RESEARCH REPORT

Robustness of individualized inferences from longitudinal resting state EEG dynamics

Maximilian Hommelsen[1] | **Shivakumar Viswanathan[1]** | **Silvia Daun[1,2]**

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

[2]Institute of Zoology, University of Cologne, Cologne, Germany

Correspondence
Silvia Daun, Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Forschungszentrum Jülich, Jülich 52425, Germany.
Email: s.daun@fz-juelich.de

Abstract

Tracking how individual human brains change over extended timescales is crucial to clinical scenarios ranging from stroke recovery to healthy aging. The use of resting state (RS) activity for tracking is a promising possibility. However, it is unresolved how a person's RS activity over time can be decoded to distinguish neurophysiological changes from confounding cognitive variability. Here, we develop a method to screen RS activity changes for these confounding effects by formulating it as a problem of change classification. We demonstrate a novel solution to change classification by linking individual-specific change to inter-individual differences. Individual RS-electroencephalography (EEG) was acquired over 5 consecutive days including task states devised to simulate the effects of inter-day cognitive variation. As inter-individual differences are shaped by neurophysiological differences, the inter-individual differences in RS activity on 1 day were analysed (using machine learning) to identify distinctive configurations in each individual's RS activity. Using this configuration as a decision rule, an individual could be re-identified from 2-s samples of the instantaneous oscillatory power spectrum acquired on a different day both from RS and confounded RS with a limited loss in accuracy. Importantly, the low loss in accuracy in cross-day versus same-day classification was achieved with classifiers that combined information from multiple frequency bands at channels across the scalp (with a concentration at characteristic fronto-central and occipital zones). Taken together, these findings support the technical feasibility of screening RS activity for confounding effects and the suitability of

1 | INTRODUCTION

Tracking the changes to a person's brain over extended timescales (e.g., days to years) is crucial to understand neural plasticity processes in numerous clinically relevant scenarios, from stroke recovery (Bonkhoff et al., 2020; Giaquinto et al., 1994; Grefkes & Fink, 2020; Rehme et al., 2011; Saes et al., 2020; van der Vliet et al., 2020; Wu et al., 2016) to healthy aging (Boersma et al., 2011; Cabeza et al., 2018; Cassani et al., 2018). One promising strategy for individualized tracking is with repeated measurements of resting state (RS) activity: the ongoing neural oscillatory dynamics over a few minutes of wakeful rest (Carino-Escobar et al., 2019; Gordon et al., 2017; Guerra-Carrillo et al., 2014; Hohenfeld et al., 2018; Laumann et al., 2015; Newbold et al., 2020; Pritschet et al., 2020; Saes et al., 2020; Vecchio et al., 2013; Wu et al., 2015). Despite being 'task free', the organization of RS activity has revealed a close relationship to underlying neurobiological organization and integrity (Biswal et al., 1995; Buckner & DiNicola, 2019; Damoiseaux & Greicius, 2009; Hermundstad et al., 2013; Hoenig et al., 2018; Mišic et al., 2016; van den Heuvel et al., 2009). The apparent informativeness of RS activity coupled with its convenient and inexpensive acquisition (e.g., with electroencephalography [EEG]) has suggested its suitability to track a person's changing neurophysiology. Nevertheless, strategies to translate the information in a person's multi-day RS activity into *inferences* specific to that person's changing brain remain poorly understood.

For population inferences, each individual is a sample from the population of interest and the focus is on observed individual effects that generalize to the population (e.g., is there a mean effect for group *X*?) (Gratton et al., 2018; Poldrack, 2017). However, for an individual-specific inference, the individual is both the source of information and the target for the inference. For RS-based tracking, this could entail the use of ~5–10 min of RS activity acquired from a person on different days to make a diagnostic classification about that person (e.g., has person *X* undergone a neuroplastic change?). Furthermore, this inference procedure might itself have to be adapted to the person's unique characteristics. This narrowed scope of individualized inferences presents challenging constraints and uncertainties without a simple equivalent in population inference methods. Towards specifying and addressing these challenges, in the current study, we investigate an inference problem posed by multi-day RS tracking, namely, of classifying inter-day activity changes.

1.1 | Individualized change classification: Problem specification

Suppose a person's RS activity is acquired on two different days. In the period between these two measurements, a person's underlying neurophysiology organization might have changed in a variety of possible ways. However, it is also possible that the person's neurophysiology remained relatively unchanged over this period. These diverse possibilities can be simplified into two categories: (i) change present (i.e., positive instances of neurophysiological change of any kind, denoted as $NP+$), or (ii) change absent (i.e., negative instances, $NP-$). This scenario now poses the diagnostic problem of *change classification* (Figure 1a): How can a person's RS activity from Days 1 and 2 be used to decide whether that person's neurophysiological organization might have undergone a change of any kind ($NP+$) or not ($NP-$)?

One simple but erroneous criterion for change classification is the presence/absence of an inter-day difference in RS activity. The diverse variety of possible neurophysiological changes across the brain (i.e., $NP+$) could produce correspondingly diverse differences in inter-day RS activity. However, the absence of neurophysiological changes ($NP-$) does not imply the absence of inter-day RS differences due to at least two sources of 'nuisance' variability.

Firstly, the repeated measurements on different days can affect inter-day differences in RS activity due to measurement-related factors, as demonstrated by numerous studies of test–retest reliability (Bijsterbosch et al., 2017; Brandmaier et al., 2018; Cox et al., 2018; Noble et al., 2019; Postema et al., 2019). Such spurious activity changes can lead to misclassifications of $NP-$ as $NP+$ and also $NP+$ as $NP-$. Secondly, 'rest' is an under-constrained cognitive state that can introduce incidental,
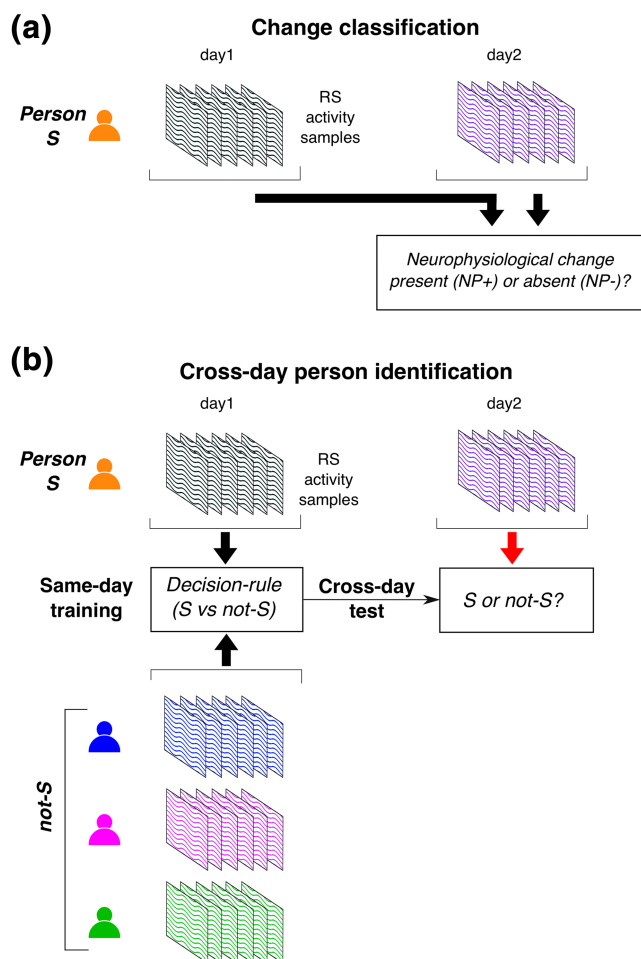
FIGURE 1 Change classification and person identification. (a) Change classification: Use resting state (RS) activity samples of person S from Day 1 (black wavy lines) and Day 2 (magenta lines) to decide if S underwent a neurophysiological change between days (NP+) or not (NP−). (b) Cross-day person identification: Train a decision rule to identify S from RS activity samples of S and a collection of others (not-S) from Day 1. Use decision rule to decide whether samples from Day 2 are from S or not?

idiosyncratic variations in a person's neural state (e.g., Day 1: mind-wandering; Day 2: sleepiness; Day 3: recalling emotional memories) (Benjamin et al., 2010; Diaz et al., 2013; Duncan & Northoff, 2013; Gonzalez-Castillo et al., 2021; Kawagoe et al., 2018). Such cryptic differences in the cognitive state between days could produce differences in RS neural activity even when underlying neurophysiology is unchanged. This can lead to false positives (Type-I errors) where the absence of change (NP−) is misclassified as a neuroplastic change (NP+). Inter-day variability in cognitive states is a particular concern because it is difficult to eliminate and involves true neural activity differences that could confound classification even if measurement-related variability were to be low and highly controlled.

A desirable solution to change classification would be a *robust* decision rule that enables accurate classification <u>despite</u> the presence of measurement and cognitive state variability between days. 'Robustness' is used here in the sense of maintaining performance despite deviations and uncertainties about model assumptions (e.g., Box & Andersen, 1955; Huber, 1981; Xu & Mannor, 2012) and is not used as a synonym for measurement reliability (e.g., Brandmaier et al., 2018). In the current study, we used multivariate pattern analysis (MVPA; Haynes, 2015; Varoquaux et al., 2017) to investigate whether and how such robust decision rules (classifiers) might be obtained. A decision rule's ability to *generalize* (Bishop, 2006), that is, to classify information from a different setting than the information used to train the decision rule, provided a quantitative measure of robustness.

1.2 | Proposed approach: Change classification formulated as cross-day person identification

Identifying a robust decision rule with supervised machine learning involves a critical constraint. For multi-day RS-tracking, the equivalent of one 'trial' is the change related to a single pair of RS measurements from different days. Due to practical constraints of RS tracking, only a single trial (i.e., one pair of RS measurements) might be available from a person and the status of this one trial (NP+/NP−) might be unknown and to be determined by classification. Therefore, training a classifier to categorize inter-day differences as either NP+ or NP− using a set of prior examples of these categories from that person would be infeasible. To accommodate this severe constraint, we propose a strategy to use the *distinctiveness* of individual RS activity as an alternative source of information for change classification.

Our proposed strategy is based on the observation that RS change classification is qualitatively isomorphic to the well-studied problem of RS-based *person identification*. Numerous studies demonstrate that individual RS activity is highly distinctive to the extent that a person can be identified relative to others solely from their RS activity (Campisi & Rocca, 2014; Finn et al., 2015; see, e.g., Huang et al., 2012; Pani et al., 2020; Valizadeh et al., 2019). In that framework, identification is a form of population inference with a focus on multivariate relationships in a person's RS activity that generalize to samples of the person's own activity but <u>not</u> to the activity of others. These distinctive characteristics are seemingly like the person's biometric signature (or 'fingerprint'). Crucially, to identify a person across different days, these RS 'fingerprints' have to be robust to measurement-

related and cognitive state variability (see below). Therefore, we reasoned that procedures used for person identification could be relevant to robust change classification.

The qualitative similarities between change classification and person identification are illustrated in Table 1 (also see Figure 1b). Each row of the table shows the inter-sample relationship of interest, the associated sources of inter-sample differences (measurement, cognitive state, neurophysiological), and the idealized categorization (right column). For idealized change classification (Rows 1 and 2), only the presence/absence of inter-day neurophysiological differences are of relevance to the desired categorization (NP−/NP+), although the other differences are 'nuisance' factors to be ignored. This is also the case for idealized *cross-day* person identification (Rows 3 and 4), where measurement and cognitive state differences are irrelevant for accurate identification. Furthermore, idealized classification of NP− (Row 1) and person S (Row 3) has a similar structure.

For (non-robust) change classification based on simple activity differences, cognitive state differences could lead to a misclassification of NP− (Row 1) as NP+. This nuisance factor can also lead to a misidentification of person S (Row 3) as being not-S. A misidentification could also theoretically occur if the same person undergoes a neurophysiological change across days (Row 3, absent*). This error would be equivalent to the classification of NP+ (Row 2).

Based on this qualitative isomorphism (i.e., between Rows 1-2 and 3-4), a question is whether a *same-day* classifier trained to identify a particular person (Rows 5 and 6) might, <u>under suitable conditions</u>, be usable for cross-

day identification and hence possibly change classification. In prior studies of person identification, same-day classification (Rows 5 and 6), involved samples of a person's RS activity that were obtained from the same RS session of the same day. Thus, the inter-sample differences from the same person (Row 5) can be assumed to have small measurement and cognitive differences, and no neurophysiological differences. However, inter-sample differences from different persons (Row 6) would be associated with both measurement and cognitive state differences, as well as differences in the neurophysiological phenotype. Consequently, a decision rule trained for same-day person identification might not be robust for cross-day person identification unless it represents information about that person's unique neurophysiology in some manner.

