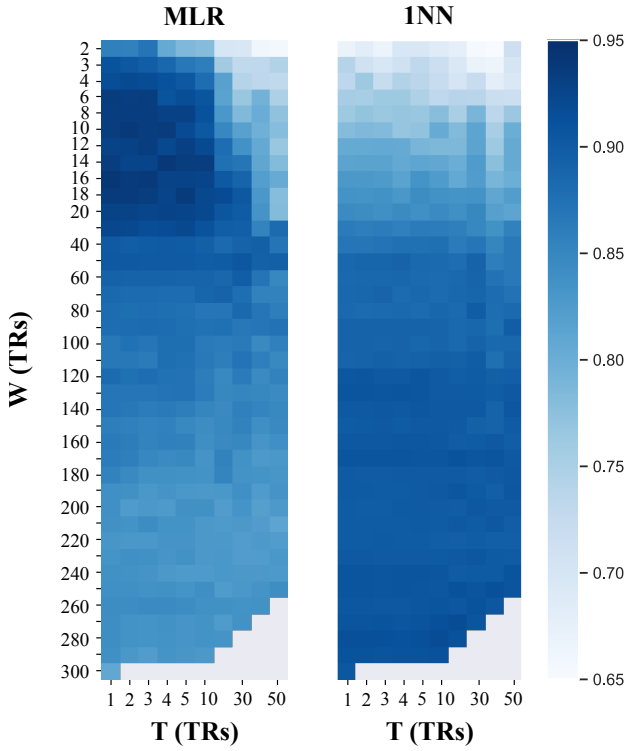

Figure 3.4: Comparison on the W-T space of MLR and kNN classifiers on Dataset C. Classification performance on the parameter space of W and T for MLR (in blue) and kNN (in red) classifiers. Darker blue corresponds to higher accuracy. Note that both plots have been zoomed in the W axis, between W=2 and W=20.

this result may be explained because of the different nature of both classifiers. Complexity of kNN classifier is directly related to the number of samples, whereas in logistic regression it is defined by the number of free parameters (here equal to the number of links). Therefore, augmenting the number of data samples directly increases the complexity of the kNN –a given test sample might be assigned to a larger number of neighbors because the train set is larger. Conversely, MLR complexity remains constant when the number of samples increases. For the case of using wFC, the accuracy reached is higher for kNN than for MLR.

As explained before, increasing the displacement T is equivalent to subsampling the augmented dataset, i.e. keeping only those FC_i that are a number T of time points apart. This operation could reduce the computational load by removing highly correlated samples that do not increase the classification accuracy. In practice, we can see from Fig. 3.4 that increasing T has no effect on the classification accuracy of the kNN classifier. This suggests that is how the samples distribute over the data space what explains the classification with kNN more than the number of samples *per se*. With MLR, conversely, there is an effect on the accuracy, which starts to decrease after T=10. This points to a combination of number of samples and their distribution as an explanation of the improvement in performance.

Effect of data augmentation projected onto the PC space

In order to better understand the improvement achieved by the MLR classifier after applying data augmentation, we show the process graphically in the plane defined by the first two principal components (PCs). Fig. 3.5 shows the whole Dataset C projected onto the two first PC after applying PCA. A total of 22 subjects $\times$ 5 sessions per subject yields 110 sample dots (66 for MV and 44 for RS). Two main clusters can be observed that represent the two conditions (MV in orange and RS in blue in the figure). These condition clusters dominate the data space: sessions from individual subjects are not distinguishable from each other, i.e. there are no subject-specific clusters.

The samples from a given subject have been highlighted in Fig. 3.5A. These five samples comprise the train set that would be feed into the model. It can be observed that there is a large empty region where the separatrix could be placed. This freedom could lead to a suboptimal decision boundary that misclassifies some of the test samples.
Figure 3.5B shows how data augmentation has created new samples from one selected training session (marked with a circle). The new samples were obtained by computing its dFC ($W = 100$, $T = 1$), and projecting them onto the same PC1-PC2 plane. Figure 3.5C shows the result of the same operation applied to all the five sessions that comprised the train set. As it can be seen in the figure, the new samples are now defining two separate clouds corresponding to each of the two classes. These two clusters describe the data space in a more precise way: samples from the same class are close enough to each other, delimiting better the region between the clusters. After doing this, the solution provided by a linear classifier will lie in this constrained region. Interestingly, when subsampling the augmented space by using $T = 7$ (with crosses in Fig. 3.5C) the shaped of the clusters is preserved. This could explain how the high accuracy is maintained until $T > 10$. These results indicate that data augmentation could be a good option for problems where training samples are scarce but the classes are separable enough.
Finally, it is important to note that while the samples in Fig. 3.5 are represented after projecting the data onto the first two principal components, the actual classification reported in Fig. 3.4 is using the whole multidimensional space.

3.3.2. Prediction of four drug conditions

The dataset used can be classified into four categories: PLAC, MTH, ATM and CIT (see appendix A.D for details). A total of 26 participants underwent scanning and performed a sustained-attention task. Each participant was scanned six consecutive times on four different days, corresponding to the four drug conditions. Data augmentation was again applied only on the train set as explained for Dataset C.

Figure 3.5: Visualization of data augmentation on the PC space.

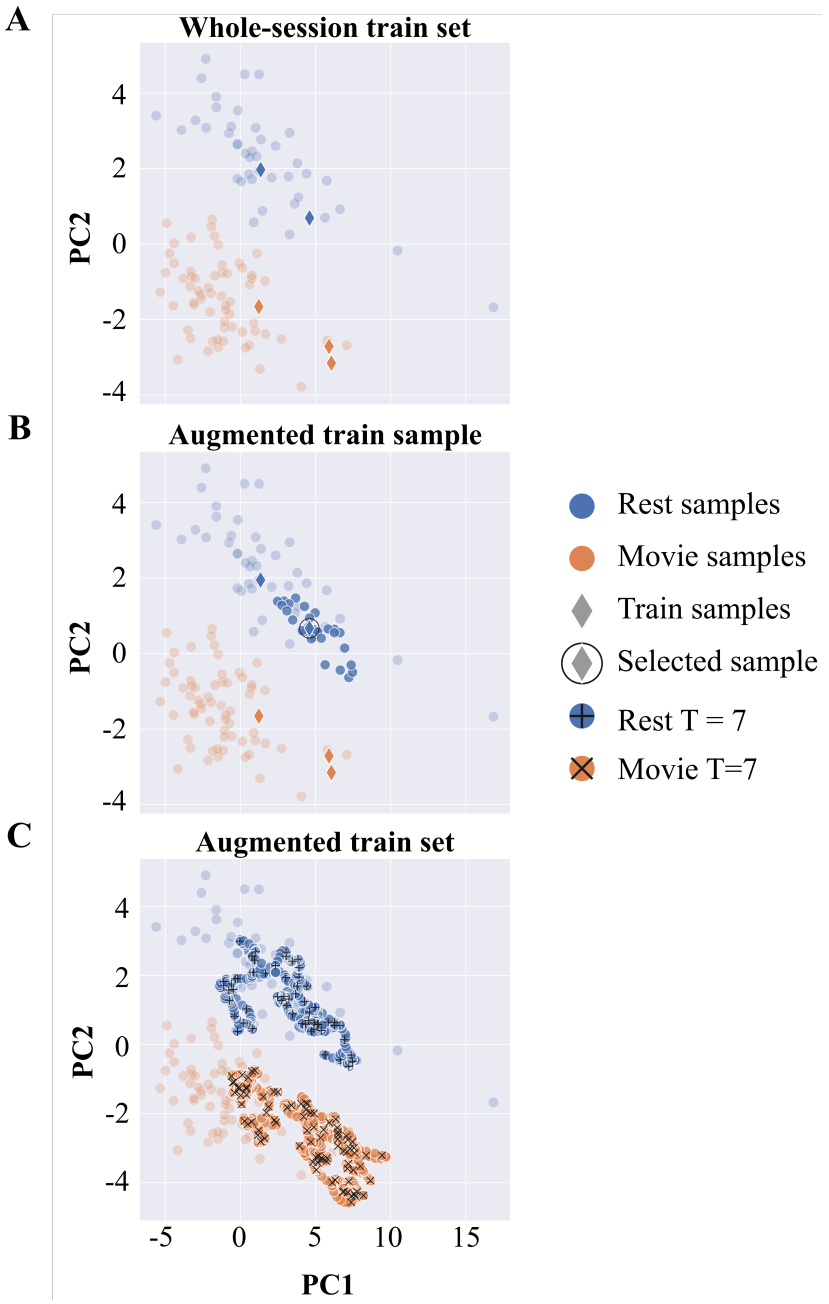

Figure 3.5: ***(preceding page):*** **A)** The whole Dataset C projected onto its two first principal components with five train samples from the same subject (2 RS in blue, 3 MV in orange). **B)** The same two-dimensional plane after the projection of the created samples (W=100, T=1) using the selected (with a circle) session. **C)** The PC1-PC2 plane after augmenting (W=100, T=1) and projecting the five samples in the train set highlighted in **A)** . The classifier may find a better linear solution between the two clouds. The crosses show the augmented samples after applying a T=7. It can be noticed that the two clouds are subsampled when T=7 but the shape is maintained, without modifying the region between the classes.

