Robustness of individualized inferences from longitudinal resting state dynamics

Maximilian HOMMELSEN[1], Shivakumar VISWANATHAN[1], Silvia DAUN[1,2*]

* Corresponding author

Running title: Individualized inference for longitudinal RS

Affiliations:

1. Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Forschungszentrum Jülich, Jülich, Germany

2. Institute of Zoology, University of Cologne, Cologne, Germany

Corresponding author:

Silvia Daun, PhD

Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Forschungszentrum Jülich, 52425, Jülich, Germany

Email: s.daun@fz-juelich.de, *Tel*: +49-2461-61 8638, *Fax*: +49-2461-61 1518

ABSTRACT

Individual human brains plastically reorganize over extended timescales ranging from days to years, which makes these changes difficult to track. One promising tracking indicator is the neural activity at rest. However, the rest state (RS) is vulnerable to incidental inter-day cognitive variability that can confound the detection of changes to underlying neurophysiology. Here we show that this confounding is minimized by tracking changes to the distinctiveness of an individual's RS activity, which is shaped by individual neurophysiology. Using longitudinal (5-day) RS acquired with EEG, we devised empirical simulations of distinctiveness changes confounded by cognitive variation. These inter-day changes were correctly classified with over 96% accuracy from 2-second snapshots of instantaneous oscillatory power from RS and confounded-RS. Surprisingly, the individual indicators of distinctiveness were concentrated at characteristic fronto-central and occipital zones. These findings support the suitability of longitudinal RS for individualized inferences about neurophysiological change in health and disease.

INTRODUCTION

Tracking how individual human brains change over extended time-scales (e.g., days to years) is crucial to identify the mechanisms of neural plasticity in scenarios ranging from healthy aging (Boersma et al. 2011; Cabeza et al. 2018; Cassani et al. 2018) to stroke recovery (Giaquinto et al. 1994; Rehme et al. 2011; Wu et al. 2016; Bonkhoff et al. 2020; Saes et al. 2020; van der Vliet et al. 2020). Given the many practical challenges that long-term tracking entails, one promising strategy to sample an individual's neural state repeatedly over time is with measures of resting state (RS) activity (Vecchio et al. 2013; Carino-Escobar et al. 2019; Newbold et al. 2020; Pritschet et al. 2020; Saes et al. 2020). RS activity refers to the ongoing neural activity while individuals maintain a minimally instructed state of "rest" over several minutes. Despite being an ill-defined state, oscillatory neural dynamics at rest have been shown to be reliable indicators of neurophysiological organization and integrity (Biswal et al. 1995; Damoiseaux and Greicius 2009; Van Den Heuvel et al. 2009; Hermundstad et al. 2013; Miŝic et al. 2016; Hoenig et al. 2018; Buckner and DiNicola 2019). However, the suitability of longitudinal RS to enable individual-specific inferences is poorly understood. Specifically, can inter-day differences in an individual's RS activity be translated into unambiguous inferences about that individual's changing brain? One important potential confound in drawing such inferences with RS activity is the ill-defined rest task itself.

The rest task has two dimensions: (i) a *behavioral* state typically specified by instructions to stay still and keep eyes open (or closed) (Barry et al. 2007); and (ii) a *cognitive* state typically specified by instructions to relax and avoiding thinking of anything specific. Although the behavioral state can be objectively verified, the criteria to verify the cognitive state are ill defined (Benjamin et al. 2010; Duncan and Northoff 2013; Kawagoe et al. 2018). Consequently, a person's RS activity on multiple days could theoretically be associated with different neural states depending on the person's incidental cognitive state on each day (e.g., session 1: free mind-wandering, session 2: struggling to stay awake, session 3: replaying emotional memories) (Diaz et al. 2013). This potential for <u>true</u> (i.e., non-artifactual) inter-day RS activity differences without changes to

underlying neurophysiology is a serious confounding factor when the objective is to detect and track individual-specific neurophysiological change. As this confound originates from the rest task itself, rather than a particular brain imaging modality or quantitative measure of RS activity (e.g., related to connectivity), the conundrum is whether this confound can be addressed without changing the rest task given its many other merits.

Here, we investigated whether RS activity inherently has properties that can help resolve this inferential confound. To identify these hypothesized properties, we devised empirical simulations of confounding inter-day RS activity changes. Then, using a machine-learning approach, we evaluated the robustness of RS-based inferences in the presence of these simulated confounds.

The problem of disambiguating the neural origin of longitudinal RS differences can be formulated as a classification problem. Suppose a person has the RS activity patterns A_p and A_q on days p and q respectively. If A_p and A_q differ from each other, the problem is to decide whether this difference is attributable to (i) a possible neurophysiological change (abbreviated as *NP*+), rather than (ii) an incidental difference in cognitive state (i.e., *NP*-). Hypothetically, if these two categories of activity differences were to have distinctive signatures, then a procedure to detect these signatures would enable an accurate categorization of an activity difference as either *NP*+ or *NP*- and thus resolve the confound.

The plausibility of such signatures is suggested by two differing kinds of constraints on RS activity. The RS activity evoked by the rest task can be formulated in terms of the task's unique information processing demands that differentiate it from other tasks, for example, memorizing a stimulus or executing a physical action. Alternatively, RS activity can be formulated as an activity pattern that is structured by each person's unique neurophysiological phenotype in a characteristic form that is also shared by other tasks. As these constraints, i.e., information processing demands versus individual-specific neurophysiology, emphasize different aspects of RS activity, we reasoned

that longitudinal changes to these constraints might have distinctive signatures in RS activity. We tested this hypothesis by empirically simulating longitudinal changes to these constraints.

Scalp-EEG at rest was longitudinally acquired on five consecutive days from a group of healthy, young individuals who were assumed to be neurophysiologically stable over this period. We additionally acquired EEG while these individuals performed two additional tasks on each day that were designed to produce *pseudo*-rest states, that were matched to rest in behavior but not in cognitive state (Figure 1). These pseudo-rest states served to simulate the confounding role of incidental cognitive state variation on inter-day RS changes (*NP*-). To simulate RS activity changes due to verifiable neurophysiological changes (*NP*+), we used a "crowdsourcing" approach where RS activity from *other* individuals served to empirically simulate an individual's changed neurophysiology across days.

With the above formulation, the problem of distinguishing *NP*+ from *NP*- was equivalent to the machine-learning problem of person identification (Valizadeh et al. 2019). Adopting this approach, classifiers were trained to distinguish each person from all other individuals based on two-second samples of the instantaneous power spectrum at rest on a single day. We then tested whether each individual's RS-based identity from one day was sufficient to re-identify that same individual from activity samples acquired on a different day. If *NP-/NP*+ have distinctive signatures, then an RS-based classifier for person *P* from one day should correctly re-identify *P* even from samples of pseudo-RS activity on a different day, namely, ignoring incidental inter-day variations in cognitive state (*NP*-). Conversely, *P*'s classifier should classify RS-activity samples from other individuals as not-*P*, namely, indicating a true change in individual-neurophysiology (i.e., *NP*+).

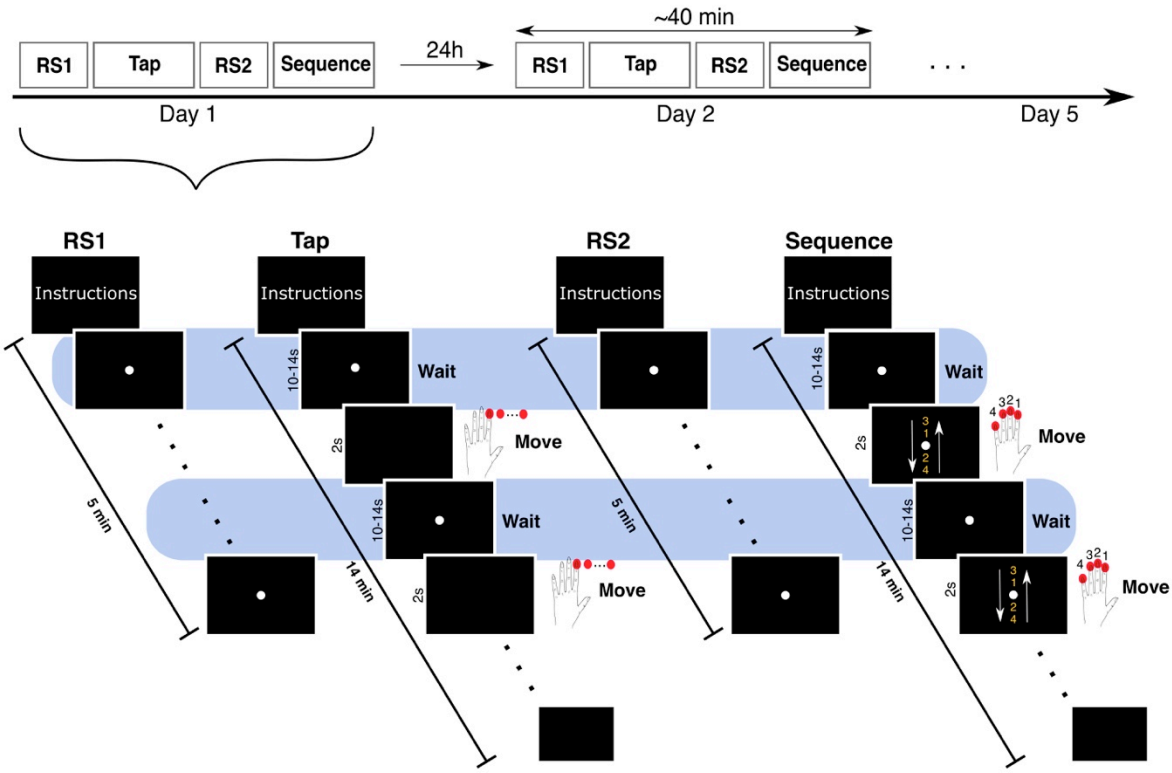


Figure 1: Experimental paradigm. Four tasks (*RS1, Tap, RS2, Sequence*) were performed daily in the same fixed order on five consecutive days. The details of each task on one day are schematically illustrated. A white fixation point was continuously displayed during the *RS1* and *RS2* task periods, and during extended "waiting" periods in the *Tap* and *Sequence* tasks (highlighted in blue). In the *Tap* task, a blank screen cued a 2s movement interval requiring left index-finger movements to repeatedly press a button (shown as red dots). In the *Sequence* task, the movement cue was an image depicting four numbers between two arrows (not drawn to scale). The number-to-finger mapping is shown with the cartoon (left) hand. The arrows indicate the required cyclical sequence of buttons to be pressed repeatedly, for example, 3-1-2-4-4-2-1-3-3-1-2-4, etc.

MATERIALS & METHODS

Participants

Twenty seven healthy volunteers (11 female, age (mean ± sd): 27.9 years ± 3.4, range: 22-34 years) participated in the study and received monetary compensation. Participants had normal or corrected-to-normal vision; no history of neurological or psychiatric disease; were not under medication at that time; and had no cranial metallic implants (including cochlear implants). Handedness was not an inclusion criterion. Based on the Edinburgh Handedness Inventory (Oldfield 1971), 22 participants were right handed (score > 50), 2 were left handed (score < -50) and 3 had intermediate scores. The study was approved by the Ethics Commission of the Faculty of Medicine, University of Cologne (Zeichen: 14-006). All participants provided their written informed consent before the start of the experiment.

Datasets from 24 (of the 27) participants were used for statistical analyses (see section: Data quality assessment).

Apparatus and EEG data acquisition

Stimuli were displayed using the software Presentation (v. 20.2 Build 07.25.18, Neurobehehavioral Systems, Inc.) on an LCD screen (Hanns-G HS233H3B, 23-inch, resolution: 1920 x 1080 pixels). Movement responses were recorded with the fMRI Button Pad (1-Hand) System (LXPAD-1x5-10M, NAtA Technologies, Canada).

Scalp-EEG was acquired with a 64-channel active Ag/AgCl electrode system (actiCap, Brain Products, Germany) having a standard 10-20 spherical array layout (ground electrode at AFz, reference electrode on the left mastoid). Three electrodes (FT9, FT10, TP10) were used to record electrooculographic (EOG) activity: one below the left eye to record vertical movements and the other two near the left and right lateral canthi to record horizontal movements. During acquisition, measured voltages (0.1µV resolution) were amplified by a BrainAmp DC amplifier (BrainProducts GmbH, Germany) at a sampling rate of 2.5 kHz and filtered (low cutoff: DC, high cutoff: 250 Hz).

To ensure reliable positioning of the EEG cap across sessions, a stereotactic neuronavigation system (Brainsight v. 2.3, Rogue Research Inc, Canada) was used on each session to co-register the spatial coordinates of five selected electrodes (AFz, Cz, POz, C5, C6) to their coordinates on the first session (see section: Procedure).

Experiment protocol and paradigm

Participant completed five sessions of approximately 40 minutes each (Figure 1, upper panel) where each session was scheduled at the same time on consecutive days (Monday to Friday). Sessions took place at three possible times: morning (6 x 9AM), noon (9 x 12PM) and afternoon (12 x 3PM). Due to technical problems during the scheduled recording, for one participant, the fifth session was re-acquired after a gap of three days.

Each session (Figure 1, lower panel) consisted of two resting state recordings (*RS1* and *RS2*) interleaved with two non-rest tasks (referred to as *Tap* and *Sequence*) that required participants to press buttons in response to visual cues. These four task periods had the same order (i.e., *RS1*, *Tap, RS2, Sequence*) for all participants on all days. Each task period began with an instruction screen describing the task to be performed. Following task completion, another instruction screen informed participants to take a short break and to press a button to initiate the next part when they were ready.

As described in detail below, the *Tap* and *Sequence* tasks had relatively long and variable inter-stimulus-intervals (10-14s) where participants had to fixate on the screen as they "waited" for a visual cue to initiate a response. The cognitive states during these waiting periods (referred to as *TapWait* and *SeqWait*) were the primary focus of these tasks. The behavioral demands of the *Tap* and *Sequence* tasks was designed to modulate the cognitive states during these pre-movement wait periods, for example, related to movement preparation during *TapWait* and covert rehearsal of a movement sequence during *SeqWait.* With this covert modulation, the *TapWait* and *SeqWait* could be considered pseudo-rest states as they were matched to *RS1* and *RS2* in behavioral state but not in cognitive state. Furthermore, the *Tap* task was intended to produce cognitive states that were similar

within and between days while the *Sequence* task was designed to elicit cognitive states that could systematically change across days. This was implemented by inducing subjects to learn a difficult motor sequence where performance could improve with increasing practice across days. We now describe the different task periods in detail.

1: Resting State (RS1). During this period lasting ~5minutes, a small white dot was continuously displayed at the center of the screen. Participants were instructed to keep their eyes open, fixate on the displayed white dot, relax and avoid movements (also see section: Procedure).

2: Tap task. In this task-period, a small white dot was centrally displayed on the screen (as in *RS1*). However, after variable intervals of 10-14 seconds, this dot disappeared for a 2 second period before reappearing. The offset of the dot was the cue for participants to repeatedly and rapidly press a button with their left index finger until the dot reappeared on the screen. The total task (duration ~14 minutes) consisted of 60 movement periods (dot absent) interleaved with 60 waiting periods (dot present). These waiting periods are referred to as *TapWait* and the response execution periods are referred to as *TapMov*.

3: Resting State (RS2). A second resting state recording was acquired with all task parameters being identical to *RS1*. This recording is referred to as *RS2*.

4: Sequence task. As with the *Tap* task, the sequence task consisted of 60 waiting periods of 10-14s each (i.e., *SeqWait*) where a small white dot was centrally displayed on the screen interleaved with 60 movement periods of 2s duration (i.e., *SeqMov*). Unlike the *Tap* task, each movement period was cued by a centrally displayed visual stimulus consisting of four vertically displayed digits (3-1-2-4) between two vertical arrows. Each number was mapped to a different button on the response pad. The vertical ordering of the numbers indicated the sequence in which the indicated buttons had to

be pressed using fingers of the left hand. The arrows indicated that this sequence had to be executed rapidly and repeatedly in a cyclical manner starting from top to bottom and back. For example, following stimulus onset, the required sequence of button-presses was 3-1-2-4-4-2-1-3-3-1-2-4-… and so on. This continuing sequence had to be executed until the offset of the stimulus. No performance feedback was provided during the task. This particular sequence of digits was selected as it was challenging to execute rapidly. To promote learning of this sequence across trials and days, the same sequence of numbers and number-to-finger-mapping was used on all sessions. The same sequence and number-to-finger mapping was also used for all participants.

Handedness was not an inclusion criterion in our experiment. To obtain a degree of homogeneity in the task-related neural activity, all participants used the fingers of their left hand to execute the button-press responses in the *Tap* and *Sequence* tasks.

Procedure

Prior to the start of the recordings on each of the five days, participants completed the Positive and Negative Affect Schedule (PANAS) (Watson et al. 1988) and completed brief questionnaires to report the caffeine consumption on that day and the amount and quality of sleep on the previous night. On the first day, participants received detailed instructions about the experiment. For the resting state periods, participants were instructed to keep their eyes open, fixate on the displayed white dot and to avoid movements. Additionally, they were also asked to relax, stay awake and not think of anything in particular. For the *Tap* task, participants were instructed to press the buttons as rapidly as possible without causing discomfort. For the *Sequence* task, participants were familiarized with the task and the mapping of the number to finger. They practiced performing the task using a different digit sequence from the one used in the main experiment. Furthermore, they were explicitly instructed to try to improve their performance (particularly the number of buttons pressed during each response period) on each session. Finally, on all sessions, we repeatedly emphasized the importance of minimizing eye-blinks, maintaining

fixation at all times during the recording, and the avoidance of all unnecessary movements of the fingers, head and body.

As the study's objective was to relate the spatio-temporal organization of neural activity across days, minimizing inter-day variation in the EEG cap's position was an important priority. We therefore implemented an additional spatial registration procedure on each day after the EEG cap was secured to the participant's head. Using a stereotactic neuronavigation system, the participant's head was registered to the Montreal Neurological Institute (MNI) space using standard cranial landmarks. The positions of five selected electrodes along the midline and lateral axis (AFz, Cz, POz, C5, C6) were then localized using the neuronavigation software. The electrode locations obtained on the first day were used as the reference for the remaining sessions. On each subsequent session, the positioning of the cap was interactively adjusted so that each electrode's coordinates closely matched its reference location. Due to scheduling constraints, this spatial registration procedure was not performed for 7 participants.