We investigated the conditions required for robust inter-day person identification using longitudinal resting EEG activity acquired on 5 consecutive days from healthy, young participants who were assumed to be neurophysiologically stable over this period. A decision rule to identify a person S was trained (using machine learning) to distinguish between (i) examples of RS activity from S on a single day and (ii) examples of single day RS activity from a diverse pool of other individuals (i.e., not-S) (Figure 1b). Each example was a brief 2 s activity sample representing a dynamically variable neural state during an RS session (Calhoun et al., 2014; Hutchison et al., 2013). The decision rule for same day S/not-S classification is used to classify samples of S/not-S from a different day. A robust decision rule should be insensitive to inter-day cognitive variability when a person's

TABLE 1 Relationship between change classification and person identification

Classification problem	Inter-sample relationship (input)	Sources of inter-sample variation			Category (output)
		Measurement differences	Cognitive state differences	Neurophys. differences	
Change classification (cross-day)	Person S (Day 1 vs. Day 2)	Present	Present	**Absent**	*NP−*
	Person S (Day 1 vs. Day 2)	Present	Present	**Present**	*NP+*
Person identification (cross-day)	Person S (Day 1 vs. Day 2)	Present	Present	**Absent***	**Same-person**
	Person S vs. not-S	Present	Present	**Present**	**Different person**
Person identification (same-day)	Person S (Day 1 vs. Day 1)	Low	Low	**Absent**	**Same-person**
	Person S vs. not-S	Present	Present	**Present**	**Different person**

Note: Idealized solution concepts for binary change classification (Rows 1 and 2) person identification: Cross-day (Rows 3 and 4) and same-day (Rows 5 and 6). Each row shows the inter-sample comparisons (left column), the inter-sample differences due to measurement factors (blue), cognitive state (magenta), neurophysiological factors (orange), and the idealized categorization (green). For details, see text.

neurophysiology is unchanged. This predicted robustness was tested on pseudo-rest states that were designed to challenge cross-day classification with a diverse range of cognitive variation (Figure 2a).

2 | MATERIALS AND METHODS

2.1 | Participants

Twenty seven healthy volunteers (11 female, age [mean $\pm$ SD]: 27.9 years $\pm$ 3.4, range: 22–34 years)

participated in the study after providing their written informed consent. Participants had normal or corrected-to-normal vision, had no history of neurological or psychiatric disease, were not under medication and had no cranial metallic implants (including cochlear implants). Handedness was not a selection criterion (right handed: 22; left handed: 2; intermediate: 3, based on the Edinburgh Handedness Inventory; Oldfield, 1971). The participants received monetary compensation on completion of all sessions. The study complied with the Declaration of Helsinki and was approved by the Ethics Commission of the Faculty of Medicine, University of Cologne (Zeichen: 14-006).
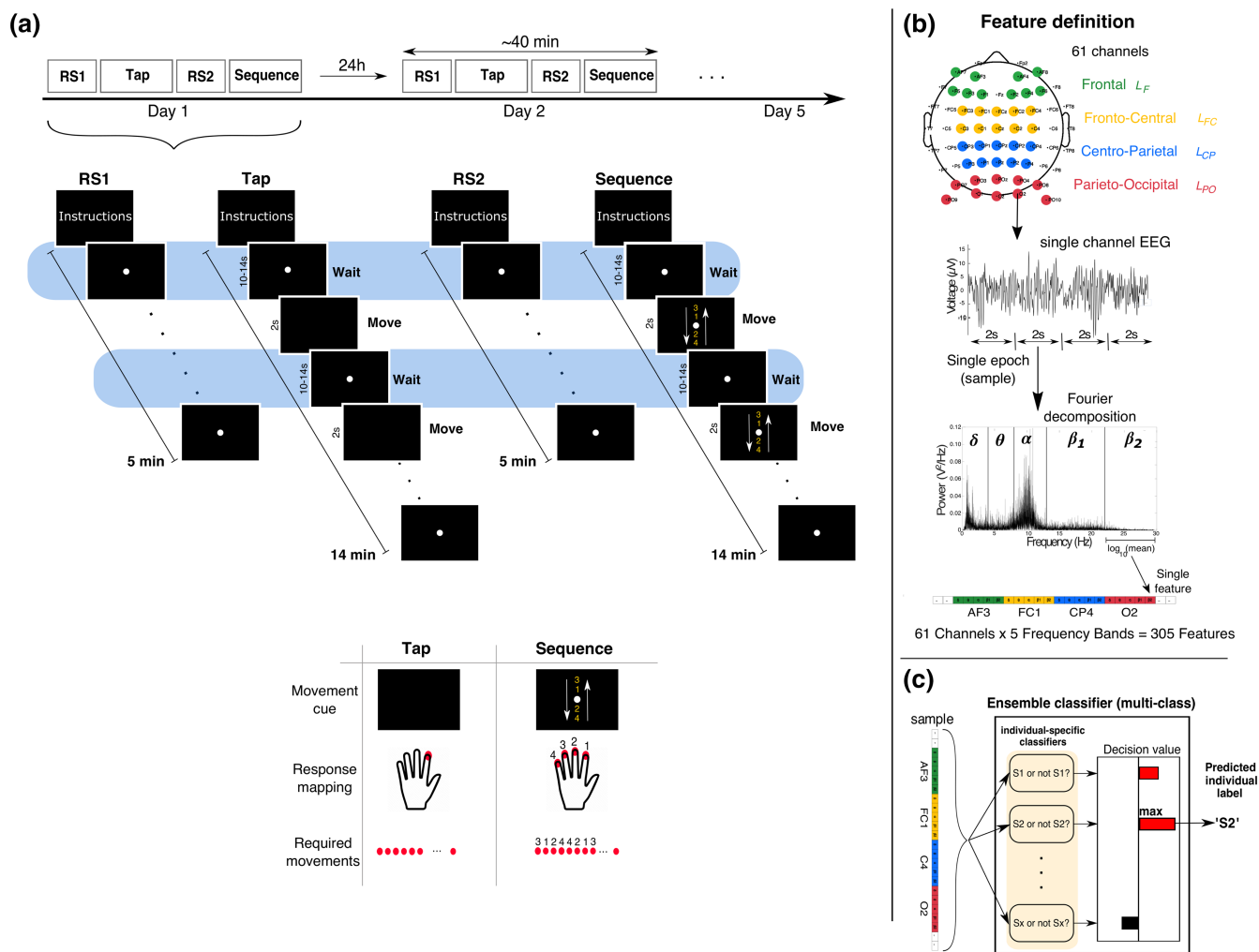
FIGURE 2 Paradigm and processing pipeline. (a) The protocol involved four tasks (*RS*1, *Tap*, *RS*2, *Sequence*) performed in the same order on 5 consecutive days. Task details for a single day are illustrated. A white fixation point was continuously displayed during *RS*1 and *RS*2, and during 'waiting' periods in the *Tap* and *Sequence* tasks (highlighted in blue). The Move cue in the *Tap* task (bottom panel) was a blank screen that required repeated button presses using the left index-finger (red dots). The Move cue in the *Sequence* task depicted four numbers between two arrows (not drawn to scale) and required buttons to be pressed in a continuous cyclical manner according to the number-to-finger mapping shown. (b) Schematic of feature vector specification. Channels in each mono-location subset are identified by colour (green: L_F; yellow: L_{FC}; blue: L_{CP}; red: L_{PO}). The continuous signal from each channel was segmented into 2-s 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. (c) Schematic of a multiclass decision with an ensemble of individual-specific binary classifiers. Each classifier evaluates the sample (S_X or not-S_X) 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)

Datasets from 24 (of the 27) participants were used for statistical analyses (see Section 2.6).

2.2 | 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 × 1080 pixels). Behavioural responses were recorded with the fMRI Button Pad (1-Hand) System (LXPAD-1 × 5-10 M, NAtA Technologies, Canada).

Scalp-EEG was acquired with a 64-channel active Ag/AgCl electrode system (actiCap, Brain Products, Germany) with a standard 10–20 spherical array layout (ground electrode at AFz, reference electrode on the left mastoid). Electrooculographic (EOG) activity evoked by horizontal and vertical eye movements was recorded with three electrodes (FT9, FT10 and TP10) placed at the left and right lateral canthi and below the left eye respectively. During acquisition, measured voltages (0.1 μV resolution) were amplified, filtered (low cut-off: DC, high cut-off: 250 Hz) and digitized at a sampling rate of 2.5 kHz (BrainAmp DC, BrainProducts GmbH, Germany).

The positioning of the EEG cap was registered with a stereotactic neuronavigation system (Brainsight v. 2.3, Rogue Research Inc, Canada) (details below).

2.3 | Experiment protocol and paradigm

Each participant completed five sessions (~40 min each) that were scheduled at the same time on consecutive days (Monday to Friday) at one of three possible time slots: morning (9 AM × 6), noon (12 PM × 9) or afternoon (3 PM × 12). For one participant, the fifth session was re-acquired after a gap of 3 days due to technical problems during the scheduled recording. Each session consisted of two RS recordings (RS1 and RS2) interleaved with two non-rest tasks (referred to as *Tap* and *Sequence*) in the same fixed order, namely, RS1, *Tap*, RS2 and *Sequence*. A schematic of the protocol and the different tasks is shown in Figure 2a.

2.3.1 | Task rationale

Despite being a task-free state, RS acquisition involves the task-like specification of (i) a *behavioural* state specified by instructions to stay still and keep eyes open (or closed) (Barry et al., 2007) and (ii) a *cognitive* state that is typically specified by instructions to relax and avoiding thinking of anything specific. This under-specified cognitive state is a source of uncertainty about the associated neural state, which is compounded in multi-day RS as this cognitive state might differ between days. All tasks that followed RS1 were included to test the consequences of this uncertainty.

RS2 activity was acquired in the same session on the same day without any changes to the measurement setup (e.g., removal of the EEG cap). However, RS2 occurred ~15–20 min after RS1 and always followed the *Tap* task, which could lead RS2 to deviate from RS1 in cognitive state due to potential carry-over effects from the *Tap* task (e.g., see Lim et al., 2010). Thus, RS1 and RS2 were different measurement instances of resting activity from the same day with relatively low measurement-driven differences but possible differences in cognitive state.

The non-rest tasks (*Tap* and *Sequence*) were designed to produce *pseudo* rest states that deviated from RS1 to differing extents in cognitive demands and longitudinal properties. Both tasks required participants to press buttons in response to visual cues. These visual cues were separated by relatively long and variable inter-stimulus intervals (10–14 s) where participants had to monitor the screen as they waited for the visual cue. Unknown to participants, this 'waiting' state (referred to as *TapWait* and *SeqWait*) was the primary focus of these tasks. These waiting states were similar to the rest state in behaviour (i.e., eyes open and no movement) but not in the accompanying cognitive state, which was determined by the task. The *Tap* task required the execution of a simple repetitive movement in response to the visual cue, which was assumed to produce a waiting state (viz., covert movement preparation) that was relatively similar between days. However, the *Sequence* task required the execution of a difficult multi-movement sequence where behavioural performance could improve with repeated practice across days. The *Sequence* task thus served to elicit a waiting state (viz., covert rehearsal of the movement sequence) that could systematically change across days with learning.

2.3.2 | Task details

Each task period began with an instruction screen describing the task to be performed and ended with a screen that instructed participants to take a short break and press a button to initiate the next task period when they were ready.

Resting state (RS1)
A white dot was continuously displayed at the centre of the screen over this period (duration: ~5 min).

Participants were instructed to remain still, keep their eyes open, fixate on the displayed white dot and relax by not thinking of anything in particular while remaining awake.

Tap task

As in *RS*1, a white dot was centrally displayed on the screen during this task period. However, after variable intervals of 10–14 s, this dot disappeared for a 2-s period. The offset of the dot was the cue for participants to use their left index finger to repeatedly press a button as rapidly as comfortably possible until the dot reappeared on the screen. The task (duration: ∼14 min) consisted of 60 movement periods (referred to as *TapMov*) interleaved with 60 waiting periods (i.e., *TapWait*).

Resting state (RS2)

A second RS recording (referred to as *RS*2) was acquired with the same task parameters as *RS*1.

Sequence task

Similar to the *Tap* task, a white dot was centrally displayed on the screen over 60 waiting periods of 10–14 s each (i.e., *SeqWait*) that were interleaved with 60 movement periods of 2-s duration (i.e., *SeqMov*). Unlike the *Tap* task, each movement period was cued by a centrally displayed visual stimulus depicting four digits (3-1-2-4) that were vertically ordered between two vertical arrows (see Figure 2a). Each number was mapped to a different button on the response pad. The ordering of the numbers indicated the sequence in which the corresponding buttons had to be pressed using fingers of the left hand. The arrows indicated that this sequence had to be executed in a cyclical manner starting from top to bottom and back. Here, the required sequence of button presses following the cue was 3-1-2-4-4-2-1-3-3-1-2-4- ... and so on. This continuing sequence had to be executed rapidly until the offset of the stimulus. The sequence used here was selected to be challenging for rapid execution. To promote learning of the sequence across trials and days, the same sequence and number-to-finger mapping was used on all sessions (and for all participants). No performance feedback was provided during the task. However, on each session, participants were encouraged to improve their task performance, namely, increase the number of sequences executed during the response periods.