Figure 3.6 shows the mean accuracy in the classification of the four drug conditions as a function of both W and T (chance level is in this case 0.25). A pattern emerges very similar to that shown in Fig.3.4. The model was trained using the 70% of the samples, with (Fig. 3.6A) and without (Fig. 3.6B) controlling for subjects' identity.
As shown for Dataset C, MLR again differentiates better across the four classes after applying data augmentation to the train set. The highest accuracy is reached within the region defined by $10 < W < 70$ and $1 < T < 10$. Similar to the previous dataset, the classifier yields better results with shorter time windows and small values of T, starting to decrease after $T = 10$. This points again to a combination of distribution and number of samples as a reason for the improvement in accuracy. As seen for Dataset C, the augmented sample might delimit a narrower region of the high dimensional space compared with wFC, helping so the model to find a better solution among all the possibilities.

3.3.3. Condition classification controlling for subject

We repeated the same analysis but now controlling for subjects' identity, as done previously on Dataset C. We used 70% of the data for training and measured the performance over the remaining 30% of the sessions, whose subjects were not in the train set (Figure 3.6B).

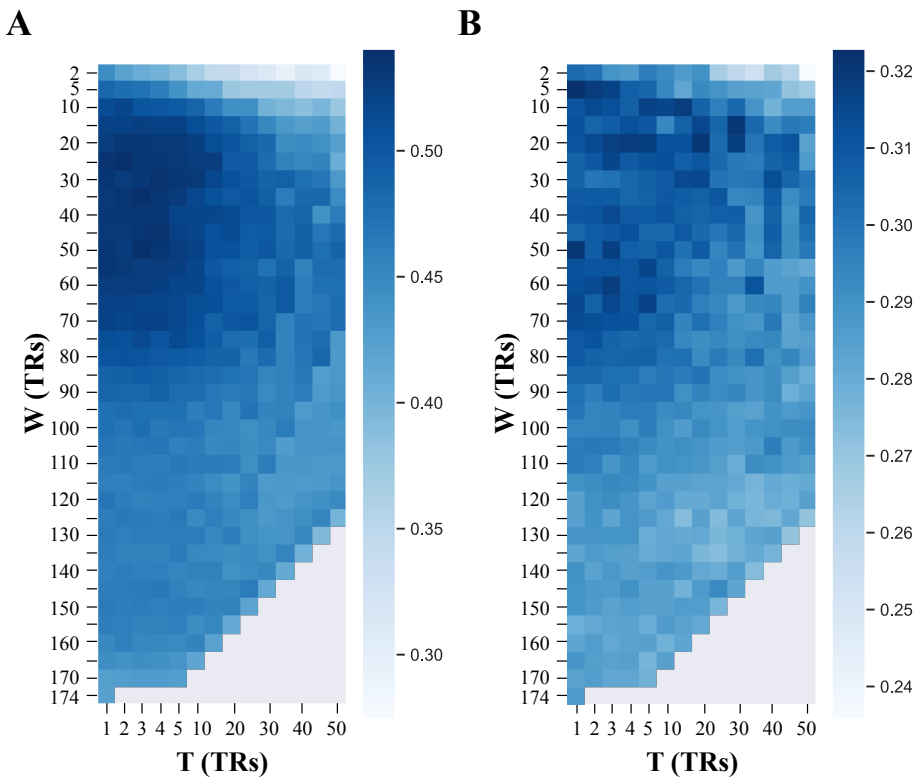

Figure 3.6: Classification performance for MLR classifier on Dataset D. Mean accuracy in the classification of four drug conditions after applying data augmentation. Plots show the performance achieved depending on the W-T values without **(A)** and with **(B)** the application of controlling for subject cross-validation.

Overall, there is a general drop in the accuracy, which reflects the fact that within-subject similarity was overestimating the classification of conditions shown in Fig. 3.6A. Although less clear, Fig. 3.6B shows a pattern similar to Fig. 3.6A, i.e. the highest accuracy reaches a plateau for $10 < W < 70$ and $1 < T < 10$.

Interestingly, when we ensure that samples of the same subject cannot be both in training and test sets –i.e. we removed the within-subject similarity–, the maximum accuracy drops from 0.55 to 0.32; still above the chance level 0.25 but considerably lower. This suggest that in this dataset with less separable classes compared to the previous one, the classification of subjects is interfering with the classification of conditions when observations of the same group (subject) are not properly controlled.

To further study how these subject-condition dimensions interact, we performed the classification of subjects' identity on the same dataset, and exploring the again the parameter space for W and T.

We used the MLR classifier, with one train sample per subject and repeating 50 times. Fig.3.7 show that the highest accuracy is reached for shortest values of W (between 2 and 5) and largest number of samples ($T = 1$).

Remarkably, the obtained patterns are not completely overlapping when comparing Fig.3.6B and Fig.3.7: i.e. classification of conditions controlling for subjects and classification of subjects. The area showing the highest accuracy when classifying subjects (Fig. 3.7) is defined by short window lengths and low values of subsampling ($2 < W < 5$ and $1 < T < 4$). This suggests that, even if the subject-based clusterization of the data is dominant over condition, different values for W and T may emphasize either one or the other classification.

3.4. Discussion

In this study, we evaluated the possible application of data augmentation to enhance classification accuracy when using connectivity patterns as a cognitive biomarker. Data augmentation was performed by splitting fMRI BOLD signals into shorter time windows before the calculation of FC

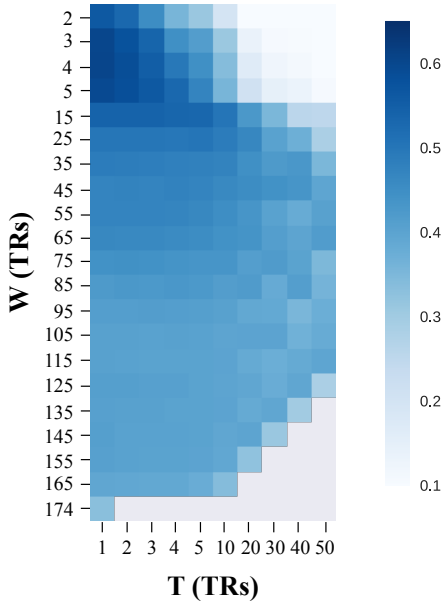

Figure 3.7: MLR accuracy on subject classification on Dataset D. Performance of MLR classifier on the parameter space comprised by W and T. The plot shows the mean accuracy when classifying subjects on Dataset D.

estimates, exploiting the dFC technique. We assessed data augmentation on two different classification problems using two datasets: two types of classifiers –kNN and MLR– were trained with Dataset C to differentiate across two tasks (movie viewing and resting state); and, Dataset D was used to discriminate across four different drug conditions (MPH, ATM, CIT and placebo).

We showed that both classifiers yielded different results on dataset C. Applying data augmentation to MLR improved the accuracy in the classification compared to using wFC. Conversely, kNN performance decreased when augmenting the sample. This differences in performance can be explained by the increase in the kNN complexity after augmenting the sample, while the number of parameters of the MLR model –and therefore its complexity– remains constant. As seen in Fig. 3.5, projecting the data onto the two first principal components can help us to understand the reasons that support MLR's improvement: Applying data augmentation shrinks the region of the space where the model can fit the linear solution. On the contrary, the open region barely defined by a few whole-session FC samples is too wide to fit a line that can generalize well to unseen samples. Moreover, the improvement is not purely due to the greater number of samples reached, but also to the spatial distribution of these samples.

We also applied the same methodology to Dataset D using the MLR classifier. The accuracy in the prediction of the four classes was again higher after applying data augmentation compared to wFC. In this case, when samples from the same subject could fall both in training and test sets, the accuracy was overestimated due to the within-subject similarity. We repeated the analysis but now keeping entire subjects either in train or in test sets. When doing this, the accuracy dropped in the entire W-T parameter space. However, short time windows still yielded better results than using wFC. To study the effect of within-subject similarity on the accuracy, we performed a subject classification on the same data. Interestingly, the region of the W-T plane with higher accuracy did not match exactly the W-T plane obtained with the classification of conditions. This suggests that the subject and condition dimensions interact in different timescales that can be determined in terms of window length and time shift.

Generalization when working with connectivity data may be complicated because different sources of variability will affect our samples, making difficult the prediction of new incoming data. Addressing this issue by averaging sessions can destroy informative dynamic patterns. Although shorter FC estimates might be noisier [Hindriks et al., 2016, Leonardi and Van De Ville, 2015], the dynamic information provided –lost when using standard wFC– may overcome the increase in the estimation noise. The method presented here tries to take advantage of this information and use it for the identification of cognitive states.

Many works have tackled the issue of finding meaningful differences across connectivity states to better understand cognition or mental disease. Among all the techniques used, machine learning approach provides a framework, where these differences can be assessed according to performance criteria. Basing the method on the increasing of performance is important because a more precise model may describe better the problem, and will be, therefore, more generalizable. In addition, accuracy can be also used in many other different ways, for instance, as a criterion for identifying the specific subnetworks that support a given cognitive condition.

The results presented here show that data augmentation, applied through dFC, may improve the accuracy on the classification of cognitive conditions/states. FC data usually comprise a number of dimensions –one per connectivity link–, which is extremely large in comparison with the number of samples. In these kind of situations, regularization methods may be applied to control for overfitting. This is what data augmentation does by guiding the model to find a more generalizable solution, very important in those contexts where training samples are scarce.