The application of electrode gel followed after cap positioning. Skin-electrode impedance was brought below 10 kΩ before starting the recording. Recordings were acquired in a light-dimmed and acoustically shielded EEG chamber. Participants were seated in a comfortable chair with their heads stabilized with a chinrest in front of the computer screen at a viewing distance of ~65cm. The response pad was placed in a recess under the table so that participants could not see their hands during the task-periods (especially while pressing the buttons). During the recording, participants were monitored via a video camera to ensure that they maintained fixation, minimized eye-blinks, and stayed awake.

EEG preprocessing

The EEG data were preprocessed using the EEGLAB software (Delorme and Makeig 2004) and custom scripts in a MATLAB environment (R2016b, MathWorks, Inc., Natick, MA).

The continuous recordings were down-sampled to 128Hz, and then band-pass filtered to the

range 1Hz-40Hz with a Hamming windowed sinc FIR filter (high pass followed by low pass). The continuous recordings then underwent an artifact correction process to remove oculomotor activity related to eye-blinks and saccades.

Eye blink removal was performed separately for each day's dataset (including all task periods) using the procedure described by Winkler et al. (2015). Following this procedure, a copy of a dataset was first created which was then filtered with a high-pass 2 Hz filter. This duplicate dataset was visually inspected to remove data segments and EEG channels with artifacts related to repeated paroxysmal amplitudes changes (> 50μV), electromyographic contamination, electrical noise and signal loss. Next, the artifact-free data from all task-periods were segmented into 2s epochs. These epochs were then submitted to an Independence Components Analysis (ICA) decomposition using the infomax-ICA algorithm (implemented as *runica* in EEGLAB). To minimize inter-condition biases, ICA was performed on a balanced mixture of epochs from *RS1, TapWait, RS2* and *SeqWai*t. Epochs from the *TapMov* and *SeqMov* periods were excluded from this step to avoid movement-specific biases. The ICA weights obtained with the duplicate dataset were then transferred and applied to the original, non-filtered dataset. ICA components related to eye-blinks and saccades were then identified and removed using an automatic detection algorithm ADJUST (Mognon et al. 2011).

Following eye-blink correction, the original dataset was then again visually inspected to remove time periods and channels with artifacts. The signals in rejected channels were replaced with signals interpolated from other channels using spherical spline interpolation. All channels were then re-referenced to the Common Average Reference. Finally, the visually inspected continuous data were segmented into 2s epochs according to the six different experimental states: *RS1*, *RS2*, *TapWait*, *TapMov*, *SeqWait* and *SeqMov.* For the two movement-related states (*TapMov* and *SeqMov),* epochs were defined from +0.25s to +2.25s following the visual cue to exclude initial transients and response-time delays following cue onset and to include residual movements in the period immediately following the cue offset. To avoid any carry-over effects from movement into

the *TapWait* and *SeqWait* epochs, a time interval of 500ms immediately prior to cue onset and 1000ms immediately following cue offset were excluded before segmenting the *TapWait* and *SeqWait* epochs. Furthermore, *TapWait* and *SeqWait* epochs that contained any button presses were excluded.

To establish face-validity of the task states based on their time-courses, we created a separate set of epochs from -1 to +3s relative to the onset of the visual cue. The signals were band-pass filtered in the β frequency band (14-30Hz) and the signal amplitude was extracted using the Hilbert transform. After removing edge artifacts, the signal was normalized by calculating the percentage change in the signal relative to the mean amplitude in the pre-stimulus period [-898ms, 0ms]. After normalization, the signals were averaged across epochs, days and subjects.

Data quality assessment

Preprocessing resulted in 135 datasets (27 subjects x 5 days). To be included in our analysis, each subject had to have completed the first three of the four tasks on all sessions and have at least 4 (out of 5) session-datasets that met the following data-quality criteria. For analysis, we required a preprocessed dataset to have (i) less than seven rejected channels, (ii) ≥ 90 artifact-free epochs from both resting state periods (i.e., *RS1* and *RS2*), and (iii) ≥ 90 artifact-free epochs from the available resting-matched conditions (i.e., *TapWait, SeqWait*). Note that the number of epochs for *TapMov* and *SeqMov* were necessarily ≤ 60 as each task only had 60 response periods of 2s duration.

Datasets from 24 out of 27 subjects met these data-quality criteria: 18 (of 24) had completed all 4 task-periods on each session and the remaining 6 (of the 24) subjects had completed only the first 3 (of the 4 parts). To maintain uniformity in the statistical analyses, final analyses were performed only on the best 4 of the 5 session-datasets. For participants where all 5 datasets were of high quality, we excluded the first day's dataset as it might involve effects of initial familiarization. To maximize the use of the available data after these exclusions, analyses involving only *RS1* and *RS2* included data from 24 participants, while analyses involving any of the non-rest tasks used data

from 18 participants. For these 18 participants, the mean number of epochs per day in *TapMov* was 52.7 (min: 45.3, max: 57.7; SD = 2.9; minimum/day = 36) and in *SeqMov* was 53.4 (min: 49.2, max: 57.7; SD = 2.5; minimum/day = 42).

Feature specification: Oscillatory power spectrum

All classification analyses were based on a description of the oscillatory power spectrum on each 2-second epoch. Each epoch's power spectrum was described using 305 features that specified the power in five canonical frequency bands (δ: 1-3.5 Hz; θ: 4-7.5 Hz; α: 8-13.5 Hz; β_1 (low β): 14-22.5; β_2 (high β): 23-30 Hz) at each of the 61 channels.

These features were extracted with the procedure schematically displayed in Figure 2A. For each 2s epoch of EEG activity, the oscillatory power spectrum at each channel over the range of 1 to 30 Hz (0.5Hz resolution) was computed using the Fast Fourier Transform (FFT). The power at all frequencies within each band's frequency range was averaged to obtain the mean power per frequency band. The mean power per band was then logarithmically transformed (base 10) so that the resulting distribution across epochs had an approximate normal distribution. These five features (one per band) provided a minimal description of each channel's power spectrum. Finally, these five features from each channel were concatenated to obtain a single vector with 305 feature values (5 frequency bands x 61 channels). This extended feature set describing an epoch's power spectrum across the scalp was used for the classification analyses.

For detailed analyses, we defined subsets of the full feature set referred to here as the (i) mono-band and (ii) mono-location feature sets. Each *mono-band* feature set (B_f) consisted of features belonging to only one frequency band *f*. The five mono-band feature sets (each with 61 features) were B_δ, B_θ, B_α, $B_{\beta 1}$ and $B_{\beta 2}$. Each *mono-location* feature set (L_z) (Figure 2A, top panel) consisted of features from 10 bilaterally symmetric channels in the spatial zone *z* on the scalp along the anterior-posterior axis. The four mono-location sets were defined at the frontal (L_F); fronto-central (L_{FC}), centro-parietal (L_{CP}) and parieto-occipital (L_{PO}) zones respectively.

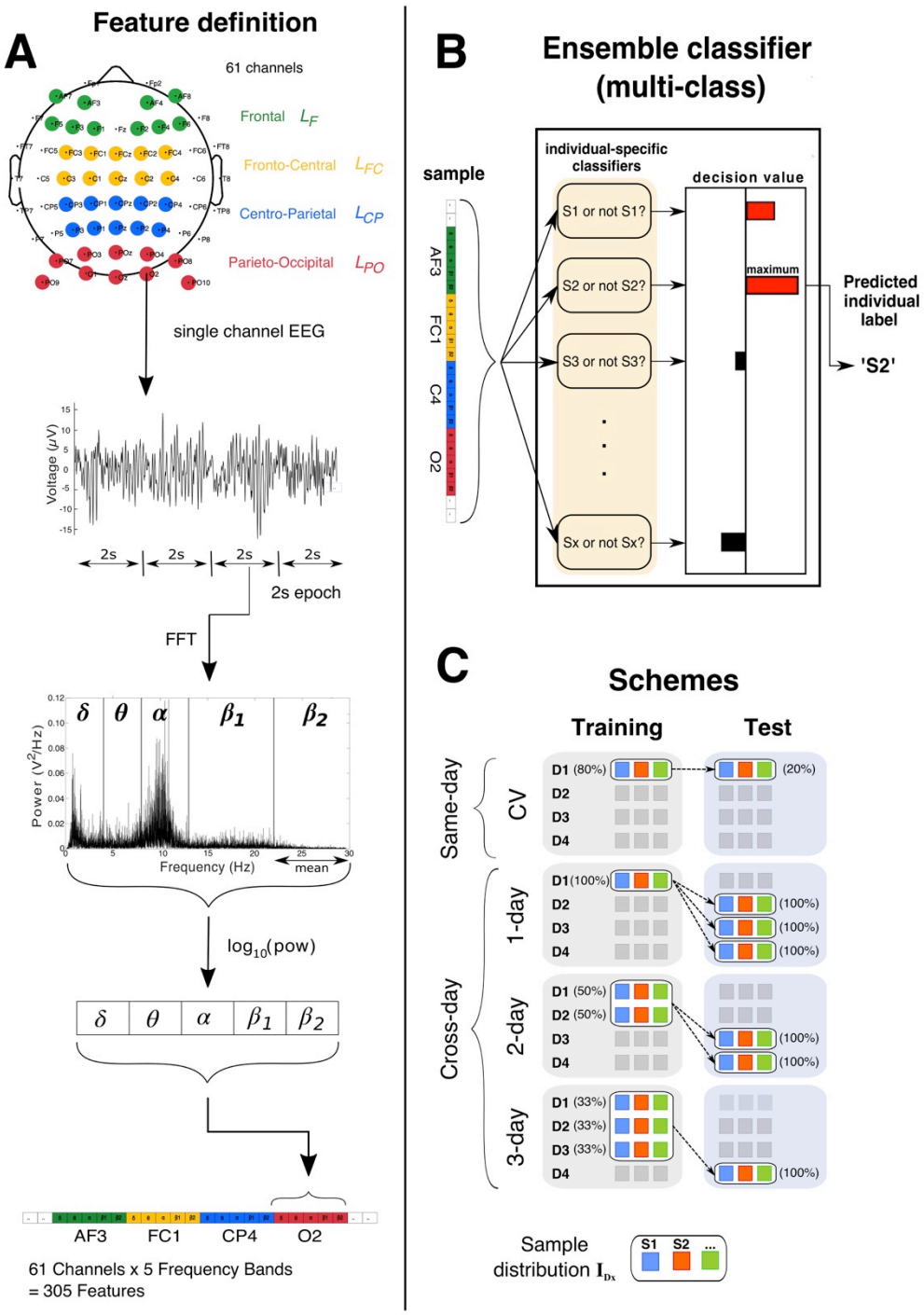


Figure 2: Classification procedure. (A) Feature definition pipeline**.** Channels in each mono-location subset are identified by color (green: L_F, yellow: L_{FC}, blue: L_{CP}, red: L_{PO}). The continuous signal from each channel was segmented into 2s epochs followed by an estimation of the frequency spectrum with the Fast Fourier Transform (FFT). The mean power within each of the five bands was log transformed (base 10) and concatenated with corresponding values from all other channels to obtain a feature vector. **(B)** Schematic of a multiclass decision with an ensemble of individual-specific binary classifiers. Each classifier evaluates the sample (*Sx* or not-*Sx*) to output a decision-value (red bars > 0, black bars < 0) and the classifier with the maximum decision value was the predicted label (here, S2). **(C)** Classification schemes ${}^{A}\mathbf{I}_{Dx} \rightarrow {}^{A}\mathbf{I}_{Dy}$ (rows) were defined by the configuration of training (left column) and test sets (right column) (where **Di** denotes samples from day *i*). The sample distribution ($\mathbf{I}_{Dx}$) had samples from all individuals (multi-colored boxes). Percentages indicate the proportion of each day's samples used for training/testing. Same-day identification was estimated with 5-fold cross-validation (CV). The training set for cross-day aggregation had an equal proportion of samples from each day and the total number of training samples was the same across aggregation levels.

Multi-class classification: Definition and accuracy

All classification models were numerically estimated using a soft-margin linear Support Vector Machine (SVM) algorithm as implemented by the *LinearSVC* package in the *scikit-learn* library (Pedregosa et al. 2011) implemented in Python 3.6. SVM learning was initialized with parameters (tolerance = 10^{-5}, max iterations =10^4, hinge loss, balanced class weighting, $C = 1$).

As defined above, each epoch was a 2-second sample of the ongoing oscillatory activity from one person (of 24) on one specific day (of 4) engaged in a particular task state (of 6 possible states: true rest {*RS1*, *RS2*}, pseudo-rest {*TapWait, SeqWait*}, non-rest {*TapMov, SeqMov*}). The classification analyses involved predicting an epoch's origin either by (i) a person's identity or (ii) task-state. Multi-class classifiers (using an ensemble of binary classifiers) were used for person identification as described below. Standalone binary classifiers were used to identify alternative task-states of the same person.

The input to a multi-class classifier (see Figure 2B) was a single sample (i.e., epoch) from an unspecified person S_x in the studied group and the required output was the predicted identity of that person (e.g., S_2). The multi-class classifiers used here employed a *one-vs-all* scheme (as implemented by *scikit-learn*). Specifically, an N-class classifier ($N \geq 2$) consisted of an ensemble of N binary-classifiers where each was independently trained to distinguish whether a sample was from one specific person (e.g., S_2) or from any of the other N-1 persons (i.e., not S_2). With such an ensemble, each sample was separately evaluated by each of the N binary-classifiers to obtain a decision value from each classifier (i.e., the signed distance to the separation hyperplane (Rifkin and Klautau 2004)). These decision values were compared and the final classification was assigned to the binary classifier with the maximum decision value.

To obtain a measure of classification accuracy of each individual classifier from the ensemble classification accuracy, we defined the accuracy a_i of the classifier for person S_i as

$$a_i = \frac{1}{2}\,(H_i + CR_i)$$

where H_i denotes the hit rate (i.e., positive identification rate) of the classifier and CR_i denotes the correct rejection rate. The hit rate H_i was the proportion of instances where samples from S_i were correctly predicted as being from S_i by the ensemble (i.e., the classifier S_i had a larger decision value than the competing classifiers). Correct rejection was defined based on the pair-wise relationship of S_i to each of the other classifiers S_j. If the ensemble (incorrectly) predicts S_i for a sample from a different person S_j then it implies that the classifier S_i (incorrectly) had a larger decision value than the competing classifiers, i.e., a false positive. The false positive rate $FP_{i,j}$ denotes the proportion of instances where samples from S_j were incorrectly predicted as being from S_i by the ensemble. The correct rejection $CR_{i,j}$ was defined as $CR_{i,j} = 1 - FP_{i,j}$. Based on this rationale, the overall correct rejection CR_i for S_i was defined as the mean of the pair-wise correct rejection rates

$$CR_i = \frac{1}{N-1}\sum_{j=1}^{N} CR_{i,j} \text{ where } j \neq i$$

With this formulation, random chance for each classifier was 50% even though random chance for the entire ensemble was (100/*N)%*.

Multi-class classification: Training and testing schemes

Classification was defined by the samples used for training and testing. Irrespective of classifier type (multi-class or standalone binary classifier), the training data were always balanced, i.e., having an equal number of samples per class, to avoid biases arising from imbalanced classes (Abraham and Elrahman 2013).

Person (multi-class) identification was organized into two schemes based on whether the training and test samples belonged to the (i) same day (namely, same-day vs cross-day identification) and the (ii) same task (namely, same-task vs cross-task identification). A schematic of the same-day/cross-day schemes are shown in (Figure 2C). For convenience, we use the following notational convention to describe these classification schemes. As multi-class

classification involves an ensemble decision, it involves the conjoint influence of the sample distributions from multiple persons. This combined distribution on a particular state (e.g., *RS1*) on day *d* is denoted as ${}^{RS1}\mathbf{I}_d$. A classification scheme where a decision-rule is trained on samples from ${}^{A}\mathbf{I}_p$ (i.e., from task state *A* on day *p*) and tested on samples from ${}^{B}\mathbf{I}_q$ (i.e., from state *B* on day *q*) is denoted as ${}^{A}\mathbf{I}_p \rightarrow {}^{B}\mathbf{I}_q$. Similarly, a classification scheme where a decision-rule was trained on samples aggregated from different days (e.g., ${}^{A}\mathbf{I}_p$ and ${}^{A}\mathbf{I}_q$) and tested on ${}^{B}\mathbf{I}_r$. is denoted as ${}^{A}\mathbf{I}_p \circ {}^{A}\mathbf{I}_q \rightarrow {}^{A}\mathbf{I}_r$. (see below).

<u>*Same-day/same-task identification:*</u> The accuracy of same-day person identification in task state *A* (${}^{A}\mathbf{I}_p \rightarrow {}^{A}\mathbf{I}_p$) was estimated using a 5-fold cross-validation (CV) procedure (Blum et al. 1999). Specifically, the set of samples from state *A* on one day (for example, day D1 in Figure 2C, upper row), were partitioned into 5 equal folds. Training was performed on four folds (80% of the sample set) and tested on the left-out fifth fold (the remaining 20%). This training-testing procedure was repeated so that each fold was used as a test-set once. The mean classification accuracy across folds was defined as the same-day identification accuracy for that day. In this manner, the CV accuracy was estimated separately for each of the four days and the mean CV accuracy across days was denoted as the same-day accuracy for task state *A*.

<u>*Cross-day/same-task identification:*</u> For cross-day identification in task state *A* (${}^{A}\mathbf{I}_p \rightarrow {}^{A}\mathbf{I}_q$), samples in the test set were from a different day than the samples in the training set. We modulated the training set's day-specificity by aggregating samples from different days in a stratified manner. In an *n*-day training set, the *k* training samples per person consisted of *k*/*n* samples from each of *n* different days. Here, *n* could take three possible values, namely, 1,2 or 3 (see Figure 2C, first column). The number of samples per person, *k*, was held constant to enable comparison of classification accuracy across all values of *n*. Unlike the training set, samples in the test-set were always from the same day. Mean identification accuracy for a particular *n*-day aggregation scheme (e.g., ${}^{A}\mathbf{I}_{d1} \circ {}^{A}\mathbf{I}_{d2} \ldots {}^{A}\mathbf{I}_{dn} \rightarrow {}^{A}\mathbf{I}_r$.) was obtained by (i) independently estimating the accuracy for each

possible training/test-set combinations that satisfied the day constraints (e.g., day $p \neq$ day $q \neq$ day r) and then (ii) averaging these accuracy values.

Cross-task identification: This was treated as a special instance of cross-day identification. For example, the accuracy for the configuration ${}^{A}\mathbf{I}_p \rightarrow {}^{B}\mathbf{I}_q$ was estimated by replacing the test set with samples from state *B* while retaining all other day-related constraints as in cross-day/same-task identification. Unless specified otherwise, cross-task identification was always tested across days, that is, the training and test sets were always from different days. This was done to exclude potential inter-state similarities that might be present due to the conjoint preprocessing of data from all states within the same day (see section: EEG preprocessing).

