Title: Behavioural relevance of spontaneous, transient brain network interactions in fMRI

Authors: D. Vidaurre*, A. Llera, S.M. Smith and M.W. Woolrich

Abstract: How spontaneously fluctuating functional magnetic resonance imaging (fMRI) signals in different brain regions relate to behaviour has been an open question for decades. Correlations in these signals, known as functional connectivity, can be averaged over several minutes of data to provide a stable representation of the functional network architecture for an individual. However, associations between these stable features and behavioural traits have been shown to be dominated by individual differences in anatomy. Here, we show that time-varying fMRI functional connectivity, detected at time-scales of several seconds, has associations with behavioural traits that are not dominated by anatomy. We found that this was the case for several trait groups, despite the fact that average functional connectivity accounts for the largest proportion of variability in the fMRI signal between individuals. The finding that time-varying fMRI functional connectivity has a unique relationship to population behavioural variability implies that it might reflect transient neuronal communication fluctuating around a stable neural architecture.

Introduction

The emergence of large-scale distributed networks in spontaneous brain activity as measured by functional magnetic resonance imaging (fMRI) is a widely studied phenomenon (Biswal et al., 1995; Fox and Raichle, 2007). These networks have been consistently identified using cross-regional temporal correlations – referred to as functional connectivity (FC) (Beckmann et al., 2005; Damoiseaux et al., 2006; Smith et al., 2013; Hipp and Siegel, 2015). Typically, FC is estimated by averaging over several minutes of data (e.g. across a scanning session) to provide a stable representation of the functional network architecture for an individual (Finn et al., 2015). This *average FC* has previously been associated with mental performance (Hampson et al., 2006; Hasson et al., 2009) and, more generally, to widespread behavioural phenotypes (Smith et al., 2015). However, there is evidence that these associations are to a large extent driven by structural differences between subjects (Bijsterbosch et al., 2018; Llera et al., 2019). We hypothesised that, while average FC appears to be dominated by structural information, the temporal deviations of FC might not, and could thereby have a distinct relationship with behaviour. This would provide evidence that the time-varying fMRI FC can reflect momentary neuronal communication fluctuating around a stable functional architecture, and might be related to dynamic elements of cognition such as attention and thinking (Smallwood and Schooler, 2015; Kucyi, 2017).

While there is clear evidence that electrophysiologically-derived FC relates to momentary mental states (Palva and Palva, 2012; Hipp et al., 2011; O'Neill et al., 2017; Quinn et al., 2018), whether or not dynamic changes in fMRI-derived FC reflects distinct and transient patterns of modes of communication between neuronal populations is still controversial (Gratton et al., 2018). In the absence of stimuli, known actions or any ground truth, discerning whether time-varying FC carries biological meaning in the resting state is not straightforward (Lurie et al., 2018; Kucyi et al., 2018). One possibility is to use indirect behavioural correlates, for example, by assessing the extent to which FC prior to task onset influences task performance (Sadaghiani et al., 2015), quantifying how the execution of a task induces differences in subsequent resting-state FC (Waites et al., 2005), or using a low demanding task with well-defined behavioural information as a surrogate of actual resting-state (Kucyi et al., 2017). However, these are

normally subtle effects, and other researchers have reported little or no differences in FC between task and rest (Hampson et al., 2006; Gratton et al., 2018).

Here, we take a different route, by relating time-varying FC to population variability in behavioural traits. For this purpose, we combined subject-specific estimates of time-varying fMRI FC, average FC and structural connectivity – for the first time considering these three elements in one single analysis. Using different groups of behavioural traits, we explored the relationship between time-varying FC and population behaviour, after accounting for the explanatory power of the stable subject-specific average FC and structural connectivity features. We reasoned that if time-varying fMRI FC represents biologically meaningful communication between neuronal populations, then it should be capable of accounting for aspects of the subjects' behavioural phenotypes not explained by the average FC or the structural connectivity information. We also examined the relationship between time-varying FC and a separate group of anatomical traits, further confirming that time-varying FC was fundamentally unrelated to the anatomy.

To measure time-varying fMRI FC, we used a state-based model where each state is associated with a specific pattern of FC (Vidaurre et al., 2017), such that instantaneous changes in FC are manifested by a change of state. This approach is based on a version of the Hidden Markov model (HMM) that, in comparison to previous versions of the model used on fMRI (Vidaurre et al., 2017), emphasises changes in FC over changes in amplitude. To measure structural connectivity, we used fractional anisotropy (FA; Basser and Pierpaoli, 1996), mean diffusivity (MD; Basser et al., 1994) and anisotropy mode (MO; Ennis and Kindlmann, 2006). Using the time-varying FC estimated separately for each subject, we devised a test specifically targeted to answer the proposed question. This is an adaptation of the Mantel test, a statistical procedure widely used in the field of ecology (Mantel, 1967). In particular, we explicitly tested the hypothesis: "Time-varying FC explains aspects of the subject traits that cannot be explained by average FC or structural connectivity information alone".

Results

We used 1003 subjects' resting-state fMRI data from the Human Connectome Project (HCP; Smith et al, 2013), where each subject underwent four 15-min sessions (two per day). TR was 750ms. We used a data-driven parcellation obtained through spatial independent component analysis (ICA), and extracted 50 components (Beckmann et al., 2009). All the time series were then fed to a Hidden Markov model (HMM), which provided a group-level description of the time varying FC in the data (Vidaurre et al, 2017; Vidaurre et al, 2018b). Five repetitions of the inference were conducted. In this case, the HMM was designed to emphasise periods in time with distinct FC. Specifically, the HMM described the fMRI time series using a discrete number of states, each endowed solely with a covariance matrix, and time series of state activation probabilities, referred to as the state time courses. Eight states were used in this case. In this model, the states are shared across subjects, whereas the state time courses are subject-specific. In order to get a fully subject-specific description of time-varying FC, we then performed a process of *dual-estimation* to obtain subject-specific versions of the group-level HMM state estimations (see Methods), where each subject has their own set of state descriptions (e.g., FC matrices) as well as their own state time courses.

Fig. 1A illustrates the time-varying FC variability as described by the HMM at the scale of seconds from a given 15min scanning session, including the FC estimation for two of the states. **Fig. 1B** shows the distribution of dwell (or life) times for two of the states. As represented conceptually in **Fig. 1C** and shown in the Methods section, the HMM contains information about both the average FC (stable, subject specific variability) and time-varying FC (within-

session variability). The group-level HMM contains part of the average FC information, whereas the subject-specific, dual-estimated HMMs (having overall many more parameters) contains the entirety of the average FC information as well as information of its time-point by time-point variability. Both the group-level and subject-specific aspects of the modelling contain information about the time-varying FC.

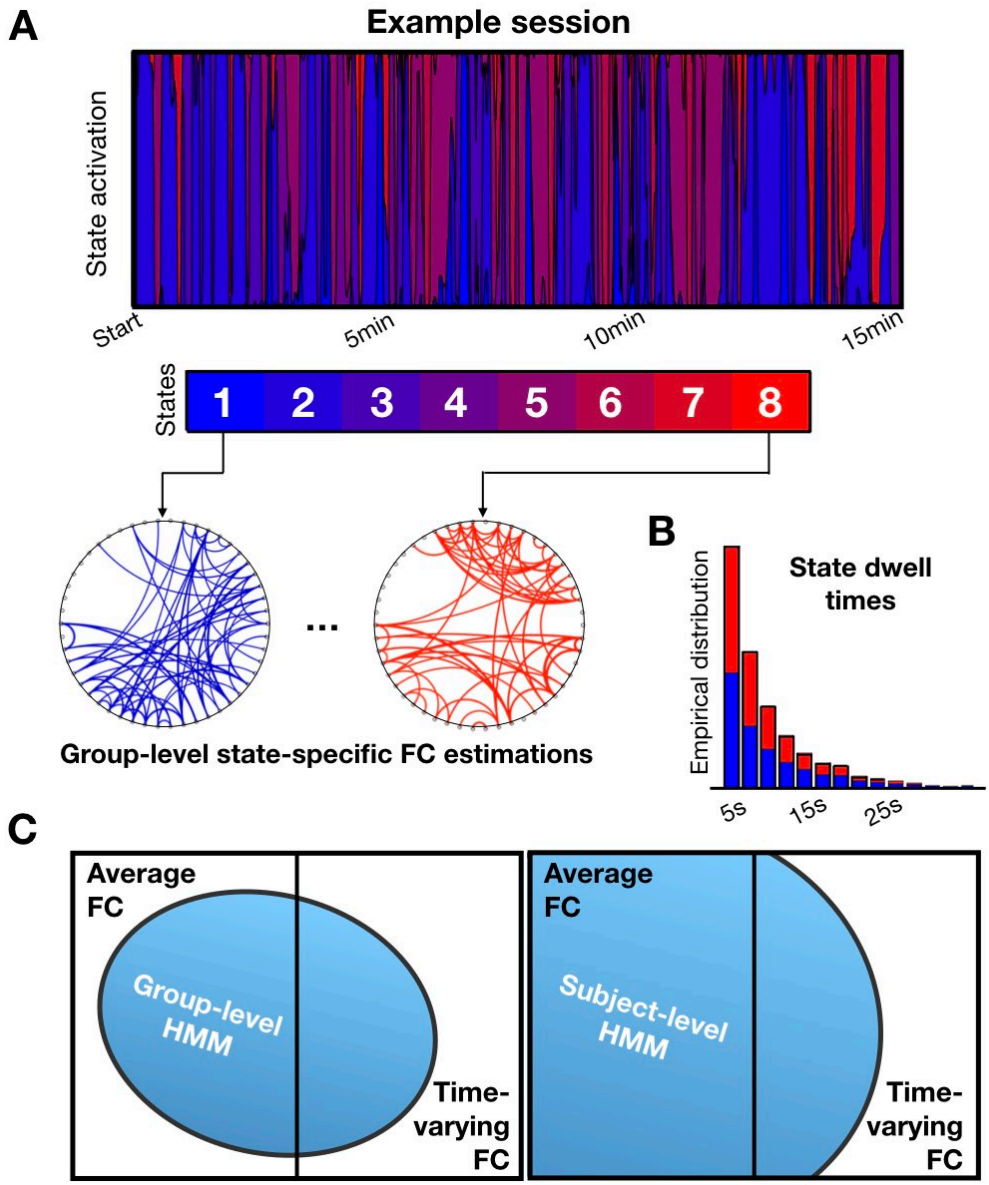

Fig. 1. **(A)** Illustration of the HMM description of time-varying FC, for an example session; states are represented as connectivity matrices. **(B)** Depiction of the distribution of state dwell times for two of the states, at the group level. **(C)** Venn diagram conceptually illustrating the information contained in the group-level and subject-level HMM estimations with respect to average and time-varying FC; the dual-estimated (subject) HMM estimations contain all the average FC information, whereas the group-level HMM contains only part.