Another advantage provided by data augmentation is also the possibility to study the timescales and trajectories of dFC across conditions. Some methods attempt to extract, directly from the dFC, the so-called brain states –consistent connectivity patterns that reoccur in time [Haimovici et al., 2017, Cabral et al., 2017, Hansen et al., 2015]. These brain states might be more easily interpretable, in terms of their association with behavior, than the dFC itself. However, the specific nature of each method makes difficult the comparison between them. Moreover, dFC is some-

times obtained separately, before including any behavioral information. The dFC parameters –such as window length– are, in these cases, chosen without taking into consideration the cognitive application. On the contrary, the nature of our approach combines the creation of the dFC with the question being evaluated on the same problem.
Dynamic FC has been already used to find differences across cognitive tasks in healthy individuals using pattern classification and accuracy criteria [Shirer et al., 2012, Gonzalez-Castillo et al., 2015]. Yet, this is the first time, to our knowledge, that classification accuracy is used in a cross-validation manner as a criterion to choose the parameters of dFC. After finding the optimal ranges for W and T, we propose to define the FC states in a reverse way: from the structure shaped by the augmented samples. These samples can draw meaningful trajectories in the data space, akin to those shown in Fig. 3.5C, that might be linked to behavior. There is, indeed, some evidence that brain states can be similar across groups, but it is in the dynamics between the states –i.e. the trajectories– where those differences between conditions are more manifest [Cohen, 2018]. Understanding the outlined trajectories across sessions or subjects might help to grasp the association between FC and behavior. Achieving this result would confirm both the value of dFC as a cognitive biomarker and the validity of the method.

CHAPTER 4

General discussion

One of the main goals in brain neuroimaging research is the understanding of how specialized brain areas cooperate to support cognitive functions. The analysis of brain connectivity has become a widespread approach to tackle this issue, especially using fMRI data. The present work aimed to evaluate the suitability and possible advantages of machine learning in the analysis of brain connectivity, which allows for the study of cognition at the whole-brain level. The guiding line of this work was the comparison of several connectivity measurements, such as FC, EC and dFC. We assessed to what extent these measures are capable of capturing informative spatiotemporal patterns from fMRI time series.

As a conceptual framework, we can represent connectivity estimates from fMRI data in a three-dimensional space where each axis corresponds to one source of variability: subjects, cognitive condition and time. In chapter 2, we compared the stability across sessions of two connectivity measures, Pearson correlation FC and MOU-based EC. We assessed the relationship between the predictability of subjects' identity and the session-to-session variability. Results confirmed previous evidence in the literature

of high prediction accuracy using FC. However, we saw that EC measures overcome the performance of FC in classifying subjects. This suggests that EC may better capture individual traits by including temporal information via the model-based estimation of EC, unlike FC. Finally, we studied how different cognitive conditions, e.g. task and rest, might interfere with the identification of subjects using EC. We proposed a method for feature selection that allowed us to compare and differentiate between the brain subnetworks specific to either subject or condition information. We called these subnetworks *signatures* of subject or condition.

In chapter 3, we focused on the differentiation across cognitive conditions using FC: First, we compared rest to task and, then, we classified connectivity data into four different drug conditions. In this case, we changed the temporal scale of interest from between- to within-session variability. We extracted the dynamic FC from fMRI sessions by segmenting the time series into shorter signals. The instantaneous FC samples created by means of this sliding-window technique were fed into the model, which was used to predict the class of the remaining test data. The novelty of the method lies in the application of a typical machine learning technique, known as data augmentation, to the classification of functional connectomes. It permitted us to regularize the classifier, diminishing so the overfitting commonly observed when dealing with such high-dimensional data. In practice, this approach showed an improvement in the classification performance compared with whole-session FC. These results also indicate that the data comprise informative fluctuations that are related to cognitive states, which can be captured using dFC. Interestingly, the optimal timescale found for subject classification was different from the timescale for condition classification. This suggests that temporal fluctuations of subjective traits and cognitive conditions have different scales that might be considered as temporal signatures, similar to the distinct spatial signatures discussed in chapter 2.

An overarching goal of the thesis was the comparison across different machine learning methodologies applied to connectivity data. The same analyses were carried out with two classifiers, remarkably different in their approach: k-nearest-neighbors and multinomial logistic regression (MLR).

kNN is a non-linear classifier that predicts based on the similarity between each test sample and the train set. Conversely, MLR tries to fit a linear model that best separates each class with respect to the rest. These differences may explain their diverse performance after data augmentation. By increasing the sample we were also increasing the complexity of kNN, whereas MLR can take more advantage of having a larger number of samples. In short, kNN predicts on global connectome measures by means of a given metric, –here Pearson correlation coefficient. In contrast, MLR focuses on local characteristics by tuning a weight for each individual link. Results point to the potential of MLR as an adequate tool to study brain connectivity due to its better performance and interpretability in terms of studying the local/specific interactions.

4.1. Machine learning techniques for brain connectivity applications

The connectivity measures used in this work can be interpreted as different approaches to characterize the information carried by fMRI signals. Typical fMRI data consist of a set of voxel-wise BOLD time series of a certain length. The number of voxels required to cover the whole brain is usually large. This high-dimensional information contained in fMRI data can be converted into a more amenable structure both in time –i.e. through Pearson correlation– and in space –by grouping subsets of voxels. There are different criteria to group voxel-wise time series, either by applying a three-dimensional atlas or by extracting some temporal structure, e.g. via ICA [Calhoun et al., 2004]. However, a general-purpose or group-wise parcellation may not capture well individualized characteristics, a key aspect in personalized medicine [Kong et al., 2018]. A problem-tailored approach might define a space more suitable to this end [Kottaram et al., 2018], similar to the method presented in chapter 3 to extract the timescale.

As explained also in chapter 3, the parcellation scheme, and, in particular, the number of ROIs, determines the size –number of columns– of the

design matrix used to train the classifier, which is a critical issue when applying machine learning tools. Although steps have been taken to study the influence of the parcellation scheme in the prediction accuracy [Finn et al., 2015, Airan et al., 2016, Eickhoff et al., 2015], further research is needed. Due to its impact on the created data space, different parcellation schemes should be compared, ranging from coarse to spatially more detailed. There is a current trend in neuroimaging towards the application of finer and more detailed parcellation schemes [Glasser et al., 2016, Airan et al., 2016]. The main reason is because they are thought to capture better the complexity of brain function. Yet, a larger number of ROIs entails an increase in the number of links in the connectivity matrix. This operation may accordingly increase the complexity of the data. Shen parcellation [Shen et al., 2013], for instance, widely used in the literature, comprises more than 250 functional ROI, which yields about 26,000 non-duplicate links. Such a large number of links will produce an unbalanced design matrix, comprising much less samples than parameters. This is an undesired situation because it increases the risk for overfitting. Classifiers like MLR are suitable to this kind of problems because they may include a regularization term that controls for overfitting. This term penalizes those features that are less important to the model, which can operate as a feature selection technique. As shown in chapter 3, data augmentation may also act as a different regularization method.

The EC-based approach presented in chapter 2 presents two main advantages in this context. First, MOU-EC requires a lower number of classifier parameters compared to FC because the all-to-all FC matrix is constrained through the SC. We preferred coarse parcellations schemes, such as AAL (116) or Hagmann (66), due to their lower number of ROIs. A possible limitation of too coarse parcellations is that differences in condition (e.g. many tasks) may be to subtle to be detected in some cases. This points to a trade-off between a robust measure and a more detailed biomarker. Interestingly, the MOU-EC is including temporal information directly through the model, in the tau constant and in the time-lagged covariances. The obtained results on the classification of subjects using EC suggest that this information is somehow related to individual traits. This may indicate that

this information is more efficiently exploited by the classifier without the necessity of increasing the model complexity, i.e. with more parameters to fit.
Another approach to include temporal information in the measures is by means of dFC. We hypothesized that increasing the number of samples could aid the classifier to generalize on the training set. Results indicate that this is indeed the case, even if samples obtained from shorter signals are necessarily noisier. Other types of window when computing dFC, as well as other approaches of noise addition or jittering to increase the number of samples may be also considered.

4.2. Clinical relevance

During last years a growing body of literature has applied pattern recognition techniques to the clinical domain. Most of these studies aim to extract imaging biomarkers with the purpose of diagnosing accurately neurological or psychiatric diseases. These methods are not intended to substitute for existing diagnostic tools, but to complement them and, importantly, help to unveil the neurobiological basis of the disorder. In addition to diagnosis, other applications include risk assessment, early detection, differential diagnosis, subtyping of patients and predicting treatment response [Woo et al., 2017].
Even though many of the published works are based on structural/anatomical imaging, functional-based methods are gaining prominence. Functional biomarkers are especially important for neural diseases that do not strongly affect the brain structure, like depression, or schizophrenia. Resting-state fMRI has been already applied to predict neurological and psychiatric disorders. The majority of them are based either on task or on resting state fMRI, but combining them may yield more information about the patients' condition.
Initially, these methods adopted a group-wise diagnosis perspective, i.e. patients vs. controls. However, more personalized methods that point to subject-specific and stable measures have emerged during last years.

Subject-based variability in functional connectivity measures has also shown to be related to increased clinical symptoms of psychiatric disorders [Kaufmann et al., 2017]. The authors of this paper decided to average the identification accuracy across different tasks into one unique score. From our perspective, this operation is equivalent to collapsing the three-dimensional space represented in Fig 1.3 into one point. It may be interesting to see how these scores change with more stable and higher subject-specific measures, such as MOU-EC.

Another study performed on the same dataset [Xia et al., 2018] suggests that the co-morbidity among psychiatric diagnosis could be also addressed by using functional networks. Their approach includes a psychological score in the method for the extraction of the relevant connections. In this sense, it is similar to the RFE technique described in chapter 2, applied to go beyond the whole-brain connectome approach to the specific subnetworks that support the prediction. This is a relevant subfield to be explored, especially in those cases where the subnetworks supporting the clinical classification overlap with task-related features observed in healthy individuals. Disentangling the coexistence of healthy/disease and task-related subnetworks using classification accuracy may help to understand what differentiates patients from control participants [Cohen, 2018].