Participants used fingers of their left hand to execute the button-press responses in the *Tap* and *Sequence* tasks. Note that handedness was not an inclusion criterion in our experiment,

2.4 | Procedure

On the first day, participants received detailed instructions about the different task periods of the experiment. They were familiarized with the number-to-finger mappings required for the *Sequence* task and practiced the task on a sequence that was different from the one used in the main experiment. Each session included reminders to minimize blinking, maintain fixation at all times during the EEG recording in all task periods, and avoid unnecessary movements of the fingers, head and body. At the start of each session, participants completed the Positive and Negative Affect Schedule (PANAS) (Watson et al., 1988) and brief questionnaires about their caffeine consumption on that day and the amount and quality of sleep on the previous night.

We used a prospective strategy to minimize inter-day variation in the positioning of the EEG cap with a spatial registration procedure on each day. 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 registered in this space. The electrode locations from the first day were used as a spatial reference for the remaining sessions. On each subsequent session, the cap's position was adjusted to align these selected electrodes to their reference locations. Due to scheduling constraints, this spatial registration procedure was not performed for seven 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 at a comfortable chair with their heads stabilized with a chinrest in front of the computer screen at a viewing distance of ∼65 cm. The response pad was placed in a recess under the table to prevent visibility of the hands during the task periods. Participants were monitored during the recording via a video camera to ensure that they maintained fixation, minimized eye-blinks and stayed awake.

2.5 | EEG preprocessing

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

2.5.1 | Artefact correction/rejection

The continuous recordings were down-sampled to 128 Hz, and then band-pass filtered to the range 1–40 Hz with a Hamming windowed since FIR filter (high pass followed by low pass). Artefacts due to oculomotor activity were corrected in the continuous recordings with independent component analysis (ICA) using a procedure described by Winkler et al. (2015) (see Supplementary Methods 1). This artefact correction was performed separately for each day's dataset to maintain their independence. The recordings were then visually inspected to reject time periods and channels with artefacts related to repeated paroxysmal amplitudes changes (>50 µV), electromyographic contamination, electrical noise and signal loss. Signals at rejected channels were replaced by interpolation from other channels (using EEGLAB's implementation of spherical spline interpolation). All channels were then re-referenced to the Common Average Reference.

2.5.2 | Epoch definition

The artefact free continuous data were segmented into 2-s epochs according to the six different experimental states: *RS*1, *RS*2, *TapWait*, *TapMov*, *SeqWait* and *SeqMov*. The epoch duration of 2 s was heuristically selected to (i) be short enough to obtain a sufficiently large number of samples for the classification analyses (see below) while (ii) being long enough to obtain a suitable estimate of the power spectrum. A 2-s duration also allowed epochs from the non-movement periods to be matched to the duration of the task-defined movement periods.

To exclude carry-over effects from the movement periods into the *TapWait* and *SeqWait* epochs, a time interval of 500 ms immediately prior to cue onset and 1000 ms immediately following cue offset was excluded before segmenting the *TapWait* and *SeqWait* epochs. All *TapWait* and *SeqWait* epochs that contained button presses were also excluded. For the two movement-related states (*TapMov* and *SeqMov*), epochs were defined from +0.25 to +2.25 s following the visual cue to exclude initial transients and response-time delays following cue onset and to include any residual movements in the period immediately following the cue offset.

2.6 | Data quality assessment

Preprocessing resulted in 135 session datasets (27 participants × 5 days). For inclusion in the final analysis, each participant had to have completed the first three of the four tasks on all sessions and have at least four (out of 5) session datasets that met the following data-quality criteria. We required a preprocessed session dataset to have (i) less than seven rejected channels, (ii) ≥ 90 artefact-free epochs from both RS periods (i.e., *RS*1 and *RS*2) and (iii) ≥ 90 artefact-free epochs from the available resting-matched conditions (i.e., *TapWait* and *SeqWait*). Note that the number of epochs for *TapMov* and *SeqMov* were necessarily ≤ 60 as each task only had 60 response periods of 2-s duration. To maintain uniformity across participants, final analyses were performed only on the best four of the five session datasets from each participant. If all five session datasets were of high quality, the first day's dataset was excluded as it might involve effects of initial familiarization.

Datasets from 24 out of 27 participants met the above data-quality criteria: 18 (of 24) had completed all four task periods on each session while the remaining six (of the 24) participants had completed only the first three (of the four parts). Only two of the 24 participants required the use of the first day's dataset. To maximize the use of the available data, analyses involving only *RS*1 and *RS*2 included data from 24 participants, while analyses of the non-rest tasks used data from 18 participants.

For the 24 participants, the mean number of epochs per day per participant for *RS*1 was 137.81 (min: 125.7, max: 146.5, $SD = 5.28$, min/day = 96), and for *RS*2 was 138.11 (min: 116.5, max: 148.0, $SD = 6.89$, min/day = 91). For the 18 participants used to analyse the non-rest tasks, the mean number of epochs per day per participant for *TapWait* was 160.6 (min: 146.0, max: 175.0, $SD = 7.23$, min/day = 129), and for *SeqWait* was 172.29 (min: 156.5, max: 186.5, $SD = 9.13$, min/day = 129). The corresponding values for lower for *TapMov* = 55.62 (min: 49.0, max: 59.0, $SD = 2.62$, min/day = 41), and for *SeqMov* = 56.74 (min: 53.25, max: 59.5, SD = 1.78, min/day = 43).

2.7 | Classifier specification

Each epoch was a 2-s sample of the ongoing activity from one person (of 24) on one specific day (of 4), although in a particular task state (of six possible states: true rest [*RS*1, *RS*2], pseudo-rest [*TapWait*, *SeqWait*] and non-rest [*TapMov*, *SeqMov*]). For our analyses, the basic classification problem was whether an epoch's activity could be used to identify (i.e., classify) that epoch's origin either by (i) a person's identity (using a multi-class classifier) or (ii) task state within the same person (using a standalone binary classifier).

2.7.1 | Feature specification: Power spectrum

The power spectrum on each 2-s epoch was the basis for all classification analyses. 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 as schematically displayed in Figure 2b. For each 2-s epoch of EEG activity, the power spectrum at each channel over the range of 1–30 Hz (0.5-Hz resolution) was computed using the fast Fourier transform (FFT). The power at all frequencies within each frequency band 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 × 61 channels). Because the shape of the power spectrum (i.e., relative power in the different bands) as well as the power in each band at different channels might themselves be individual-specific characteristics, no additional normalization was applied to the feature values.

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, $B_{\beta1}$ and $B_{\beta2}$. Each *mono-location* feature set (L_z) (Figure 2b, 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.

2.7.2 | Machine learning algorithm

The classifiers were numerically estimated (or learned) using a machine learning algorithm operating on a collection of samples (i.e., training set). For this purpose, we used a soft-margin linear support vector machine (SVM, with $L2$ regularization) algorithm (Boser et al., 1992) as implemented by the *LinearSVC* package in the *scikit-learn* library (Pedregosa et al., 2011) in Python 3.6. The SVM algorithm was pragmatically selected for being a commonly used, standard algorithm. SVM learning was initialized with parameters: tolerance $= 10^{-5}$, max iterations $= 10^4$, hinge loss, and balanced class weighting. The hyper-parameter C had a value of 1, which has been shown to be a reasonable default for M/EEG classification (Varoquaux et al., 2017). Tuning C's value to our data only marginally changed the classification accuracies obtained with $C = 1$ (results not shown). For all classifier estimations, the training data were always balanced (i.e., having an equal number of samples per class).

2.8 | Multi-class classification

Numerous prior studies demonstrate that RS activity can serve as a 'fingerprint' for person identification (Campisi & Rocca, 2014; Finn et al., 2015; Huang et al., 2012; Pani et al., 2020; Valizadeh et al., 2019). Although our focus was not on the neural basis of individual differences and trait identification (Demuru et al., 2017; Finn et al., 2017; Gratton et al., 2018; Smit et al., 2005, 2006), a person identification approach, using multi-class classifiers, provided a convenient technical platform for our test of individual-specific change classification.

2.8.1 | Definition

An N-class classifier ($N \geq 2$) in our analyses consisted of an ensemble of N binary classifiers employing a *one*-vs-*all* scheme (as implemented by *scikit-learn*). The input to such a multi-class classifier is a single sample (i.e., epoch) from an unspecified person S_X in the studied group and the output is the predicted identity of that person (e.g., S_2) (Figure 2c). Each person is associated with a unique classifier in the ensemble. Specifically, the binary classifier for each person (e.g., S_2) was independently trained to decide whether a sample was from that person or from all of the other $N - 1$ persons (i.e., not S_2). Therefore, to predict a person's identity with the entire ensemble, the input sample is separately evaluated by the decision rules of each of the N binary classifiers to obtain a decision value from each classifier (i.e., the signed distance to the separation hyperplane; Rifkin & Klautau, 2004). These decision values are compared, and the final classification is assigned to the binary classifier with the maximum decision value. Therefore, for successful classification, the competing decision rules have to differ from each other to avoid persistent ties between multiple decision rules.

For compactness, we use the following notational convention to describe the multi-class classifiers. A

multi-class classification is an ensemble statistical decision that involves the conjoint influence of sample distributions from multiple persons. This combined distribution for 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 task 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 a collection of 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 $^{A}\mathbf{I}_p \circ {}^{A}\mathbf{I}_q \rightarrow {}^{B}\mathbf{I}_r$ (see below for details).

2.8.2 | Accuracy scoring

Even though an ensemble was used for multi-class classification, our focus was on the accuracy of each individual-specific binary classifier in the ensemble. This individualized accuracy a_i of the classifier for person S_i was defined as

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

The hit rate H_i (or positive identification rate) was the proportion of instances where samples from S_i were correctly predicted as being from S_i by the ensemble (i.e., a true positive where the classifier S_i had a larger decision value than the competing classifiers). The correct rejection rate CR_i was defined based on the pairwise relationship of S_i to each of the other classifiers S_j. If the ensemble 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, that is, a false positive. The false positive rate $FP_{i,j}$ denotes the proportion of instances where the ensemble incorrectly predicted that samples from S_j were from S_i. The correct rejection $CR_{i,j}$ was defined as $CR_{i,j} = 1 - FP_{i,j}$. The overall correct rejection CR_i for S_i was defined as the mean of the pairwise 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 accuracy for each classifier was 50% even though random chance for the entire ensemble was $(100/N)\%$. Furthermore, the mean accuracy for a particular classification scheme (e.g., $^{RS1}\mathbf{I}_p \rightarrow {}^{RS1}\mathbf{I}_q$) as used here refers to the mean accuracy of the individual classifiers in the ensemble as calculated by the above procedure.

The accuracy score can have different contributions from the hit rate (e.g., high false negatives) and the correct rejection rate (e.g., high false positives). To disentangle these contributions, we also estimated the *recall* and *precision* scores (Davis & Goadrich, 2006) for each individual S_i as defined below:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}},$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}.$$

The recall score for S_i would be low if samples from S_i are misclassified as belonging to another individual (i.e., high false negatives). However, the precision score for S_i would be low if samples from other individuals are misclassified as belonging to S_i (i.e., high false positives).

Confusion matrices were used to visualize which individuals were misclassified (i.e., confused) with each other. The rows of a confusion matrix represent the true label of a sample and the columns indicate the predicted label for that sample by the ensemble. The value at the row corresponding to S_i and column corresponding to S_j indicated the proportion of samples from S_i that were classified as S_j. The rows/columns of the matrices were re-organized to cluster together individuals who were confused with each other. This was implemented with the so-called Louvain method to maximize modularity (Blondel et al., 2008), implemented in the Community Detection Toolbox (Kehagias, 2021).

2.9 | Evaluation of classifier robustness

The robustness (or generalization) of person identification with a particular classifier was evaluated by the quality of identification decisions on test samples with a *different* origin from the samples used to obtain the classifier (i.e., the training samples). This evaluation was organized into two schemes based on whether the training and test samples belonged to the (i) same day versus a different day (Figure 3a) and the (ii) same task state versus a different task state.