Time-varying FC explains behaviour not explained by average FC

Given that the HMM also contains information about the average FC, neither proving that the FC temporal variability is distinct from the average FC (see SI), nor finding that the HMM parameters can explain some aspects of behaviour, are sufficient evidence of time-varying FC

being behaviourally informative. In order to unambiguously discern whether or not time-varying FC explains behaviour in a unique manner, we computed (N by N, where N is the number of subjects) distance matrices (DM), where the distances are calculated according to how different a specific descriptor (e.g., average FC or dual-estimated HMM) is between each pair of subjects, yielding an average FC-DM and an HMM-DM. Next, from the battery of variables provided by the HCP, we selected six different groups of behavioural traits: demographic, cognitive, affective, personality, anatomical and sleep-related (**Table SI-1**). After adjusting for sex and motion, and imputing missing values, we computed behavioural-DM for each of these groups. Finally, we used a partialled version of the Mantel algorithm (Mantel, 1967) to assess: (i) whether each of the behavioural variable groups has a significant association with time-varying FC above and beyond the variance explained by average FC; and, conversely, (ii) whether average FC explains behaviour above and beyond time-varying FC. This is done by regressing the DM which we wish to correct for (i.e. the average FC-DM) out of the DMs that are to be compared (i.e. the HMM-DM and the behavioural-DMs, see Methods).

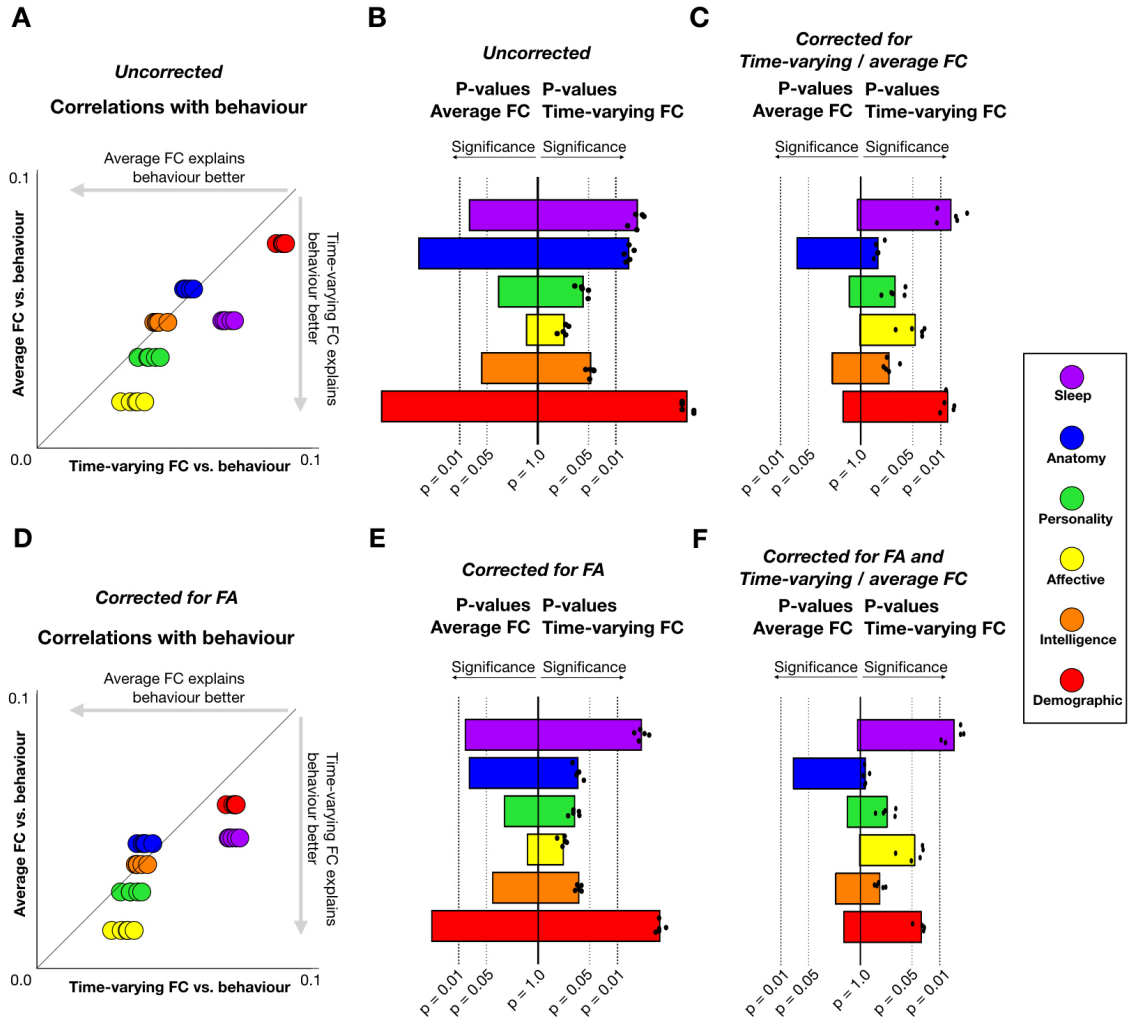

Fig 2. Time-varying FC explains behaviour above and beyond average FC. **(A)** Unadjusted correlations between average FC and behaviour, against correlations between time-varying FC and behaviour. **(B)** Statistical tests on whether time-varying FC explains behaviour (right bars), and whether average FC explains behaviour (left bars); dots represent independent runs of the HMM and lines represent significance levels. **(C)** Statistical tests on whether time-varying FC explains behaviour above and beyond average FC (right bars); and on whether average FC explains behaviour above and beyond time-varying FC (left bars). **(D)** Same as in A, but now corrected for structural connectivity information in the form of fractional anisotropy (FA). **(E)** Statistical tests corrected for FA. **(F)** Same tests than in C, but now corrected by FA.

Fig. 2 presents how the between-subject differences in time-varying FC compare to between-subject differences in average FC in terms of explaining behaviour. **Fig. 2A** shows this comparison in terms of the (uncorrected) correlations between the brain measures DM's and the behavioural-DMs, with one marker per repetition of the HMM inference; note that there is no variability in the Y-axis because the average FC estimation is deterministic. This panel shows that time-varying FC correlates more strongly than the average FC with all the groups of behavioural (non-anatomical) traits, except for the intelligence-related traits, where it performs similarly. **Fig. 2B** presents the results of the Mantel test for these correlations, for time-varying FC (right side of the bars; dots correspond to individual runs of the HMM and bars are averages), and average FC (left side of the bars). The vertical black dotted lines represent significance levels. Using the partialled Mantel algorithm, **Fig. 2C** shows the extent to which there is significant behavioural variability explained by the time-varying FC estimations above and beyond average FC (right), and the extent to which there is significant behavioural variability explained by the average FC descriptor above and beyond time-varying FC (left). These results confirm that time-varying FC carries information relevant to behaviour on top of the stable subject FC traits, although the extent to which it does depends on the specific aspect of behaviour being examined. For example, none of the tests achieved significance for the intelligence-related traits, which seemed to correlate similarly with average and time-varying FC. The demographic traits, on the other hand, were strongly represented by the unique information contained in the time-varying FC descriptors. The affective traits were also more correlated with time-varying FC, which was able to achieve (moderate) significance once the average FC was discounted. Interestingly, time-varying FC was extremely informative for the traits relating to sleep quality, suggesting that the symptoms of insomnia and poor sleep quality are well described in terms of time-varying FC in ways that are not evident in the average (see Stevner et al. 2019 for related discussion on data-driven models of sleep using similar models).

An important aspect known to confound the relationship between FC and behavioural traits is the anatomical features of the subjects. First, variability in the specific anatomical arrangement of the different brain regions is known to significantly influence the estimation of variability across subjects in FC (Bijsterbosch et al. 2018). Second, much of the covariation between average FC and behaviour (as Smith et al., 2015) can be explained by structural variations (Llera et al., 2019). In order to assess the actual influence of the anatomy on the different brain activity descriptors, we specifically related these to a DM constructed using all the FreeSurfer summary statistics provided by the HCP. These are denoted as the "anatomy" group. **Fig. 2C** shows that the average FC is comparably more related to the anatomy than the time-varying FC estimation. This suggests that the temporal aspects of network activity that are captured by the HMM (but not by the average FC) are less confounded by anatomical differences than the spatial aspects.

Time-varying FC explains aspects of behaviour not explained by structural connectivity

In a recent study, Llera et al. (2019) suggested that structural variability across subjects may underlie the main axis of covariation between average FC and population behaviour previously reported by Smith et al. (2015). Fractional anisotropy (FA; Basser and Pierpaoli, 1996), a partial measure of structural connectivity, was found to be among the most explanatory structural features. In order to discern how much of the relationship between time-varying FC and behaviour is ultimately mediated by structural connectivity information, we deconfounded FA from the behavioural traits (see Methods) and repeated the above analyses in **Fig. 2A-C**. **Fig. SI-2** shows the amount of (cross-validated) variance explained by FA for each trait: FA

was particularly explanatory of several demographic and intelligence-related traits, as well as most traits within the anatomical group. **Fig. 2D-F**, with the same interpretation as **Fig. 2A-C**, show the results of relating the FA-deconfounded behavioural traits to average and time-varying FC. As observed, the relationship of time-varying FC with the affective, demographic, and sleep-related traits remains significant, yet with some loss of power for the demographic behavioural group. Likewise, the average FC descriptor decreases its correlation with the anatomical DM, without much alteration of the corresponding p-values.

We also repeated the same analysis for mean diffusivity (MD) and anisotropy mode (MO), which offer a complementary perspective to FA with regards to white matter organisation (Basser et al., 1994; Ennis and Kindlmann, 2006). As shown in **Fig. SI-3** and **Fig. SI-4**, MD and MO explain an amount of behavioural variance that is similar to FA, but with some small differences (for example, FA explains the intelligence-related traits more effectively). Even though these differences are reflected in the corresponding statistical tests (as depicted in **Fig. SI-5**), the overall conclusions drawn from FA remain valid for MD and MO. In the following, we will refer only to FA for simplicity.