Classification schemes: Interpretation of accuracy relationships

The same-day accuracy for a particular state was treated as a pre-condition to estimate the cross-day identification accuracy for that state. If same-day accuracy was greater than random chance, it would confirm that the distribution from which the training set was drawn contained sufficient information to allow identification in the absence of potential inter-day changes. Cross-day accuracy is reported and interpreted here only if this pre-condition was satisfied.

Based on this pre-condition, a reduction in cross-day (1-day) accuracy (e.g., ${}^{A}\mathbf{I}_p \rightarrow {}^{A}\mathbf{I}_q$) relative to same-day accuracy (e.g., ${}^{A}\mathbf{I}_p \rightarrow {}^{A}\mathbf{I}_p$) can be attributed to a systematic difference in the distributions ${}^{A}\mathbf{I}_p$ and ${}^{A}\mathbf{I}_q$ between days (red arrow, Figure 3A). Aggregation was used to evaluate the source of this cross-day accuracy reduction by varying the statistical properties of the training set (i.e., by aggregating samples across days) while holding the properties of the test set constant. Specifically, we assumed aggregation would lead to decision-rules that discount day-specific properties in favor of day-general properties. Therefore, depending on the relative roles of day-specific/general properties in the classification decision, the cross-day accuracy might stay constant, increase or decrease with increasing aggregation (Figure 3A).

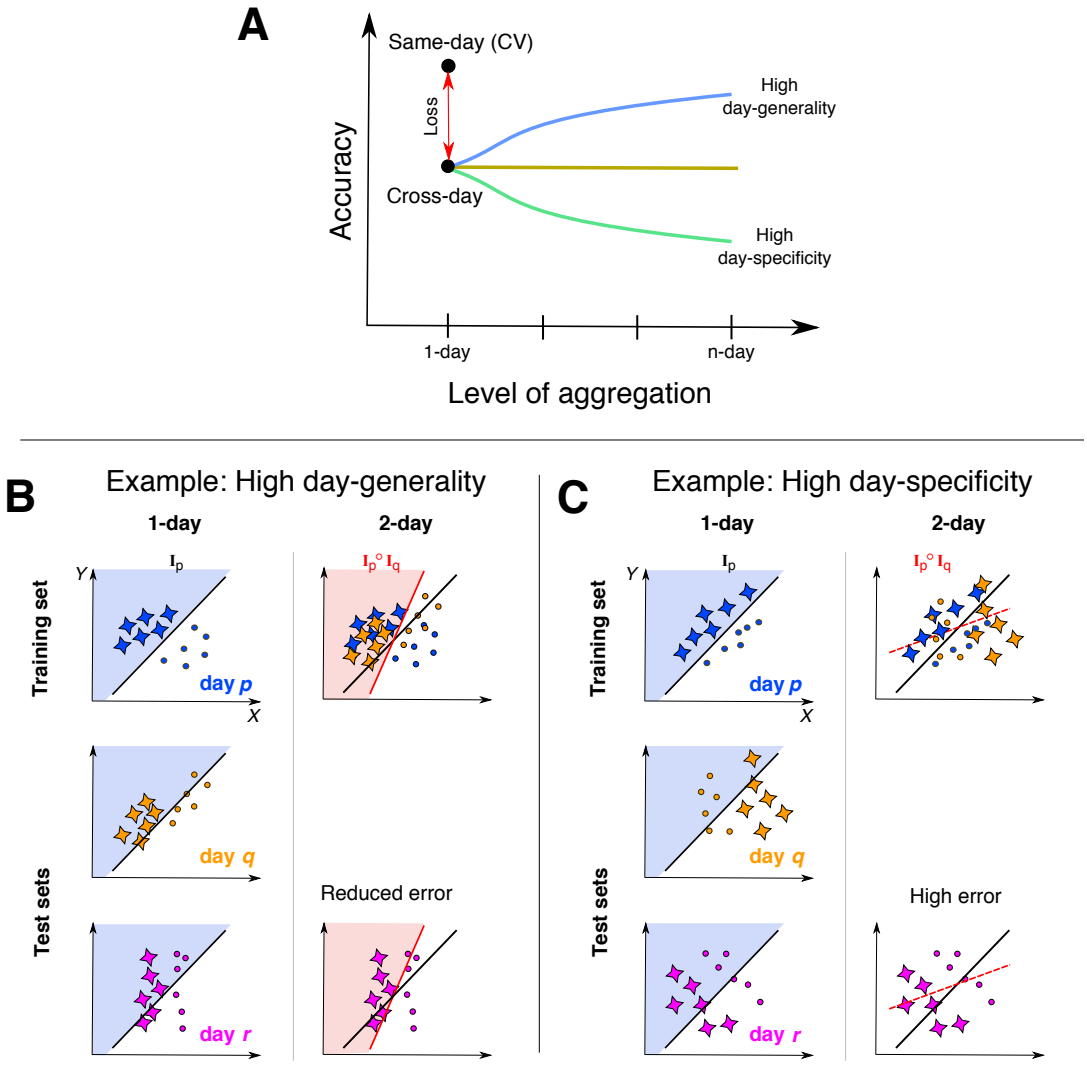


Figure 3: Effect of aggregation. (A) Schematic of relationship between same-day and cross-day identification accuracy. Cross-day (1-day) accuracy can be lower than same-day accuracy (red-arrow) due to day-specificity of the decision-rule. Training decision-rules on aggregated samples (y-axis) can change cross-day accuracy, which could increase (blue, see panel **B**), or stay constant (dark green), or even decrease (light green, see panel **C**). (**B**) Idealized example of how cross-day accuracy (column 1) can increase with aggregation (column 2) due to day-general information. Samples from two classes (stars, circles) are shown along two features (day-general: *X*, day-specific: *Y*) with each day's samples shown in different colors (*p*: blue, *q*: orange, *r:* purple). The 1-day decision-rule ($\mathbf{I}_p$) (top left panel) is depicted with thick black line and shaded areas. This decision-rule can successfully classify samples from days *q* and *r* but with some errors. However, a decision-rule trained on data from days *p* and *q* ($I_p \circ I_q$) (thick red line, red shaded area) reduces cross-day classification errors (lower right). **(C)** Idealized example of high day-specificity. Even though the classes are separable within each day, the 1-day decision-rule ($\mathbf{I}_p$) has a poor cross-day accuracy (column 1). 2-day training (column 2) produces a decision-rule with worse classification both on the training set itself (dotted red line) as well as across days (lower right).

Figures 3B and 3C show idealized examples of how aggregation could both increase as well as decrease cross-day accuracy. In the example shown in Figure 3B, the two classes systematically differ on feature *X* (x-axis) but with an inconsistent role for feature *Y* (y-axis). Due to incidental day-specific variation, feature *Y* has a role in distinguishing the classes on day *p* but not on other days. Consequently, a decision-rule trained on day *p* does not effectively separate the classes on other days (column 1). However, training on aggregation samples from day *p* and *q* (column 2) reduces *Y*'s role in the aggregated decision-rule leading to an improved separation of the classes across days. Figure 3C illustrates an extreme example of day-specificity where the two features have a conjoint relationship allowing classification within each day but with low generality across days. Therefore, training on samples aggregated from day *p* and *q* leads to an overall reduction in accuracy on the training set itself as well across days.

Multi-class classification: Weights and normalized weights

The characteristic weights for a particular classification scheme (e.g., ${}^{A}\mathbf{I}_p \rightarrow {}^{A}\mathbf{I}_q$) were obtained by averaging the weights across all training sets. In a multiclass classifier, the decision-rules are organized in a winner-take-all competition to label each sample (Figure 2B). Therefore, for each sample to be uniquely assigned to only one person, the person-specific classifiers in the ensemble necessarily require different decision-rules. This difference in decision-rules might only be in the sign (positive/negative) assigned to the weights. Therefore, for all weight-related analyses, the absolute values of the weights were used in order to allow inter-individual comparisons.

The distribution of high-valued weights in the mono-band/location subsets was estimated as follows. For each subject, the absolute weights of all features were first sorted. The weights in the kth percentile of the sorted weights were then identified and the relative proportion of these selected high-valued weights contained in each of the mono-band sets was then calculated. This procedure was repeated for different values of *k*. These proportions were separately calculated for the mono-

location sets. Since the temporal channels were not included in any of the mono-location subsets, these channels were excluded for the proportion calculations.

To identify the high-consistency weights, the absolute weights were z-scored across all features for each subject to retain information about inter-feature weight differences in the statistical tests.

For features i and j, the weight $|w_i|$ might be greater than $|w_j|$ while the power $|P_i|$ might be less than $|P_j|$. Consequently, neither the relationship between the weights nor the power are reliable indicators of the relative influence of i and j on the eventual classification decision. Therefore, we defined a feature i's unit weight $\overline{w}_i$ as the idealized weight value such that $\overline{w}_i P_i = 1$. The normalized weight was thus defined as the ratio $w_i/\overline{w}_i$, which was effectively equal to $w_i P_i$. Due to the large differences in power between bands, for statistical comparisons, the absolute normalized weights (i.e., |w*P|) were z-scored <u>within</u> each band for each subject.

Statistical Analysis

The relative differences in the accuracy of different classification schemes were assessed by performing paired t-tests, repeated measures one-way or two-way analysis of variance (ANOVA) implemented by the pingouin python package (version 0.3.2) (Vallat 2018).

The random chance accuracy for the multi-class and standalone binary classifier was 50% and accuracy deviations from random chance were evaluated with one-sample t-tests. The Bonferroni threshold was used to correct for multiple comparisons. Due to the sequential relationship between the different multiclass classification schemes, following Figure 3A, the tests for same-day accuracy (CV) and cross-day accuracy were planned tests that were considered significant at a threshold of $p < 0.05$. The tests for 2-day and 3-day aggregation were evaluated at a threshold of $p < (0.05/2)$. For tests for the mono-band and mono-location sets, the thresholds were further corrected for the number of feature sets.

Two kinds of error-bars are used in the plots. For plots depicting variable changes due to a

single-factor, error bars indicate the standard deviation (SD). Plots depicting multi-factor changes use error bars displaying the within-subject standard error (s.e.m.) (O'Brien and Cousineau 2014). The type of error-bar is explicitly noted in the figure caption.

RESULTS

Face-validity of individual power spectra

The individual-identification approach assumes that (i) there are likely to be inter-day differences in an individual's power spectrum at rest, (ii) there are systematic inter-individual differences in the power spectra at rest. Qualitatively, these assumptions were readily evident in the full spectrum (Figure 4A) before its reduction into the minimal description used for the classification analyses. The power spectrum at channel Oz is shown for all individuals on all days and task-states (Figure 4B). The power spectrum in *RS1* showed a typical shape with high power in the delta and alpha frequency bands, with a prominent peak at the alpha band for several individuals. There were qualitative differences in the power spectra between days (vertical) and between states (horizontal). For example, the prominent peak in the alpha frequency band was greatly reduced for *SeqMov* relative to all other states. Figure 4C shows the mean power spectrum (averaged across days) in *RS1* across channels for a few selected individuals (diagonal). Each individual's spectra showed a characteristic form across channels with a characteristic higher power in the delta and alpha bands and a higher overall power in the posterior and anterior channels relative to the central channels. This pattern was evident in the systematic inter-individual differences in the power spectra.

However, what was less obvious was whether an inter-day change in person S_i's power spectrum could lead S_i being mis-identified as S_j. Here we sought a quantitative answer to this question.

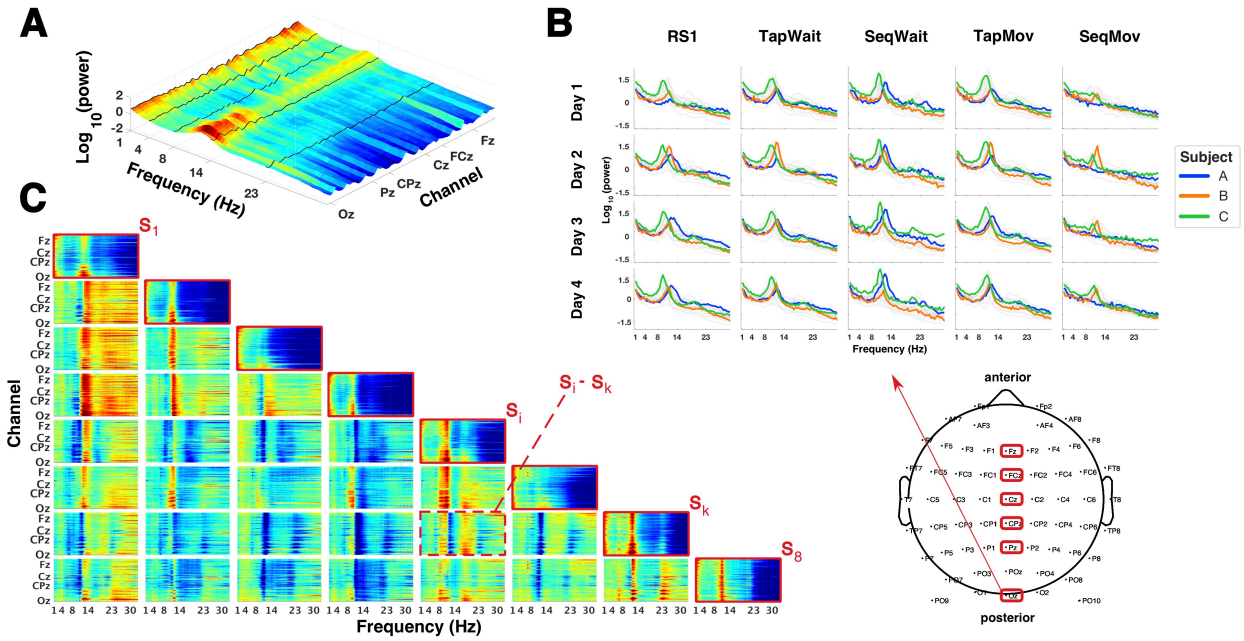


Figure 4: Inter-day and inter-individual differences. (A) Representative example of individual oscillatory power spectrum in *RS1* (averaged across days) at all channels across the scalp (posterior-to-anterior ordering). Dark black lines indicate frequency band boundaries. At all channels, power was higher in the δ and α bands, and higher in posterior and anterior channels while being lower at the central channels. **(B)** Frequency spectra at channel Oz for all subjects (gray lines) by day and task-state (blue, orange, green lines indicate three representative subjects). Individual frequency spectra have a characteristic shape across task-states but with modulations in power across days and task-states, especially in *TapMov* and *SeqMov*. **(C)** Mean cross-day oscillatory power in *RS1* for 8 selected subjects (diagonal, red boundary) and their pair-wise inter-individual differences. For each pair of individual S_i and S_k, the difference in power S_i - S_k is shown at row *i*, column *k* (example shown with dotted red line). Inter-individual power differences are confined to narrow frequency bands (e.g. α-band in S2-S3) while being broadly distributed for others (e.g. S4-S7).

Robust identification of individuals from RS activity within and across days

To identify a person from a 2s sample of RS activity with an ensemble classifier, a decision-rule was numerically estimated to represent each person's unique RS characteristics. The decision-rules estimated for each day could identify each person (of 24) from a sample acquired on the same day (i.e., according to the scheme ${}^{RS1}\mathbf{I}_p \rightarrow {}^{RS1}\mathbf{I}_p$) with a mean cross-validated (CV) accuracy of 99.98 ± 0.04% (mean ± sd) that was significantly larger than the theoretically expected accuracy for random guessing [> 50%: $t_{23} = 5596.13$, $p < 0.00001$]. However, for longitudinal RS, the key issue was whether decision-rules from one day could identify a person from a sample acquired on a different day (i.e., ${}^{RS1}\mathbf{I}_p \rightarrow {}^{RS1}\mathbf{I}_q$). The same-day decision-rules identified individuals across days with a mean accuracy of 92.10% ± 6.8% that was higher than random chance [$t_{23} = 30.14$, $p < 0.00001$] but less accurate than same-day identification by ~8% [paired $t_{23} = 5.64$, $p = 0.00001$]. This reduction in identification accuracy across days illustrated the challenge of longitudinal RS. Errors in identifying a person *P* across days seemingly suggest that *P's* unique activity characteristics had changed across days. However, such inter-day neurophysiological changes were unlikely to have occurred in the individuals studied here.

In numerical terms, the cross-day loss in accuracy implies that certain properties of each day's decision-rules were of predictive relevance to same-day samples but of limited generality to other days. To discount the role of these day-specific properties in favor of day-general properties, the decision-rules were trained using samples aggregated from *multiple* days (i.e., ${}^{RS1}\mathbf{I}_p \circ {}^{RS1}\mathbf{I}_q \ldots \rightarrow {}^{RS1}\mathbf{I}_s$)(see Methods). The mean cross-day accuracy *increased* from 92.10% ± 6.8% without aggregation (1-day) to 95.92 ± 3.63% with 2-day aggregation, with an additional increase to 97.39% ± 2.65% with 3-day aggregation [one-way ANOVA, $F_{2,46} = 28.83$, $p < 0.00001$] (Figure 5, Table 1). Following aggregation, the cross-day accuracy was a mere ~2% lower than the same-day accuracy consistent with the absence of large-scale neurophysiological changes between days.