Unsupervised approaches could be a different way of studying co-morbidity and heterogeneity within each diagnostic category. This is the case of depression, which can manifest in different combinations of symptoms. [Drysdale et al., 2017] characterized depression into four different subtypes using resting state and functional connectivity measures. Other works, based also on unsupervised methods, were able find different number of subtypes, and sometimes the subtype information did not improve the accuracy in the diagnosis (ref Feder). These inconsistencies and also issues with reproducibility [Dinga et al., 2018] underscore the importance of establishing robust methodologies, and the potential benefit of having large datasets publicly available [Woo et al., 2017].

Previous examples show reliable differences between patient and control groups in static FC. Other studies show also differences based on dynamic FC estimates [Rashid et al., 2018]. Reocurring functional brain states

seem to be often similar across conditions or groups in dFC. Interestingly, the critical difference appears not to be in brain state FC organization itself, but in the dynamics of –and transitions between– the brain states [Cohen, 2018]. This fact could be exploited by using our data augmentation technique proposed in chapter 3. Results suggest that new connectivity samples define trajectories that may be further studied to understand the differences between the cognitive states.

The dynamical approach can be adopted in a spatio-temporal manner, as shown in [Kottaram et al., 2018]. The authors developed a method, based on machine learning, to predict schizophrenia diagnosis. They show the highest accuracy when both the temporal states and brain regions are adjusted dynamically –temporally and spatially– according to a classification criterion. This result points to the convenience of adapting –instead of imposing– the design matrix to the problem, similarly as we did in chapter 3.

On the other hand, dFC-based methods also allow us to test for hypotheses related to rapid changes in FC. For example, a relationship between attention lapses and FC in attention deficit hyperactivity disorder (ADHD) has been proposed [Cohen, 2018]. This relationship between sustained-attention and fluctuations in FC might be assessed dynamically. Dataset D, used in chapter 3, comprises fMRI scans of participants performing an attention task. The four cognitive conditions used are defined by three drugs related to sustained-attention (plus placebo). Data augmentation was calculated through dFC, and each condition might define different trajectories and states. Participants' behavior –such as response time, awareness– could be compared with these trajectories over time. DFC could be used to predict fluctuations of behavioral scores, e.g. by using regression methods. These methods would serve to understand better the connectivity patterns of healthy subjects. Moreover, it might be applied as well to attention-related disorders, such as ADHD. At the same time, this finding would validate the association between dFC and the ongoing cognitive processes.

4.3. Closing remarks

The work presented in this thesis can be seen as a benchmark of time series modeling and machine learning tools applied to the neuroimaging field. The methodology, as well as the results shown, are a first step towards the study of cognition and neuropathologies. This is a very active field nowadays, but more robust approaches that elicit interpretable results are needed. Exploring the high-dimensional space created by brain measurements –fMRI but also, EEG, MEG– requires both powerful prediction methods and a deeper understanding of their relationship with the data structure. Theory (model-based approaches) is essential to refine connectivity measures and understand where the information is beyond the brute-force application of machine learning tools. In addition, machine learning techniques are also necessary to properly disentangle noise sources and make accurate prediction beyond the beauty of mathematical models. Without this synergistic effort, the analysis of large datasets would remain limited. As an example about how to treat variability –and how its definition depends on the particular problem– time was seen as undesired noise in chapter 2, but as a source of information in chapter 3. This shows that further study is required to determine the optimal temporal and spatial scales going back and forth between specializing the prediction tools for a given problem, and generalizing uncovered principles to the whole field.

APPENDIX A

Datasets

Four datasets acquired at different locations were used in this work:

- **Dataset A** was acquired for the Day2day project [Filevich et al., 2017] at the Max Planck Institute for Human Development (Berlin, Germany) of 40-50 resting-state sessions recorded from the 6 subjects over 6 months. The uniqueness of this data lies in the capability to have statistically valid evaluation of the session-to-session variability for single subjects.

- **Dataset B** is publicly available and is part of the Consortium for Reliability and Reproducibility (CoRR) [Zuo et al., 2014]. We used this dataset to generalize the results of the discriminability when increasing the number of subjects (up to 30 subjects).

- **Dataset C** [Mantini et al., 2012] was recently analyzed using our model-brain dynamic model to extract effective connectivity [Gilson et al., 2017]. We used it to perform a twofold classification with respect to both subjects and conditions (resting-state versus movie

viewing) in chapter 2. We used it again in chapter 3 to evaluate the impact of applying data augmentation to the classification of conditions..

- **Dataset D** was already analyzed and published in [Nandam et al., 2014,Hester et al., 2012]. It was used in chapter 3 to confirm the improvement in the accuracy achieved after applying data augmentation. Subjects were grouped –and classified– into four different drug conditions: methylphenidate (MTH), atomoxetine (ATM), citalopram (CIT) and placebo (PLAC).

In this section, we provide details about the acquisition of the blood-oxygen-level dependent (BOLD) signals for the four datasets.

A.1. Dataset A

This dataset has two parts. The first part (A1) is longitudinal and consists of resting-state fMRI sessions from eight subjects (8) (age 24-32, 6 female). Two subjects (one male, one female) were not able to continue the study and were discarded. The other six subjects (6) underwent scanning between 40 and 50 times along a period of six months. The second part of the dataset (A2) was acquired during the same period of time. A total fifty subjects (50) (age 18-32, all female) were scanned during a single fMRI session each using the same MRI sequences. Participants were free of psychiatric disorder according to a personal interview —mini-international neuropsychiatric interview [Sheehan et al., 1998]— and had never suffered from a mental disease. The study was approved by the local ethics committee (Charité University Clinic, Berlin). Participants were instructed to remain with their eyes closed and data acquisition had to be constrained to 5 min per scan due to experimental limitations.
Images were acquired on a 3 T Magnetom Trio MRI scanner system (Siemens Medical Systems, Erlangen, Germany) using a 12-channel radiofrequency head coil. Functional images were collected using a T2*-weighted echo planar imaging (EPI) sequence sensitive to BOLD contrast (TR = 2000 ms, TE = 30 ms, image matrix = 64×64, FOV =

$216 \times 216 \times 129$ mm^3, flip angle = 80°, bandwidth=2042 Hz/pixel, voxel size=$3 \times 3 \times 3$ mm^3, 36 axial slices using GRAPPA acceleration factor. Structural images were obtained using a three-dimensional T1-weighted magnetization-prepared gradient-echo sequence (MPRAGE) based on the ADNI protocol (www.adni-info.org): repetition time (TR) = 2500 ms; echo time (TE) = 4.77 ms; TI = 1100 ms, acquisition matrix = $256 \times 256 \times 192$ mm^3, flip angle = 7°; bandwidth=140 Hz/pixel, $1 \times 1 \times 1$ mm^3 voxel size.

Pre-processing

The data was preprocessed using SPM5 (Wellcome Department of Cognitive Neurology, London, UK) and DPARSF/DPABI [Yan et al., 2016] after discarding the first 10 volumes of each session. It included: slice timing and head-motion correction (6 parameters spatial transformation), spatial normalization to the Montreal Neurological Institute (MNI) template, and spatial filtering of 4 mm FWHM. Linear trends were removed from the fMRI time courses before band-pass filtering (0.01-0.08 Hz). The data was parcellated using the automated anatomical labeling (AAL) atlas [Tzourio-Mazoyer et al., 2002] into 116 regions of interest (ROIs), which includes the whole cortex and the cerebellum.

A.2. Dataset B

This dataset consists of ten (10) fMRI resting-state sessions acquired from thirty healthy participants (30) every three days for one month [Zuo et al., 2014]. Each session lasted 10 minutes. To minimize head movement, straps and foam pads were used to fix the head snugly during each scan. The participants were instructed to relax and remain still with their eyes open, not to fall asleep, and not to think about anything in particular. The screen presented a fixation point and after the scans, all the participants were interviewed, and none of them reported to have fallen asleep in the scanner. The time of day of MRI acquisition was controlled within participants.

Recording sessions were performed using a GE MR750 3.0 Tesla scanner (GE Medical Systems, Waukesha, WI) at CCBD, Hangzhou Normal University. T2-weighted echo-planar imaging (EPI) sequence was performed to obtain resting state fMRI images for 10 minutes using the following setup: TR = 2000 ms, TE = 30 ms, flip angle = 90°, field of view = 220×220 mm^2, matrix = 64×64, voxel size = $3.4 \times 3.4 \times 3.4$ mm^3, 43 slices. A T1-weighted fast spoiled gradient echo (FSPGR) was used with the following protocol: TR = 8.1 ms, TE = 3.1 ms, TI = 450 ms, flip angle = 8°, field of view = 256×256 mm^2, matrix = 256×256, voxel size =$1 \times 1 \times 1$ mm^3, 176 sagittal slices) was carried out to acquire a high-resolution anatomical image of the brain structure.

Pre-processing

Dataset B was preprocessed with SPM12 (Wellcome Trust Centre for Neuroimaging, London, UK) and DPARSF/DPABI [Yan et al., 2016]. The first 5 fMRI volumes were discarded in order to let the BOLD signal reach stability. The pre-processing pipeline included: slice-timing correction, realignment for motion correction, co-registration of the T1 anatomical image to the mean functional image, detrending, regression of 6 movement parameters, 5 principal component analysis (PCA) white matter and CSF Compcorr, and spatial normalization to MNI coordinates. Scrubbing with Power 0.5 and linear interpolation was applied. This data was also parcellated into 116 ROIs using the AAL parcellation [Tzourio-Mazoyer et al., 2002] and time courses were band-pass filtered between 0.01 and 0.08 Hz, as done with Dataset A. No further global signal regression and spatial smoothing were applied.