Temporal and spatial components of time-varying FC explain different aspects of behaviour

The (dual-estimated) HMM descriptor of time-varying FC contains several features, including average FC. In the previous section, we assessed the strictly time-varying aspects of FC by using a version of the Mantel algorithm that regresses out the average FC information before testing. Here, we further dissect which specific aspects of time-varying FC are more related to behaviour. In particular, we considered the fractional occupancy (FO), defined as the amount of time spent by each subject in each state. As well as information about the average FC (which we regressed out), the FO contains purely temporal aspects of time-varying FC without being affected by how each of the identified networks varies spatially across subjects. We then considered the same dual-estimated HMM descriptor that we used before, but while correcting for FO (and average FC). With this descriptor, we targeted the spatial aspects of time-varying FC; that is, above and beyond how much time each subject spends in each network state, this examines the extent to which specific spatial differences in the visited transient networks relate to behaviour. **Fig. 3** presents these results; as before, the dots represent individual realisations of the HMM inference, and the bars represent the average. In all panels, the left side (or Y-axis in the scatter plots) refers to the temporal aspects of time-varying FC (i.e., the FO after correcting for average FC), whereas the right side refers to the spatial aspects (i.e., the dual-estimated HMM after correcting for average FC and FO). The bottom panels show the results after additionally correcting for FA.

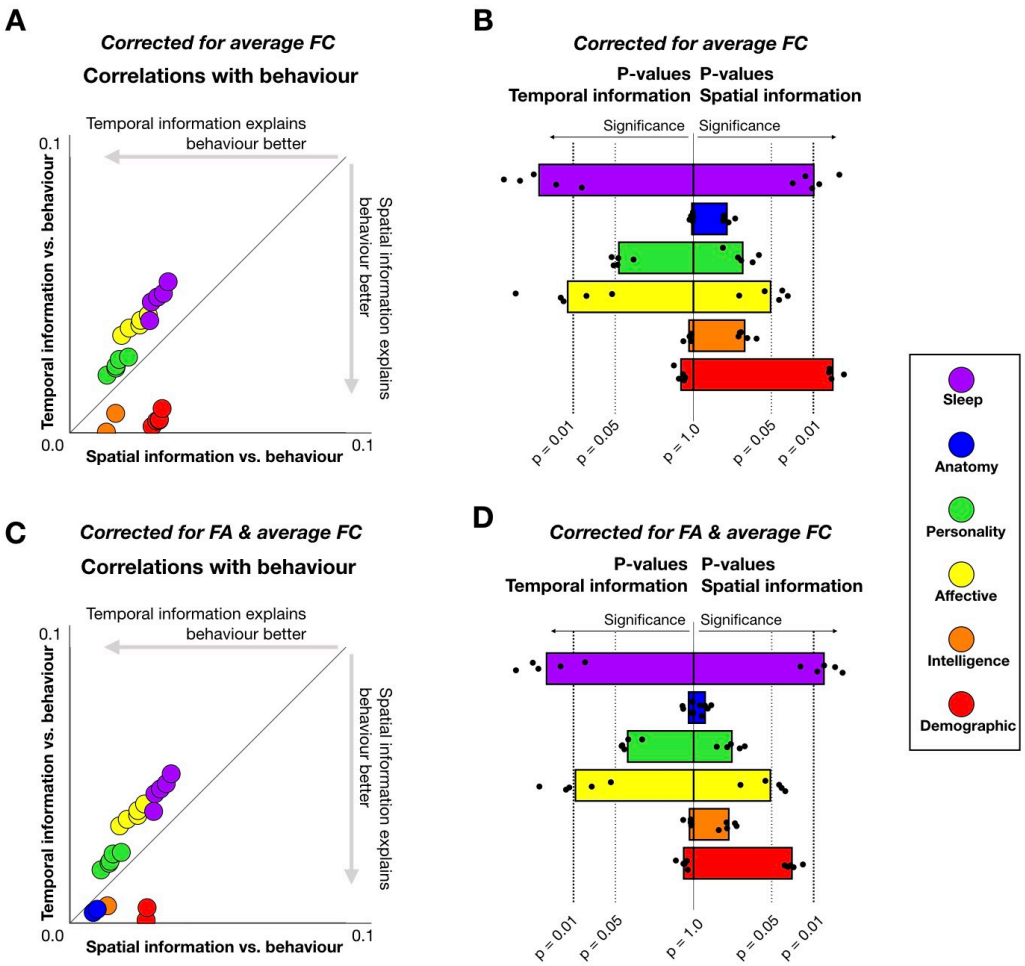

Fig 3. The temporal and spatial aspects of time-varying FC explain distinct aspects of behaviour. The left bars refer to statistical testing on the fractional occupancy (FO, see main text), which contains purely temporal information. The right bars refer to statistical testing on the purely spatial aspects of the time-varying FC estimations. Interpretation of panels is as in **Fig. 3**.

As observed in **Fig. 3**, sleep is well represented by both the temporal aspects of time-varying FC, but better represented by the temporal aspects, whereas the demographic variables are much better represented by the spatial information. The affective and personality traits are also better represented by the temporal components, and intelligence better represented by the spatial components yet without reaching significance. Strikingly, the affective traits become highly significant when looking at the FO only, indicating that the temporal dimension is particularly effective in relating to this type of behaviour. Although the difference is only moderate and the correlations are quite low, the spatial aspects of time-varying FC are more related to the anatomy than the temporal features. The demographic and the anatomical traits are the only ones affected by regressing out FA.

Changes in variance and amplitude of the signal do not explain behaviour

In order to rule out the possibility that the reported significance is driven by within-session changes in the variance or amplitude of the signal, rather than in FC, we also ran all HMM estimations with two different configurations: in the first, the states were characterised by distinct patterns of signal variance, without modelling the off-diagonal elements of the covariance matrix; in the second, the states had distinct patterns of signal amplitude and a common full covariance matrix shared across states, including diagonal and off-diagonal

elements. Therefore, FC was not allowed to vary between states in neither of these variants. In this case, we did not partial the tests with regard to average FC, given that these models do not explicitly contain any subject-specific information about FC. The results of this are shown in **Fig. SI-6**. The tests exhibited no significance for any of the behavioural traits, except for moderate significance on the affective traits for the amplitude-based HMM model.

Reproducibility of DMs

The HCP data contains four sessions per subject, with the first two (1 and 2) being acquired on one day and the last two (3 and 4) on the following day. In order to quantify the reproducibility of the estimations, we estimated separate HMM models for the first day and for the second day, i.e. for sessions 1 and 2, and then for sessions 3 and 4. We also estimated HMM models for the first session of the day, and then separately for the second session of the day. For each of these, we also computed subject-specific FC (correlation) matrices, representing average FC in the traditional manner. For each estimation, (HMM- or average FC-related), we then computed distance matrices (DMs). Similar to the previous analyses, we derived DMs for the entire HMM model as well as for the FO. **Fig. SI-7** presents a quantitative assessment of the reproducibility of the estimations in terms of how similar their respective DMs were across half-splits of the data. Here, the dots represent a measure of distance between one pair of subjects. As observed, the average FC description (being a simpler quantity to estimate) exhibits the highest reproducibility. The difference with the HMM descriptors is however not large. As expected, the reproducibility within day is considerably larger than between days.

Discussion

In the resting state, the quantification of time-varying functional connectivity (FC) has elicited considerable interest and controversy: that is, to what extent can we measure and interpret within-session changes in the pattern of FC between areas? Whereas many studies rely on the average magnitude of activation that is evoked by a task or stimulus, FC is a second-order statistic and therefore is harder to estimate accurately. Similarly, it is unclear whether or not FC can reflect changing patterns of communication between distant neuronal populations, and therefore be meaningful for investigating cognition. Even though the total amount of variability attributed to stable subject FC features is considerably higher than the variability contained in within-session variations (Gratton et al., 2018), here we show that fMRI-derived FC indeed contains both stable and time-varying behaviourally meaningful information, and that time-varying FC can explain behavioural variability that is less likely to be mediated by structural connectivity and other anatomical features. This suggests that time-varying FC may represent meaningful neuronal dynamics related to certain aspects of behaviour. As a consequence, the study of FC fluctuations remains promising for the understanding of transient cognition.

In order to shed light on this question, it is informative to disentangle the different mechanisms by which time-varying FC computed from fMRI data could be non-informative: first, the characterisation of time-varying FC may be limited by fundamentally technical issues; second, the actual amount of information contained in the time-variations, when assessed unbiasedly, may be negligible; and third, even if we can prove that there is information in time-varying FC that can be reliably quantified, it may still not be cognitively significant. We argue that certain technical limitations do not apply to all methods of estimating time-varying FC equally. In the case of the HMM, for example, the technical limitation of having a statistically unstable estimation due to limited amounts of data is overcome by using large amounts of data in the estimation of each state (on average, 125h per state).

In this study, we showed that it does not necessarily follow that time-varying FC is deficient in explaining behavioural traits from the fact that time-varying FC explains considerably less variance of the brain data than stable (subject-specific) FC features. We consider that discussing physiological relevance in terms of explained variance is not appropriate for two reasons: (i) that "physiological relevance" must be connected to a specific scientific question – e.g. relevant to the study of sleep; and (ii) that, provided such a question, there is not prior evidence to assume that the most informative aspect of the signal (for that question) is the one that explains the most variance in the data. For example, in the context of prediction it is a well-known phenomenon that the first principal components of the predictor data are not necessarily the most explanatory to predict the target variable (Frank and Friedmann, 1993). As an example closer to neuroscience, electrophysiological signals hold most of their variance at lower frequencies. In comparison, only a small fraction of variance is contained e.g. in the gamma frequencies (>40 Hz). These, however, have been demonstrated to be essential to behaviour (Jensen et al., 2007). In summary, the argument that there is considerably more variability in the between-subject than in the within-subject differences cannot be used to claim the lack of biological relevance of these features.