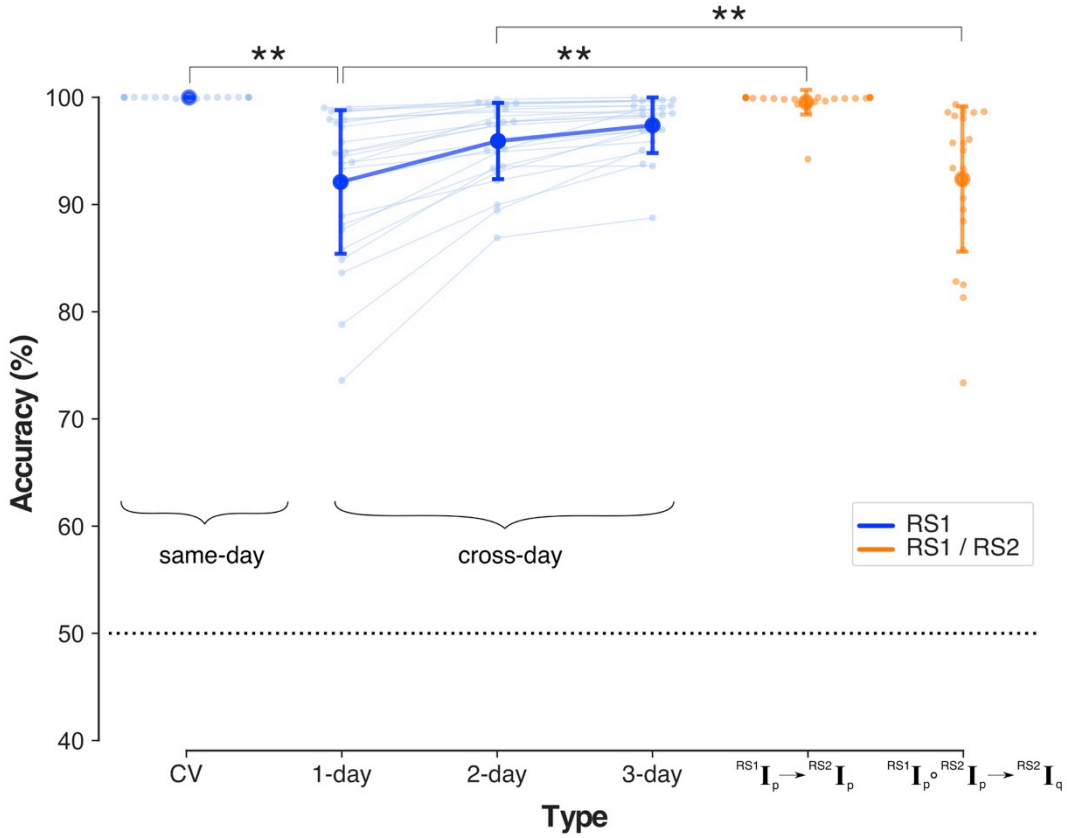


Figure 5: Identification accuracy at rest. Mean identification accuracy with *RS1* (blue) on same-day (CV), across days (1-day, 2-day, 3-day), and schemes relating *RS1* and *RS2* (orange). Light colored dots/lines depict individual accuracies (*N*=24). Horizontal dotted line depicts random chance accuracy (50%). Error bars: SD (** = $0.00001 \leq p < 0.001$).

Table 1: Identification accuracies in different experimental states reported as Mean % (SD) All values were significantly above random chance (50%) (see Suppl. Table 1**).**

States	Type			
	CV	**1-day**	**2-day**	**3-day**
RS1 (*N*=24)	99.98 (0.04)	92.10 (6.84)	95.92 (3.63)	97.39 (2.65)
RS1 (*N*=18)	99.98 (0.06)	92.79 (6.76)	96.61 (3.29)	97.53 (2.51)
RS2 (*N*=24)	99.99 (0.04)	91.59 (7.49)	95.86 (4.18)	96.99 (3.50)
TapWait (*N*=18)	99.99 (0.02)	92.58 (6.38)	96.36 (3.42)	97.61 (2.77)
SeqWait (*N*=18)	99.99 (0.02)	93.67 (7.35)	97.12 (4.36)	98.03 (3.80)
TapMov (*N*=18)	99.94 (0.12)	92.39 (6.72)	96.12 (3.34)	97.29 (2.41)
SeqMov (*N*=18)	99.97 (0.10)	93.47 (8.41)	96.67 (4.64)	97.94 (2.99)

The effects of aggregation with *RS1* were replicated with samples from *RS2*, which was the second RS measurement on each day (Table 1) [two-way ANOVA, Condition {*RS1, RS2*} x Type {1-day, 2-day, 3-day}, Type*Condition: $F_{2, 46} = 0.56$, $p = 0.57$; Type: $F_{2, 46} = 31.29$, $p < 0.00001$; Condition: $F_{1, 23} = 0.38$, $p = 0.54$]. *RS2* also confirmed the role of *day-specific* properties of the decision-rules (Figure 5). Same-day decision-rules from *RS1* classified samples of *RS2* from the same day (${}^{RS1}\mathbf{I}_p \rightarrow {}^{RS2}\mathbf{I}_p$) with a mean accuracy of 99.54 ± 1.17% that was significantly greater than the accuracy in classifying *RS1* across days (${}^{RS1}\mathbf{I}_p \rightarrow {}^{RS1}\mathbf{I}_q$) (92.10 ± 6.84%) [paired $t_{23} = 5.19$, $p = 0.00003$]. Furthermore, *RS2* validated the relevance of aggregating samples from different days (rather than different measurements) to reduce day-specificity. Decision-rules trained on aggregated

same-day samples from *RS1* and *RS2* (${}^{RS1}\mathbf{I}_p \circ {}^{RS2}\mathbf{I}_p \rightarrow {}^{RS1}\mathbf{I}_r$) had a *lower* cross-day accuracy (92.38 ± 6.92%) than decision-rules trained on aggregated *RS1* samples from two days (${}^{RS1}\mathbf{I}_p \circ {}^{RS1}\mathbf{I}_q \rightarrow {}^{RS1}\mathbf{I}_r$) (95.92 ± 3.63 %) [paired t_{23}=-4.83, p = 0.00007].

In summary, the cross-day loss in accuracy without aggregation was indicative of inter-day variations in RS activity. However, the high cross-day accuracy (with aggregation) was evidence for an individual-specific neural pattern that was stable across days <u>despite</u> inter-day variations in RS activity. A possible (but unverified) source of this inter-day variation was the person's incidental cognitive state on each day. Before testing the stability of the hypothesized neural signature to experimentally verified perturbation in cognitive state, we assessed the information sources that contribute to this stable neural pattern.

Low cross-day identification with information from only one frequency or one location

Each sample was a snapshot of RS activity described by 305 informational features (5 bands x 61 channels). To assess the informational role of these different features, we evaluated whether identification comparable to the full feature-set was possible with subsets of features that were defined either by frequency band (i.e., mono-band sets) or spatial location (i.e., mono-location sets).

Each *mono-band* feature set (B_f) consisted of features from one frequency band *f* at all 61 channels. For all five mono-band sets (Figure 6A, Table 2), same-day identification had a mean accuracy greater than 95%. However, the size of the cross-day loss in accuracy was band-dependent and ranged from ~14% for B_α to nearly ~32% for B_δ [ANOVA, Type {CV, 1-day} x Band {B_δ, B_θ, B_α, $B_{\beta1}$, $B_{\beta2}$}, Type*Band: $F_{4,92}$ = 24.77, p < 0.00001; Type: $F_{1,23}$ = 231.51, p < 0.00001; Band: $F_{4,92}$ = 40.40, p < 0.00001]. The divergence in cross-day losses for B_α and B_δ was striking as these two bands had a characteristically higher power relative to the other bands (Figure 4). Training with multi-day aggregation (Figure 6B) increased cross-day accuracy by differing amounts for each band by, for example, +10% for $B_{\beta2}$ but only +6% for B_δ [ANOVA, Band {B_δ, B_θ, B_α, $B_{\beta1}$, $B_{\beta2}$}x Type {*1-day, 2-day, 3-day*}, Type*Band: $F_{8,184}$ = 9.19, p < 0.00001; Type: $F_{2,46}$ = 146.28, p < 0.00001;

Band: $F_{4, 92} = 43.20$, $p < 0.00001$]. However, even with 3-day aggregation, the residual difference between cross-day and same-day accuracy (minimum: ~7% for B_α, maximum: ~26% for B_δ) was larger than the ~2% difference with the full feature-set.

Table 2: Identification accuracies for *RS1* with mono-band and mono-location feature subsets reported as Mean % (SD). All values were significantly above random chance (50%) (see Suppl. Table 2).

Subset	**Type**			
(*N*=24)	**CV**	**1-day**	**2-day**	**3-day**
$\boldsymbol{B_\delta}$	96.09 (2.55)	64.67 (7.92)	67.88 (8.11)	70.11 (8.02)
$\boldsymbol{B_\theta}$	97.57 (1.53)	76.99 (7.69)	81.76 (7.11)	83.70 (6.94)
$\boldsymbol{B_\alpha}$	98.51 (1.18)	84.19 (7.74)	88.38 (6.33)	89.59 (5.67)
$\boldsymbol{B_{\beta 1}}$	99.64 (0.59)	81.40 (10.44)	87.04 (9.03)	88.92 (8.14)
$\boldsymbol{B_{\beta 2}}$	99.72 (0.35)	76.37 (9.98)	83.21 (8.99)	86.67 (7.95)
$\boldsymbol{L_F}$	98.04 (1.79)	82.68 (8.90)	87.29 (7.46)	88.88 (6.57)
$\boldsymbol{L_{FC}}$	98.57 (1.54)	86.92 (9.38)	90.39 (7.45)	91.77 (6.12)
$\boldsymbol{L_{CP}}$	98.00 (1.78)	85.28 (8.58)	89.44 (7.30)	90.37 (6.53)
$\boldsymbol{L_{PO}}$	97.92 (2.02)	81.02 (8.32)	86.47 (7.35)	87.97 (7.06)

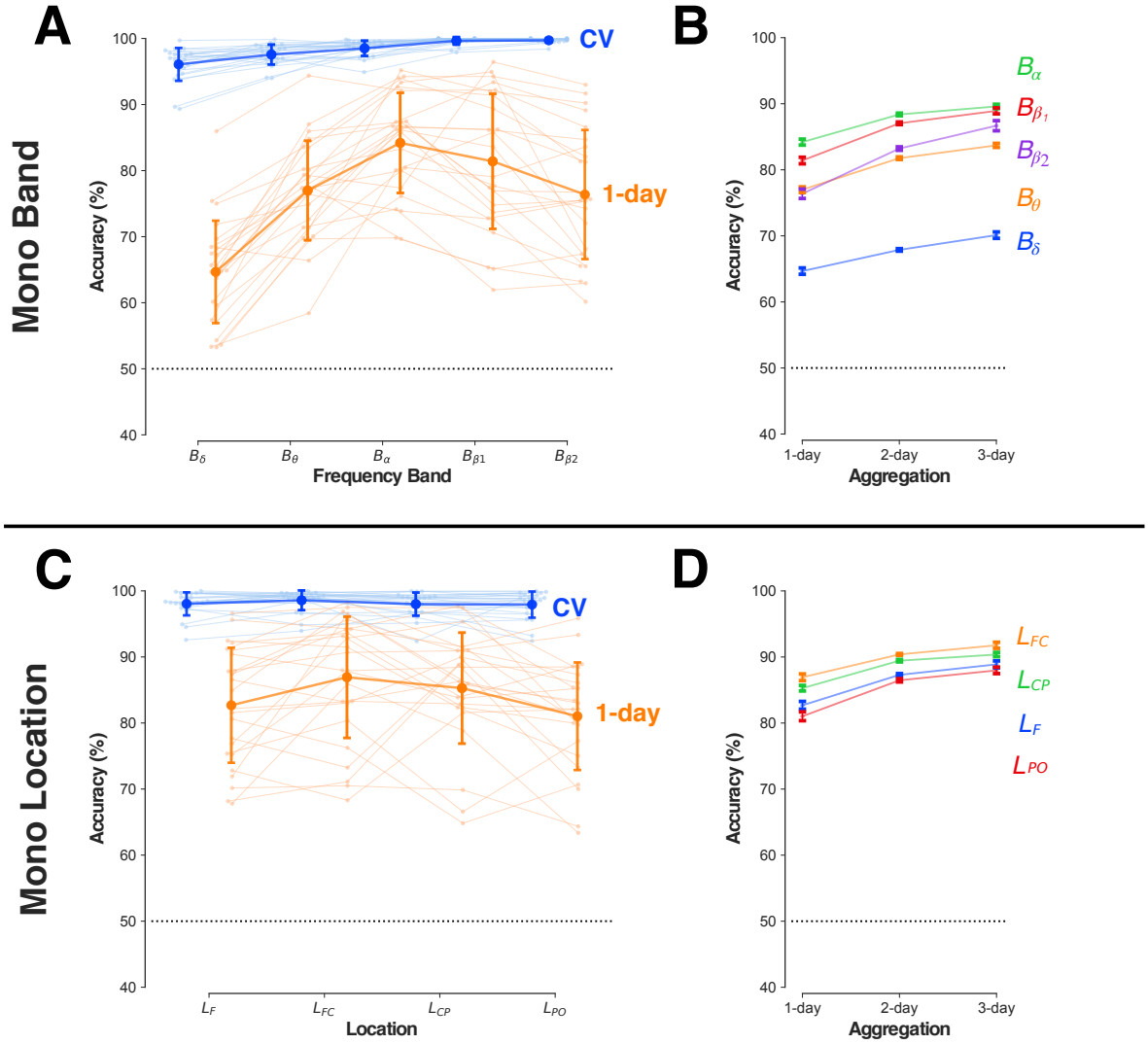


Figure 6: Identification at rest with mono-band/location feature subsets. (A) Mean identification accuracy for *RS1* with mono-band feature sets of increasing frequency (x-axis) on the same-day (blue, CV) and across-days (orange, 1-day). Light-colored dots/lines depict individual accuracies (*N*=24*)*. Error bars: SD. **(B)** Change in cross-day identification with increasing aggregation (x-axis) for different mono-band feature subsets (colored lines). Error bars: Within-subject s.e.m. **(C)** Mean identification accuracy for mono-location feature sets (x-axis, from anterior to posterior) with graphical representation and error bars as in panel **A**. **(D)** Change in cross-day identification with increasing aggregation (x-axis) for different mono-location feature subsets (colored lines). Error bars: Within-subject s.e.m. Horizontal dotted line depicts the random chance accuracy (50%) in all panels.

Each *mono-location* feature set (L_A) consisted of 50 features (5 bands x 10 channels) at location *A* (Figure 2A). The mean same-day accuracy was greater than 95% for all mono-location feature sets (Figure 6C, Table 2). However, the mean cross-day (1-day) accuracy showed reductions of ~12%-16% for all locations [ANOVA, Type {CV, 1-day} x Location {L_F, L_{FC}, L_{CP}, L_{PO}}, Type*Location: $F_{3,69} = 3.65$, $p = 0.016$; Type: $F_{1,23} = 109.36$, $p < 0.00001$; Location: $F_{3,69} = 5.56$, $p = 0.0018$]. The mean cross-day accuracy for the fronto-central (L_{FC}) and centro-parietal (L_{CP}) sets were marginally higher than for the parieto-occipital (L_{PO}) and frontal (L_F) sets. This zonal accuracy difference was notable as the mean power for all bands was typically higher over the posterior and anterior channels than the centrally located channels (Figure 4C). Aggregation increased cross-day accuracy by ~6% for all four location sets (Figure 6D) [ANOVA, Location: {L_F, L_{FC}, L_{CP}, L_{PO}} x Type {*1-day, 2-day, 3-day*} [Type*Location: $F_{6,138} = 2.07$, $p = 0.06$; Type: $F_{2,46} = 115.45$, $p < 0.00001$; Location: $F_{3,69} = 4.78$, $p = 0.0043$]. Nevertheless, the residual ~7%-10% cross-day loss in accuracy was larger than with the full feature-set.

In summary, all the mono-band and mono-location sets enabled same-day identification with nearly error-free accuracy. However, this information content did not confer robustness to the inter-day variation in RS activity. Even with aggregation, these feature sets had a much lower cross-day accuracy than the full feature-set that combined these feature sets together. This divergence suggests that the high cross-day robustness with the full feature-set might involve a crucial role for relationships between different frequency bands (i.e., unlike the mono-band subsets) at spatially distributed channels (i.e., unlike the mono-location subsets). To assess how multi-feature relationships might confer cross-day robustness, we evaluated the pattern of weights associated with the different features.

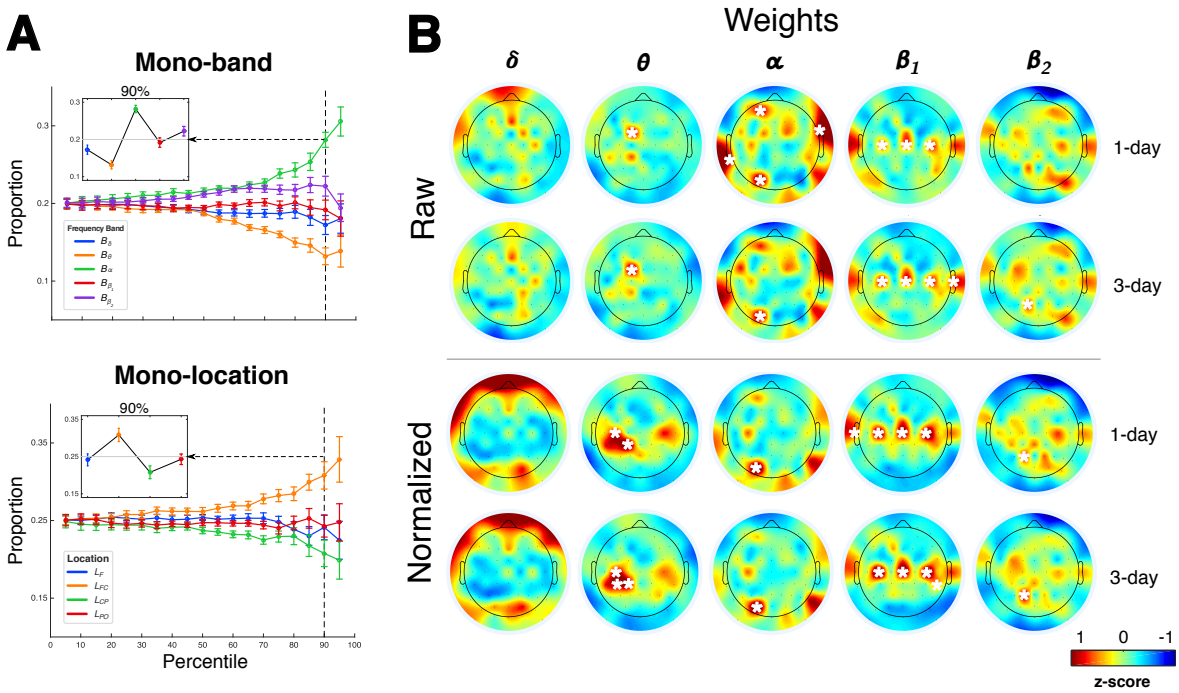


Figure 7: Decision-rule consistency at rest. (**A**) Mean proportions of high-valued (absolute) weights in the mono-band subsets (upper panel) and mono-location subsets (lower panel) at increasing percentiles (x-axis). Proportions (y-axis) are shown as deviations from an equal distribution across subsets (1/5 = 0.2, upper panel; and 1/4 = 0.25; lower panel). Insets show weight distribution in the 90th percentile (vertical dotted line). Error bars: Within-subject s.e.m. **(B)** Spatial distribution of high-valued weights (absolute) per frequency band. Upper two rows show the mean raw weights (z-scored across all features) and their changes with aggregation (1-day, 3-day). Lower two rows show the corresponding distributions for the mean normalized weights (z-scored per band due to the large power differences between bands). Weights in each scalp map that were significantly greater than zero are indicated with a white asterisk ($p < 0.05/61$, see Suppl. Figure 1).