A.3. Dataset C

We used a third dataset to study the discrimination between subjects and conditions. In this case, a total of 22 subjects (age 20-31, 15 females) were scanned during rest with eyes opened and natural viewing condition. Volunteers were informed about the experimental procedures, which were

approved by the Ethics Committee of the Chieti University, and signed a written informed consent. In the resting state, participants fixated a red target with a diameter of 0.3 visual degrees on a black screen. In the natural viewing condition, subjects watched and listened to 30 minutes of the movie 'The Good, the Bad and the Ugly' in a window of 24×10.2 visual degrees. Visual stimuli were projected on a translucent screen using an LCD projector, and viewed by the participants through a mirror tilted by 45 degrees. Auditory stimuli were delivered using MR-compatible headphones. For each subject, 2 and 3 scanning runs of 10 minutes duration were acquired for resting state and natural viewing, respectively.
The BOLD signals were acquired with a 3T MR scanner (Achieva; Philips Medical Systems, Best, The Netherlands) at the Institute for Advanced Biomedical Technologies in Chieti, Italy. The functional images were acquired using T2*-weighted echo-planar images (EPI) with BOLD contrast using SENSE imaging. EPIs comprised of 32 axial slices acquired in ascending order and covering the entire brain with the following protocol: TR = 2000 ms, TE = 3.5 ms, flip angle = 90°, in-plane matrix = 230×230, voxel size = $2.875 \times 2.875 \times 3.5$ mm^3. For each subject, 2 scanning sessions of 10 minutes duration were acquired for resting state and 3 for natural viewing. Each run included 5 dummy volumes, allowing the MRI signal to reach steady state and the subsequent 300 functional volumes were used for the analysis. Eye position was monitored during scanning using a pupil-corneal reflection system at 120 Hz (Iscan, Burlington, MA, USA). A three-dimensional high-resolution T1-weighted image was acquired for anatomical referencing using an MPRAGE sequence with TR = 8.1 ms, TE = 3.7 ms, voxel size = $0.938 \times 0.938 \times 1$ mm^3 at the end of the scanning session.

Pre-processing

Data[1] were preprocessed using SPM8 (Wellcome Department of Cognitive Neurology, London, UK), including slice-timing and head-motion correction (see Methods in [Gilson et al., 2017] for details), co-registration

[1]The data are available at `github.com/MatthieuGilson/EC_estimation`.

between anatomical and mean functional image, and spatial normalization to MNI stereotaxic space (Montreal Neurological Institute, MNI) with a voxel size of $3 \times 3 \times 3$ mm^3. Mean BOLD time series were extracted from $N = 66$ regions of interest (ROIs) of the brain atlas used in [Hagmann et al., 2008] for each recording session. The discarded subjects in the present study are 1, 11 and 19, among the 22 subjects available online (numbered from 0 to 21). The same subjects were discarded in our recent study [Gilson et al., 2017] because of abnormally high BOLD variance.

A.4. Dataset D

This dataset was already analyzed in [Hester et al., 2012, Nandam et al., 2014]. A total of twenty-seven healthy (27), right-handed, 18-35 year old male participants underwent scanning, one testing session per week, for four consecutive weeks. Subjects took at each session —identical capsules in a randomized order— either: methylphenidate 30 mg (MTH), atomoxetine 60 mg (ATM), citalopram 30 mg (CIT) or placebo (dextrose). Participants completed a go/no-go task (GNG) that consisted of six blocks of 225 trials during the fMRI acquisition. Responses were recorded and controlled for each of the six blocks/runs, using the E-prime software (version 1.1; Psychology Software Tools). For further details see Methods in [Hester et al., 2012, Nandam et al., 2014].
The BOLD signals were acquired using a 1.5T Siemens Sonata scanner (Wesley Hospital, Brisbane, Australia) using a gradient-echo pulse sequence and sequential slice acquisition (TR = 2000 ms, TE = 30 ms, flip angle = 90°, 29 contiguous slices of 3 mm thickness, 10 %gap, in-plane resolution of 3.6×3.6 mm in a FOV of 384 mm). A high-resolution T1-weighted isotropic (1 mm^3) structural MPRAGE image was also acquired.

Pre-processing

Dataset D was preprocessed with SPM12 (Wellcome Trust Centre for Neuroimaging, London, UK) and DPARSF/DPABI [Yan et al., 2016]. The

pre-processing pipeline included the same steps applied to Dataset B. One subject was discarded after further analysis due to high movement.

APPENDIX B

Brain parcellations

ROI index	Brain region	Subsystem
1, 2	Calcarine fissure	occipital
3, 4	Cuneus	
5, 6	Lingual gyrus	
7, 8	Superior occipital gyrus	
9, 10	Middle occipital gyrus	
11, 12	Inferior occipital gyrus	
13, 14	Fusiform gyrus	
15, 16	Heschl gyrus	temporal
17, 18	Superior temporal gyrus	
19, 20	Superior temporal pole	
21, 22	Middle temporal gyrus	
23, 24	Middle temporal pole	
25, 26	Inferior temporal gyrus	
27, 28	Precentral gyrus	central
29, 30	Postcentral gyrus	
31, 32	Paracentral lobule	
33, 34	Superior parietal gyrus	parietal
35, 36	Inferior parietal gyrus	
37, 38	Supramarginal gyrus	
39, 40	Angular gyrus	
41, 42	Precuneus	
43, 44	Superior frontal gyrus, dorsolateral	frontal
45, 46	Superior frontal gyrus, orbital part	
47, 48	Middle frontal gyrus	
49, 50	Middle frontal gyrus, orbital part	
51, 52	Inferior frontal gyrus, opercular part	
53, 54	Inferior frontal gyrus, triangular part	
55, 56	Inferior frontal gyrus, orbital part	
57, 58	Rolandic operculum	
59, 60	Supplementary motor area	
61, 62	Olfactory cortex	
63, 64	Superior frontal gyrus, medial	
65, 66	Superior frontal gyrus, medial orbital	
67, 68	Gyrus rectus	
69, 70	Insula	cingulate
71, 72	Anterior cingulate	
73, 74	Median cingulate	
75, 76	Posterior cingulate	
77, 78	Hippocampus	
79, 80	Parahippocampal gyrus	
81, 82	Amygdala	
83, 84	Caudate nucleus	
85, 86	Putamen	
87, 88	Pallidum	
89, 90	Thalamus	
91, 92	Cerebellum crus 1	cerebellum
93, 94	Cerebellum crus 2	
95, 96	Cerebellum 3	
97, 98	Cerebellum 4-5	
99, 100	Cerebellum 6	
101, 102	Cerebellum 7	
103, 104	Cerebellum 8	
105, 106	Cerebellum 9	
107, 108	Cerebellum 10	
109, 110	Vermis 1-2	
111, 112	Vermis 4-5	
113, 114	Vermis 7	
115, 116	Vermis 9	

116 ROI from AAL parcellation

ROI index	Abbreviations	Brain region	Subsystem
1, 2	CUN	Cuneus	occipital
3, 4	PCAL	Pericalcarine cortex	
5, 6	LING	Lingual gyrus	
7, 8	LOCC	Lateral occipital cortex	
9, 10	FUS	Fusiform gyrus	
11, 12	ST	Superior temporal cortex	temporal
13, 14	TT	Transverse temporal cortex	
15, 16	MT	Middle temporal cortex	
17, 18	IT	Inferior temporal cortex	
19, 20	PREC	Precentral gyrus	central
21, 22	PSTC	Postcentral gyrus	
23, 24	PARC	Paracentral lobule	
25, 26	SP	Superior parietal cortex	parietal
27, 28	IP	Inferior parietal cortex	
29, 30	TP	Temporal pole	
31, 32	SMAR	Supramarginal gyrus	
33, 34	BSTS	Bank of the superior temporal sulcus	
35, 36	PCUN	Precuneus	
37, 38	FP	Frontal pole	frontal
39, 40	CMF	Caudal middle frontal cortex	
41, 42	RMF	Rostral middle frontal cortex	
43, 44	PTRI	Pars triangularis	
45, 46	PORB	Pars orbitalis	
47, 48	POPE	Pars opercularis	
49, 50	SF	Superior frontal cortex	
51, 52	LOF	Lateral orbitofrontal cortex	
53, 54	MOF	Medial orbitofrontal cortex	
55, 56	ENT	Entorhinal cortex	cingulate
57, 58	PARH	Parahippocampal cortex	
59, 60	CAC	Caudal anterior cingulate cortex	
61, 62	RAC	Rostral anterior cingulate cortex	
63, 64	PC	Posterior cingulate cortex	
65, 66	ISTC	Isthmus of the cingulate cortex	

66 ROI from Hagmann parcellation

Bibliography

[Airan et al., 2016] Airan, R. D., Vogelstein, J. T., Pillai, J. J., Caffo, B., Pekar, J. J., and Sair, H. I. (2016). Factors affecting characterization and localization of interindividual differences in functional connectivity using MRI. *Human Brain Mapping*.

[Allen et al., 2014] Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., and Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex*, 24(3):663–676.

[Amunts et al., 2014] Amunts, K., Hawrylycz, M. J., Van Essen, D. C., Van Horn, J. D., Harel, N., Poline, J.-B., De Martino, F., Bjaalie, J. G., Dehaene-Lambertz, G., Dehaene, S., Valdes-Sosa, P., Thirion, B., Zilles, K., Hill, S. L., Abrams, M. B., Tass, P. A., Vanduffel, W., Evans, A. C., and Eickhoff, S. B. (2014). Interoperable atlases of the human brain. *Neuroimage*, 99:525–532.