Some of the conclusions of this study relate to the recent work from Liégeois et al. (2019), who found that the autoregressive model (which linearly represents how on average the signal at time point t depends across regions on the signal at time point t-1) was often more explanatory of behavioural variability than the standard average FC estimation. Because the autoregressive model is known to describe the dynamics of the signal well (Liégeois et al., 2017), the conclusion of this study is that the dynamic aspects of the data can often explain behaviour better than (average) instantaneous fMRI correlations. However, the autoregressive model focuses on different aspects of the HMM. For example, the autoregressive model includes spectral information (Vidaurre et al., 2016), as well as the directionality of the connections, which might be useful for the prediction in different ways, but also which might be driven by haemodynamic (and not neural) factors. Furthermore, it is unclear the extent to which the information contained in an autoregressive model reflects structural connectivity, and whether it explains behaviour above and beyond the anatomical information. The autoregressive model does not therefore specifically answer the question posed in this work: do variations in the fMRI correlations have biological significance above and beyond the average, and also the structural information? Can they represent instantaneous neural communication? These questions require a model that explicitly considers variations around the average FC without inclusion of these other elements, such as the version of the HMM used here. Whether the autoregressive model is more or less influenced by anatomical variability should also be a subject of future research. Other data descriptions capturing related or different aspects would be worth considering in this context, such as those based on signal events (Allan et al., 2014) or quasi-periodic patterns (Thompson et al., 2014).

It is important to appreciate that neither the HMM nor other commonly used time-varying FC estimators are explicitly biophysical models. Decisions about the appropriate number of states and other parameters are useful insofar as they affect the extent to which we can address the specific question at hand. For example, estimating more states will offer a more fine-grained representation of the data, which might be necessary in certain applications but cannot be interpreted as more or less faithful to the biology. In general, different parametrisations just offer different perspectives of the data, and, assuming model identifiability, the HMM is not more or less valid than other models. We also acknowledge that, while the version of the HMM used in this work is designed to emphasise time-varying FC, it could also be sensitive to changes in amplitude (Duff et al., 2018). However, we have explicitly tested that a version of the HMM only based on changes in amplitude is unable to explain behaviour to the same extent, emphasising the importance of time-varying FC. Other aspects of the data that can influence

the HMM estimation are long-range temporal dependencies, which are not explicitly modelled by the HMM. A quantitative assessment of the long-term dependencies in the data and how they affect the HMM estimation will be subject of future investigations.

It is also important to note that, because the estimation of the brain measures' (e.g. average and time-varying FC) DMs is unsupervised with no consideration of the behavioural information, the Mantel method is a conservative test in the sense that information that is important to behaviour is not necessarily well represented in the DM. For the same reason, the partialled version of the test is a first order approximation and needs to be interpreted with caution. Again, this is because there could be some information of the to-be-deconfounded data that is important for behaviour, but is not well represented in the corresponding DM. However, this issue is not specific to the Mantel test, and affects any testing procedure aiming to remove the influence of high-dimensional confounds of a given statistical association.

Methods

We provide some details on the nature of the Hidden Markov model estimation, the computation of the distances between each pair of subjects for each of the considered measures or subject variables, and the application of the Mantel algorithm to test the relation between imaging and non-imaging variables.

An FC-focused Hidden Markov model

The Hidden Markov model (HMM) is a probabilistic framework used to model time series using a finite number of recurring patterns that succeed each other in some order (Rabiner, 1989). Each of these patterns or states are an instantiation of a certain probability distribution. The HMM is generic in the sense that it can accommodate different state probability distributions, depending of the type of data we are processing and the features that we wish to model. A suitable choice for fMRI data is a Gaussian distribution (Vidaurre et al., 2017), where state, indexed by k, is modelled by a certain "mean activity″ parameter μ_k and a covariance matrix Σ_k. The probability density function that describes a certain time point x_t, assuming that state k is active at time t, is given by

$$2\,\pi^{-J/2}\,|\Sigma_k|^{1/2}\exp\frac{-(x_t-\mu_k)\Sigma_k^{-1}(x_t-\mu_k)'}{2},$$

where J is the number of brain regions (here ICA components), $|\Sigma_k|$ is the determinant of the state-specific covariance matrix Σ_k and exp is the exponential function. Here, Σ_k represents the covariance of the residuals after we account for the mean parameter μ_k.

Here, in order to focus the HMM decomposition on the FC changes, and in order to make the HMM estimation more comparable to standard analyses of time-varying FC, we enforced $\mu_k = 0$, so that the probability density function for state k is now described as

$$2\,\pi^{-J/2}\,|\Sigma_k|^{1/2}\exp\frac{-x_t\,\Sigma_k^{-1}x_t'}{2},$$

As illustrated in **Fig. 1B**, this model carries information of both average FC and time-varying FC.

Another important element of the HMM, also estimated from data, is the transition probability matrix (TPM), encoding the probability of transitioning from one state to another at any time

point. Practically speaking, the TPM serves two purposes: it identifies the transitions that are more probable, and it regularises the state switching to minimise the amount of spurious transitions. In particular, in more persistent (temporally regularised) solutions, the diagonal elements will be comparably larger than the off-diagonal elements of the TPM.

The estimation of the HMM, carried out through a procedure of Bayesian variational inference (Wainwright and Jordan, 2008; Vidaurre et al., 2018a), was first computed at the group level, such that the state probability distributions were shared across subject – though the state time courses and the time spent in each state were subject-specific. Next, we computed subject-specific HMMs using a process that we refer to as dual-estimation (in analogy to dual-regression in ICA; Beckmann et al., 2009). In order to do this, we simply used the subject-specific state time courses to compute an individualised estimation of each state; then, leaning on these state estimations, we recomputed the state time courses and the TPM for each subject.

The HMM contains average FC information

Given *N* subjects and *K* states, the group level HMM estimation represents subject-specific average (or static) FC (avFC) information according to the following expression:

$$\text{avFC}_i \simeq \text{avFC}_i^{HMMgroup} = \text{sum}_k \, w_k^i \, \Sigma_k,$$

where *i* indexes subjects and w_k^i represents the fractional occupancy for subject *i* and state *k*. Given that the number of states is lower than the number of subjects ($K < N$), this is an approximation, and therefore there is at least some avFC information that is not captured by the HMM. Likewise, the HMM has information (for example in the TPM) that is not captured by a standard average FC estimation; we are referring to this information here as time-varying FC.

In order to quantify how much average FC variance is explained by the HMM, we used Riemannian distances as outlined above. With this metric, we can readily derive a FC-specific expression of explained variance as:

$$r^2 = 1 - \frac{\text{sum}_i \left(d\,(\text{avFC}_i^{HMMgroup}, \text{avFC}_i\right)^2}{\text{sum}_i \left(d\,(\text{avFC}^{ave}, \text{avFC}_i\right)^2},$$

where avFC^{ave} is the group-average FC matrix.

The dual-estimated HMM estimation, on the other hand, captures all the average FC information

$$\text{avFC}_i = \text{avFC}_i^{HMMd-e} = \text{sum}_k \, w_k^i \, \Sigma_k^{d-e},$$

where *d-e* denotes dual-estimated. This is because:

$$\text{avFC}_i^{HMMd-e} = \text{sum}_k \, w_k^i \, \frac{\text{sum}_t \, g_{tk}^i \, x_t x_t'}{\text{sum}_t \, g_{tk}^i} = \frac{1}{T} \, \text{sum}_k \, \text{sum}_t \, g_{tk}^i \, x_t x_t' =$$
$$\frac{1}{T} \, \text{sum}_t \, x_t x_t' \, \text{sum}_k \, g_{tk}^i = \frac{1}{T} \, \text{sum}_t \, x_t x_t' = \text{avFC}_i \, ,$$

where g_{tk}^{i} is the probability for subject i to be in state k at time point t, $\text{sum}_k\, g_{tk}^{i} = 1$ by the definition of a probability, and $w_k^i = \frac{1}{T}\, \text{sum}_t\, g_{tk}^{i}$. Whereas, strictly speaking, the HMM model contains the state descriptions and the TPM but not the state probabilities g_{tk}^{i}, this expression holds true no matter the specific values of g_{tk}^{i}.

The HMM contains time-varying FC information

In **Fig. SI-1, we show** that the average FC contains information that is essentially uncorrelated to the FC temporal variability. To compute a measure of the extent to which there is time-varying FC for each pair of regions, we first constructed an instantaneous estimate of FC at each time point, using a weighted sum of the dual-estimated HMM states' FC, weighted by the assigned HMM state probabilities at that time point. We then took the variance of these instantaneous estimates of FC across time to produce a (regions-by-regions, i.e. a ICA-components-by-ICA-components) matrix of estimated FC temporal variability for each given subject.

In the previous section, we have also shown that the dual-estimated HMMs contain all the information there is about average FC. Having K FC descriptions per subject instead of one, plus a TPM ruling the transitions between these states, it is apparent that the HMM contains additional information beyond average FC. An important question is then what that additional information represents. There are three possible sources of variability: actual within-session changes of FC (i.e. time-varying FC), within-session changes in the variance of the signal, or pure estimation noise. By meaningfully relating the HMM information to behaviour above and beyond the influence of average FC (see below, and Results) we can rule out the possibility that the HMM extra parameters are purely noise-driven. Now, given that both variances and correlations (i.e. FC) are part of the state descriptions, there is no analytical way to disambiguate the extent to which each of these elements drove the inference of the HMM. In order to prove the relevance of pure time-varying FC in the HMM estimation, we took an indirect route by obtaining alternative HMM estimations where the states are purely derived by changes in the amplitude of the signal. The probability density function of this model is given by

$$\prod_j \left(\frac{1}{\sqrt{2\,\pi\,\sigma_{jk}^2}} \exp \frac{-x_{tj}^2}{2\,\sigma_{jk}^2} \right),$$

where $\prod_j(\cdot)$ represents multiplication across regions, σ_{jk}^2 is the variance for region j and state k, and x_{tj} is the value of the signal for region j at time point t.

The fact that HMM using full covariances matrices can explain aspects of behaviour that this model was unable to explain (see **Fig. SI-5**) suggests that there is relevant information in the HMM that is not related to changes in variance.

Even though the mean parameter of the Gaussian distribution (which reflects the amount of activity of each state with respect to the average) was not included in the model in the first place, we estimated a third HMM model where the states were solely defined by the mean, i.e. without state-specific covariances. The purpose of this analysis is to rule out the possibility that this type of information, though not explicitly included in the HMM description used here, permeated into the state covariance matrices and indirectly drove the HMM inference. The probability density function of this model is given by

$$C\ |\Sigma|^{1/2} \exp \frac{-(x_t - \mu_k)\Sigma^{-1}(x_t - \mu_k)'}{2},$$

where, in this case, there one single covariance matrix Σ for all states. Note that this model holds important similarities with ICA, in the sense that each state or component is represented by a map of activation. We opted to use the HMM instead of ICA to make the results fully comparable with the other HMM runs.