Concentration of high-relevance features in B_α and L_{FC}

Each individual's linear decision-rule in the ensemble classifier was defined by the configuration of weights assigned to the different features, where weights with a larger magnitude (irrespective of sign) have a larger role in the classification decision even if in an indirect manner (Haufe et al. 2014). With an ensemble classifier, for a sample to be uniquely associated with a single person with high accuracy, the decision-rules in the ensemble need to have non-identical weight configurations. Therefore, a key consideration was whether the individual weight configurations differ from each other in an idiosyncratic manner with little consistency between individuals.

As a first test of consistency, we assessed whether differences in cross-day accuracies between the mono-band and mono-location subsets were an indicator of their relevance in the full feature set. For example, if subset S_x in isolation had a higher cross-day accuracy than S_y then S_x might have a larger concentration of high-weighted features than S_y as part of the full feature-set. To test this simplistic prediction, we evaluated the relative concentration of features with large weights (specified by percentile) in the different mono-band (Figure 7A, upper panel) and mono-location subsets (Figure 7A, lower panel).

The relative proportion of high-weighted features in the different mono-band subsets diverged at higher percentiles [ANOVA, Band {B_δ, B_θ, B_α, $B_{\beta1}$, $B_{\beta2}$} x Percentile {5%, 10%…95%}, Band*Perc: $F_{72, 1656} = 5.08$, $p < 0.00001$; Band: $F_{4, 92} = 7.26$, $p = 0.00003$; Perc: $F_{18, 414} < 0.001$, $p = 1$]. B_α contained the largest proportion of high-valued weights in the 90th percentile (inset Figure 7A, upper panel). This ordering was qualitatively similar to the cross-day accuracies for the mono-band feature sets (Figure 6A), where B_α had the highest mean cross-day accuracy. However, the proportion of high weights in B_δ was comparable to the other subsets despite having a lower cross-day accuracy than the other subsets. The relative proportion of high-weighted features in the mono-location subsets also diverged at higher percentiles [ANOVA, Location: {L_F, L_{FC}, L_{CP}, L_{PO}} x Percentile {5%, 10%…95%}, Location*Perc: $F_{54, 1242} = 2.00$, $p = 0.00003$; Location: $F_{3, 69} =$

3.71, p = 0.015; Type: $F_{18, 414} < 0.001$, p = 1]. The fronto-central L_{FC} subset contained a higher proportion of high-valued weights in the 90th percentile than L_{CP} that was immediately posterior to L_{FC} (inset Figure 7A, lower panel). By contrast, the mean cross-day accuracies for L_{FC} and L_{CP} in isolation were comparably similar (Figure 6C).

The concentration of high-valued weights in the different feature subsets was not a simple indicator of how these feature sets might contribute to high cross-day accuracy. However, it revealed consistencies in the distribution of relevant features across individuals, notably in B_α and L_{FC}. We therefore evaluated whether there were any specific features in each of the feature subsets that were consistently assigned a high relevance across individuals.

Distinct spatial concentrations of band-specific features

Figure 7B shows the spatial distribution of high-consistency features i.e., with mean z-scored weight significantly greater than zero (also see Suppl. Figure 1). Since the classification decision is based on summing the weighted power at each of the features, the influence of a feature on the decision is linked to the power at that feature. Therefore, we evaluated the high-consistency "raw" weights (Figure 7B, upper panels) as well as weights after normalization for power magnitude (Figure 7B, lower panels).

The distribution of weights was not idiosyncratic as might have been expected for predictors of inter-individual differences. Instead there were concentrations of high-consistency weights at distinct zones in each frequency band set. For both raw and normalized weights, the weights obtained with 3-day aggregation were similar to 1-day (without aggregation) with small changes in the relative weighting of the features.

With the exception of B_δ, all the other sets (B_θ, B_α, $B_{\beta 1}$, $B_{\beta 2}$) contained at least one high-consistency feature. For B_θ, the raw weights revealed a high-consistency feature at FC1 (and at CP1 at uncorrected thresholds (Suppl. Figure 1)). When normalized with power, there was a concentration at a slightly posterior zone at CP1 and C3, with the addition of CP3 with aggregation.

There was a corresponding but weaker concentration observable over the right hemisphere at uncorrected thresholds. Unlike the central concentration of features in B_θ, the consistent features in B_α were broadly distributed across the scalp at channels P03, TP7, AF3, FT8. However, with aggregation, the raw and normalized weights revealed a single, strongly consistent feature at PO3.

The high-consistency features in $B_{\beta 1}$ showed a striking bilaterally symmetric configuration along the transverse midline at channels C3, Cz, C4 with the aggregation-dependent roles for T7, T8 and CP6. Unlike $B_{\beta 1}$, the high-frequency range of the β band ($B_{\beta 2}$) only had a single consistent feature at P1 with a more diffuse configuration of weights at uncorrected thresholds.

In summary, the distribution of the high-consistency features over the sensorimotor cortex and the occipital cortex was prima facie support for the neurophysiological basis of the signals being used to distinguish between individuals. To assess the relevance of these distributed high-consistency zones to cross-day accuracy and the specificity of these zones to resting activity, we next turned to the relationship of *RS1* to the non-rest task states (evaluated in a subgroup of N=18 subjects (see section: Material and methods)).

Neural activity during *Tap* and *Sequence* verifiably deviate from *RS1*

A potential, but unverifiable, source of inter-day variation in *RS1* was the person's cognitive state during this period on each day. Therefore, before testing the effect of cognitive state variation on RS-based identification, we sought to explicitly verify the relationship of *RS1* to the non-rest cognitive states during the *Tap* and *Sequence* tasks. The behavioral demands during *TapMov* and *SeqMov* were designed to modulate the cognitive states during the *TapWait* and *SeqWait* periods and produce neural activity deviations from *RS1* in the absence of behavioral differences. Furthermore, the *Tap* and *Sequence* tasks were designed to elicit neural states that varied between days for *Sequence* (low cross-day similarity) but remained constant for *Tap* (high cross-day similarity).

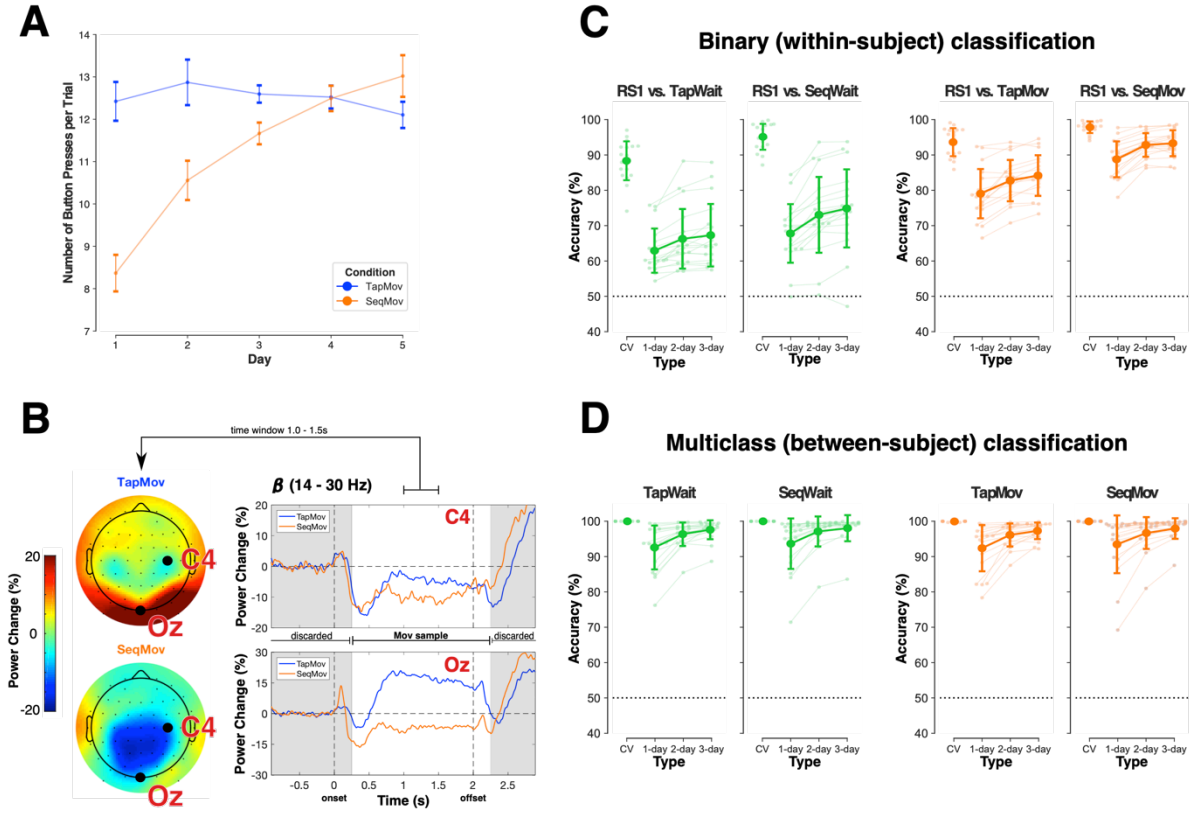


Figure 8: Difference between *RS1* and task-states. (A) Change in behavior indexed by mean number of button-presses across days (x-axis) during *TapMov* (blue) and *SeqMov* (orange). Error bars: Within-subject s.e.m. **(B)** Movement-related power dynamics in the β band (14-30 Hz) in *TapMov* (blue) and *SeqMov* (orange) at channels C4 (upper right) and Oz (lower right) averaged across participants and days. Intervals marked in gray were discarded from the *TapWait* and *SeqWait* samples used for classification to avoid movement-related carry over effects into the waiting periods. Scalp plots (left panel) show the mean power distribution over the period [+1s, +1.5s] following onset of the movement cue. **(C)** Same-day/cross-day accuracy in distinguishing *RS1* vs pseudo-rest states (green) and *RS1* vs movement states (orange) using within-subject binary classifiers. Cross-day differences to *RS1* were lowest for *TapWait* (far left) and highest for *SeqMov* (far right). Error bars: SD. (**D**) Person identification accuracy (multiclass) when the training/test sets were from the same task state (green: pseudo-rest states, orange: movement states). Error bars: SD.

The inter-day changes in behavior during the *TapMov* and *SeqMov* periods were consistent with the experimental assumptions (Figure 8A). During *TapMov*, the mean number of button presses per trial (~12-13) remained effectively constant across days [one-way ANOVA, $F_{4, 68}$ = 0.55, p = 0.70]. In contrast, during *SeqMov*, the mean number of button-presses increased from ~8 on the first day to ~13 on the fifth day [one-way ANOVA, $F_{4, 68}$ = 21.35, p < 0.00001]. This inter-day change in motor performance in *SeqMov* was systematically different from *TapMov* as confirmed by the statistically significant interaction in an ANOVA with factors Condition {*TapMov*, *SeqMov*} x Days {D1,..,D5} [Condition*Days: $F_{4, 68}$ = 12.38, p < 0.00001; Condition: $F_{1, 17}$ = 3.13, p = 0.09472; Days: $F_{4, 68}$ = 10.71, p < 0.00001].

The neural state during the movement-period (*TapMov, SeqMov)* showed typically expected dynamic states (Figure 8B). Changes in the mean β power at channel C4 (contralateral to the moved fingers) were in line with the Event-Related De-synchronization/Synchronization (ERD/ERS) phenomenon for repetitive movements (Pfurtscheller and Lopes da Silva 1999; Cassim et al. 2000; Alegre et al. 2004; Erbil and Ungan 2007), namely, a power reduction at the onset of movement execution (i.e., ERD) with an increase after the termination of all movements (i.e., ERS). Furthermore, the β power changes at Oz showed a task-dependent neural response consistent with differing visual stimulation, that is, an increase for *TapMov* (blank screen) but a decrease for *SeqMov* (image depicting the sequence). These movement-vs-wait differences were validated in the samples used for classification. A within-subject binary classification of *TapWait* vs *TapMov* had a mean cross-validated accuracy of 85.91 ± 7.23% [> 50%: t_{17} = 21.06, p < 0.00001]*;* and *SeqWait* vs *SeqMov* had a mean CV accuracy of 94.58 ± 3.20% [> 50%: t_{17} = 59.02, p < 0.00001]

The critical verification for our study was the relationship between *RS1* and the pseudo-rest states (*TapWait*, *SeqWait*). Samples from *TapWait* and *SeqWait* were distinguishable from *RS1* on the same day with high cross-validated accuracy (*RS1* vs *TapWait*: 88.35 ± 5.66%; *RS1* vs *SeqWait*: 95.13 ± 3.77%) (Figure 8C left panels, Table 3). However, the cross-day accuracy (without aggregation) for both *RS1* vs *TapWait* (62.91 ± 6.44%) and *RS1* vs *SeqWait* (67.79 ± 8.53%) was

substantially lower than the same-day accuracy by more than ~25%. Nevertheless, the cross-day accuracy for *RS1* vs *SeqWait* was marginally higher than for *RS1* vs *TapWait* with increasing aggregation [ANOVA: Condition {*RS1* vs *TapWait*, *RS1* vs *SeqWait*} x Type {1-day, 2-day, 3-day}, Condition*Type: $F_{2, 34} = 6.21$, $p = 0.00585$; Condition: $F_{1, 17} = 8.37$, $p = 0.01009$; Type: $F_{2, 34} = 38.88$, $p < 0.00001$].

TapMov and *SeqMov* were also distinguishable from *RS1* on the same-day with high (cross-validated) accuracy (*RS1* vs *TapMov:* 93.62 ± 4.09%; *RS1* vs *SeqMov*: 97.84 ± 1.70%) (Figure 8C, right panel, Table 3). Similar to the wait periods, the cross-day accuracy for *RS1* vs *SeqMov* was higher than for *RS1* vs *TapMov* across aggregation levels [ANOVA: Condition {RS vs *TapMov*, RS1 vs *SeqMov*} x Type {1-day, 2-day, 3-day}, Condition*Type: $F_{2, 34} = 0.61$, $p = 0.55$; Condition: $F_{1, 17} = 30.91$, $p = 0.00003$; Type: $F_{2, 34} = 69.47$, $p < 0.00001$].

Unlike the unverified inter-day cognitive state differences during rest, the above findings verified the neural activity differences in the task-states in *Tap* and *Sequence* to each other and to *RS1*. We next turned to the key issue of cross-task person identification.

Table 3: Classification accuracy of *RS1* vs task state (binary, within-subject) reported as Mean % (SD). All values were significantly above random chance (50%) (see Suppl. Table 3).

***RS1* vs** (*N*=18)	**Type**			
	CV	**1-day**	**2-day**	**3-day**
TapWait	88.35 (5.66)	62.91 (6.44)	66.26 (8.69)	67.28 (9.11)
SeqWait	95.13 (3.77)	67.79 (8.53)	73.05 (11.01)	74.86 (11.37)
TapMov	93.62 (4.09)	79.05 (7.17)	82.75 (5.99)	84.18 (5.92)
SeqMov	97.84 (1.70)	88.77 (5.21)	92.81 (3.43)	93.32 (3.77)

Robust identification of individuals from *Tap* and *Sequence* activity within and across days

A key measure of day-specificity was the loss in accuracy between cross-day and same-day identification. By analogy, we evaluated the relationship of *RS1* to the task-states by the difference in accuracy between same-task (i.e., with the scheme ${}^{X}\mathbf{I}_p \rightarrow {}^{X}\mathbf{I}_q$) and cross-task identification (i.e., ${}^{X}\mathbf{I}_p \rightarrow {}^{Y}\mathbf{I}_q$). Before evaluating cross-task identification with *RS1*, we evaluated whether the different *Tap* and *Sequence* task-states contained sufficient information for person identification in a same-task classification scheme (Figure 8D).

The same-day accuracy for both *TapWait* and *SeqWait* was ~99% (Figure 8D left panels, Table 1). The mean cross-day accuracy (without aggregation) for *TapWait* (92.58 ± 6.38%) was lower than its corresponding same-day accuracy by only ~7% [$t_{17} = 4.92$, $p = 0.00013$]. Similarly, for *SeqWait*, the mean cross-day (1-day) (93.67 ± 7.35%) accuracy was lower than the same-day accuracy by ~6% [$t_{17} = 3.65$, $p = 0.00197$]. Furthermore, the effect of aggregation on mean cross-day accuracy for *TapWait* and for *SeqWait* were statistically indistinguishable [ANOVA: Condition {*TapWait*, *SeqWait*} x Type {1-day, 2-day, 3-day} [Condition*Type: $F_{2,\ 34} = 0.88$, $p = 0.42$; Condition: $F_{1,17} = 1.35$, $p = 0.26$; Type: $F_{2,\ 34} = 21.30$, $p < 0.00001$].

Despite the deviations of *TapMov* and *SeqMov* along both the behavioral and cognitive dimensions of rest and their differences with each other, the accuracies of individual identification across days for *TapMov* and *SeqMov* were greater than 90% for all levels of aggregation and were not statistically distinguishable from each other (Table 1, Figure 8D right panels) [ANOVA: Condition {*TapMov*, *SeqMov*} x Type {1-day, 2-day, 3-day} [Condition*Type: $F_{2,\ 34} = 0.86$, $p = 0.43$; Condition: $F_{1,\ 17} = 1.26$, $p = 0.27$; Type: $F_{2,\ 34} = 14.50$, $p = 0.00003$].

Remarkably, despite the extensive differences between *TapWait. SeqWait*, *TapMov* and *SeqMov*, the consistency in weight patterns and weighted power for these states were qualitatively similar to each other (Figure 9, Suppl. Figure 2). Importantly, these weight patterns also showed similarities with the corresponding patterns for *RS1* (Figure 7) thus suggesting the a priori plausibility of cross-task generalization.