2.9.1 | Same-day identification

Person identification in task state A from samples on a single day p (i.e., $^{A}\mathbf{I}_p \rightarrow {}^{A}\mathbf{I}_p$) was a baseline indicator of whether $^{A}\mathbf{I}_p$ contained individual-defining information in the absence of (i) inter-day differences or (ii) cognitive state differences. Same-day identification accuracy for a particular task state was estimated using a fivefold cross-validation (CV) procedure (Blum et al., 1999). Specifically, the set of samples from state A on 1 day (e.g., day
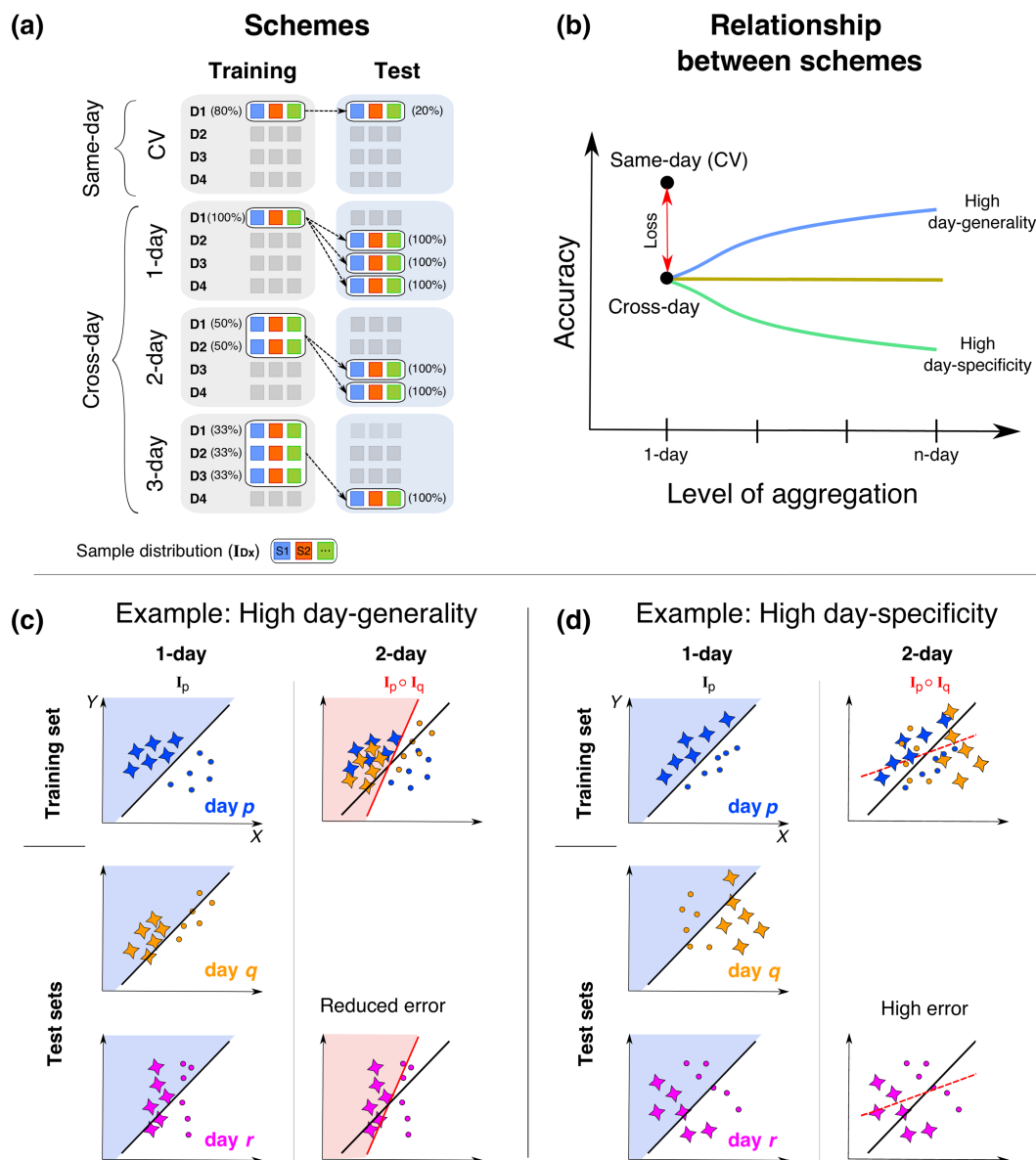
FIGURE 3 Person classification schemes and aggregation. (a) Classification schemes $^{A}\mathbf{I}_{Dx} \rightarrow {}^{A}\mathbf{I}_{Dy}$ (rows) were defined by the configuration of training (left column) and test sets (right column) across days (D_i). The sample distribution ($\mathbf{I}_{Dx}$) included samples from all individuals (multi-coloured boxes). Percentages indicate the proportion of each day samples used for training/testing. Same-day identification was estimated with fivefold cross-validation (CV). Cross-day aggregation training used an equal proportion of samples from each day and the total number of training samples was the same across aggregation levels. (b) 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), or stay constant (dark green), or even decrease (light green). (c) Idealized example of how 1-day cross-day accuracy (Column 1) can increase with aggregation (Column 2). Samples from two classes (stars, circles) are shown along two features (day-general: X, day-specific: Y) with each days samples shown in different colours (p: blue; q: orange; r: purple). The 1-day decision rule ($\mathbf{I}_p$) (top left panel) is depicted with a thick black line and shaded areas. This decision rule accurately classifies samples from days q and r but with errors. However, a decision rule trained on samples from days p and q ($\mathbf{I}_p \circ \mathbf{I}_q$) (thick red line, red shaded area) reduces cross-day classification errors (lower right). (d) 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). Two-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)

D1 in Figure 3a, upper row), were partitioned into five 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 identification accuracy across folds was defined as the same-day identification accuracy for that day (e.g., D1). 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.

2.9.2 | Cross-day identification

To evaluate cross-day identification in task state A ($^A\mathbf{I}_p \rightarrow {}^A\mathbf{I}_q$), a classifier trained on samples from day p was used to identify persons from samples from a different day q. A reduction in cross-day accuracy relative to same-day accuracy (e.g., $^A\mathbf{I}_p \rightarrow {}^A\mathbf{I}_p$) (red arrow, Figure 3b) is an indicator of reduced robustness to day-dependent differences between $^A\mathbf{I}_p$ and $^A\mathbf{I}_q$. In this framing, cross-day accuracy was only interpreted if same-day accuracy was greater than random chance.

We additionally sought to evaluate whether differences in cross-day accuracy relative to same-day accuracy were truly due to 'day'-specific factors. For this purpose, the day-specific properties of the training set were systematically varied (using an *aggregation* procedure) while holding the test set constant. The day specificity of training set was modulated by including samples from different days (e.g., $^A\mathbf{I}_{d1} \circ {}^A\mathbf{I}_{d2}... {}^A\mathbf{I}_{dn} \rightarrow {}^A\mathbf{I}_r$). In an n-day training set, the k training samples per person was an aggregation of k/n samples from each of n different days. Here, n could take the value $= 1$, 2 or 3 (see Figure 3a, first column). The number of samples per person, k, was held constant to enable comparison of classification accuracy across all values of n. Samples in the test set were never aggregated from different days. Mean identification accuracy for a particular n-day aggregation scheme was obtained by (i) independently estimating the accuracy for each possible training/test set combination that satisfied the day constraints (e.g., day $p \neq$ day $q \neq$ day r) and then (ii) averaging these accuracy values.

We assume that increasing aggregation (i.e., by increasing n) would discount day-specific properties in favour of day-general properties during training. Therefore, cross-day accuracy might change with increasing aggregation depending on the relative balance of day-specific versus day-general properties in the samples (Figure 3b). Figure 3c,d shows idealized examples of how aggregation could change cross-day accuracy. In the example shown in Figure 3c, the two classes systematically differ on feature X (x-axis) in a similar manner across days (i.e., X's role has high day generality). However, feature Y (y-axis) has a role in distinguishing the classes on day p but not on other days (i.e., Y's role has low day generality). Therefore, a decision rule trained on day p has a low accuracy in classifying samples from other days (Column 1). However, training on aggregated samples from day p <u>and</u> q (Column 2) discounts the role

of Y in the decision rule, which improves cross-day classification. Figure 3d illustrates an extreme example of day specificity where the classes systematically differ on features X and Y within each day but the relative roles of X and Y differ greatly across days (i.e., high day specificity and low day generality). In this scenario, training a decision rule on aggregated samples from days p and q <u>reduces</u> the accuracy of cross-day classification.

2.9.3 | Cross-task identification

To evaluate cross-task identification of task state A, a classifier trained on samples from day p was used to identify persons from samples from a different task state B <u>on</u> <u>a</u> different day q (i.e., $^A\mathbf{I}_p \rightarrow {}^B\mathbf{I}_q$). Cross-task identification was treated as a special instance of cross-day identification (i.e., training and test samples from different days) to conservatively exclude any inter-task similarities produced by the joint preprocessing of all tasks from the same day. Therefore, a reduction in cross-task accuracy relative to same-day/same-task accuracy (e.g., $^A\mathbf{I}_p \rightarrow {}^A\mathbf{I}_p$) was an indicator of reduced robustness to differences in the task states A and B compounded by inter-day differences. In this framing, cross-task accuracy was only interpreted if cross-day/same-task accuracy was greater than random chance.

2.10 | Weights and normalized weights

Each person's linear decision rule was defined by a configuration of weights assigned to the different features. With these feature weights, a sample could be assigned a decision value as follows:

$$\text{Decision value} = \sum_i w_i P_i - b,$$

where, for each feature i, w_i denotes the feature weight, P_i denotes the sample power, and b is a general bias term.

From this expression, a feature weight that differs from zero irrespective of sign (i.e., $|w_i| > 0$) is an indicator that the corresponding feature was relevant for classification even if only indirectly (Haufe et al., 2014; Schrouff & Mourao-Miranda, 2018). However, the relative contribution of a feature i to the final decision value also depends on $|w_i P_i|$. For features i and j, the weight $|w_i|$ might be greater than $|w_j|$, while the sample power $|P_i|$ might be less than $|P_j|$. Consequently, neither the raw absolute weights nor power is an unambiguous guide to the *relative* influence of features i and j on the 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 $w_i P_i$.

A feature's characteristic weight was obtained by averaging the feature's weight over all decision rules consistent with a particular scheme (e.g., $^A\mathbf{I}_p \rightarrow {}^A\mathbf{I}_q$) (Figure 3a). Similarly, a feature's characteristic power was the grand average of the power across samples and days. The absolute normalized weights (i.e., $|wP|$) were z-scored <u>within</u> each band for each subject to retain information about band-specific, inter-feature weighting differences.

2.11 | Statistical analysis

Statistical tests were performed using the *pingouin* package (version 0.3.2) (Vallat, 2018) and MATLAB Statistics toolbox. 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. Correlations between individual accuracy values were evaluated using Spearman's rank correlation due to the focus on relative ordering rather than a strict cardinal relationship. Tests with a p value < alpha = .05 were deemed to be statistically significant. For tests involving multiple comparisons, p values were evaluated against a Bonferroni-corrected alpha threshold. Due to the sequential relationship between the different multi-class classification schemes, the planned tests on the same-day accuracy (CV) and cross-day accuracy were evaluated at an alpha threshold of .05. However, tests on 2- and 3-day accuracy were evaluated at a threshold of alpha = (.05/2).

For plots depicting mean values at different levels of a single factor, error bars indicate the standard deviation (SD). For plots depicting the effects of multiple factors, error bars displaying the within-subject standard error (s.e.m.) (O'Brien & Cousineau, 2014). The type of error bar used is explicitly noted in the figure caption.

3 | RESULTS

3.1 | Face validity of individual power spectra

Our investigation assumed that an individual's power spectrum at rest can systematically (i) differ between days and also (ii) differ from the spectra of other individuals. We first confirmed the face validity of these assumptions in our data.

The presence of structured inter-individual differences during *RS*1 was qualitatively evident in the mean (full) power spectrum at different channels (Figure 4a) before its reduction to the minimal description used for the classification analyses. As shown for one example individual S_1, individual power spectra had a similar form across channels with a higher power in the δ and α bands and a higher overall power in the posterior and anterior channels relative to the central channels. The diversity of pairwise differences between individual spectra highlights the difficulty of representing an individual's unique properties. For example, the combination of channels and frequencies (i.e., features) at which S_2 and S_3 showed prominent differences were not the same features at which S_2 differed from S_5. Nonetheless, the required decision rule to identify S_2 was a <u>single</u> feature configuration capable of distinguishing S_2 from *all* others while allowing S_2 to be robustly re-identified across days.

Systematic inter-day differences were evident from the dissimilarity between samples from all participants and all days (90 samples per participant per day) (Figure 4b). The dissimilarity between any two samples was described by their correlation distance ($= 1 - r$, where r is the Pearson's correlation coefficient) (Diedrichsen & Kriegeskorte, 2017; Dimsdale-Zucker & Ranganath, 2019; Pani et al., 2020). For all 24 participants, the mean dissimilarity between samples from the *same day* was <u>lower</u> than between samples from different days (*cross-day*) ($t_{23} = -6.74$, $p < .0001$). However, the dissimilarity between same-day and cross-day samples varied from person to person suggesting their possible confusability with samples from other individuals. This was the critical issue to be resolved with an appropriate decision rule, to be identified using machine learning.