As shown in **Fig. SI-5**, this model was also unable to explain aspects of behaviour that were explained by the original model. By elimination, these analyses proved that the HMM contains (behaviourally-relevant) variability that is necessarily time-varying FC.

Measuring dissimilarities between subjects

The tests employed in this paper are based on comparing distance matrices (DM) containing the dissimilarities between each pair of subjects according to different criteria. In this section, we detail how to compute these for the different imaging-based and non-imaging variables.

We computed DMs for seven groups of (non-imaging) behavioural variables: demographic, intelligence, affective, personality, anatomy, and sleep-related subject traits. **Table SI-1** lists which variables were included in each group. We first standardized these matrices such that all traits had mean equal to zero and variance equal to one. Then, we used PCA to extract enough principal components to explain 95% of the variance. Because some of these variables had some missing values, we used a version of PCA able to impute the missing values (Roweis, 1998). In short, this algorithm makes iterative estimations of the principal components and the missing values until reaching convergence. PCA was also necessary because some of these traits are highly correlated and it is undesirable to overrepresent their information; for example, if two traits are 0.99 correlated, they are likely to be noisy replications of each other, in which case keeping the two will overrepresent the information they carry and, therefore, result in an underrepresentation of other traits. By using principal components instead of the raw values, these correlations are automatically accounted for. Finally, for each group of variables, the respective DMs were computed using Euclidean distances between the corresponding principal components (i.e. using both the eigenvectors and eigenvalues) for each pair of subjects.

We then computed DMs for three different imaging-based features. Two of them are related to the HMM.

The first one used the entire HMM model, which, as discussed above, contains information of both average FC and time-varying FC. In particular, we computed the so-called symmetric Kullback-Leibler divergence between each pair of (dual estimated) subject HMMs, denoted as M^1 and M^2.

$$\mathrm{Dist}^{HMM}(M^1, M^2) = 0.5\,\mathrm{KL}(M^1||M^2) + 0.5\,\mathrm{KL}(M^2||M^1),$$

where $\mathrm{KL}(M^1 \parallel M^2)$ represents the standard (non-symmetric) Kullback-Leibler divergence between probabilistic models M^1 and M^2. In statistical science, the Kullback-Leibler divergence represents how much information a probability distribution contains in relation to a second reference probability distribution. Whereas the Kullback-Leibler divergence has a closed form for many well-known distributions (e.g. the Gaussian distribution), this is not the case for more complex models such as the HMM. For this reason, we adapted the mathematical approximation proposed by Do (2003) for discrete state distributions to the Gaussian case:

$$\mathrm{KL}\,(M1 \,||\, M2) \;= \mathrm{sum}_k\,(\nu_k\,\mathrm{KL}(P_k^1, P_k^2) \;+\; \nu_k\,\mathrm{KL}(G_k^1, G_k^2)),$$

where P_k^i represents the (Dirichlet-distributed) probabilities of transitioning from state k to any of the other states according to model i (i.e. the k-th row of the TPM); G_k^1 is the state Gaussian distribution for state k and model i; and ν_k is a factor representing the weight of state k in M^l. Given the initial probabilities π^l of the HMM state time courses for model M^l (which are computed from data together with the TPM), ν can be numerically computed such that it holds the necessary criteria (Do, 2003):

$$\nu\, P^1 \;=\; \nu,$$
$$\lim_{n\to\infty} \pi^1\, P^{1^n} \;\;=\; \nu,$$

The expressions for $\mathrm{KL}(P_k^1, P_k^2)$ and $\mathrm{KL}(G_k^1, G_k^2)$ are standard and can be found elsewhere (MacKay 2003). Code to compute the symmetric Kullback-Leibler divergence between two HMM models can be found in[1]. Note that these expressions require the states to be matched between HMM models; i.e. the first state of one model must correspond to the first state of the other model. This is here guaranteed by the fact that the dual-estimated HMMs are derived from the same group-level HMM.

The other time-varying FC-based feature is based on the fractional occupancies (FO). Since the FOs are not explicitly modelled as a probability distribution, we simply used Euclidean distances:

$$\mathrm{Dist}^{FO}(M^1, M^2) \;=\; \sqrt{\mathrm{sum}_k\,(w_k^1 \;-\; w_k^2)^2}\;,$$

where w_k^i represents how much time the subject i spends on state k, i.e. the FO, which was also extracted from the dual estimation process.

Finally, the last DM was computed from a conventional estimation of average FC, based on the (averaged) Pearson correlation matrix for each subject. Given that the time series were standardised for each subject, this is equivalent to assume a single Gaussian distribution for each subject, having a zero mean and a covariance matrix with ones in the diagonal. We therefore used the symmetric Kullback-Leibler divergence. Note also that the information of the average FC is contained within the dual-estimated HMM, given that the average Pearson correlation matrix can be recovered from the FO information and the dual-estimated states.

Comparing subject variables using the Mantel algorithm

The Mantel algorithm (Mantel, 1967) is a non-parametric algorithm based on permutation testing (Nichols and Holmes, 2002) used to compare DMs between subjects. Here, we used it to test the relationship between imaging-based and non-imaging variables. The Mantel algorithm takes a pair of DMs, extracts their upper-diagonal elements into two vectors of $N\,(N-1)/2$ values, and computes the correlation between these two vectors. Then, for a large number of permutations (here 10000), the rows and columns of one of the DMs are randomly shuffled before vectorising, and a surrogate distribution of DM-to-DM correlations is produced. The final p-value is calculated as the proportion of surrogate correlations that are higher than the original (unpermuted) correlation. Importantly, the permutations need to be

[1] https://github.com/OHBA-analysis/HMM-MAR/blob/master/utils/math/hmm_kl.m

chosen such that they respect the family structure of the subjects, for example by permuting between families with the same number of members or by permuting siblings within the family (Winkler et al., 2015).

A partial version of the Mantel algorithm can be used to control for a third DM, if this DM is considered to be a confound or if we wish to assess the relation between the two original DMs above and beyond the influence of the third DM. This is done simply by vectorising the confounding DM and regressing it out of the non-permuted vectorised DM. We used this approach to test the relation between time-varying FC (as expressed by the HMM) and the behavioural DMs above and beyond the influence of average FC (as expressed by the subject-specific Pearson correlation matrices).

The advantage of the Mantel algorithm is that we can dispense with the specific representation of each of these variables, which otherwise could be difficult to compare, and work only with between-subject distances. Otherwise, variables with different representations could be difficult to compare directly; e.g. a collection of subject-specific HMMs on one side and a set of demographical variables on the other are not directly comparable unless we somehow convert each HMM into a vector of values where the principles of Euclidean geometry apply.

It is worth noting that the Mantel test is heavily criticised for its use in spatial analyses, for example in studies examining the relation between the geographical location of different populations of individuals and their phylogenetic distances (Guillot and Rousset, 2013). Essentially, the Mantel algorithm is an invalid test if the permutations break the structure in the data, producing a higher number of false positives than a valid test would do; this would be the case if the permutation scheme violates the spatial autocorrelations in the case of a spatial study, or the family structure in our analysis. These objections however do not apply here insofar as our permutation approach was designed respect the family structure of the subjects (Winkler et al., 2015).

For simplicity, in the figures we summarise the statistical tests on each of the five repetitions of the HMM inference taking the average of the p-values. The mean is however not a proper p-value, and alternative testing methods would need to be undertaken for this to be the case (Vidaurre et al., 2018c).

Deconfounding fractional anisotropy and anisotropy mode from behaviour

We applied independent component analysis (implemented by the Melodic tool in FSL; Jenkinson et al., 2012) on the fractional anisotropy (FA) or anisotropy mode (MO) values for each subject across the whole brain (2mm resolution) in order to get 50 independent components of FA or MO variability across subjects. We then used cross-validated regression in order to deconfound these FA or MO components from each of the behavioural traits. More specifically, we estimated regularised regression coefficients (Hoerl and Kennard, 1970) on the training folds and applied them on each testing fold, in turn, to produce a vector of residuals, which we then used as FA- or MO-deconfounded traits. Within the training fold, the value of the regularisation parameter (which determines the complexity of the estimation) was selected in a nested cross-validation loop. We chose this strategy in light of previous work revealing that deconfounding based on cross-validation is less aggressive and biased than standard deconfounding (Snoek et al., 2019).

Acknowledgements

DV is supported by a Wellcome Trust Strategic Award (098369/Z/12/Z). The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust (203139/Z/16/Z). MWW's research is supported by the NIHR Oxford Health Biomedical Research Centre, by the Wellcome Trust (106183/Z/14/Z), and the MRC UK MEG Partnership Grant (MR/K005464/1).

Contributions

DV conceived and designed the study. DV and AL ran the analyses. DV, MW and SM wrote and edited the manuscript.

References

1. T.W. Allan, S.T. Francis, C. Caballero-Gaudes, P.G. Morris, E.B. Liddle, P.F. Liddle, M.J. Brookes and P.A. Gowland (2015). Functional Connectivity in MRI Is Driven by Spontaneous BOLD Events. *PLoS One*, e0124577.
2. P.J Basser and C. Pierpaoli (1996). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of Magnetic Resonance* **213**, 560-570.
3. C.F. Beckmann, M De Luca, J.T. Devlin and S.M. Smith (2005). Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 1001-1013.
4. C.F. Beckmann, C.E. Mackay, N. Filippini and S.M. Smith (2009). Group comparison of resting-state fMRI data using multi-subject ICA and dual regression. *Neuroimage* **47**, 148-156.
5. J.D. Bijsterbosch, M.W. Woolrich, M.F. Glasser, E.C. Robinson, C.F. Beckmann, D.C. Van Essen, S.J. Harrison and S.M. Smith (2018). The relationship between spatialconfiguration and functional connectivityof brain regions. *eLife* **7**, e32992.
6. B. Biswal, F. Zerrin-Yetkin, V.M. Haughton and J.S. Hyde (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine* **34**, 537-541.
7. J.S. Damoiseaux, S.A.R.B. Rombouts, J. Barkhof, P. Scheltens, C.J. Stam, S.M. Smith & C.F. Beckmann. Consistent resting-state networks across healthy participants. *Proceedings of the National Academy of Sciences of the USA* **103**, 13848-13853 (2006).
8. E. Duff, T. Makin, M. Cottaar, S.M. Smith and M.W. Woolrich (2018). Disambiguating brain functional connectivity. *NeuroImage* **173**, 540-550.
9. M. N. Do (2003). Fast approximation of Kullback–Leibler distance for dependence trees and Hidden Markov Models. *IEEE Signal Processing Letters* **10**, 115-118.
10. D.B. Ennis and G. Kindlmann (2006). Orthogonal tensor invariants and the analysis of diffusion tensor magnetic resonance images. *Magnetic Resonance in Medicine* **55**, 136-146.
11. E.S. Finn, X. Shen, D. Scheinost, M.D. Rosenberg, J. Huang, M.M. Chun, X. Papademetris and R.T. Constable (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nature neuroscience 18, 1664–1671.
12. M.D. Fox and M.E. Raichle. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience* **8**, 700-711.
13. L.L.E. Frank and J.H. Friedman (1993). A statistical view of some chemometrics regression tools. *Technometrics* **7**, 109-135
14. C. Gratton, T.O. Laumann, A.N. Nielsen, D.J. Greene, E.M. Gordon, A.W. Gilmore, S.M. Nelson, R.C. Coalson, A.Z. Snyder, B.L. Schlaggar, N.U.F. Dosenbach and S.E. Petersen (2018). Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron* **98**, 439-452.
15. G. Guillot and F. Rousset (2013). Dismantling the Mantel tests. *Methods in Ecology and Evolution* **4**: 336-344.