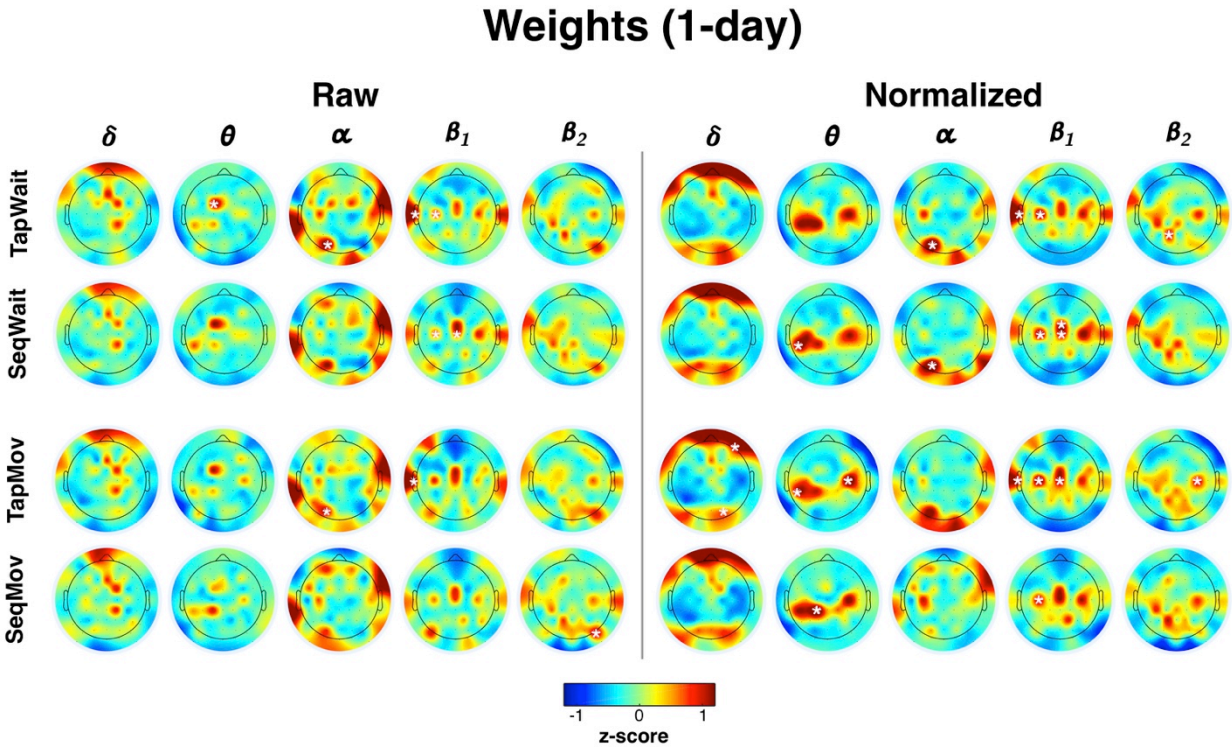


Figure 9: Spatial distribution of weights in different task-states. Mean high-valued weights (absolute, z-scored) for the different frequency bands (raw: left panel, normalized: right panel) without aggregation. Weights in each scalp map that were significantly greater than zero are indicated with a white asterisk ($p < 0.05/61$, see Suppl. Figure 2). For both raw and normalized weights, each frequency band (column) had a characteristic spatial distribution of high weighted channels that was qualitatively similar across task-states and also to *RS1* (Figure 7B).

Robust cross-task identification with RS1 with full feature-set

If person identification with *RS1* was based on a neural signature related to an individual's neurophysiological state then identification should be possible despite cognitive state variations. Therefore, decision-rules trained on *RS1* should be capable of accurate person identification with samples acquired from the pseudo-rest states (*TapWait* and *SeqWait*) and the movement states (*TapMov* and *SeqMov*).

We used the cross-task scheme ${}^{RS1}\mathbf{I}_p \rightarrow {}^{X}\mathbf{I}_q$ to test the invariance of *RS1*-based identification to inter-day cognitive state variations (i.e., task states *X*) (Figure 10A, Table 4**)**. Increasing deviations from *RS1* solely due to cognitive state differences (*X = {RS1, TapWait, SeqWait}*) did not produce comparable, statistically distinguishable reductions in mean identification accuracy (*RS1*: 92.79 ± 6.76%, *TapWait*: 91.90 ± 6.46%; *SeqWait*: 90.81 ± 7.09%) [one-way ANOVA, $F_{2, 34} = 2.06$, $p = 0.14$]. However, increasing deviations from *RS1* due to cognitive <u>and</u> behavioral state differences (*X = {RS1, TapMov, SeqMov}*) produced significant reductions in identification accuracy most notably for *SeqMov* (*TapMov*: 88.79 ± 7.57%; *SeqMov*: 83.85 ± 10.35%)[one-way ANOVA, $F_{2, 34} = 14.07$, $p = 0.00003$].

To disentangle the role of cross-task from cross-day effects, we compared cross-task (${}^{RS1}\mathbf{I}_p \rightarrow {}^{X}\mathbf{I}_q$) and same-task identification (${}^{X}\mathbf{I}_p \rightarrow {}^{X}\mathbf{I}_q$) across days (Table 1,4**)**. For the pseudo-rest states (*X = {TapWait, SeqWait}*), cross-task accuracy with *RS1* decision-rules produced a small but statistically significant reduction relative to same-task identification [ANOVA, Train {*RS1*, *Same*} x Condition {*TapWait*, *SeqWait*}, Train*Condition: $F_{1,17} = 4.14$, $p = 0.06$; Train: $F_{1,17} = 10.01$, $p = 0.00566$; Condition: $F_{1, 17} = 0.00001$, $p = 0.99$]. The cross-task accuracy reduction was significantly larger for the movement-states (*X = {TapMov, SeqMov}*) with a larger loss for *SeqMov* [ANOVA, Train{*RS,Same*} x Condition {*TapMov*, *SeqMov*}, Train*Condition: $F_{1,17} = 9.15$, $p = 0.00764$; Train: $F_{1,17} = 43.94$, $p < 0.00001$; Condition: $F_{1, 17} = 2.51$, $p = 0.13$].

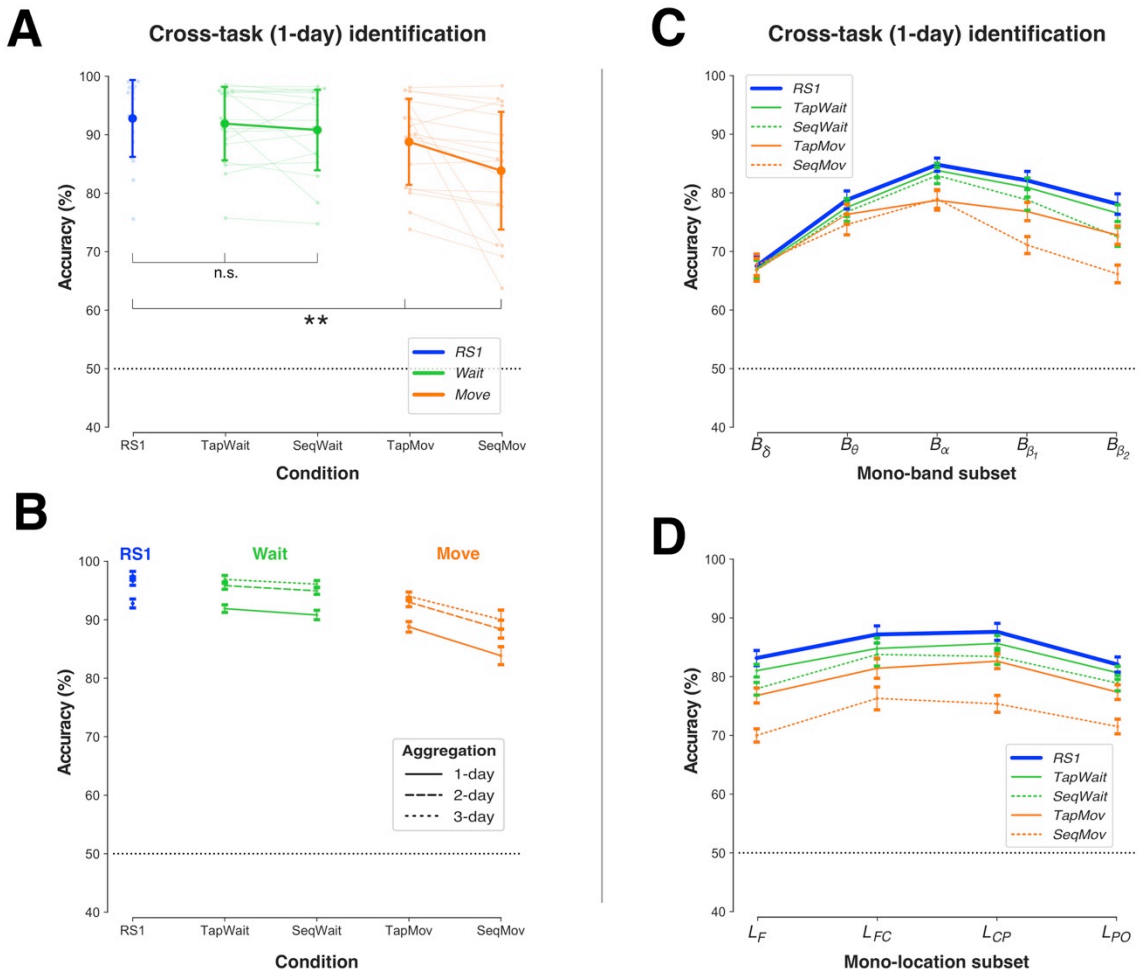


Figure 10: Cross-task identification with RS1. (A) Mean accuracy of decision-rules trained on *RS1* (1-day) and tested across days on *RS1* (blue), pseudo-rest states (green) and movement states (orange). Light colored dots/lines indicate individual accuracies. Error bars: SD. Accuracy differences between *RS1* and pseudo-rest states were not statistically significant (n.s.), but were between *RS1* and movement states (** = $0.001 \leq p < 0.00001$). (**B**) Decision-rules trained on *RS1* with different levels with aggregation (dotted lines) increased cross-day accuracy for all task-states. Error bars: Within-subject s.e.m. **(C)** Cross-task/day accuracy of *RS1* with mono-band subsets. Deviations from cross-day accuracy for *RS1* were larger for the movement states (orange) than the pseudo-rest states (green) and deviations increased with the frequency (lowest for B_{δ}, highest for $B_{\beta 2}$). Error bars: Within-subject s.e.m. **(D)** Cross-task/day accuracy of *RS1* with mono-location subsets. Deviations from *RS1* were larger for movement states (orange) than pseudo-rest states (green). Error bars: Within-subject s.e.m.

Table 4: Accuracy of cross-task ${}^{RS1}\mathbf{I}_p \rightarrow {}^{X}\mathbf{I}_q$ identification reported as Mean % (SD). All values were significantly above random chance (50%) (see Suppl. Table 4).			
Test states	**Type**		
(*N = 18*)	**1-day**	**2-day**	**3-day**
TapWait	91.90 (6.46)	95.84 (3.30)	96.90 (2.44)
SeqWait	90.81 (7.09)	94.95 (4.39)	96.09 (3.40)
TapMov	88.79 (7.57)	93.02 (5.49)	94.01 (4.51)
SeqMov	83.85 (10.35)	88.39 (9.28)	90.03 (8.87)

To disentangle the role of day-specificity in ${}^{RS1}\mathbf{I}_p \rightarrow {}^{X}\mathbf{I}_q$, we used multi-day aggregation (${}^{RS1}\mathbf{I}_p \circ {}^{RS1}\mathbf{I}_q \ldots \rightarrow {}^{X}\mathbf{I}_r$). Although aggregation reduced day-specificity with *RS1* (Figure 5), it could nevertheless increase specificity to the properties of *RS1*. If so, aggregation might lower the accuracy of cross-task identification. Contrary to this possibility, aggregation *increased* cross-task accuracy to the pseudo-rest states (*TapWait*, *SeqWait*) in a comparable manner to same-task accuracy (Figure 10B) [ANOVA: Condition {*RS1, TapWait, SeqWait*} x Type {1-day, 2-day, 3-day} [Condition*Type: $F_{4, 68} = 0.52$, $p = 0.72$; Condition: $F_{2, 34} = 2.44$, $p = 0.10$; Type: $F_{2, 34} = 21.63$, $p < 0.00001$]. This was particularly striking because aggregation (i.e., related to day-specificity) produced a relatively larger increase in cross-task accuracy than a change in task-specificity. Following aggregation, the mean residual cross-task/day accuracy loss relative to same-task/day identification with *RS1* was only ~3%. Aggregation also *increased* cross-task accuracy to the movement states (*TapMov, SeqMov*) [ANOVA: Condition {*RS1, TapMov*, *SeqMov*} x Type {1-day, 2-day, 3-day} [Condition*Type: $F_{4, 68} = 1.35$, $p = 0.26$; Condition: $F_{2, 34} = 13.04$, $p = 0.00006$;

Type: $F_{2, 34} = 29.33$, $p < 0.00001$]. Following aggregation, the mean residual cross-task/day difference was less than ~10% for the movement states.

The full-feature set has a crucial role in limiting the cross-day loss in accuracy (Figure 6). Appling the cross-task scheme ${}^{RS1}\mathbf{I}_p \rightarrow {}^{X}\mathbf{I}_q$ to the mono-band (Figure 10C) and mono-location (Figure 10D) feature sets provided further evidence of the importance of the full feature-set to enable robust cross-task identification.

For the mono-band feature sets (Figure 10C), increasing deviations from *RS1* in cognitive state (*X = {RS1, TapWait, SeqWait}*) lead to state-related accuracy reductions that were also larger for the higher frequency bands [ANOVA: Condition {*RS1, TapWait, SeqWait*} x Band {B_δ, B_θ, B_α, $B_{\beta1}$, $B_{\beta2}$} [Condition*Band: $F_{8, 136} = 2.58$, $p = 0.01181$; Condition: $F_{2, 34} = 6.41$, $p = 0.00434$; Band: $F_{4, 68} = 18.37$, $p < 0.00001$]. In a similar manner, increasing deviations from *RS1* for the movement states (*X = {RS1, TapMov, SeqMov}*) produced state-related accuracy reductions that were greater for *SeqMov* than for *TapMov* particularly at the higher frequencies [ANOVA: Condition {*RS1, TapMov, SeqMov*} x Band {B_δ, B_θ, B_α, $B_{\beta1}$, $B_{\beta2}$} [Condition*Band: $F_{8, 136} = 8.97$, $p < 0.00001$; Condition: $F_{2, 34} = 20.60$, $p < 0.00001$; Band: $F_{4, 68} = 12.18$, $p < 0.00001$]. These task-linked accuracy reductions were notably absent at B_δ.

The pattern of cross-task accuracy deviation from *RS1* took a different form for the mono-location feature sets (Figure 10D). For the pseudo-rest states (*X = {RS1, TapWait, SeqWait}*), increasing deviations from *RS1* lead to increasing accuracy reductions (largest for *SeqWait*) that were relatively uniform at all the locations [ANOVA: Condition {*RS1, TapWait, SeqWait*} x Location {L_F, L_{FC}, L_{CP}, L_{PO}} [Condition*Location: $F_{6, 102} = 0.98$, $p = 0.44$; Condition: $F_{2, 34} = 9.00$, $p = 0.00073$; Location: $F_{3, 51} = 4.34$, $p = 0.0084$]. This pattern of reduction was similar for the movement states (*X = {RS1, TapMov, SeqMov}*), where deviations from *RS1* lead to accuracy reductions that were largest for *SeqMov* and relatively uniform at all locations [ANOVA: Condition {*RS1, TapMov, SeqMov*} x Location {L_F, L_{FC}, L_{CP}, L_{PO}} [Condition*Location: $F_{6, 102} = 0.99$, $p = 0.43$; Condition: $F_{2, 34} = 32.90$, $p < 0.00001$; Location: $F_{3, 51} = 4.95$, $p = 0.00433$].

In summary, decision-rules trained on *RS1* on a single day could identify individuals from samples from states that verifiably differed from *RS1* to differing extents. Identification was almost indistinguishable from *RS1* when the task-states did not involve explicit movement. Surprisingly, a significant portion of the cross-task/day accuracy reduction was accounted for by inter-day variations in *RS1*. The full feature-set was crucial to this high identification accuracy as suggested by the large accuracy reductions with band/location-defined feature subsets. The high accuracy of person identification in the presence of explicit cognitive variation was direct evidence of individual-specific neural patterns with high specificity to detect an individual's neurophysiological organization.

DISCUSSION

The main question was whether the confounding role of the ill-defined rest task when making inferences about longitudinal RS could be resolved without changing the rest task itself. To this end, we evaluated whether RS activity might have intrinsic properties that allow an analysis-based resolution of this confound. Individual identification across days of both *NP*+ and *NP*- was possible with high accuracy in the presence of simulated confounding inter-day differences. These properties of the longitudinal dynamics of EEG oscillatory power spectrum at rest provide an existence proof for distinctive measurable signatures of neurophysiological difference that are robust to fluctuations in cognitive state. Consistent with a signature based in whole-brain neurophysiology, identification was contingent on a full feature-set that allowed inter-band relationships distributed across the scalp with a reduced specificity with band- and location-specific features.

Empirical simulations of cognitive and neurophysiological variation

A methodological novelty here was our use of empirical "simulations". Although ad hoc, these simulations provided a valuable means to obtain verifiable *true* instances of cognitive state variation and neurophysiological change relative to RS.

The potential for arbitrary cognitive state variation during the rest task is related to experimental context. As previous studies have demonstrated, instructions can influence variability (Duncan and Northoff 2013; Kawagoe et al. 2018). However, beyond the assumption that participants were awake, we did not explicitly model the participant's cognitive state, for example, using participant's self-reported subjective assessments of their cognitive state during RS measurement (Diaz et al. 2013). Since the cognitive state and the extent of its fluctuation during rest are difficult to establish for each individual, it could have been the case that the high identification accuracy with *RS1* was more indicative of having a group of highly motivated and instruction-compliant participants rather than the neural characteristics in the resting state. Therefore, the *Tap* and *Sequence* tasks provided verifiable within-subject examples of deviations from rest to assess the generality of RS-based inferences.

In a longitudinal setting, the classification problem of interest requires a decision between *NP+* and *NP-* <u>within</u> the same individual. However, here *NP+* was defined based on samples of RS activity from *other* individuals. This use of inter-individual differences provided a pragmatic means to simulate a diverse range of possible changes to an individual's neurophysiology, with the assumption that detecting true within-subject neurophysiological change would possibly be far more challenging. For example, in the *Sequence* task, the motor learning across the five days in our experiment involved neuroplastic changes (Wymbs et al. 2012; Wymbs and Grafton 2014; Bassett et al. 2015) and the accompanying changes in *SeqWait* and *SeqMov* over the duration of the experiment (Figure 8) could be considered as consequence of this learning-induced neuroplasticity. However, due to the unclear carryover effects of these plastic changes on *RS1* over this five day period, we instead used the *SeqWait* and *SeqMov* to simulate incidental cognitive-state variations

(*NP*-) with high inter-day variance, where the neural dynamics on each day was a poor model of the dynamics on other days.