3.2 | Individual identification from RS activity within and across days

3.2.1 | High same-day accuracy but reduced cross-day accuracy of individual decision rules

To identify a person from a 2-s 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 $99.98 \pm 0.04\%$ (mean $\pm$ SD) that was significantly larger than the theoretically expected accuracy for random guessing (> 50%: $t_{23} = 5596.13$, $p < .00001$) (Figure 5a, Table A.1). However, for longitudinal tracking, a key
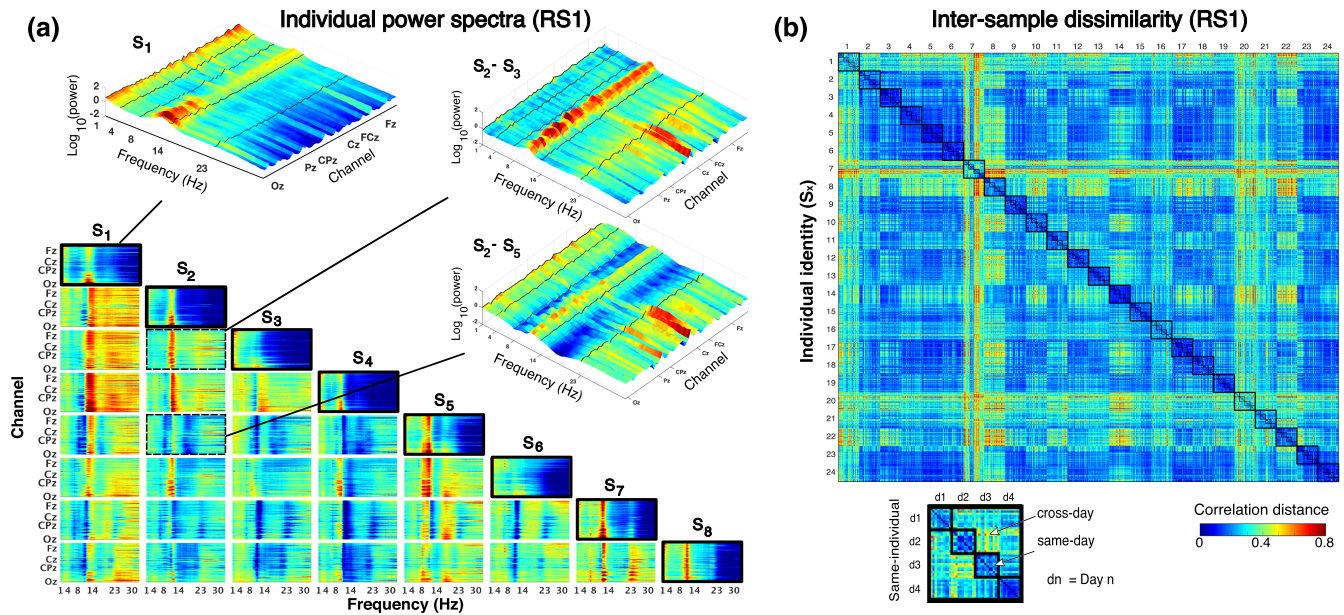
FIGURE 4 Inter-individual and inter-day differences. (a) Matrix showing the (full) power spectra in *RS*1 at all channels (averaged across samples and days) for eight selected individuals (S_i, diagonal, thick black boundary) and their pairwise differences. The difference in power spectra for each pair of individuals S_i and S_j (i.e., S_i and S_j) is shown at row i, column j of matrix. In each spectrogram, channels have a posterior-to-anterior ordering. Insets show magnified view of the power spectrum for S_1 (left upper) and differences for S_2 and S_3 (right upper) and S_2–S_5 (right lower), with frequency band boundaries marked with black lines. (b) Inter-sample dissimilarity matrix for *RS*1 (90 samples per individual per day, each sample was defined by 305 features = 61 channels × 5 bands). The dissimilarity of two samples was defined by their correlation distance (= $1 - r$, where r is the Pearson's correlation coefficient). Large black squares on diagonal contain values from the same individual, and the four smaller squares each contain same-day values

demand is that decision rules from 1 day should identify a person from samples acquired on a different day (i.e., $^{RS1}I_p \rightarrow {}^{RS1}I_q$). The same-day decision rules identified individuals across days with a mean accuracy of 92.10% ± 6.8% that was higher than random chance ($t_{23} = 30.14$, $p < .00001$) but less accurate than same-day identification by ~8% (paired $t_{23} = 5.64$, $p = .00001$).

The confusion matrix (Figure 5b) of how individuals were misclassified during cross-day (1-day) identification revealed four clusters of individuals who were confused with each other. Notably, the individuals with the lowest cross-day accuracies (namely, S_2, S_{11}, S_{15} and S_{24}) belonged to different clusters rather than being solely confused with each other. The clustering of misclassified individuals suggested that errors in identifying an individual S_X were due to a combination of (i) changes to S_X's RS activity between days (i.e., false negatives) and (ii) changes to other individuals who were then misclassified as S_X (i.e., false positives). Nevertheless, the increased errors in individual identification illustrate the challenge of *NP*+/*NP*− decisions. Errors in identifying a person S_X across days seemingly imply that S_X's unique identifying characteristics had changed across days even though the individuals here were unlikely to have changed in their underlying neurophysiology over the 5-day testing period.

3.2.2 | Aggregated training increases cross-day accuracy

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 favour of day-general properties, the decision rules were trained using samples aggregated from *multiple* days (i.e., $^{RS1}I_p \circ {}^{RS1}I_q \dots \rightarrow {}^{RS1}I_s$) (Figure 5a). The mean cross-day accuracy *increased* from 92.10% ± 6.8% without aggregation (1-day) to 95.93 ± 3.63% with 2-day aggregation, with an additional increase to 97.39% ± 2.65% with 3-day aggregation (one-way analysis of variance [ANOVA], $F_{2,46} = 28.83$, $p < .00001$). Following aggregation, the cross-day accuracy was a mere ~2% lower than the same-day accuracy. The effects of aggregated training on individual-specific identification errors are shown in Figure 5c. The decision rules obtained with 3-day aggregation were associated with fewer false negatives (indexed by the higher recall score) especially for individuals with the lowest 1-day accuracies, that is, S_2, S_{11}, S_{15} and S_{24}. This was associated with interrelated changes in errors in individuals who belonged to the same cluster. For example, there was a prominent reduction in false positives (indexed by
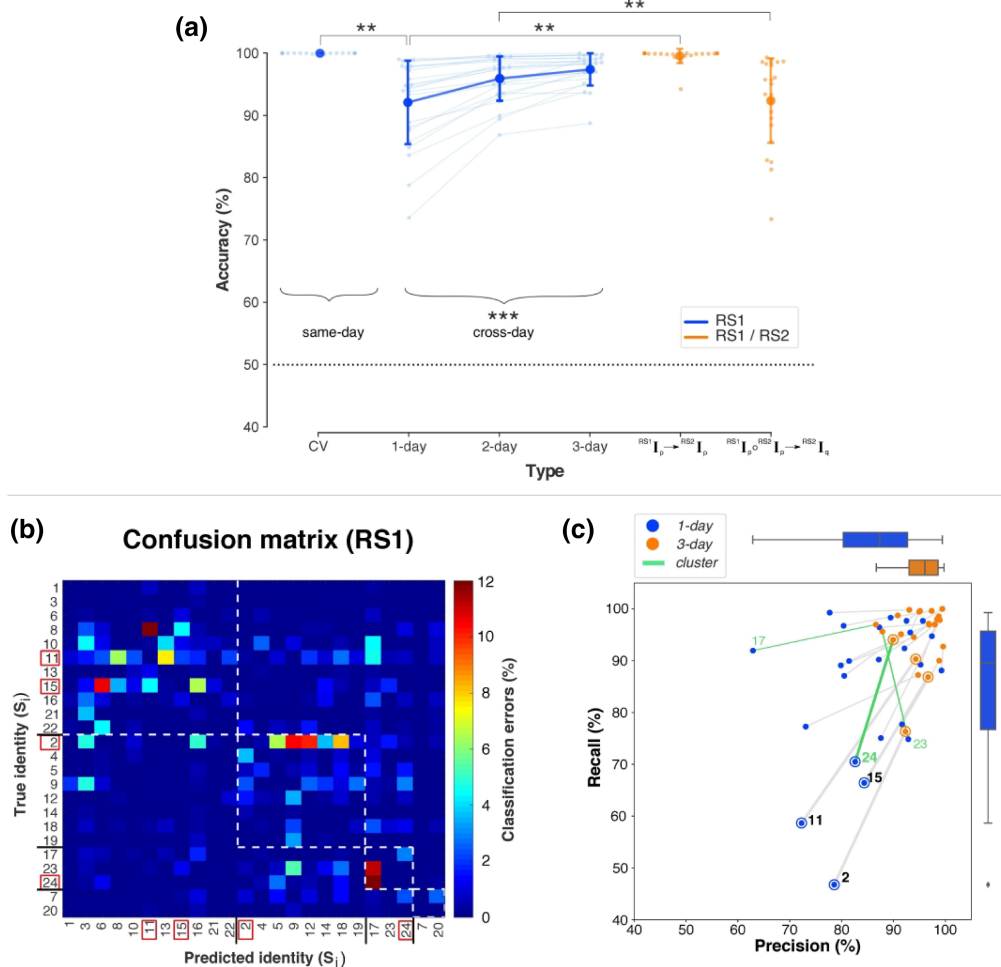
FIGURE 5 Identification accuracy at rest. (a) Mean identification accuracy with *RS*1 (blue) on same-day (cross-validation [CV]), across days (1-day, 2-day, 3-day), and schemes relating *RS*1 and *RS*2 (orange). Light coloured dots/lines depict individual accuracies ($N = 24$). Horizontal dotted line depicts random chance accuracy (50%). Error bars: standard deviation (SD) (** $= .00001 \leq p < .001$; *** $= p < .00001$). (b) Confusion matrix for cross-day (1-day) identification (only errors are shown). Dotted squares indicate clusters of individuals who are more confused with each other. Identities of individuals with the lowest cross-day accuracies are highlighted with red squares. (c) Changes to precision and recall scores with aggregation for the whole group (shown with boxplots) and for individuals (1-day: blue dots; 3-day: orange dots). Individuals with lowest 1-day accuracy are indicated with rings and thick grey lines. Green lines highlight S_{24}, S_{23} and S_{17} who belong to the same cluster, shown in (b)

the higher precision score) for S_{17} who was in the same cluster S_{24} and S_{23} (highlighted in green). The increased accuracy with aggregation despite the true inter-day differences in RS activity was consistent with the presence of *day-general* properties (Figure 3).

3.2.3 | Cross-day and cross-measurement identification are not equivalent

We next assessed whether the above accuracy relationships across days (with and without aggregation) was related to a difference in *days* rather than simply a difference in *measurements*.

In our experimental protocol (Figure 2a), *RS*2 was the second RS measurement on each day. The effects of aggregation on cross-day identification with *RS*1 were successfully replicated on *RS*2 without statistically detectable differences (Table A.1) (two-way ANOVA, condition [*RS*1, *RS*2] × type [1-day, 2-day, 3-day], type * condition: $F_{2, 46} = 0.56$, $p = .57$; type: $F_{2, 46} = 31.31$, $p < .00001$; condition: $F_{1, 23} = 0.38$, $p = .54$]. Importantly, *RS*2 validated the *day-specific* properties of the decision rules (Figure 5a). Same-day decision rules from *RS*1 classified samples of *RS*2 from the same day ($^{RS1}\mathbf{I}_p \rightarrow {}^{RS2}\mathbf{I}_p$) with a mean accuracy of $99.55 \pm 1.15\%$ that was significantly greater than the accuracy in classifying *RS*1 across days ($^{RS1}\mathbf{I}_p \rightarrow {}^{RS1}\mathbf{I}_q$) ($92.10 \pm 6.84\%$) (paired $t_{23} = 5.19$,

$p = .00003$). Furthermore, *RS*2 validated the importance of aggregating samples from different days (rather than different measurements) to reduce day specificity. Decision rules trained on aggregated same-day samples from *RS*1 and *RS*2 ($^{RS1}\mathbf{I}_p \circ {}^{RS2}\mathbf{I}_p \to {}^{RS1}\mathbf{I}_r$) had a *lower* cross-day accuracy ($92.38 \pm 6.92\%$) than decision rules trained on aggregated *RS*1 samples from two different days ($^{RS1}\mathbf{I}_p \circ {}^{RS1}\mathbf{I}_q \to {}^{RS1}\mathbf{I}_r$) ($95.93 \pm 3.63\%$) (paired $t_{23} = -4.83, p = .00007$).

In summary, the reduction in cross-day accuracy without aggregation was indicative of *inter-day* (rather than inter-measurement) variations in RS activity. Despite this inter-day variation in RS activity, the cross-day accuracy increased with aggregation revealing the existence of *day-general* properties in RS activity that were unchanged across days. These properties were consistent with an activity configuration that was putatively defined by individual-specific neurophysiological constraints.

3.3 | Information organization in RS activity for individual identification

The hypothesized configuration in RS activity was suggestive of a multivariate relationship between distributed features. However, the accuracy relationships described above do not indicate whether such a distributed configuration was necessary to enable individual identification. Therefore, we evaluated the information organization required for individual identification.