16. M. Hampson, N.R. Driesen, P. Skudlarski, J.C. Gore, and R.T. Constable (2006). Brain connectivity related to working memory performance. *The Journal of Neuroscience* **26**, 13338-13343.
17. U. Hasson, H.C. Nusbaum and S.L. Small (2009). Task-dependent organization of brain regions active during rest. *Proceedings of the National Academy of Sciences of the USA* **106**, 10841-10846.
18. J.F. Hipp, A.K. Engel and M. Siegel (2011). Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron* **69**, 387-396.
19. J.F. Hipp and M. Siegel (2015). BOLD fMRI correlation reflects frequency-specific neuronal correlation. *Current Biology* **25**, 1368-1374.
20. A.E. Hoerl and R.W. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**, 55-67.
21. O. Jensen, K. Kaise and J.P. Lachaux (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Cognitive Sciences* **30**, 317-324.
22. M. Jenkinson, C.F. Beckmann, T.E.J. Behrens, M.W. Woolrich and S.M. Smith (2012). FSL. *NeuroImage* **62**, 782-790.,
23. A. Kucyi, M.J. Hove, M. Esterman, R.M. Hutchison and E.M. Valera (2017). Dynamic brain network correlates of spontaneous fluctuations in attention. *Cerebral Cortex* **25**, 1831-1840.
24. A. Kucyi, A. Tambini, S. Sadaghiani, S. Keilholz and J.R. Cohen (2018). Spontaneous cognitive processes and the behavioral validation of time-varying brain connectivity. *Network Neuroscience* **2**, 397–417.
25. A. Kucyi (2017). Just a thought: How mind-wandering is represented in dynamic brain connectivity. *NeuroImage* **180**, 505-514.
26. R. Liégeois, T.O. Laumann, A.Z. Snyder, J. Zhou and B.T. Yeo (2017) Interpreting temporal fluctuations in resting-state functional connectivity MRI. *NeuroImage* **163**, 437-455.
27. R. Liégeois, J. Li, R. Kong, C. Orban, D. Van De Ville, T. Ge, M.R. Sabuncu and T. Yeo (2019). Resting brain dynamics at different timescales capture distinct aspects of human behavior. *Nature Communications* **10**, 2317.
28. A. Llera, T. Wolfers, P. Mulders and C.F. Beckmann (2019). Inter-individual differences in human brain structure and morphometry link to variation in demographics and behavior. *eLife* (in press).
29. D. Lurie, D. Kessler, … and V. Calhoun (2019) On the nature of time-varying functional connectivity in resting fMRI. *PsyArXiv*. DOI: 10.31234/osf.io/xtzre
30. D.J.C. MacKay (2003). Information Theory, Inference, and Learning Algorithms. *Cambridge University Press*.
31. N. Mantel (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209-220.
32. T.E. Nichols and A.P. Holmes (2002). Nonparametric Permutation Tests For Functional Neuroimaging: A Primer with Examples, *Human Brain Mapping*, **15**, 1-25.
33. S. Palva and J.M. Palva (2012). Discovering oscillatory interaction networks with M/EEG: challenges and breakthroughs. *Trends in Cognitive Sciences* **16**, 219-230.
34. P.J. Basser, J. Mattiello, D. Le Bihan (1994). Estimation of the effective self-diffusion tensor from the NMR spin-echo. *Journal of Magnetic Resonance, Series B* **103**, 247-254.
35. G.C. O'Neill, P.K. Tewarie, G.L. Colclough, L.E. Gascoyne, B.A.E. Hunt, P.G. Morris, M.W. Woolrich, M.J. Brookes (2017). Measurement of dynamic task related functional networks using MEG. *NeuroImage* **146**, 667-678.
36. A.J. Quinn, D. Vidaurre, R. Abeysuriya, R. Becker, A.C. Nobre and M.W. Woolrich (2018). Task-Evoked Dynamic Network Analysis Through Hidden Markov Modeling. *Frontiers in Neuroscience* **12**, 603.
37. L.R. Rabiner (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* **77**, 257-286.
38. S. Roweis (1998). EM algorithms for PCA and SPCA. *Advances in neural information processing systems* 10: 626-632; Jordan, M.I., Kearns, M.J., Solla, S.A., Eds.; MIT Press: Cambridge, MA, USA.
39. S. Sadaghiani, J.B. Poline, A. Kleinschmidt, and M. D'Esposito (2015)). Ongoing dynamics in large-scale functional connectivity predict perception. *Proceedings of the National Academy of Sciences of the USA* **112**, 8463-8468.
40. J. Smallwood and J.W. Schooler (2015). The science of mind wandering: empirically navigating the stream of consciousness. *Annual Review of Psychology* **66**, 487-518.

41. Smith S.M., D. Vidaurre, C.F. Beckmann, M.F. Glasser, M. Jenkinson, K.L. Milller, T.E. Nichols, E.C. Robinson, G. Salimi- Khorshidi, M.W. Woolrich, D.M. Barch, K. Ugurbil and D.C. Van Essen (2013). Functional connectomics from resting-state fMRI. *Trends in Cognitive Sciences* **17**, 666-682.
42. S.M. Smith, C.F. Beckmann, J. Andersson, E.J. Auerbach, J. Bijsterbosch ,G. Douaud, E. Duff, D.A. Feinberg, L. Griffanti, M.P. Harms, M. Kelly, T. Laumann, K.L. Miller, S. Moeller, S. Petersen, J. Power, G. Salimi-Khorshidi, A.Z. Snyder, A.T. Vu, M.W. Woolrich, J. Xu, E. Yacoub, K. Ugurbil, D.C. Van Essen and M.F. Glasser (2013). Resting-state fMRI in the Human Connectome Project. *Neuroimage* **80**, 144-168.
43. S.M. Smith, T.E. Nichols, D. Vidaurre, A.M. Winkler, T.E.J. Behrens, M.F. Glasser, K. Ugurbil, D.M. Barch, D.C. Van Essen and K.L. Miller (2015). A positive-negative mode of population covariation links brain connectivity, demographics and behaviour. *Nature Neuroscience* **18**, 1565-1567.
44. L. Snoek, S. Miletić and S. Scholte (2019). How to control for confounds in decoding analyses of neuroimaging data. *NeuroImage* **184**, 741-760.
45. A. Stevner, D. Vidaurre, J. Cabral, K. Rapuano, S.F.V. Nielsen, E. Tagliazucchi, H. Laufs, P. Vuust, G. Deco, M.W. Woolrich, E. Van Someren and M.L. Kringelbach (2019). Discovery of key whole-brain transitions and dynamics during human wakefulness and non-REM sleep. *Nature Communications* **10**, 1035.
46. G.J. Thompson, W.J. Pan, M.E. Magnuson, D. Jaeger and S.D. Keilholz (2014).Quasi-periodic patterns (QPP): Large-scale dynamics in resting state fMRI that correlate with local infraslow electrical activity. *NeuroImage* **84**, 1018–1031.
47. D. Vidaurre, A.J. Quinn, A.P. Baker, D. Dupret, A. Tejero-Cantero and M.W. Woolrich (2016). Spectrally resolved fast transient brain states in electrophysiological data. *NeuroImage* **126**, 81-95.
48. D. Vidaurre, S.M. Smith and M.W. Woolrich (2017). Brain networks are hierarchically organised in time. *Proceedings of the National Academy of Sciences of the USA* **114**, 12827-12832.
49. D. Vidaurre, R. Abeysuriya, R. Becker, A.J. Quinn, F. Alfaro-Almagro, S.M. Smith and M.W. Woolrich (2018a). Discovering dynamic brain networks from Big Data in rest and task. *NeuroImage* **180**, 646-656.
50. D. Vidaurre, L.T. Hunt, A.J. Quinn, B.A.E. Hunt, M.J. Brookes, A.C. Nobre and M.W. Woolrich (2018b). Spontaneous cortical activity transiently organises into frequency specific phase-coupling networks. *Nature Communications* **9**, 2987.
51. D. Vidaurre, M.W. Woolrich, A.M. Winkler, T. Karapanagiotidis, J Smallwood and T.E. Nichols (2018c). Stable between-subject statistical inference from unstable within-subject functional connectivity estimates. *Human Brain Mapping* **40**, 1234-1243.
52. M.J. Wainwright & M.I. Jordan (2008). Graphical models, exponential families and variational inference. *Foundations and Trends in Machine Learning* **1**, 1-305.
53. A.B. Waites, A. Stanislavsky, D.F. Abbott and G.D. Jackson (2005). Effect of prior cognitive state on resting state networks measured with functional connectivity. *Human Brain Mapping* **24**, 59-68.
54. A. Winkler, M.A. Webster, D. Vidaurre, T.E. Nichols & S.M. Smith (2015). Multi-level block permutation. *NeuroImage* **123**, 253-268.

Supplemental Information

FC temporal variability is distinct from average FC

In order for FC temporal variability and average FC to explain distinct aspects of behaviour, these descriptors must contain non-shared elements of information of brain function. This amounts to showing that there is unique, subject-specific behaviourally-relevant information in the time-varying FC, which is not contained in the average FC. To do this, we computed the average FC for each subject and compared it with a measure of FC temporal variability for each subject (see Methods). **Fig. SI-1** shows that the amount of average FC for any pair of regions is unrelated to amount of FC temporal variability for such a pair. **Fig. SI-1A** shows an example for one subject, where the upper triangular matrix represents average FC and the lower triangular matrix represents FC temporal variability. **Fig. SI-1B** represents the relation between average FC and FC temporal variability for that same subject as a scatter plot, where each dot corresponds to a pair of regions. Although the null hypothesis cannot be proved in this way, we note that the correlation between these measures is 0.01 and non-significant. **Fig. SI-1C** shows a histogram of correlations between average FC and FC temporal variability across subjects; the mean correlation is 0.0015 and is non-significantly positive.