[Anderson et al., 2011] Anderson, J. S., Ferguson, M. A., Lopez-Larson, M., and Yurgelun-Todd, D. (2011). Reproducibility of single-subject functional connectivity measurements. *American Journal of Neuroradiology*.

[Bastos-Leite et al., 2015] Bastos-Leite, A. J., Ridgway, G. R., Silveira, C., Norton, A., Reis, S., and Friston, K. J. (2015). Dysconnectivity within the default mode in first-episode schizophrenia: a stochastic dynamic causal modeling study with functional magnetic resonance imaging. *Schizophr Bull*, 41:144–153.

[Betti et al., 2013] Betti, V., Della Penna, S., de Pasquale, F., Mantini, D., Marzetti, L., Romani, G. L., and Corbetta, M. (2013). Natural scenes viewing alters the dynamics of functional connectivity in the human brain. *Neuron*, 79:782–797.

[Birn et al., 2013] Birn, R. M., Molloy, E. K., Patriat, R., Parker, T., Meier, T. B., Kirk, G. R., Nair, V. A., Meyerand, M. E., and Prabhakaran, V. (2013). The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *NeuroImage*.

[Biswal et al., 1995] Biswal, B., Yetkin, F., Haughton, V., and Hyde, J. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn Reson Med*, 34:537–541.

[Bowyer, 2016] Bowyer, S. M. (2016). Coherence a measure of the brain networks: past and present. *Neuropsychiatric Electrophysiology*.

[Bressler and Tognoli, 2006] Bressler, S. L. and Tognoli, E. (2006). Operational principles of neurocognitive networks. *International Journal of Psychophysiology*.

[Brodersen et al., 2011] Brodersen, K. H., Schofield, T. M., Leff, A. P., Ong, C. S., Lomakina, E. I., Buhmann, J. M., and Stephan, K. E. (2011). Generative embedding for model-based classification of fmri data. *PLoS Comput Biol*, 7:e1002079.

[Buckner et al., 2013] Buckner, R. L., Krienen, F. M., and Yeo, B. T. (2013). Opportunities and limitations of intrinsic functional connectivity MRI.

[Cabral et al., 2017] Cabral, J., Vidaurre, D., Marques, P., Magalhães, R., Silva Moreira, P., Miguel Soares, J., Deco, G., Sousa, N., and Kringelbach, M. L. (2017). Cognitive performance in healthy older adults relates to spontaneous switching between states of functional connectivity during rest. *Scientific Reports*.

[Calhoun et al., 2004] Calhoun, V., Pearlson, G., and Adali, T. (2004). Independent component analysis applied to fMRI data: A generative model for validating results. In *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*.

[Calhoun et al., 2017] Calhoun, V. D., Lawrie, S. M., Mourao-Miranda, J., and Stephan, K. E. (2017). Prediction of individual differences from neuroimaging data. *Neuroimage*, 145:135–136.

[Calhoun et al., 2014] Calhoun, V. D., Miller, R., Pearlson, G., and Adali, T. (2014). The Chronnectome: Time-Varying Connectivity Networks as the Next Frontier in fMRI Data Discovery.

[Catani et al., 2013] Catani, M., Thiebaut de Schotten, M., Slater, D., and Dell'Acqua, F. (2013). Connectomic approaches before the connectome. *NeuroImage*.

[Chang et al., 2013] Chang, C., Liu, Z., Chen, M. C., Liu, X., and Duyn, J. H. (2013). EEG correlates of time-varying BOLD functional connectivity. *NeuroImage*.

[Chang et al., 2015] Chang, L., Gianaros, P., Manuck, S., Krishnan, A., and Wager, T. (2015). A sensitive and specific neural signature for picture-induced negative affect. *PLoS Biol*, 13:e1002180.

[Chen et al., 2015] Chen, B., Xu, T., Zhou, C., Wang, L., Yang, N., Wang, Z., Dong, H.-M., Yang, Z., Zang, Y.-F., Zuo, X.-N., and Weng, X.-C. (2015). Individual variability and test-retest reliability revealed by ten repeated resting-state brain scans over one month. *PLoS One*, 10:e0144963.

[Cohen, 2018] Cohen, J. R. (2018). The behavioral and cognitive relevance of time-varying, dynamic changes in functional connectivity.

[Cordes et al., 2000] Cordes, D., Haughton, V. M., Arfanakis, K., Wendt, G. J., Turski, P. A., Moritz, C. H., Quigley, M. A., and Meyerand, M. E. (2000). Mapping functionally related regions of brain with functional connectivity MR imaging. *Am J Neuroradiol*, 21:1636–1644.

[Da Mota et al., 2014] Da Mota, B., Fritsch, V., Varoquaux, G., Banaschewski, T., Barker, G. J., Bokde, A. L. W., Bromberg, U., Conrod, P., Gallinat, J., Garavan, H., Martinot, J.-L., Nees, F., Paus, T., Pausova, Z., Rietschel, M., Smolka, M. N., Ströhle, A., Frouin, V., Poline, J.-B., Thirion, B., and IMAGEN consortium (2014). Randomized parcellation based inference. *Neuroimage*, 89:203–215.

[Deco et al., 2011] Deco, G., Jirsa, V., and McIntosh, A. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nat Rev Neurosci*, 12:43–56.

[Deco and Kringelbach, 2014] Deco, G. and Kringelbach, M. L. (2014). Great expectations: using whole-brain computational connectomics for understanding neuropsychiatric disorders. *Neuron*, 84:892–905.

[Dinga et al., 2018] Dinga, R., Schmaal, L., Penninx, B., van Tol, M. J., Veltman, D., van Velzen, L., van der Wee, N., and Marquand, A. (2018). Evaluating the evidence for biotypes of depression: attempted replication of drysdale et.al. 2017. *bioRxiv*.

[Drysdale et al., 2017] Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., Fetcho, R. N., Zebley, B., Oathes, D. J., Etkin, A., Schatzberg, A. F., Sudheimer, K., Keller, J., Mayberg, H. S., Gunning, F. M., Alexopoulos, G. S., Fox, M. D., Pascual-Leone, A., Voss, H. U., Casey, B. J., Dubin, M. J., and Liston, C. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat Med*, 23:28–38.

[Eickhoff et al., 2015] Eickhoff, S. B., Thirion, B., Varoquaux, G., and Bzdok, D. (2015). Connectivity-based parcellation: Critique and implications. *Human Brain Mapping*, 36(12):4771–4792.

[Ekstrom, 2010] Ekstrom, A. (2010). How and when the fMRI BOLD signal relates to underlying neural activity: the danger in dissociation. *Brain Res Rev*, 62:233–244.

[Engel et al., 2013] Engel, A. K., Gerloff, C., Hilgetag, C. C., and Nolte, G. (2013). Intrinsic coupling modes: multiscale interactions in ongoing brain activity. *Neuron*, 80:867–886.

[Filevich et al., 2017] Filevich, E., Lisofsky, N., Becker, M., Butler, O., Lochstet, M., Martensson, J., Wenger, E., Lindenberger, U., and Kühn, S. (2017). Day2day: investigating daily variability of magnetic resonance imaging measures over half a year. *BMC Neurosci*, 18:65.

[Finn et al., 2017] Finn, E. S., Scheinost, D., Finn, D. M., Shen, X., Papademetris, X., and Constable, R. T. (2017). Can brain state be manipulated to emphasize individual differences in functional connectivity? *Neuroimage*, 160:140–151.

[Finn et al., 2015] Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., and Constable, R. T. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat Neurosci*, 18:1664–1671.

[Frässle et al., 2015] Frässle, S., Stephan, K. E., Friston, K. J., Steup, M., Krach, S., Paulus, F. M., and Jansen, A. (2015). Test-retest reliability of dynamic causal modeling for fmri. *Neuroimage*, 117:56–66.

[Fries, 2005] Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci*, 9:474–480.

[Friston, 2011] Friston, K. (2011). Functional and effective connectivity: A review. *Brain Connect*, 1:8.

[Friston et al., 2003] Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *Neuroimage*, 19:1273–1302.

[Friston et al., 2014] Friston, K. J., Kahan, J., Biswal, B., and Razi, A. (2014). A DCM for resting state fMRI. *Neuroimage*, 94:396–407.

[Gilson et al., 2017] Gilson, M., Deco, G., Friston, K., Hagmann, P., Mantini, D., Betti, V., Romani, G. L., and Corbetta, M. (2017). Effective connectivity inferred from fmri transition dynamics during movie viewing points to a balanced reconfiguration of cortical interactions. *NeuroImage*.

[Gilson et al., 2016] Gilson, M., Moreno-Bote, R., Ponce-Alvarez, A., Ritter, P., and Deco, G. (2016). Estimation of directed effective connectivity from fMRI functional connectivity hints at asymmetries of cortical connectome. *PLoS Comput Biol*, 12:e1004762.

[Glasser et al., 2016] Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., and Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536:171–178.

[Goebel et al., 2003] Goebel, R., Roebroeck, A., Kim, D., and Formisano, E. (2003). Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magn Reson Imaging*, 21:1251–1261.

[Gonzalez-Castillo and Bandettini, 2018] Gonzalez-Castillo, J. and Bandettini, P. A. (2018). Task-based dynamic functional connectivity: Recent findings and open questions.