### 3.3.1 | 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 × 61 channels). To test 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 A.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] × band [B_δ, B_θ, B_α, $B_{\beta1}$, $B_{\beta2}$], type * band: $F_{4,92} = 24.83$, $p < .00001$; type:

$F_{1,23} = 232.11$, $p < .00001$; band: $F_{4,\ 92} = 40.30$, $p < .00001$). The divergence in cross-day losses for B_α and B_δ was striking as these two bands have 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, $B_{\beta1}$, $B_{\beta2}$] × type [1-day, 2-day, 3-day], type * band: $F_{8,\ 184} = 9.19$, $p < .00001$; type: $F_{2,\ 46} = 146.02$, $p < .00001$'; band: $F_{4,\ 92} = 43.13$, $p < .00001$). 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.

Each *mono-location* feature set (L_z) consisted of 50 features (5 bands × 10 channels) in the spatial zone z (Figure 2a). The mean same-day accuracy was greater than 95% for all mono-location feature sets (Figure 6c, Table A.2). However, the mean cross-day (1-day) accuracy showed reductions of ~12%–16% for all locations (ANOVA, type [CV, 1-day] × location [L_F, L_{FC}, L_{CP}, L_{PO}], type * location: $F_{3,69} = 3.77$, $p = .015$; type: $F_{1,23} = 108.91$, $p < .00001$'; location: $F_{3,\ 69} = 5.45$, $p = .0020$]. 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 4a). Aggregation increased cross-day accuracy by ~6% for all four location sets (Figure 6d) (ANOVA, location: [L_F, L_{FC}, L_{CP}, L_{PO}] × type [1-day, 2-day, 3-day], type * location: $F_{6,\ 138} = 2.07$, $p = .06$; type: $F_{2,\ 46} = 115.38$, $p < .00001$; location: $F_{3,\ 69} = 4.79$, $p = .0043$). Nevertheless, the residual ~7%–10% loss in cross-day accuracy was larger than with the full feature-set.

In summary, all the mono-band and mono-location sets contained sufficient information to enable same-day identification with nearly error-free accuracy. However, this information had a low day generality. Even with aggregation, these feature sets had a lower cross-day accuracy than the full feature-set that combined these feature sets together. This divergence suggests that the higher cross-day robustness with the full feature set involves a role for multivariate 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 this multi-feature configuration might be organized, we evaluated the pattern of weights associated with the different features of the full feature set.
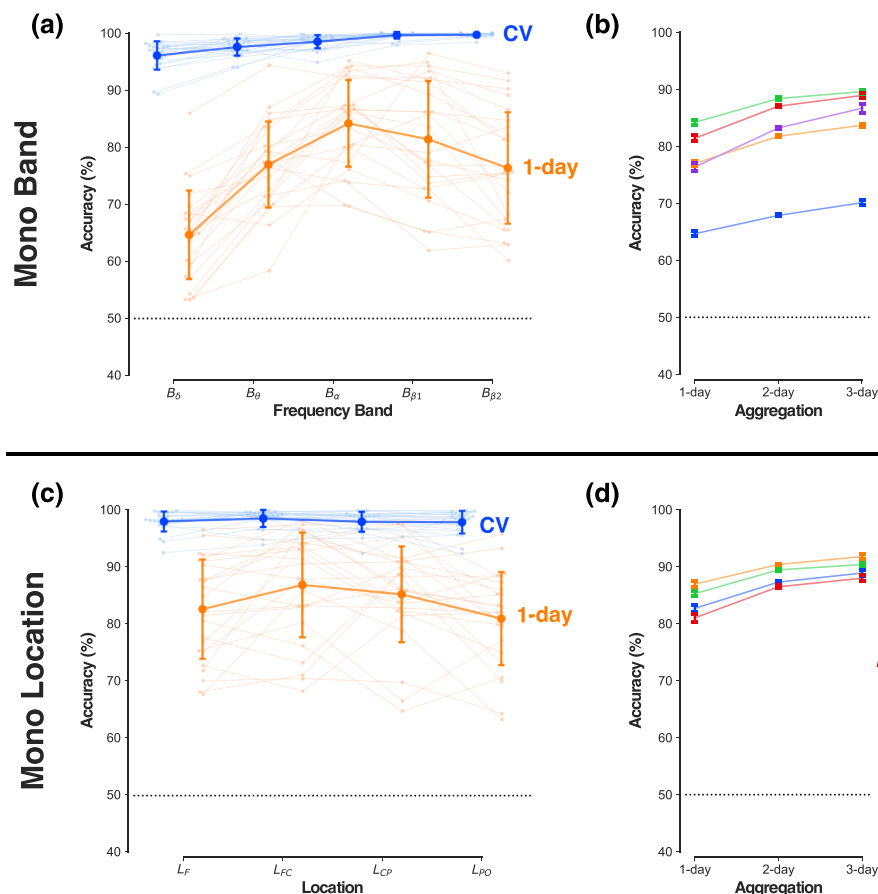
FIGURE 6 Identification at rest with mono-band/location feature subsets. (a) Mean identification accuracy for *RS*1 with mono-band feature sets of increasing frequency (*x*-axis) on the same-day (blue, CV) and across-days (orange, 1-day). Light-coloured dots/lines depict individual accuracies (*N* = 24). Error bars: standard deviation (SD). (b) Change in cross-day identification with increasing aggregation (*x*-axis) for different mono-band feature subsets (coloured lines). Error bars: Within-subject s.e.m. (O'Brien & Cousineau, 2014). (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 (coloured lines). Error bars: Within-subject s.e.m. Horizontal dotted lines depicts the random chance accuracy (50%) in all panels

3.3.2 | Concentration of high-consistency informative features at fronto-central and occipital zones

Each individual's decision rule was defined by the configuration of weights assigned to the different features. Because a decision rule uniquely identifies a person, the feature weights that define a person's decision rule would need to be different for that of all others. Despite these weighting differences, certain features might nevertheless be informative across individuals. To identify these features, we evaluated the inter-individual consistency in the influence of different features on the identification decision. A feature's influence on the individual's decision rule was quantified by the feature's normalized weight (to correct for inter-feature power differences) that was then *z*-scored (within each frequency band) to retain inter-feature relevance differences.

Figure 7 shows the topographic distribution of these high-consistency features of the full feature-set with mean normalized, *z*-scored weights (averaged across individuals) that were significantly greater than zero (see Figure S1 for corresponding non-normalized [raw] weights). At corrected thresholds (see *t* values in Figure 7, lower panels), the features associated with all frequency bands except the δ band contained at least one high-consistency feature. Rather than having an idiosyncratic organization, the high-consistency features were concentrated at distinctive zones in each frequency band.

In B_θ, there was a concentration of high consistency features at CP1 and C3, with the addition of CP3 with aggregation. There was a similar, although weaker, concentration of consistent features at corresponding channels over the right hemisphere. Showing a similar spatial organization, the high-consistency features in $B_{\beta1}$ showed a striking bilaterally symmetric configuration along the
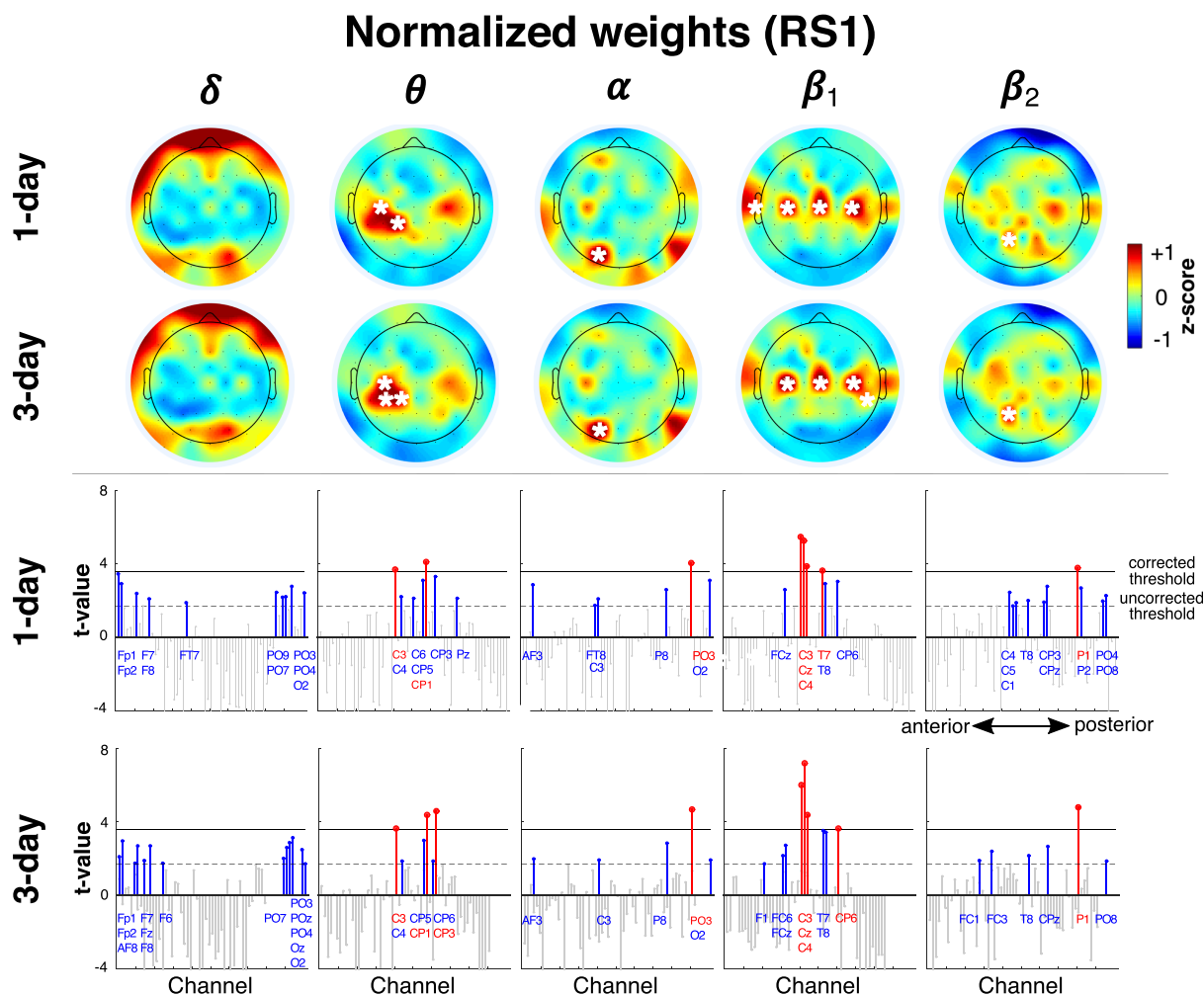
FIGURE 7 High-consistency features. Spatial distribution of high-consistency normalized weights for frequency bands of full feature set (z-scored per band) and their changes with aggregation (1-day, 3-day). Mean weights in each scalp map that were significantly greater than zero are indicated with a white asterisk ($p < .05/61$). Lower two rows show t values for the features corresponding to the upper rows. Channels have an anterior-to-posterior ordering (x-axis). Red stems indicate channels with t values higher than the corrected threshold ($p < .05/61$, horizontal black line), while blue stems show channels that only pass uncorrected thresholds ($p < .05$, dotted horizontal line). Coloured channel labels are grouped from top-to-bottom for visibility and correspond to stems from left to right

transverse midline at channels C3, Cz and C4 with an aggregation-modulated role for CP6 and T7 (and possibly T8). This similarity in organization was notable since the frequency ranges of the θ band (4–7.5 Hz) and β_1 (14–22.5 Hz) were not contiguous and were separated by the α band.

Unlike this central concentration of features in $B_{\beta1}$ and B_θ, the features in B_α contained a single, strongly consistent feature in the occipital zone at PO3. At uncorrected thresholds, there were other distributed features across the scalp that were weakly consistent for both 1- and 3-day identification, namely, at AF3, C3, P8 and O2. Similarly, the features of the high-frequency β_2 band (i.e., $B_{\beta2}$) only had a single consistent feature at P1 with a diffuse distribution of consistent features at uncorrected thresholds.

In general, the distribution of high-consistency features was by itself not a simple indicator of their contribution to cross-day accuracy. For example, the relative number of high-valued (raw) weights in the different bands and spatial locations had a low correspondence to relative accuracy of cross-day identification based solely on the mono-band/location subsets (see Figure S2). Nevertheless, the organized distribution of high-consistency features at channels over the sensorimotor cortex and the occipital cortex was prima facie support for an individual-specific configuration with a basis in neurophysiological constraints. These high-consistency zones were of particular relevance to the relationship of $RS1$ to the non-rest task states where the power over the sensorimotor and occipital zones was expected to differ from $RS1$.

3.4 | Relationship of rest to non-rest states

The behavioural demands during *TapMov* and *SeqMov* were designed to modulate the cognitive states during the *TapWait* and *SeqWait* periods and produce neural activity deviations from *RS*1 in the absence of behavioural 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). We sought to first explicitly verify that such deviations from *RS*1 were indeed present. Note that all analyses of *Tap* and *Seq* states were performed in a subgroup of $N = 18$ participants (see Section 2).