By using individual-differences in this manner, the problem of distinguishing between *NP*+ and *NP*- was equivalent to the problem of individual identification with similarities to numerous studies that have sought to use RS-EEG as an individual-specific signature for biometric identification (Campisi and Rocca 2014; Gui et al. 2015; Valizadeh et al. 2019). Although our focus was not on biometric identification, our findings are consistent with these prior studies in demonstrating the high distinctiveness of individual differences and its robust detectability even in two-second snapshots of oscillatory power spectra at rest. However, the inter-individual differences in cross-day identification with *RS1* both with and without aggregation (Figure 5,6A, 6C) also demonstrated that resting activity was not the equivalent of a fingerprint, in being immune to cognitive state or even whether a person is alive (Campisi and Rocca 2014). Even though individual identification was possible across tasks with high accuracy, the RS-based individual signatures were not completely independent of cognitive state. Large deviations from rest during *TapMov* and *SeqMov* did reduce cross-task identification accuracy even though identification was above random chance. These accuracy reductions were due to cognitive state differences with *RS1* and not merely because *TapMov* and *SeqMov* conditions lacked identifiable signatures or had more movement-related artifacts (Figure 8D).

Reliability of inference versus reliability of features

Numerous prior studies have investigated the inter-day similarity in RS activity within a test-retest framework (Bijsterbosch et al. 2017; Cox et al. 2018; Noble et al. 2019; Postema et al. 2019). In that framework, the focus is on evaluating whether a particular measure of RS activity on day *p* was reliably reproduced for RS activity on day *q* in the assumed <u>absence</u> of a true change. However, our focus is on the reliability of inferences in the assumed <u>presence</u> of true inter-day

activity changes. This focus required differing considerations about how an individual's unique identity was defined and represented as illustrated with an analogy to object recognition.

Consider images of the same object *X* from day 1 (test) and day 2 (retest) (Figure 11) represented by a list of filled pixel locations (i.e., features). With this representation, a simple measure of test-retest reliability is whether a pixel's filled state on day 1 is a reliable predictor of its state on day 2. The scenario in Figure 11A is consistent with a *high* feature-level reliability as the majority of filled pixels on day 1 are also filled on day 2. However, this high reliability is misleading about the object's unique identity. On day 1, object *X* can be readily distinguished from object *Y* based on a few critical pixels (circled). These critical pixels on object *X* are, however, unfilled on day 2. Thus, object *X* is not uniquely identifiable on day 2 as it is now confusable with object *Y*. Conversely, in the scenario shown in Figure 11B, a pixel-based test of reliability would indicate a *low* reliability due to the large number of filled pixels from day 1 that are unfilled on day 2. However, this low reliability is a limitation of how the object was represented (i.e., as a list of filled pixel locations relative to the main axes). If this representation included information about the relationships between the filled locations, then the object's defining characteristics would be deemed as being reliably conserved on day 2, e.g., a rotation of the object *X* on day 2 would bring it into correspondence with its form on day 1.

As demonstrated by this analogy, high test-rest reliability of individual features does not imply high individual identification. Furthermore, it illustrates the crucial relationship between reliability and assumptions about how identity is defined and represented. The demands for a flexible representation an individual's identity in a flexible manner were an important motivation to use a multivariate classification approach.

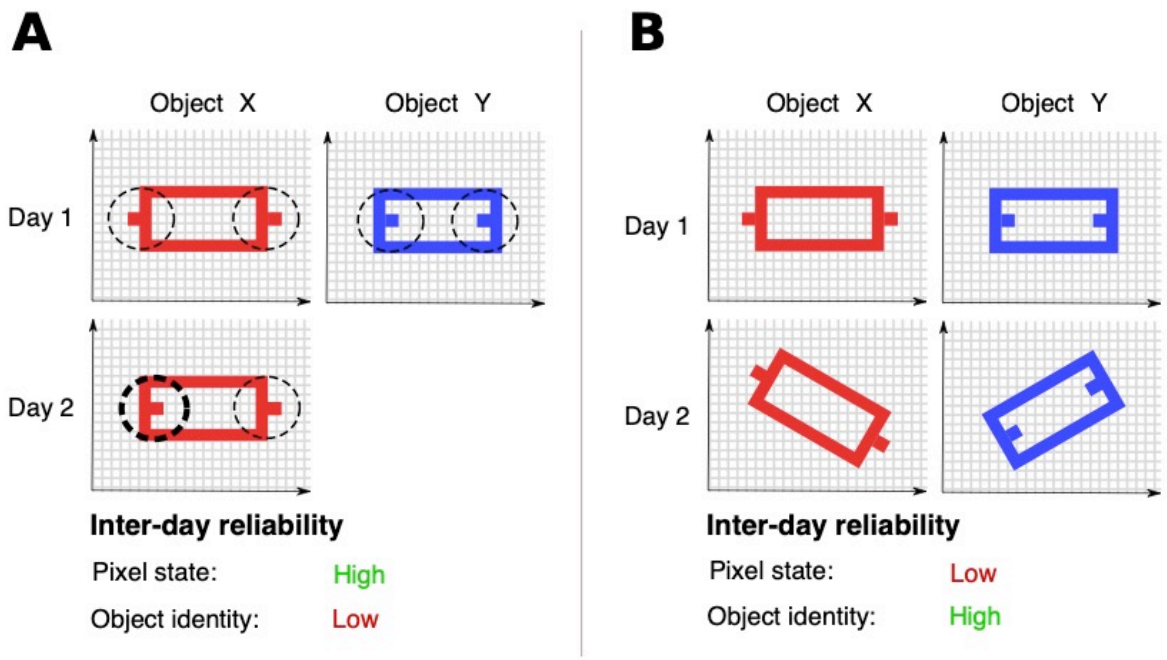


Figure 11: Test-retest reliability versus individual re-identification. (A) Objects *X* (red) and *Y* (blue) are uniquely defined by the configuration of filled and unfilled pixels. On day 1, (top row), the dotted circles indicates the critical pixels that distinguish *X* and *Y*. Most pixels of object *X* from day 1 are also filled on day 2 (lower row). However, pixels in the left dotted circle on day 2 differ in their state from day 1. Due to this difference on day 2, object *X* cannot be uniquely re-identified as being object *X* day based on its form as it is now confusable with object *Y*. High inter-day reliability in pixel state does not imply the same for object identity. (**B**) The orientation of objects *X* on day 2 is rotated relative to its orientations on day 1. If this orientation is accounted for, then object *X* can be uniquely re-identified on day 2. However, when considering individual pixels, most of the filled pixels on day 1 are not on day 2. Low inter-day reliability in pixel state does not imply the same for object identity.

With this approach, an individual's identity is defined relative to others. Therefore a tradeoff is that each individual's decision-rule could be defined differently with changes to the other individuals in the population. For a support vector machine, the decision boundary depends on the most similar samples in the distribution rather than overall characteristics of the distributions (Guyon and Elisseeff 2003). Therefore, the number of individuals is less of a concern than the characteristics of the most-similar individual. Finally, the approach here was to evaluate inferences rather than be a general test of differences. The multi-class approach involves finding features that differ between individuals to the exclusion of shared features. Therefore, an inter-day change in an individual on these shared features would not be detected.

What makes an individual signature robust to changes in cognitive state?

Our primary findings are based on black-box statistical inferences, namely, the pattern of classification accuracies obtained with different training/test sets that were selected based on experimental variables (primarily, the effect of day, conditions defining cognitive state and feature set). Therefore, an important issue is whether these statistical regularities are consistent with a neurophysiological signature in RS activity rather than a byproduct of other factors specific to our implementation.

The shape of the power spectrum in the frequency domain at rest has long been suggested as an important individual characteristic (Näpflin et al. 2007; Chiang et al. 2011; Bazanova and Vernon 2014). This shape has multiple peaks over a background of 1/f noise. The specific frequencies at which these peaks occur, particularly in the α band and in the β band have been the topic of considerable investigation (van Albada and Robinson 2013; Voytek et al. 2015). Importantly, in different cognitive states, the changes to this spectrum are not arbitrary and primarily involve changes to the power at the peaks (as well as small shifts in the peak frequency) but without considerable changes to the 1/f background (Buzsáki et al. 2012; Cole and Voytek 2017).

Therefore one possibility is that an individual's decision-rule represents the shape of their unique power spectrum. If this were the case then it would provide a plausible explanation for the observed high specificity despite cognitive state variation. In our feature representation, the power was averaged into five canonical bands rather than a more detailed frequency spectrum. Therefore, capturing the individual shape of the spectrum and approximate location of the α power peak would require a role for features representing the power in the θ, α and β_1 bands. Indeed, it was these three bands that also showed the main consistencies in term of a few, high valued weights. The classical depiction of the power spectrum is from a particular channel. Our finding suggests that representation of the individual-specific power in the different bands were distributed over the scalp with a concentration in the fronto-central and occipital zones. Although the power spectra are similar across channels, any one channel is an incomplete representation of that individual's characteristic power distribution. Consequently, it might lack the robustness to represent individual variability across days. By contrast, a decision-rule that combines each band's best representation might have a greater robustness.

With a multi-class categorization approach, the weights defining each subject's decision-rule have to be necessarily unequal. If not, a sample will be classified in the same way by different classifiers. Therefore, the presence of consistencies in the weight distribution and also in the normalized weights was noteworthy. One source of potential shared consistency was the use of a one-vs-all training procedure. Therefore, for any two persons in the group of *M* individuals, the "all" set would contain the same *M-2* individuals (for *P,* all will be *M-P,* and for *Q* it would be *M-Q*). However, another source of consistency is neuroanatomical organization, for example, the basic location and function of the cortical and sub-cortical networks.

Even though we do not use an explicit model of functional connectivity, the multivariate representation suggests a coupling between power values at these distributed locations. A distinction is often drawn between static and dynamic connectivity (Hutchison et al. 2013; Calhoun et al. 2014). Typically, this is based on how the neural time-series over the resting task is used.

Static connectivity refers to the extraction of a single measure (e.g., a graph) from the time-series. However, dynamic connectivity is based on the view that resting state refers to a collection of states that dynamically vary at different time points. In our study, the core variance model involved an assumption about time-scales. Each same-day measurement was segmented into two-second non-overlapping epochs where each epoch was treated as a sample drawn from an underlying individual-specific distribution. The dynamic variability between samples was assumed to be an important characteristic of this individual-specific distribution. Cross-day/cross-task identification was predicated on whether training classifiers based on the inter-sample variability on short time-scales (i.e., between the samples acquired within seconds/minutes of each other on the same day) was a viable model for samples obtained on long time-scales, i.e., days apart (Figure 1). Nevertheless, our approach and findings are agnostic as to whether the inter-sample differences indicate variability around a characteristic mean value (i.e., static connectivity) or characteristic transitions between distinct states (i.e., dynamic connectivity).

The implied multi-variable coupling is notable also with regard to the characteristic of general machine learning algorithms. Generalization accuracy is known to reduce with an increase in the number of features (the so called Hughes effect (Campenhout 1978; Sima and Dougherty 2008)). However, here, a feature set of 305 features showed greater cross-day generalization than small feature-sets of 50/60 features that have comparable same-day cross-validated accuracy.

Outlook

In the current study, we assume that individuals do not undergo extensive plastic changes. If individual identification was not possible even with a group of healthy individuals over a period of five days, then the value of this approach would be in question for longer periods of time and, possibly, with populations where such neuroplastic changes would be expected. Prior studies, although mainly cross-sectional, have found changes to the power spectrum with aging (van Albada et al. 2010; Chiang et al. 2011; Voytek et al. 2015; Knyazeva et al. 2018). With increasing age,

there is a reduction in the frequencies at which the power peaks in the alpha and beta bands occur. Voytek et al. (2015) suggest that this is because of a change in the 1/f baseline that indicates increased physiological noise with aging. Furthermore, systematic longitudinal changes in the power spectrum have been observed following stroke (Giaquinto et al. 1994; Saes et al. 2020). Thus, the application of this physiological signature to monitor longitudinal RS in clinical populations is an important future priority.

Acknowledgments

This work was supported by the University of Cologne Emerging Groups Initiative (CONNECT group) implemented into the Institutional Strategy of the University of Cologne and the German Excellence Initiative. SD gratefully acknowledges support from the German Research Foundation (DA 1953/5-2). We thank Hannah Kirsten, Alexandra Kurganova, and members of the INM-3 for their valuable assistance in data acquisition.

Conflict of interest

None

REFERENCES

Abraham A, Elrahman SMA. 2013. A Review of Class Imbalance Problem. J Netw Innov Comput. 1:332–340.

Alegre M, De Gurtubay IG, Labarga A, Iriarte J, Malanda A, Artieda J. 2004. Alpha and beta oscillatory activity during a sequence of two movements. Clin Neurophysiol. 115:124–130.

Barry RJ, Clarke AR, Johnstone SJ, Magee CA, Rushby JA. 2007. EEG differences between eyes-closed and eyes-open resting conditions. Clin Neurophysiol. 118:2765–2773.

Bassett DS, Yang M, Wymbs NF, Grafton ST. 2015. Learning-induced autonomy of sensorimotor systems. Nat Neurosci. 18:744–751.

Bazanova OM, Vernon D. 2014. Interpreting EEG alpha activity. Neurosci Biobehav Rev. 44:94–110.

Benjamin C, Lieberman DA, Changl M, Ofen N, Whitfield-Gabrieli S, Gabrieli JDE, Gaab N. 2010. The influence of rest period instructions on the default mode network. Front Hum Neurosci. 4:1–9.

Bijsterbosch J, Harrison S, Duff E, Alfaro-Almagro F, Woolrich M, Smith S. 2017. Investigations into within- and between-subject resting-state amplitude variations. Neuroimage. 159:57–69.

Biswal B, Zerrin Yetkin F, Haughton VM, Hyde JS. 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. Magn Reson Med. 34:537–541.

Blum A, Kalai A, Langford J. 1999. Beating the hold-out: bounds for K-fold and progressive cross-validation. Proc Annu ACM Conf Comput Learn Theory. 203–208.

Boersma M, Smit DJA, De Bie HMA, Van Baal GCM, Boomsma DI, De Geus EJC, Delemarre-Van De Waal HA, Stam CJ. 2011. Network analysis of resting state EEG in the developing young brain: Structure comes with maturation. Hum Brain Mapp. 32:413–425.

Bonkhoff AK, Hope T, Bzdok D, Guggisberg AG, Hawe RL, Dukelow SP, Rehme AK, Fink GR. 2020. Bringing proportional recovery into proportion : Bayesian modelling of post-stroke

motor impairment. 1–18.

Buckner RL, DiNicola LM. 2019. The brain's default network: updated anatomy, physiology and evolving insights. Nat Rev Neurosci. 20:593–608.

Buzsáki G, Anastassiou C a., Koch C. 2012. The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. Nat Rev Neurosci. 13:407–420.

Cabeza R, Albert M, Belleville S, Craik FIM, Duarte A, Grady CL, Lindenberger U, Nyberg L, Park DC, Reuter-Lorenz PA, Rugg MD, Steffener J, Rajah MN. 2018. Maintenance, reserve and compensation: the cognitive neuroscience of healthy ageing. Nat Rev Neurosci. 19:701–710.

Calhoun VD, Miller R, Pearlson G, Adali T. 2014. The Chronnectome: Time-Varying Connectivity Networks as the Next Frontier in fMRI Data Discovery. Neuron. 84:262–274.

Campenhout JM Van. 1978. On the Peaking of the Hughes Mean Recognition Accuracy - the Resolution of an Apparent Paradox. IEEE Trans Syst Man Cybern. SMC-8:390–395.

Campisi P, Rocca D La. 2014. Brain waves for automatic biometric-based user recognition. IEEE Trans Inf Forensics Secur. 9:782–800.

Carino-Escobar RI, Carrillo-Mora P, Valdés-Cristerna R, Rodriguez-Barragan MA, Hernandez-Arenas C, Quinzaños-Fresnedo J, Galicia-Alvarado MA, Cantillo-Negrete J. 2019. Longitudinal analysis of stroke patients' brain rhythms during an intervention with a brain-computer interface. Neural Plast. 2019.

Cassani R, Estarellas M, San-Martin R, Fraga FJ, Falk TH. 2018. Systematic review on resting-state EEG for Alzheimer's disease diagnosis and progression assessment. Dis Markers. 2018.

Cassim F, Szurhaj W, Sediri H, Devos D, Bourriez JL, Poirot I, Derambure P, Defebvre L, Guieu JD. 2000. Brief and sustained movements: Differences in event-related (de)synchronization (ERD/ERS) patterns. Clin Neurophysiol. 111:2032–2039.

Chiang AKI, Rennie CJ, Robinson PA, van Albada SJ, Kerr CC. 2011. Age trends and sex differences of alpha rhythms including split alpha peaks. Clin Neurophysiol. 122:1505–1517.

Cole SR, Voytek B. 2017. Brain Oscillations and the Importance of Waveform Shape. Trends Cogn Sci. 21:137–149.

Cox R, Schapiro AC, Stickgold R. 2018. Variability and stability of large-scale cortical oscillation patterns. Netw Neurosci. 2:481–512.

Damoiseaux JS, Greicius ÆMD. 2009. Greater than the sum of its parts : a review of studies combining structural connectivity and resting-state functional connectivity. 525–533.

Delorme A, Makeig S. 2004. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods. 134:9–21.

Diaz BA, van der Sluis S, Moens S, Benjamins JS, Migliorati F, Stoffers D, den Braber A, Poil SS, Hardstone R, Van't Ent D V., Boomsma DI, de Geus E, Mansvelder HD, Van Someren EJW, Linkenkaer-Hansen K. 2013. The Amsterdam Resting-state Questionnaire reveals multiple phenotypes of resting-state cognition. Front Hum Neurosci. 7:1–15.

Duncan NW, Northoff G. 2013. Overview of potential procedural and participant-related confounds for neuroimaging of the resting state. J Psychiatry Neurosci. 38:84–96.

Erbil N, Ungan P. 2007. Changes in the alpha and beta amplitudes of the central EEG during the onset, continuation, and offset of long-duration repetitive hand movements. Brain Res. 1169:44–56.

Giaquinto S, Cobianchi A, Macera F, Nolfe G. 1994. EEG Recordings in the course of recovery from stroke. Stroke. 25:2204–2209.

Gui Q, Jin Z, Xu W. 2015. Exploring EEG-based biometrics for user identification and authentication. 2014 IEEE Signal Process Med Biol Symp IEEE SPMB 2014 - Proc. 1–6.

Guyon I, Elisseeff A. 2003. An Introduction to Variable and Feature Selection. J Mach Learn Res. 3:1157–1182.

Haufe S, Meinecke F, Görgen K, Dähne S, Haynes JD, Blankertz B, Bießmann F. 2014. On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage. 87:96–110.