[Gonzalez-Castillo et al., 2015] Gonzalez-Castillo, J., Hoy, C. W., Handwerker, D. A., Robinson, M. E., Buchanan, L. C., Saad, Z. S., and Bandettini, P. A. (2015). Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. *Proc Natl Acad Sci U S A*, 112:8762–8767.

[Gordon et al., 2017] Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H., Hampton, J. M., Coalson, R. S., Nguyen, A. L., McDermott, K. B., Shimony, J. S., Snyder, A. Z., Schlaggar, B. L., Petersen, S. E., Nelson, S. M., and Dosenbach, N. U. F. (2017). Precision functional mapping of individual human brains. *Neuron*, 95:791–807.

[Gratton et al., 2018] Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., Nelson, S. M., Coalson, R. S., Snyder, A. Z., Schlaggar, B. L., Dosenbach, N. U., and Petersen, S. E. (2018). Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. *Neuron*.

[Greene et al., 2018] Greene, A. S., Gao, S., Scheinost, D., and Constable, R. T. (2018). Task-induced brain state manipulation improves prediction of individual traits. *Nature Communications*.

[Greicius, 2008] Greicius, M. (2008). Resting-state functional connectivity in neuropsychiatric disorders. *Curr Opin Neurol*, 21:424–430.

[Guyon et al., 2002] Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1):389–422.

[Hagmann et al., 2008] Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., and Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol*, 6:e159.

[Haimovici et al., 2017] Haimovici, A., Tagliazucchi, E., Balenzuela, P., and Laufs, H. (2017). On wakefulness fluctuations as a source of BOLD functional connectivity dynamics. *Scientific Reports*.

[Handwerker et al., 2012] Handwerker, D. A., Roopchansingh, V., Gonzalez-Castillo, J., and Bandettini, P. A. (2012). Periodic changes in fMRI connectivity. *NeuroImage*.

[Hansen et al., 2015] Hansen, E., Battaglia, D., Spiegler, A., Deco, G., and Jirsa, V. (2015). Functional connectivity dynamics: modeling the switching behavior of the resting state. *Neuroimage*, 105:525–535.

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

[Hester et al., 2012] Hester, R., Nandam, L. S., O'Connell, R. G., Wagner, J., Strudwick, M., Nathan, P. J., Mattingley, J. B., and Bellgrove, M. A. (2012). Neurochemical Enhancement of Conscious Error Awareness. *Journal of Neuroscience*.

[Hindriks et al., 2016] Hindriks, R., Adhikari, M. H., Murayama, Y., Ganzetti, M., Mantini, D., Logothetis, N. K., and Deco, G. (2016). Can sliding-window correlations reveal dynamic functional connectivity in resting-state fMRI? *NeuroImage*.

[Hipp et al., 2011] Hipp, J. F., Engel, A. K., and Siegel, M. (2011). Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron*, 69:387–396.

[Hlinka et al., 2011] Hlinka, J., Paluš, M., Vejmelka, M., Mantini, D., and Corbetta, M. (2011). Functional connectivity in resting-state fMRI: Is linear correlation sufficient? *NeuroImage*.

[Hoptman et al., 2012] Hoptman, M. J., Zuo, X.-N., D'Angelo, D., Mauro, C. J., Butler, P. D., Milham, M. P., and Javitt, D. C. (2012). Decreased interhemispheric coordination in schizophrenia: a resting state fmri study. *Schizophr Res*, 141:1–7.

[Hughes, 1968] Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14:55–63.

[Hutchison et al., 2013] Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., Della Penna, S., Duyn,

J. H., Glover, G. H., Gonzalez-Castillo, J., Handwerker, D. A., Keilholz, S., Kiviniemi, V., Leopold, D. A., de Pasquale, F., Sporns, O., Walter, M., and Chang, C. (2013). Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, 80:360–378.

[Kaufmann et al., 2017] Kaufmann, T., Alnæs, D., Doan, N. T., Brandt, C. L., Andreassen, O. A., and Westlye, L. T. (2017). Delayed stabilization and individualization in connectome development are related to psychiatric disorders. *Nat Neurosci*, 20:513–515.

[Kim et al., 2011] Kim, S., Putrino, D., Ghosh, S., and Brown, E. N. (2011). A Granger causality measure for point process models of ensemble neural spiking activity. *PLoS Computational Biology*.

[Kong et al., 2018] Kong, R., Li, J., Orban, C., Sabuncu, M. R., Liu, H., Schaefer, A., Sun, N., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., and Yeo, B. T. T. (2018). Spatial Topography of Individual-Specific Cortical Networks Predicts Human Cognition, Personality, and Emotion. *Cerebral Cortex*.

[Kottaram et al., 2018] Kottaram, A., Johnston, L., Ganella, E., Pantelis, C., Kotagiri, R., and Zalesky, A. (2018). Spatio-temporal dynamics of resting-state brain networks improve single-subject prediction of schizophrenia diagnosis. *Human Brain Mapping*.

[Kurth et al., 2015] Kurth, S., Moyse, E., Bahri, M. A., Salmon, E., and Bastin, C. (2015). Recognition of personally familiar faces and functional connectivity in alzheimer's disease. *Cortex*, 67:59–73.

[Laumann et al., 2017] Laumann, T. O., Snyder, A. Z., Mitra, A., Gordon, E. M., Gratton, C., Adeyemo, B., Gilmore, A. W., Nelson, S. M., Berg, J. J., Greene, D. J., McCarthy, J. E., Tagliazucchi, E., Laufs, H., Schlaggar, B. L., Dosenbach, N. U., and Petersen, S. E. (2017). On the Stability of BOLD fMRI Correlations. *Cerebral Cortex*.

[Leonardi et al., 2013] Leonardi, N., Richiardi, J., Gschwind, M., Simioni, S., Annoni, J. M., Schluep, M., Vuilleumier, P., and Van De

Ville, D. (2013). Principal components of functional connectivity: A new approach to study dynamic brain connectivity during rest. *NeuroImage*.

[Leonardi and Van De Ville, 2015] Leonardi, N. and Van De Ville, D. (2015). On spurious and real fluctuations of dynamic functional connectivity during rest.

[Li et al., 2012] Li, B., Wang, X., Yao, S., Hu, D., and Friston, K. (2012). Task-dependent modulation of effective connectivity within the default mode network. *Front Psychol*, 3:206.

[Liégeois et al., 2017] Liégeois, R., Laumann, T. O., Snyder, A. Z., Zhou, J., and Yeo, B. T. (2017). Interpreting temporal fluctuations in resting-state functional connectivity MRI.

[Mantini et al., 2012] Mantini, D., Hasson, U., Betti, V., Perrucci, M. G., Romani, G. L., Corbetta, M., Orban, G. A., and Vanduffel, W. (2012). Interspecies activity correlations reveal functional correspondence between monkey and human brain areas. *Nat Methods*, 9:277–282.

[Matthews and Hampshire, 2016] Matthews, P. M. and Hampshire, A. (2016). Clinical concepts emerging from fmri functional connectomics. *Neuron*, 91:511–528.

[Medaglia et al., 2015] Medaglia, J. D., Lynall, M. E., and Bassett, D. S. (2015). Cognitive network neuroscience. *Journal of Cognitive Neuroscience*.

[Messé et al., 2014] Messé, A., Rudrauf, D., Benali, H., and Marrelec, G. (2014). Relating structure and function in the human brain: relative contributions of anatomy, stationary dynamics, and non-stationarities. *PLoS COmput Biol*, 10:e1003530.

[Miranda-Dominguez et al., 2014] Miranda-Dominguez, O., Mills, B. D., Carpenter, S. D., Grant, K. A., Kroenke, C. D., Nigg, J. T., and Fair,

D. A. (2014). Connectotyping: model based fingerprinting of the functional connectome. *PLoS One*, 9:e111048.

[Mitra et al., 2015] Mitra, A., Snyder, A. Z., Tagliazucchi, E., Laufs, H., and Raichle, M. (2015). Propagated infra-slow intrinsic brain activity reorganizes across wake and slow wave sleep. *Elife*, 4.

[Mueller et al., 2015] Mueller, S., Wang, D., Fox, M. D., Pan, R., Lu, J., Li, K., Sun, W., Buckner, R. L., and Liu, H. (2015). Reliability correction for functional connectivity: Theory and implementation. *Hum Brain Mapp*, 36:4664–4680.

[Mueller et al., 2013] Mueller, S., Wang, D., Fox, M. D., Yeo, B. T., Sepulcre, J., Sabuncu, M. R., Shafee, R., Lu, J., and Liu, H. (2013). Individual Variability in Functional Connectivity Architecture of the Human Brain. *Neuron*.

[Nandam et al., 2014] Nandam, L. S., Hester, R., and Bellgrove, M. A. (2014). Dissociable and common effects of methylphenidate, atomoxetine and citalopram on response inhibition neural networks. *Neuropsychologia*.

[Orlandi et al., 2014] Orlandi, J. G., Stetter, O., Soriano, J., Geisel, T., and Battaglia, D. (2014). Transfer entropy reconstruction and labeling of neuronal connections from simulated calcium imaging. *PLoS ONE*.

[Pannunzi et al., 2017] Pannunzi, M., Hindriks, R., Bettinardi, R. G., Wenger, E., Lisofsky, N., Martensson, J., Butler, O., Filevich, E., Becker, M., Lochstet, M., Kühn, S., and Deco, G. (2017). Resting-state fmri correlations: From link-wise unreliability to whole brain stability. *Neuroimage*, 157:250–262.

[Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

[Poldrack et al., 2009] Poldrack, R. A., Halchenko, Y. O., and Hanson, S. J. (2009). Decoding the large-scale structure of brain function by classifying mental states across individuals. *Psychol Sci*, 20:1364–1372.