3.4.1 | Neural activity during *Tap* and *Sequence* verifiably deviates from *RS*1

The inter-day changes in behaviour during the *TapMov* and *SeqMov* periods were consistent with the experimental assumptions (Figure 8a). During *TapMov*, the mean number of button presses during the cued 2-s period ($\sim$10–11) remained effectively constant across days (one-way ANOVA, $F_{4, 68} = 0.50$, $p = .73$). In contrast, during *SeqMov*, the mean number of button presses increased from $\sim$7 on the first day to $\sim$11 on the fifth day (one-way ANOVA, $F_{4, 68} = 40.75$, $p < .00001$). 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
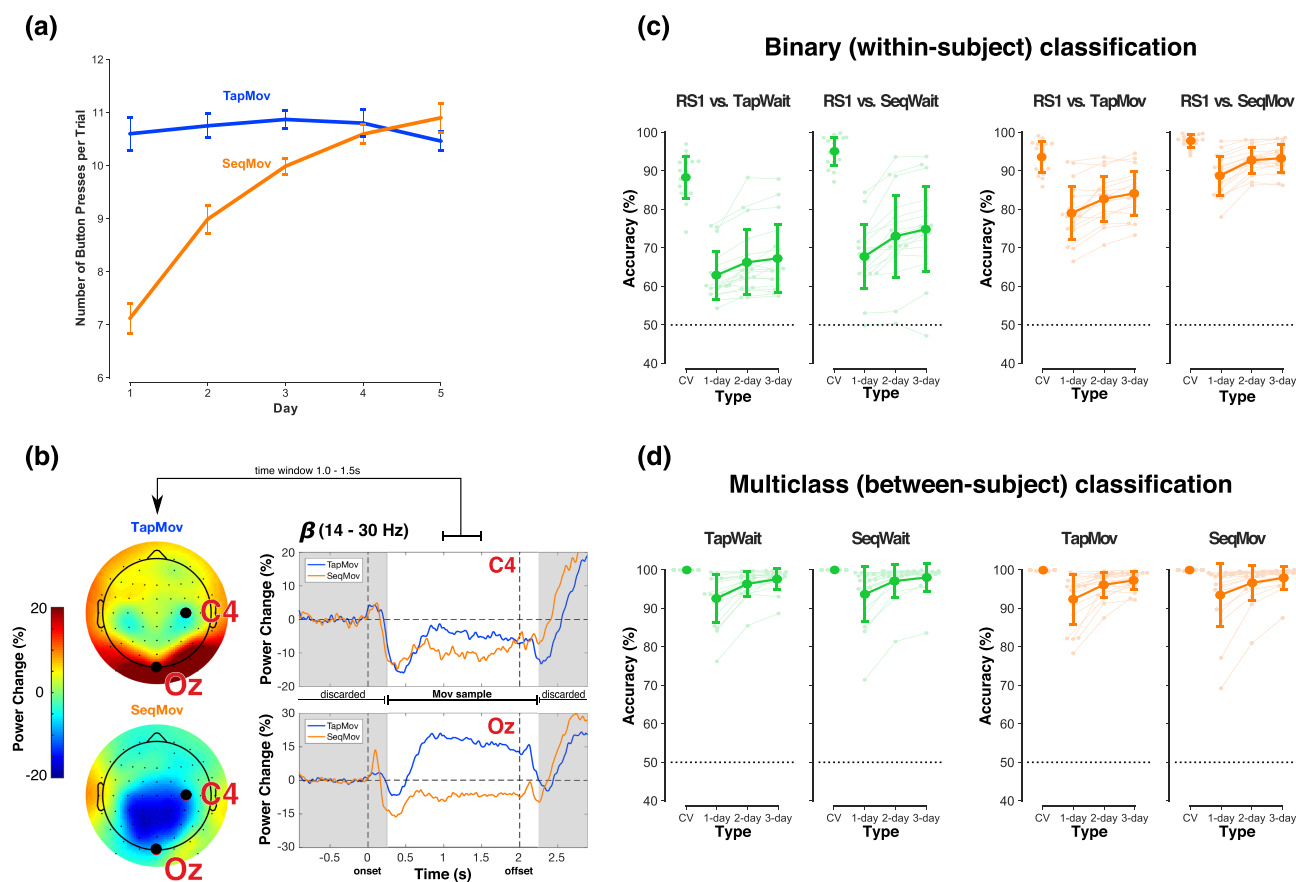
FIGURE 8 Difference between *RS*1 and task states. (a) Change in behaviour across days indexed by mean number of button-presses per response period on each day (*x*-axis) during *TapMov* (blue) and *SeqMov* (orange). Error bars: Within-subject s.e.m. (b) Movement-related power dynamics in the β band (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 grey 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 $[+1\ s, +1.5\ s]$ following onset of the movement cue. (c) Same-day/cross-day accuracy in distinguishing *RS*1 versus pseudo-rest states (green) and *RS*1 versus movement states (orange) using within-subject binary classifiers. Cross-day differences to *RS*1 were lowest for *TapWait* (far left) and highest for *SeqMov* (far right). Error bars: standard deviation (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

(condition [*TapMov*, *SeqMov*] × days [D1,..., D5]; condition * days: $F_{4, 68} = 21.00$, $p < .00001$; condition: $F_{1, 17} = 5.73$, $p = .03$; days: $F_{4, 68} = 21.53$, $p < .00001$).

The neural state during the movement period (*TapMov*, *SeqMov*) showed typically expected dynamic states (Figure 8b, Supplementary Methods 2). 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 (Alegre et al., 2004; Cassim et al., 2000; Erbil & Ungan, 2007; Pfurtscheller & da Silva, 1999), 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* versus *TapMov* had a mean cross-validated accuracy of $85.91 \pm 7.23\%$ (>50%: $t_{17} = 21.06$, $p < .00001$); and *SeqWait* versus *SeqMov* had a mean CV accuracy of $94.58 \pm 3.20\%$ (>50%: $t_{17} = 59.02$, $p < .00001$).

The critical verification for our study was the relationship between *RS*1 and the pseudo-rest states (*TapWait*, *SeqWait*). Samples from *TapWait* and *SeqWait* were distinguishable from *RS*1 on the same day with high cross-validated accuracy (*RS*1 vs. *TapWait*: $88.28 \pm 5.70\%$; *RS*1 vs. *SeqWait*: $95.12 \pm 3.74\%$) (Figure 8c, left panels, Table A.3). However, the cross-day accuracy (without aggregation) for both *RS*1 versus *TapWait* ($62.91 \pm 6.44\%$) and *RS*1 versus *SeqWait* ($67.79 \pm 8.53\%$) was substantially lower than the same-day accuracy by more than ~25%. Nevertheless, the cross-day accuracy for *RS*1 versus *SeqWait* was marginally higher than for *RS*1 versus *TapWait* with increasing aggregation (ANOVA: condition [*RS*1 vs. *TapWait*, *RS*1 vs. *SeqWait*] × type [1-day, 2-day, 3-day], condition * type: $F_{2, 34} = 6.22$, $p = .005$; condition: $F_{1, 17} = 8.37$, $p = .01009$; type: $F_{2, 34} = 38.89$, $p < .00001$).

TapMov and *SeqMov* were also distinguishable from *RS*1 on the same-day with high (cross-validated) accuracy (*RS*1 vs. *TapMov:* $93.56 \pm 4.12\%$; *RS*1 vs. *SeqMov*: $97.81 \pm 1.76\%$) (Figure 8c, right panel, Table A.3). Similar to the wait periods, the cross-day accuracy for *RS*1 versus *SeqMov* was higher than for *RS1* versus *TapMov* across aggregation levels (ANOVA: condition [RS vs. *TapMov*, *RS*1 vs. *SeqMov*] × type [1-day, 2-day, 3-day], condition * type: $F_{2, 34} = 0.61$, $p = .55$; condition: $F_{1, 17} = 30.91$, $p = .00003$; type: $F_{2, 34} = 69.47$, $p < .00001$).

The above findings verified the neural activity differences in the task states in *Tap* and *Sequence* to each other

and to *RS*1. Crucially, the structure of the same-day differences had a low cross-day generality.

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

The above differences between task states and *RS*1 raised the issue of whether the task-related functional states also disrupt the information that enables individual identification with *RS*1. To assess this possibility, we evaluated whether the different *Tap* and *Sequence* task states contained sufficient information for person identification in a same-task classification scheme (i.e., with the scheme $^{X}\mathbf{I}_p \rightarrow {^{X}\mathbf{I}_q}$ for task X) (Figure 8d).

The same-day accuracy for both *TapWait* and *SeqWait* was ~99% (Figure 8d, left panels, Table A.1). The mean cross-day accuracy (without aggregation) for *TapWait* ($92.58 \pm 6.39\%$) was lower than its corresponding same-day accuracy by only ~7% ($t_{17} = 4.92$, $p = .00013$). Similarly, for *SeqWait*, the mean cross-day (1-day) ($93.67 \pm 7.35\%$) accuracy was lower than the same-day accuracy by ~6% ($t_{17} = 3.65$, $p = .00197$). Furthermore, the effect of aggregation on mean cross-day accuracy for *TapWait* and for *SeqWait* was statistically indistinguishable (ANOVA: condition [*TapWait*, *SeqWait*] × type [1-day, 2-day, 3-day]; condition * type: $F_{2, 34} = 0.88$, $p = .42$; condition: $F_{1,17} = 1.35$, $p = .26$; type: $F_{2, 34} = 21.30$, $p < .00001$).

Despite the deviations of *TapMov* and *SeqMov* along both the behavioural 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 A.1, Figure 8d, right panels) (ANOVA: condition [*TapMov*, *SeqMov*] × type [1-day, 2-day, 3-day]; condition * rype: $F_{2, 34} = 0.86$, $p = .43$; condition: $F_{1, 17} = 1.26$, $p = .28$; type: $F_{2, 34} = 14.50$, $p = .00003$).

Thus, individual identification was robustly possible in the task states despite their differences to *RS*1. Furthermore, the identification accuracy was similar between the *Tap* and *Seq* states despite their functional differences. Two further lines of evidence supported the possibility that these similarities were based on common task-independent properties. The spatial distribution of high-consistency features for these states (Figure 9a, Figure S3) exhibited a striking qualitative similarity to each other as well as to the corresponding distribution for *RS*1 (Figure 7). Additionally, the individual cross-day (1-day) accuracy in these task states showed a striking correlation to the corresponding cross-day accuracy in *RS*1 (Figure 9b) (threshold: $p < .05/4$; *TapWait*: $r[17]$

(a)

Normalized weights (1-day)

(b)

Same-task correlation (1-day)
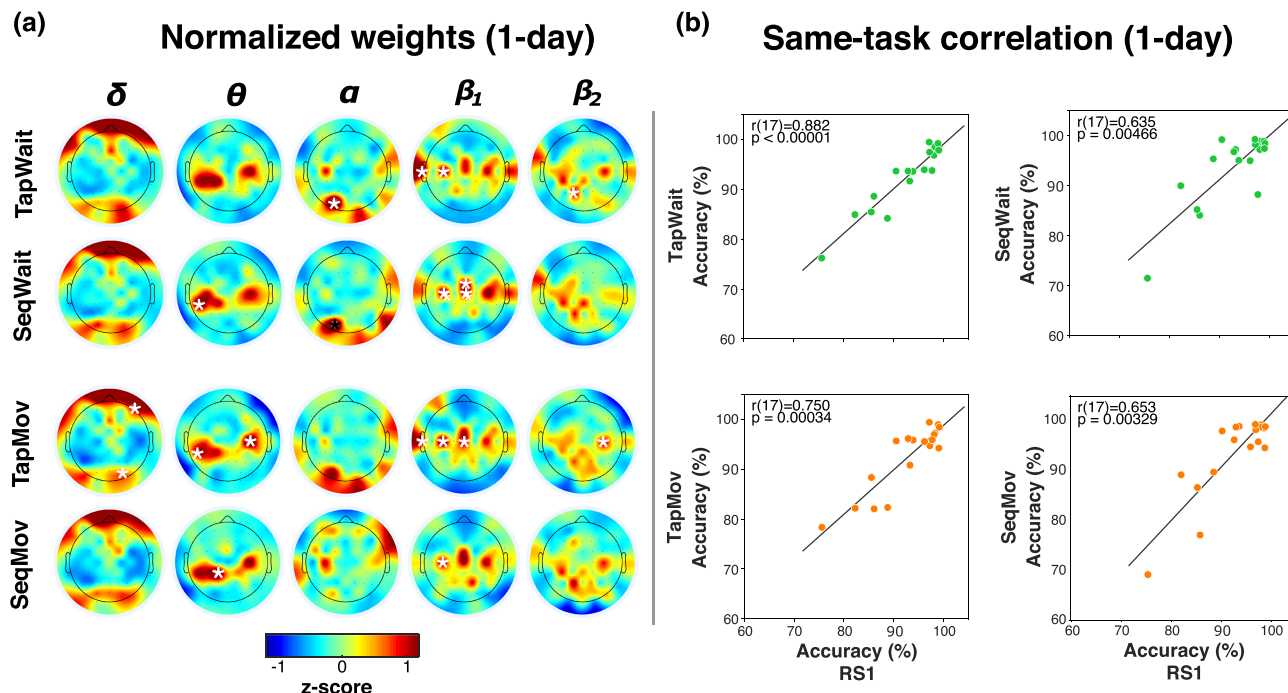
FIGURE 9 Inter-task relationships. (a) Spatial distribution of high-consistency features in different task states (absolute, z-scored) for 1-day decision rules (without aggregation). Weights in each scalp map that were significantly greater than zero are indicated with a white asterisk ($p < 0.05/61$, see Figure S3). 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 7). (b) Scatter plots of cross-day (1-day) identification accuracy in RS1 to the corresponding same-task accuracy in the pseudo-rest states (upper row) and movement states (lower row). Each dot represents one individual. correlations were assessed with Spearman's rank order correlation (threshold: $p < .05/4$)

$= .882$, $p < .00001$; *SeqWait*: $r[17] = .635$, $p = .00466$; *TapMov*: $r[17] = 0.75$, $p = .00034$; *SeqMov*: $r[17] = .653$, $p = .00329$). Thus, the inter-individual relationships revealed by the errors in cross-day classification during RS1 (Figure 5b) seemingly extended to these non-rest states as well. We next turned to a formal assessment of this cross-task relationship.