Hermundstad AM, Bassett DS, Brown KS, Aminoff EM, Clewett D, Freeman S, Frithsen A, Johnson A, Tipper CM, Miller MB, Grafton ST, Carlson JM. 2013. Structural foundations of resting-state and task-based functional connectivity in the human brain. Proc Natl Acad Sci U S A.

Hoenig MC, Bischof GN, Seemiller J, Hammes J, Kukolja J, Onur ÖA, Jessen F, Fliessbach K, Neumaier B, Fink GR, van Eimeren T, Drzezga A. 2018. Networks of tau distribution in Alzheimer's disease. Brain.

Hutchison RM, Womelsdorf T, Allen EA, Bandettini PA, Calhoun VD, Corbetta M, Della Penna S, Duyn JH, Glover GH, Gonzalez-Castillo J, Handwerker DA, Keilholz S, Kiviniemi V, Leopold DA, de Pasquale F, Sporns O, Walter M, Chang C. 2013. Dynamic functional connectivity: Promise, issues, and interpretations. Neuroimage. 80:360–378.

Kawagoe T, Onoda K, Yamaguchi S. 2018. Different pre-scanning instructions induce distinct psychological and resting brain states during functional magnetic resonance imaging. Eur J Neurosci. 47:77–82.

Knyazeva MG, Barzegaran E, Vildavski VY, Demonet J-F. 2018. Aging of human alpha rhythm. Neurobiol Aging. 69:261–273.

Mišic B, Betzel RF, De Reus MA, Van Den Heuvel MP, Berman MG, McIntosh AR, Sporns O. 2016. Network-level structure-function relationships in human neocortex. Cereb Cortex. 26:3285–3296.

Mognon A, Jovicich J, Bruzzone L, Buiatti M. 2011. ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. Psychophysiology. 48:229–240.

Näpflin M, Wildi M, Sarnthein J. 2007. Test-retest reliability of resting EEG spectra validates a statistical signature of persons. Clin Neurophysiol. 118:2519–2524.

Newbold DJ, Laumann TO, Hoyt CR, Hampton JM, Montez DF, Raut R V., Ortega M, Mitra A, Nielsen AN, Miller DB, Adeyemo B, Nguyen AL, Scheidter KM, Tanenbaum AB, Van AN, Marek S, Schlaggar BL, Carter AR, Greene DJ, Gordon EM, Raichle ME, Petersen SE, Snyder

AZ, Dosenbach NUF. 2020. Plasticity and Spontaneous Activity Pulses in Disused Human Brain Circuits. Neuron. 0:1–10.

Noble S, Scheinost D, Constable RT. 2019. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. Neuroimage. 203:116157.

O'Brien F, Cousineau D. 2014. Representing Error bars in within-subject designs in typical software packages. Quant Methods Psychol. 10:56–67.

Oldfield RC. 1971. The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia. 9:97–113.

Pedregosa F, Weiss R, Brucher M, 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, Duchesnay É. 2011. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 12:2825–2830.

Pfurtscheller G, Lopes da Silva FH. 1999. Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin Neurophysiol. 110:1842–1857.

Postema MC, De Marco M, Colato E, Venneri A. 2019. A study of within-subject reliability of the brain's default-mode network. Magn Reson Mater Physics, Biol Med. 32:391–405.

Pritschet L, Santander T, Taylor CM, Layher E, Yu S, Miller MB, Grafton ST, Jacobs EG. 2020. Functional reorganization of brain networks across the human menstrual cycle. Neuroimage. 220:117091.

Rehme AK, Fink GR, Cramon DY Von, Grefkes C. 2011. The Role of the Contralesional Motor Cortex for Motor Recovery in the Early Days after Stroke Assessed with Longitudinal fMRI.

Rifkin R, Klautau A. 2004. In Defense of One-Vs-All Classification. J Mach Learn Res. 5:2–6.

Saes M, Zandvliet SB, Andringa AS, Daffertshofer A, Twisk JWR, Meskers CGM, van Wegen EEH, Kwakkel G. 2020. Is Resting-State EEG Longitudinally Associated With Recovery of Clinical Neurological Impairments Early Poststroke? A Prospective Cohort Study. Neurorehabil Neural Repair. 34:389–402.

Sima C, Dougherty ER. 2008. The peaking phenomenon in the presence of feature-selection. Pattern Recognit Lett. 29:1667–1674.

Valizadeh SA, Riener R, Elmer S, Jäncke L. 2019. Decrypting the electrophysiological individuality of the human brain: Identification of individuals based on resting-state EEG activity. Neuroimage. 197:470–481.

Vallat R. 2018. Pingouin: statistics in Python. J Open Source Softw. 3:1026.

van Albada SJ, Kerr CC, Chiang AKI, Rennie CJ, Robinson PA. 2010. Neurophysiological changes with age probed by inverse modeling of EEG spectra. Clin Neurophysiol. 121:21–38.

van Albada SJ, Robinson PA. 2013. Relationships between electroencephalographic spectral peaks across frequency bands. Front Hum Neurosci. 7:1–18.

Van Den Heuvel MP, Mandl RCW, Kahn RS, Hulshoff Pol HE. 2009. Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. Hum Brain Mapp. 30:3127–3141.

van der Vliet R, Selles RW, Andrinopoulou ER, Nijland R, Ribbers GM, Frens MA, Meskers C, Kwakkel G. 2020. Predicting Upper Limb Motor Impairment Recovery after Stroke: A Mixture Model. Ann Neurol. 87:383–393.

Vecchio F, Babiloni C, Lizio R, De Vico Fallani F, Blinowska K, Verrienti G, Frisoni G, Rossini PM. 2013. Resting state cortical EEG rhythms in Alzheimer's disease: Toward EEG markers for clinical applications: A review. 1st ed, Supplements to Clinical Neurophysiology. Elsevier B.V.

Voytek B, Kramer MA, Case J, Lepage KQ, Tempesta ZR, Knight RT, Gazzaley A. 2015. Age-Related Changes in 1/f Neural Electrophysiological Noise. J Neurosci. 35:13257–13265.

Watson D, Clark LA, Tellegen A. 1988. Development and validation of brief measures of positive and negative affect: The PANAS scales. J Pers Soc Psychol. 54:1063–1070.

Winkler I, Debener S, Muller KR, Tangermann M. 2015. On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. Proc Annu Int Conf IEEE Eng Med Biol Soc

EMBS. 2015-Novem:4101–4105.

Wu J, Srinivasan R, Quinlan EB, Solodkin A, Small SL, Cramer SC. 2016. Utility of EEG measures of brain function in patients with acute stroke. J Neurophysiol. 115:2399–2405.

Wymbs NF, Bassett DS, Mucha PJ, Porter M a., Grafton ST. 2012. Differential Recruitment of the Sensorimotor Putamen and Frontoparietal Cortex during Motor Chunking in Humans. Neuron. 74:936–946.

Wymbs NF, Grafton ST. 2014. The Human Motor System Supports Sequence-Specific Representations over Multiple Training-Dependent Timescales. Cereb Cortex.

SUPPLEMENTARY MATERIAL

Supplementary Table 1: One-sample t-tests of mean identification accuracy in different states (see Table 1) vs random chance (50%) reported as t-value (p-value).

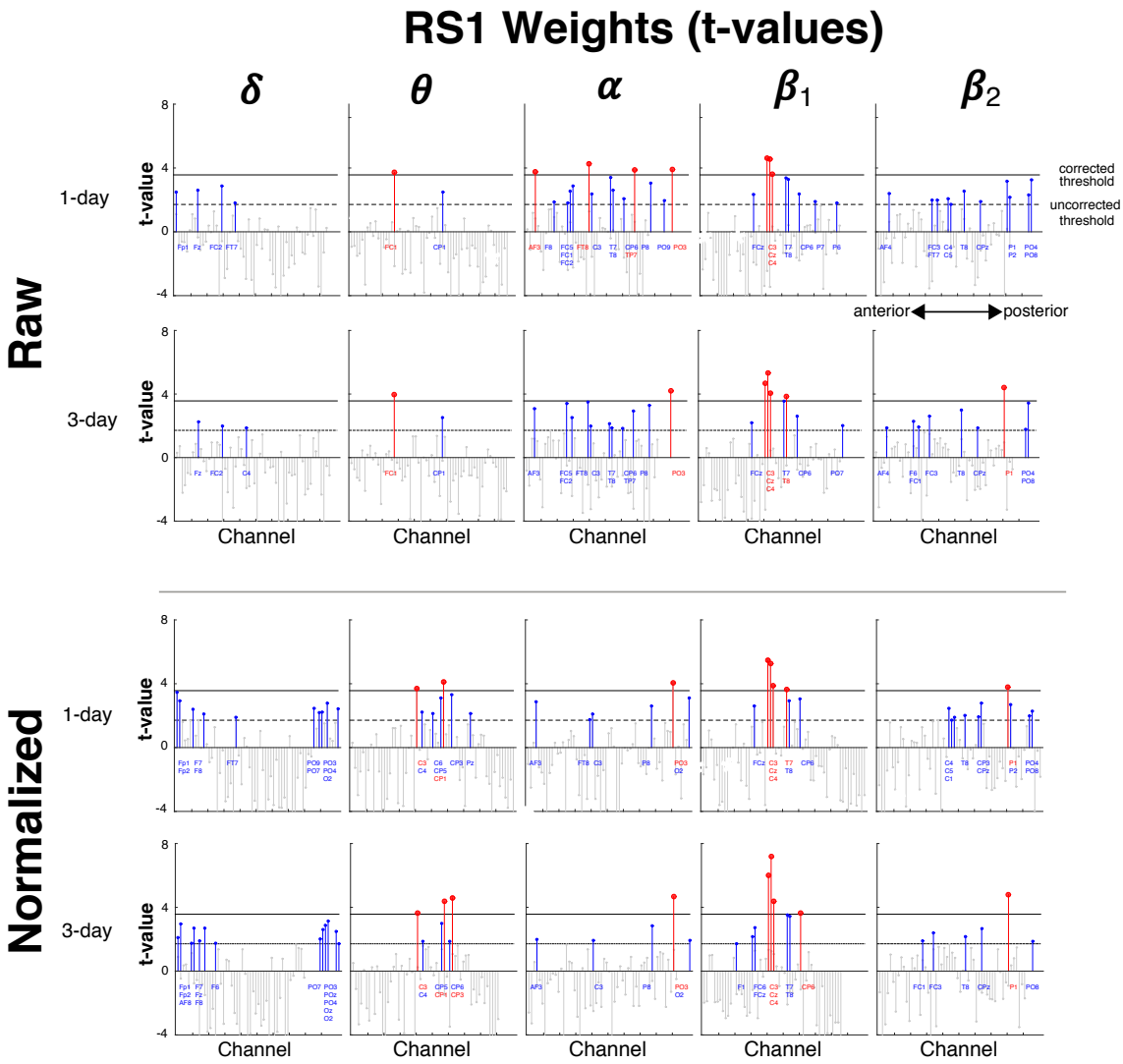
States	**Type**			
	CV	**1-day**	**2-day**	**3-day**
RS1 (*df* =23)	5596.13 (< 0.00001)	30.14 (< 0.00001)	61.95 (< 0.00001)	87.59 (< 0.00001)
RS1 (*df* =17)	3464.12 (< 0.00001)	26.85 (< 0.00001)	60.01 (< 0.00001)	80.39 (< 0.00001)
RS2 (*df* =23)	6264.07 (< 0.00001)	27.19 (< 0.00001)	53.70 (< 0.00001)	65.70 (< 0.00001)
TapWait (*df* =17)	9303.05 (< 0.00001)	28.30 (< 0.00001)	57.45 (< 0.00001)	73.02 (< 0.00001)
SeqWait (*df* =17)	9303.05 (< 0.00001)	25.21 (< 0.00001)	45.89 (< 0.00001)	53.57 (< 0.00001)
TapMov (*df* =17)	1739.06 (< 0.00001)	26.75 (< 0.00001)	58.60 (< 0.00001)	83.32 (< 0.00001)
SeqMov (*df* =17)	2215.05 (< 0.00001)	21.93 (< 0.00001)	42.65 (< 0.00001)	67.95 (< 0.00001)

Supplementary table 2: One-sample t-tests of mean identification accuracy for *RS1* with mono-band and mono-location feature subsets (see Table 2) vs random chance (50%) reported as t-value (p-value).

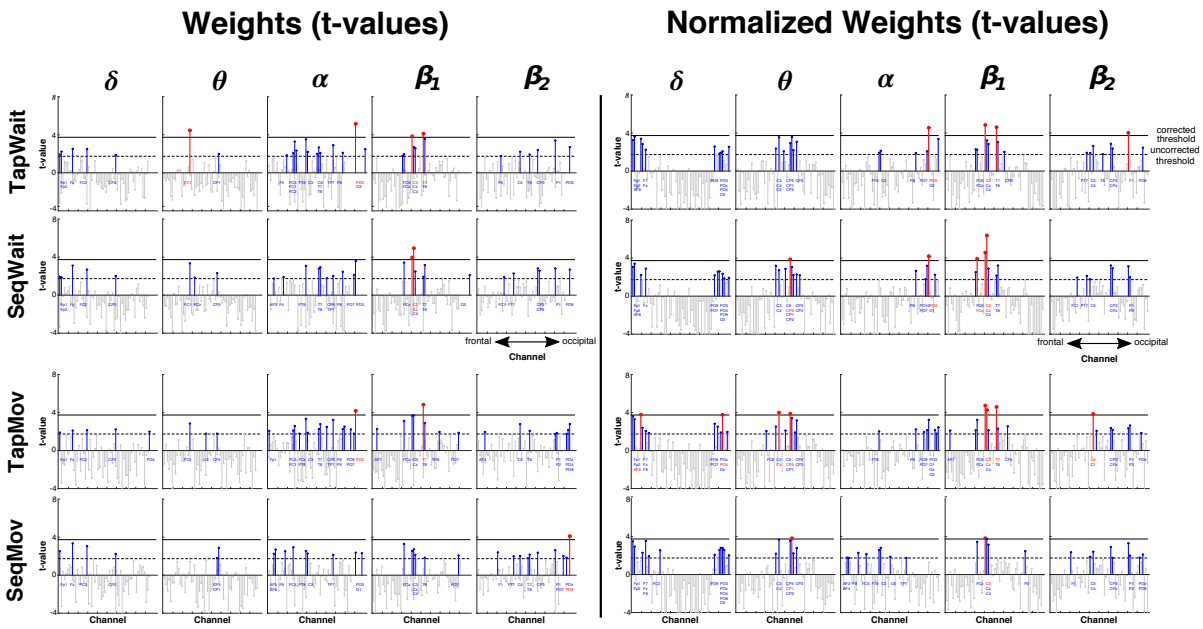
Subset (*df* =23)	**Type**			
	CV	**1-day**	**2-day**	**3-day**
$\boldsymbol{B_{\delta}}$	88.62 (< 0.00001)	9.07 (< 0.00001)	10.79 (< 0.00001)	12.29 (< 0.00001)
$\boldsymbol{B_{\theta}}$	151.87 (< 0.00001)	17.20 (< 0.00001)	21.88 (< 0.00001)	23.79 (< 0.00001)
$\boldsymbol{B_{\alpha}}$	202.00 (< 0.00001)	21.64 (< 0.00001)	29.69 (< 0.00001)	34.22 (< 0.00001)
$\boldsymbol{B_{\beta 1}}$	415.09 (< 0.00001)	14.73 (< 0.00001)	20.09 (< 0.00001)	23.42 (< 0.00001)
$\boldsymbol{B_{\beta 2}}$	703.68 (< 0.00001)	12.95 (< 0.00001)	18.10 (< 0.00001)	22.61 (< 0.00001)
$\boldsymbol{L_{F}}$	131.76 (< 0.00001)	18.00 (< 0.00001)	24.50 (< 0.00001)	29.01 (< 0.00001)
$\boldsymbol{L_{FC}}$	154.96 (< 0.00001)	19.29 (< 0.00001)	26.54 (< 0.00001)	33.46 (< 0.00001)
$\boldsymbol{L_{CP}}$	132.33 (< 0.00001)	20.15 (< 0.00001)	26.47 (< 0.00001)	30.29 (< 0.00001)
$\boldsymbol{L_{PO}}$	116.20 (< 0.00001)	18.26 (< 0.00001)	24.30 (< 0.00001)	26.33 (< 0.00001)

Supplementary table 3: One-sample t-tests of **c**lassification accuracy of *RS1* vs task state (see Table 3) against random chance (50%) reported as t-value (p-value).				
***RS1* vs**	**Type**			
(*df* =17)	**CV**	**1-day**	**2-day**	**3-day**
TapWait	28.74 (< 0.00001)	8.51 (< 0.00001)	7.94 (< 0.00001)	8.05 (< 0.00001)
SeqWait	50.81 (< 0.00001)	8.85 (< 0.00001)	8.88 (< 0.00001)	9.28 (< 0.00001)
TapMov	45.23 (< 0.00001)	17.19 (< 0.00001)	23.18 (< 0.00001)	24.49 (< 0.00001)
SeqMov	119.60 (< 0.00001)	31.55 (< 0.00001)	52.90 (< 0.00001)	48.70 (< 0.00001)

Supplementary table 4: One-sample t-tests of cross-task ${}^{RS1}\mathbf{I}_p \rightarrow {}^{X}\mathbf{I}_q$ identification accuracy (see Table 4) against random chance (50%) reported as t-value (p-value).			
Test states	**Type**		
(*df = 17*)	**1-day**	**2-day**	**3-day**
TapWait	27.52 (< 0.00001)	58.90 (< 0.00001)	81.63 (< 0.00001)
SeqWait	24.43 (< 0.00001)	43.39 (< 0.00001)	57.55 (< 0.00001)
TapMov	21.73 (< 0.00001)	33.23 (< 0.00001)	41.42 (< 0.00001)
SeqMov	13.88 (< 0.00001)	17.55 (< 0.00001)	19.14 (< 0.00001)



Supplementary Figure 1: Stem plot of t-values of high-valued weights in *RS1* per frequency band from Figure 7. Upper two rows show t-values for mean raw weights (z-scored across all features) and their changes with aggregation (1-day, 3-day). Lower two rows show the corresponding t-values for the mean normalized weights (z-scored per band). Channels have an anterior-to-posterior ordering (x-axis). Red stems indicate channels with t-values higher than the corrected threshold ($p < 0.05/61$, horizontal black line) while blue stems show channels that only pass uncorrected thresholds ($p < 0.05$, dotted horizontal line). Colored channel labels are grouped for visibility (top to bottom) correspond to stems (left to right).



Supplementary Figure 2: Stem plot of t-values of high-valued weights (1-day) in task-states per frequency band from Figure 9. All labeling conventions are as in Suppl. Figure 1.