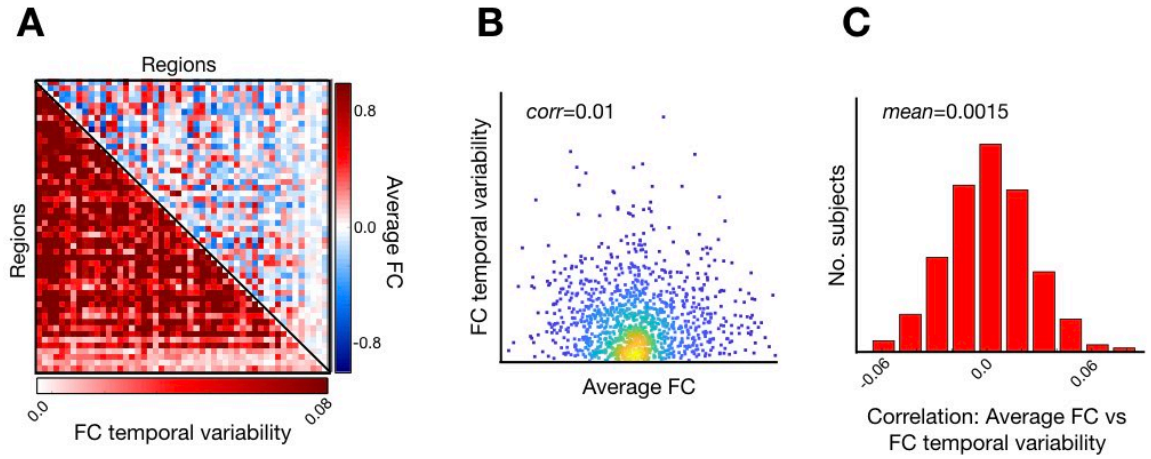

Fig. SI-1. The temporal variability of the FC within session contains information that is independent of the average FC. **(A)** For a given subject, average FC (upper triangular matrix) and FC temporal variability (lower triangular matrix). **(B)** For the same subject, scatter plot of average FC coefficients vs. FC temporal variability values. **(C)** Histogram of average FC vs. FC temporal variability correlations across subjects.

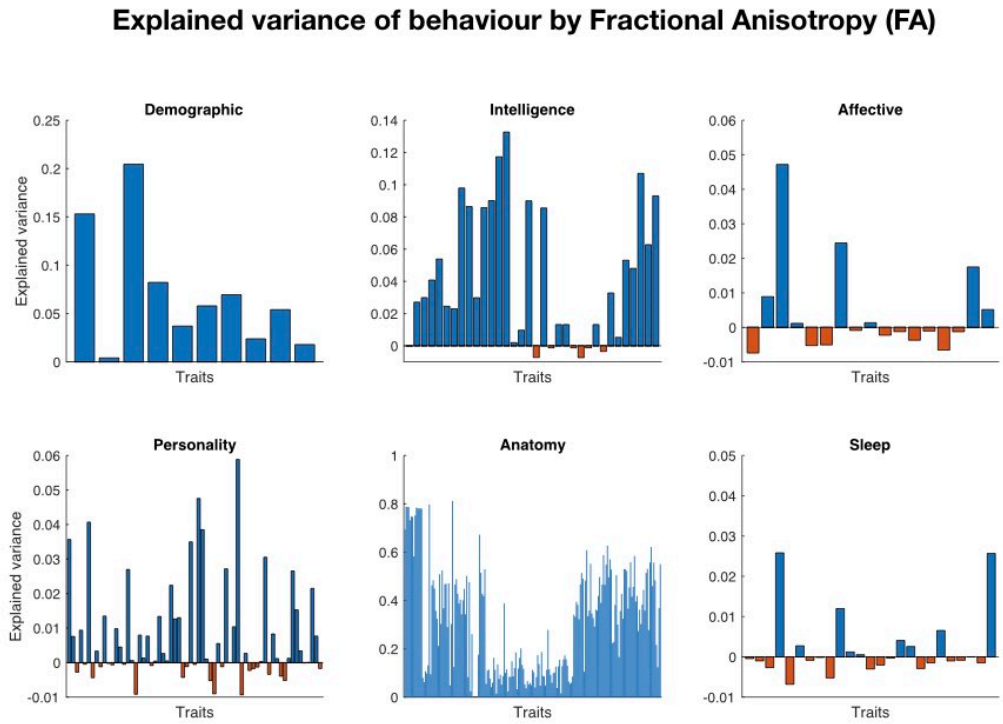

Fig. SI-2. Explained variance of each of the behavioural (and anatomical) traits by fractional anisotropy (FA).

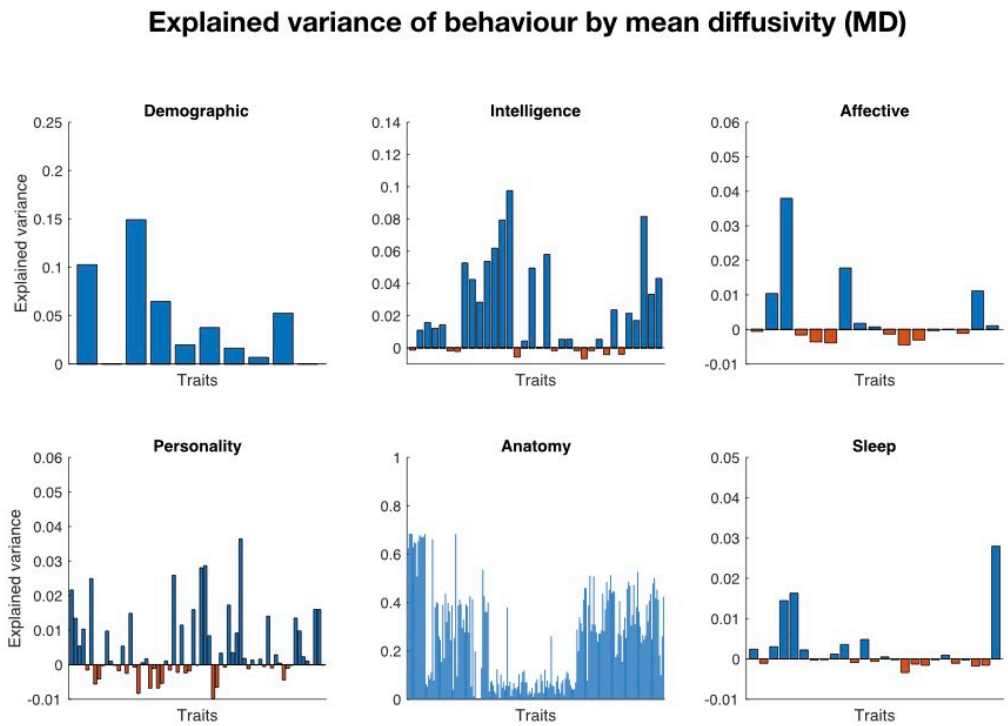

Fig. SI-3. Explained variance of each of the behavioural (and anatomical) traits by mean diffusivity (MD).

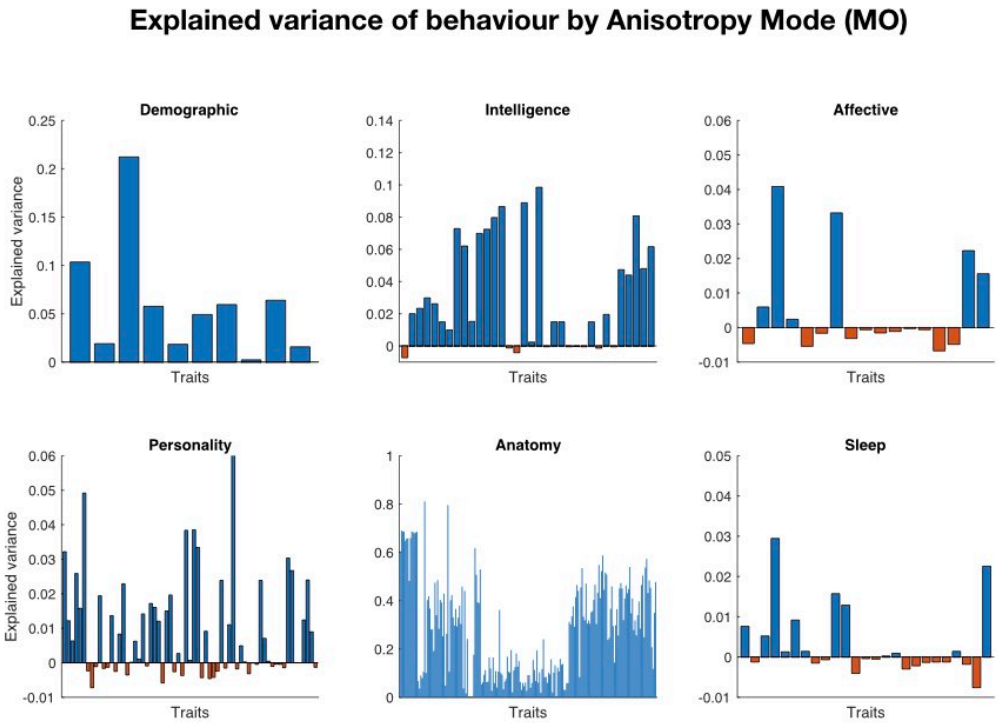

Fig. SI-4. Explained variance of each of the behavioural (and anatomical) traits by anisotropy mode (MO).