[Preti et al., 2017] Preti, M. G., Bolton, T. A., and Van De Ville, D. (2017). The dynamic functional connectome: State-of-the-art and perspectives. *Neuroimage*, 160:41–54.

[Rahim et al., 2017] Rahim, M., Thirion, B., Bzdok, D., Buvat, I., and Varoquaux, G. (2017). Joint prediction of multiple scores captures better individual traits from brain images. *Neuroimage*, 158:145–154.

[Raichle et al., 2001] Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc Natl Acad Sci U S A*, 98:676–682.

[Rashid et al., 2018] Rashid, B., Blanken, L. M., Muetzel, R. L., Miller, R., Damaraju, E., Arbabshirani, M. R., Erhardt, E. B., Verhulst, F. C., van der Lugt, A., Jaddoe, V. W., Tiemeier, H., White, T., and Calhoun, V. (2018). Connectivity dynamics in typical development and its relationship to autistic traits and autism spectrum disorder. *Human Brain Mapping*.

[Rissman and Wagner, 2012] Rissman, J. and Wagner, A. D. (2012). Distributed representations in memory: insights from functional brain imaging. *Annu Rev Psychol*, 63:101–128.

[Rosenberg et al., 2015] Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., and Chun, M. M. (2015). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*.

[Rousseeuw, 1987] Rousseeuw, P. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

[Satterthwaite et al., 2012] Satterthwaite, T. D., Wolf, D. H., Loughead, J., Ruparel, K., Elliott, M. A., Hakonarson, H., Gur, R. C., and Gur, R. E. (2012). Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. *NeuroImage*.

[Seghier and Price, 2018] Seghier, M. L. and Price, C. J. (2018). Interpreting and Utilising Intersubject Variability in Brain Function.

[Senden et al., 2017] Senden, M., Reuter, N., van den Heuvel, M. P., Goebel, R., Deco, G., and Gilson, M. (2017). Task-related effective connectivity reveals that the cortical rich club gates cortex-wide communication. *Hum Brain Mapp*.

[Shah et al., 2016] Shah, L. M., Cramer, J. A., Ferguson, M. A., Birn, R. M., and Anderson, J. S. (2016). Reliability and reproducibility of individual differences in functional connectivity acquired during task and resting state. *Brain and Behavior*.

[Shannon et al., 2013] Shannon, B. J., Dosenbach, R. A., Su, Y., Vlassenko, A. G., Larson-Prior, L. J., Nolan, T. S., Snyder, A. Z., and Raichle, M. E. (2013). Morning-evening variation in human brain metabolism and memory circuits. *Journal of Neurophysiology*.

[Sheehan et al., 1998] Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., and Dunbar, G. C. (1998). The mini-international neuropsychiatric interview (m.i.n.i.): the development and validation of a structured diagnostic psychiatric interview for dsm-iv and icd-10. *J Clin Psychiatry*, 59:22–33.

[Shehzad et al., 2009] Shehzad, Z., Kelly, A. M. C., Reiss, P. T., Gee, D. G., Gotimer, K., Uddin, L. Q., Lee, S. H., Margulies, D. S., Roy, A. K., Biswal, B. B., Petkova, E., Castellanos, F. X., and Milham, M. P. (2009). The resting brain: unconstrained yet reliable. *Cereb Cortex*, 19:2209–2229.

[Shen et al., 2013] Shen, X., Tokoglu, F., Papademetris, X., and Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage*, 82:403–415.

[Shirer et al., 2012] Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V., and Greicius, M. D. (2012). Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral Cortex*.

[Shrout and Fleiss, 1979] Shrout, P. E. and Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*.

[Siegel et al., 2017] Siegel, J. S., Mitra, A., Laumann, T. O., Seitzman, B. A., Raichle, M., Corbetta, M., and Snyder, A. Z. (2017). Data quality influences observed links between functional connectivity and behavior. *Cerebral Cortex*.

[Smith, 2012] Smith, S. M. (2012). The future of FMRI connectivity.

[Smith et al., 2011] Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., Ramsey, J. D., and Woolrich, M. W. (2011). Network modelling methods for FMRI. *NeuroImage*.

[Stephan et al., 2004] Stephan, K., Harrison, L., Penny, W., and Friston, K. (2004). Biophysical models of fMRI responses. *Curr Opin Neurol*, 14:629–635.

[Tagliazucchi et al., 2012] Tagliazucchi, E., von Wegner, F., Morzelewski, A., Brodbeck, V., and Laufs, H. (2012). Dynamic BOLD functional connectivity in humans and its electrophysiological correlates. *Frontiers in Human Neuroscience*.

[Tavor et al., 2016] Tavor, I., Jones, O. P., Mars, R. B., Smith, S. M., Behrens, T. E., and Jbabdi, S. (2016). Task-free MRI predicts individual differences in brain activity during task performance. *Science*.

[Thompson et al., 2013] Thompson, G. J., Merritt, M. D., Pan, W. J., Magnuson, M. E., Grooms, J. K., Jaeger, D., and Keilholz, S. D. (2013). Neural correlates of time-varying functional connectivity in the rat. *NeuroImage*.

[Tononi, 2005] Tononi, G. (2005). Consciousness, information integration, and the brain.

[Tononi et al., 1994] Tononi, G., Sporns, O., and Edelman, G. M. (1994). A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences*.

[Tzourio-Mazoyer et al., 2002] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15:273–289.

[van den Heuvel and Hulshoff Pol, 2010] van den Heuvel, M. P. and Hulshoff Pol, H. E. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity.

[Van Dijk et al., 2010] Van Dijk, K. R. A., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W., and Buckner, R. L. (2010). Intrinsic Functional Connectivity As a Tool For Human Connectomics: Theory, Properties, and Optimization. *Journal of Neurophysiology*.

[Vanderwal et al., 2017] Vanderwal, T., Eilbott, J., Finn, E. S., Craddock, R. C., Turnbull, A., and Castellanos, F. X. (2017). Individual differences in functional connectivity during naturalistic viewing conditions. *Neuroimage*, 157:521–530.

[Varela et al., 2001] Varela, F., Lachaux, J. P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*.

[Varoquaux et al., 2017] Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., and Thirion, B. (2017). Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *Neuroimage*, 145:166–179.

[Waller et al., 2017] Waller, L., Walter, H., Kruschwitz, J. D., Reuter, L., Müller, S., Erk, S., and Veer, I. M. (2017). Evaluating the replicability, specificity, and generalizability of connectome fingerprints. *NeuroImage*.

[Woo et al., 2017] Woo, C.-W., Chang, L. J., Lindquist, M. A., and Wager, T. D. (2017). Building better biomarkers: brain models in translational neuroimaging. *Nat Neurosci*, 20:365–377.

[Xia et al., 2018] Xia, C. H., Ma, Z., Ciric, R., Gu, S., Betzel, R. F., Kaczkurkin, A. N., Calkins, M. E., Cook, P. A., García de la Garza, A., Vandekar, S. N., Cui, Z., Moore, T. M., Roalf, D. R., Ruparel, K., Wolf, D. H., Davatzikos, C., Gur, R. C., Gur, R. E., Shinohara, R. T., Bassett, D. S., and Satterthwaite, T. D. (2018). Linked dimensions of psychopathology and connectivity in functional brain networks. *Nature Communications*.

[Xie et al., 2017] Xie, H., Calhoun, V. D., Gonzalez-Castillo, J., Damaraju, E., Miller, R., Bandettini, P. A., and Mitra, S. (2017). Whole-brain connectivity dynamics reflect both task-specific and individual-specific modulation: A multitask study. *Neuroimage*.

[Yan et al., 2016] Yan, C.-G., Wang, X.-D., Zuo, X.-N., and Zang, Y.-F. (2016). Dpabi: Data processing & analysis for (resting-state) brain imaging. *Neuroinformatics*, 14:339–351.

[Zuo et al., 2014] Zuo, X.-N., Anderson, J. S., Bellec, P., Birn, R. M., Biswal, B. B., Blautzik, J., Breitner, J. C. S., Buckner, R. L., Calhoun, V. D., Castellanos, F. X., Chen, A., Chen, B., Chen, J., Chen, X., Colcombe, S. J., Courtney, W., Craddock, R. C., Di Martino, A., Dong, H.-M., Fu, X., Gong, Q., Gorgolewski, K. J., Han, Y., He, Y., He, Y.,

Ho, E., Holmes, A., Hou, X.-H., Huckins, J., Jiang, T., Jiang, Y., Kelley, W., Kelly, C., King, M., LaConte, S. M., Lainhart, J. E., Lei, X., Li, H.-J., Li, K., Li, K., Lin, Q., Liu, D., Liu, J., Liu, X., Liu, Y., Lu, G., Lu, J., Luna, B., Luo, J., Lurie, D., Mao, Y., Margulies, D. S., Mayer, A. R., Meindl, T., Meyerand, M. E., Nan, W., Nielsen, J. A., O'Connor, D., Paulsen, D., Prabhakaran, V., Qi, Z., Qiu, J., Shao, C., Shehzad, Z., Tang, W., Villringer, A., Wang, H., Wang, K., Wei, D., Wei, G.-X., Weng, X.-C., Wu, X., Xu, T., Yang, N., Yang, Z., Zang, Y.-F., Zhang, L., Zhang, Q., Zhang, Z., Zhang, Z., Zhao, K., Zhen, Z., Zhou, Y., Zhu, X.-T., and Milham, M. P. (2014). An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci Data*, 1:140049.