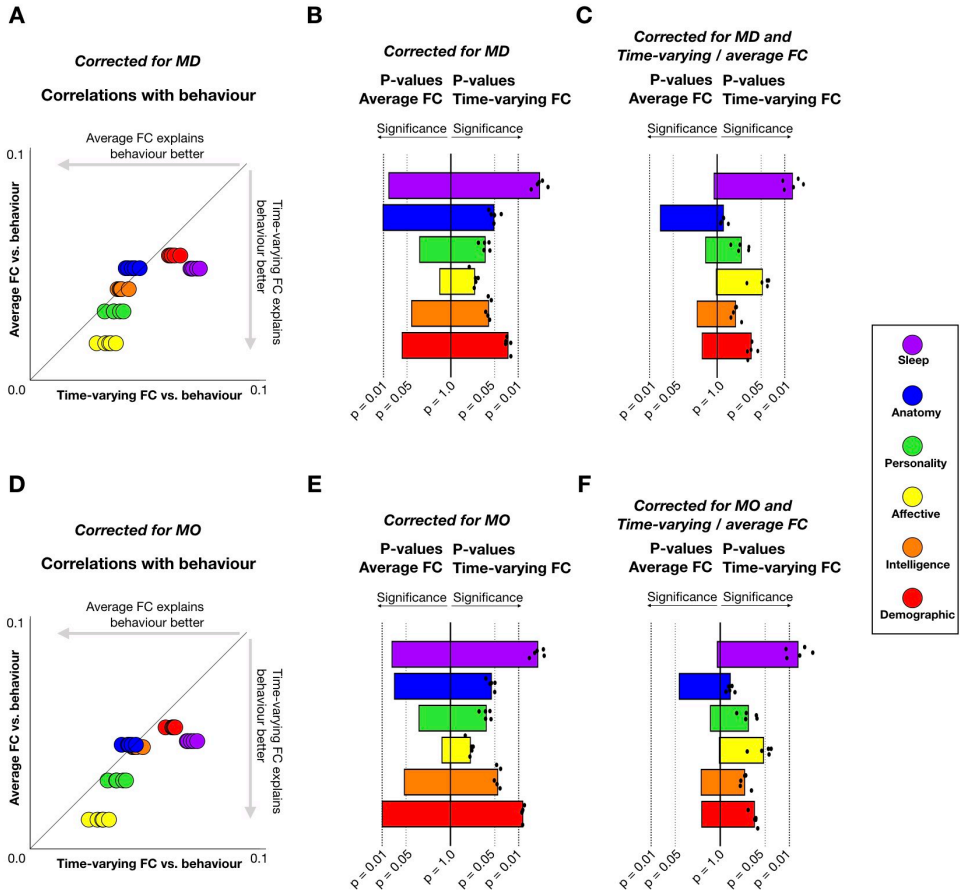

Fig. SI-5. Relation between (average and time-varying) FC and behaviour after correcting the behavioural variables for mean diffusivity (MD) and anisotropy mode (MO) **(A)** Correlations between average/time-varying

FC and behaviour, corrected by MO. **(B)** Statistical tests, after correcting for MD, on whether time-varying FC explains behaviour (right bars), and on whether average FC explains behaviour (left bars); dots represent independent runs of the HMM and lines represent significance levels. **(C)** Statistical tests, after correcting for MO, on whether time-varying FC explains behaviour above and beyond average FC (right bars), and on whether average FC explains behaviour above and beyond time-varying FC (left bars). **(D-F)** are like **(A-C)** but correcting for MO instead of MD.

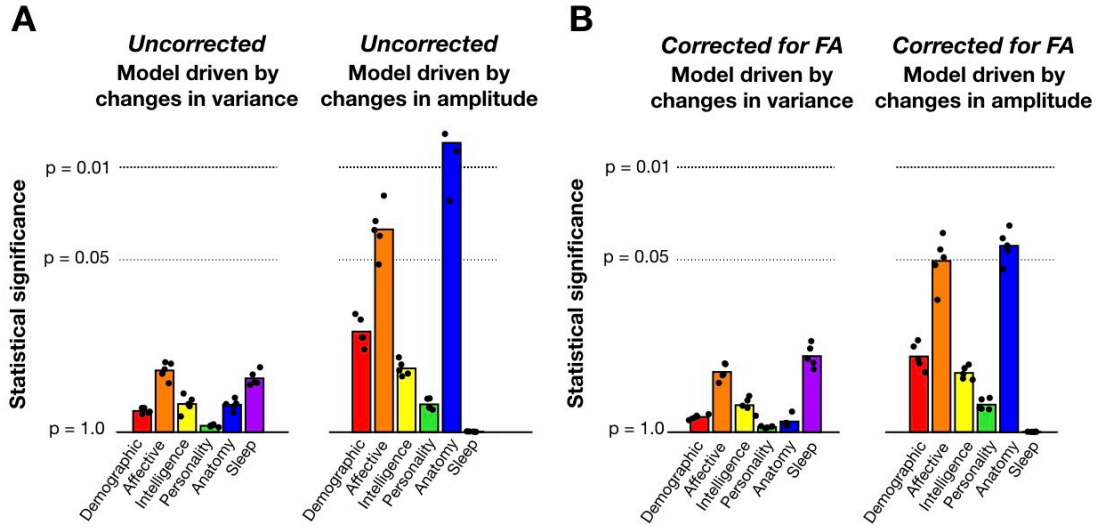

Fig. SI-6. Changes in amplitude or variance do not explain behaviour. The bars represent statistical testing on an HMM model designed to capture changes in variance (left) and an HMM model designed to capture changes in amplitude (right). **(A)** Statistical tests not corrected for FA. **(B)** Statistical tests corrected for FA. None of the tests were not corrected for average FC.

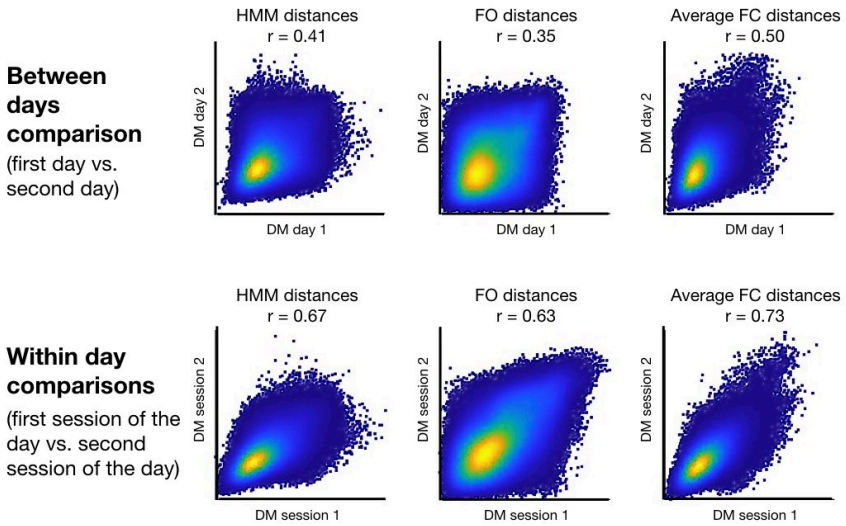

Fig. SI-7. Reproducibility of the estimations between the first and the second day of scanning (top row), and between the first session and the second session of each day (bottom row). Each dot corresponds to an element of the DM, i.e. a distance measure between a pair of subjects. Each column represents a different descriptor of the data.

Trait group	Traits
Demographic	Age, handedness, race, ethnicity, SSAGA_Employ, SSAGA_Income, SSAGA_Educ, SSAGA_InSchool, SSAGA_Rlshp, SSAGA_MOBorn
Intelligence	MMSE_Score, PicSeq_Unadj, PicSeq_AgeAdj, CardSort_Unadj, CardSort_AgeAdj, Flanker_Unadj, Flanker_AgeAdj, PMAT24_A_CR, PMAT24_A_SI, PMAT24_A_RTCR, ReadEng_Unadj, ReadEng_AgeAdj, PicVocab_Unadj, PicVocab_AgeAdj, ProcSpeed_Unadj, ProcSpeed_AgeAdj, VSPLOT_TC, VSPLOT_CRTE, VSPLOT_OFF, SCPT_TP, SCPT_TN, SCPT_FP, SCPT_FN, SCPT_TPRT, SCPT_SEN, SCPT_SPEC, SCPT_LRNR, IWRD_TOT, IWRD_RTC, ListSort_Unadj, ListSort_AgeAdj
Affective	AngAffect_Unadj, AngHostil_Unadj, AngAggr_Unadj, FearAffect_Unadj, FearSomat_Unadj, Sadness_Unadj, LifeSatisf_Unadj, MeanPurp_Unadj, PosAffect_Unadj, Friendship_Unadj, Loneliness_Unadj, PercHostil_Unadj, PercReject_Unadj, EmotSupp_Unadj, InstruSupp_Unadj, PercStress_Unadj, SelfEff_Unadj
Personality	NEOFAC_A, NEOFAC_O, NEOFAC_C, NEOFAC_N, NEOFAC_E, NEORAW_01, NEORAW_02, NEORAW_03, NEORAW_04, NEORAW_05, NEORAW_06, NEORAW_07, NEORAW_08, NEORAW_09, NEORAW_10, NEORAW_11, NEORAW_12, NEORAW_13, NEORAW_14, NEORAW_15, NEORAW_16, NEORAW_17, NEORAW_18, NEORAW_19, NEORAW_20, NEORAW_21, NEORAW_22, NEORAW_23, NEORAW_24, NEORAW_25, NEORAW_26, NEORAW_27, NEORAW_28, NEORAW_29, NEORAW_30, NEORAW_31, NEORAW_32, NEORAW_33, NEORAW_34, NEORAW_35, NEORAW_36, NEORAW_37, NEORAW_38, NEORAW_39, NEORAW_40, NEORAW_41, NEORAW_42, NEORAW_43, NEORAW_44, NEORAW_45, NEORAW_46, NEORAW_47, NEORAW_48, NEORAW_49, NEORAW_50, NEORAW_51, NEORAW_52, NEORAW_53, NEORAW_54, NEORAW_55, NEORAW_56, NEORAW_57, NEORAW_58, NEORAW_59, NEORAW_60'
Anatomy	All FreeSurfer (FS_*) traits
Sleep	PSQI_Score, PSQI_Comp1, PSQI_Comp2, PSQI_Comp3, PSQI_Comp4, PSQI_Comp5, PSQI_Comp6, PSQI_Comp7, PSQI_Min2Asleep, PSQI_AmtSleep, PSQI_Latency30Min, PSQI_WakeUp, PSQI_Bathroom, PSQI_Breathe, PSQI_Snore, PSQI_TooCold, PSQI_TooHot, PSQI_BadDream, PSQI_Pain, PSQI_Other, PSQI_Quality, PSQI_SleepMeds, PSQI_DayStayAwake, PSQI_DayEnthusiasm, PSQI_BedPtnrRmate

Table SI-1. List of behavioural and anatomical traits per group.