3.5 | Robust generalization of rest-based decisions to cross-task individual identification

If person identification with RS1 was based on a neural configuration 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 A.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 ($RS1$: $92.79 \pm 6.76\%$, *TapWait*: $91.90 \pm 6.46\%$; *SeqWait*: $90.81 \pm 7.09\%$) (one-way ANOVA, $F_{2, 34} = 2.06$, $p = .14$). However, increasing deviations from RS1 due to cognitive <u>and</u> behavioural state differences ($X = [RS1, TapMov, SeqMov]$) produced significant reductions in identification accuracy most notably for *SeqMov* (*TapMov*: $88.79 \pm 7.57\%$; *SeqMov*: $83.85 \pm 10.35\%$) (one-way ANOVA, $F_{2, 34} = 14.07$, $p = .00004$).

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 (Tables A.4 and A.1 respectively). 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*] × condition [*TapWait*, *SeqWait*], train * condition: $F_{1,17} = 4.14$, $p = .06$; train: $F_{1,17} = 10.02$, $p = .00566$; condition: $F_{1, 17} = .00001$, $p = 1.00$). 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*] × condition [*TapMov*, *SeqMov*],

(a) Cross-task (1-day) identification

(b)

(c)
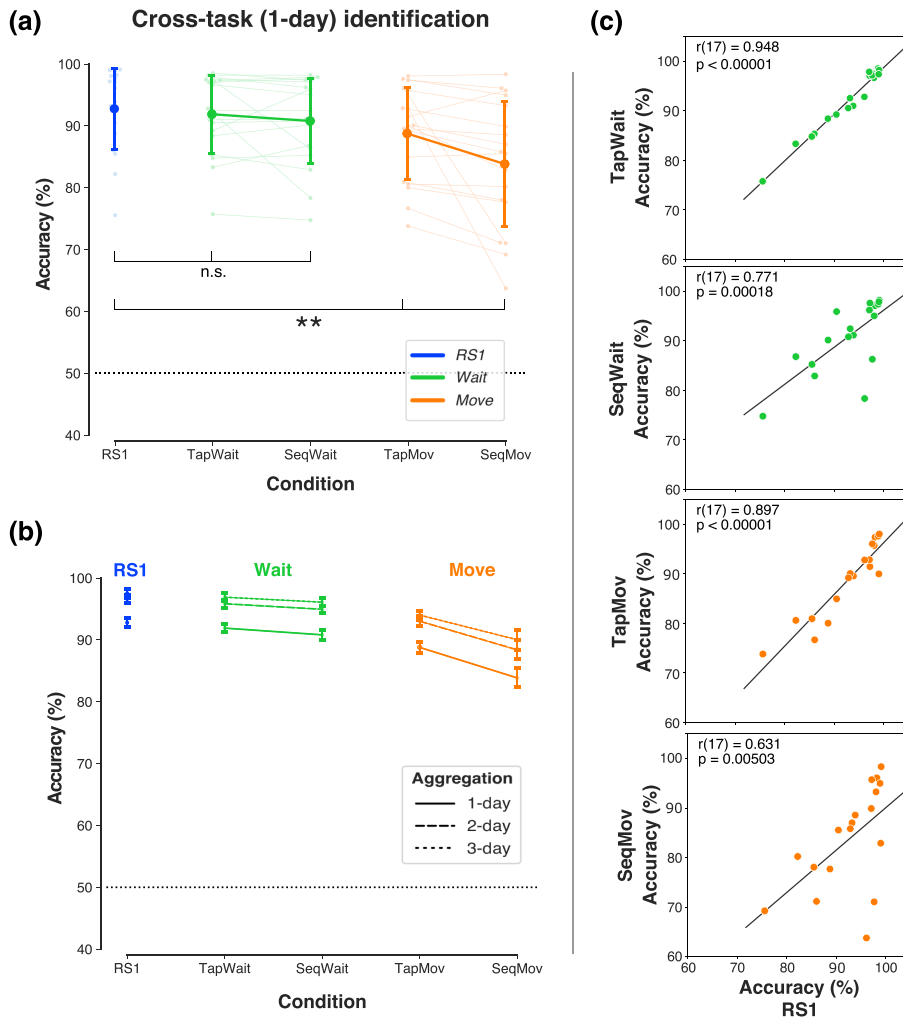
FIGURE 10 Cross-task identification with *RS*1. (a) Mean accuracy of decision rules trained on *RS*1 (1-day) and tested across days on *RS*1 (blue), pseudo-rest states (green) and movement states (orange). Light coloured dots/lines indicate individual accuracies. Error bars: standard deviation (SD). Accuracy differences between *RS*1 and pseudo-rest states were not statistically significant (n.s.), but were between *RS*1 and movement states (** = .00001 ≤ *p* < .001). (b) Decision rules trained on *RS*1 with different levels with aggregation (dotted lines) increased cross-day accuracy for all task states. Error bars: Within-subject s.e.m. (c) Scatter plots of cross-day (1-day) accuracy in *RS*1 to the corresponding cross-task accuracy in all non-rest tasks. Each dot represents one individual. Correlations were assessed with Spearman's rank order correlation (threshold: *p* < .05/4)

train * condition: $F_{1,17} = 9.15$, $p = .00764$; train: $F_{1,17} = 43.94$, $p < .00001$; condition: $F_{1, 17} = 2.51$, $p = .13$).

To disentangle the role of day specificity in $^{RS1}\mathbf{I}_p \rightarrow {}^{X}\mathbf{I}_q$, we used multi-day aggregation ($^{RS1}\mathbf{I}_p \circ {}^{RS1}\mathbf{I}_q$ ... $\rightarrow {}^{X}\mathbf{I}_r$). Although aggregation reduced day specificity with *RS*1 (Figure 5), this could have been achieved by increasing specificity to the properties of *RS*1. Such an 'overfitting' to *RS*1 (i.e., increased task specificity) might lower the accuracy of cross-task identification. Alternatively, aggregation could have reduced both day and task specificities and thus increase the accuracy of cross-task identification. Consistent with this latter 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 [*RS*1, *TapWait*, *SeqWait*] × type [1-day, 2-day, 3-day]; condition * type: $F_{4, 68} = 0.52$, $p = .72$; condition: $F_{2, 34} = 2.44$, $p = .10$; type: $F_{2, 34} = 21.63$, $p < .00001$). 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/cross-day accuracy loss relative to same-task/cross-day identification with *RS*1 was only ~3%. Aggregation also *increased* cross-task accuracy to the movement states (*TapMov*, *SeqMov*) (ANOVA: condition [*RS*1, *TapMov*, *SeqMov*] × type [1-day, 2-day, 3-day]; condition * type: $F_{4, 68} = 1.35$, $p = .26$; condition: $F_{2, 34} = 13.04$, $p = .00006$; type: $F_{2, 34} = 29.33$, $p < .00001$). Following aggregation, the mean residual cross-task/cross-day difference was less than ~10% for the movement states.

Similar to the same-task correlations described above (Figure 9b), the individual *cross-task* (1-day) accuracy in each of these task states showed a statistical significant correlation to the corresponding cross-day accuracy in *RS*1 (Figure 10c) (threshold: *p* < .05/4; *TapWait*: *r*[17] = .948, *p* < .00001; *SeqWait*: *r*[17] = .771, *p* = .00018; *TapMov*: *r*[17] = .897, *p* < .00001; *SeqMov*: *r*[17] = .631, *p* = .00503). The correlation coefficients were particularly high for both *Tap* states (*TapWait* and *TapMov*) as compared with the *Seq* states (*SeqWait* and *SeqMov*).

Furthermore, the scatter plots suggested that the relatively lower cross-task/cross-day accuracy for *SeqMov* was driven by the low generalization of a few individuals.

In summary, decision rules trained on *RS*1 on a single day could identify individuals from samples from states that verifiably differed from *RS*1 to differing extents. Importantly, aggregated training solely on *RS*1 lead to *increases* in identification accuracy on samples from these non-rest task states. Unlike the full-feature set, applying the cross-task scheme $^{RS1}\mathbf{I}_p \rightarrow {}^{X}\mathbf{I}_q$ to the mono-band and mono-location feature sets produced large accuracy reductions (Figure S4). This confirmed that the conjoint role of features from more than one frequency band and spatial zone was crucial to obtain high cross-task/cross-day identification accuracy. Taken together, the cross-task/cross-day robustness of person identification with a distributed feature-set was consistent with the properties of a configuration constrained by individual neurophysiology, that is, a critical demand for change classification.

4 | DISCUSSION

Elucidating how neural oscillatory dynamics during 'task free' rest reveal individual-specific neurophysiological organization is an important objective for cognitive neuroscience. However, despite the extended analysis of RS-EEG power and individual differences here, the narrow motivation for this study was an analytical (rather than neuroscientific) challenge posed by RS based tracking, namely, robust and individualized inferences about inter-day change that screen out irrelevant cognitive variation. With this analytical objective, we contribute a novel formulation of this general issue, namely, change classification, and a possible solution approach that leverages the individual distinctiveness of RS activity, where decision rules for person identification serve as a tool for cross-day change classification. Consistent with the goal of robust change classification despite cognitive variability, these decision rules were capable of cross-day/cross-task person identification with a low cross-day loss (Figures 5 and 10), despite inter-day cognitive variation of different magnitudes (Figure 8). Rather than being idiosyncratic effects, the information represented by the decision rules demonstrated robust day-general characteristics under aggregation (Figures 3, 5 and 10) and an organization suggesting a basis in individual neurophysiology (Figures 6, 7 and 9). These results provide a proof of concept that RS change classification might be addressable in a general manner, to complement approaches based on domain-specific RS biomarkers (Hohenfeld et al., 2018; Rashid & Calhoun, 2020; Woo et al., 2017).

Change classification requires a decision rule to translate RS activity into an *NP+/NP−* decision but does not specify how a suitable decision rule is to be obtained. Therefore, suitable decision rules could conceivably be obtained with other approaches that do not involve either person identification or machine learning (i.e., inductive inferences). It is hence worth considering when person identification might (and might not) be a relevant strategy. We suggest below that (i) change classification requires a *change-informed representation* and (ii) person identification might be particularly relevant when such a representation is not known a priori.

4.1 | Change evaluation versus change-informed representation

A plausible alternative approach to change classification is to directly *evaluate* inter-day RS activity, for instance, with a dissimilarity measure (as in Figure 4b) or an assessment of inter-day variability within a test–retest reliability framework (Bijsterbosch et al., 2017; Cox et al., 2018; Noble et al., 2019; Postema et al., 2019). The obtained score could then be converted into a change classification decision, for example, a suitably high inter-day reliability (or similarity) might suggest an *NP−* categorization (i.e., no neurophysiological change), whereas a low reliability (or similarity) would suggest an *NP+* categorization. However, such evaluation-based approaches would be incomplete without specifying how individual RS activity is to be *represented* for this evaluation, as illustrated by the following analogy to object recognition.

Consider images depicting the same object *X* from Days 1 and 2 (Figure 11a) represented by filled pixel locations (i.e., features). A simple measure of inter-day reliability is whether the filled pixels on Day 1 are also reliably filled on Day 2. Scenario *A* would have a *high* feature-level reliability as most filled pixels on Day 1 are also filled on Day 2 (suggesting *NP−*), whereas Scenario *B* would have a *low* reliability (suggesting *NP+*). However, these change classification decisions are faulty inferences about the overall shape of the depicted object. In Scenario *A*, the object's shape differs between days (i.e., analogous to *NP+*), although in Scenario *B*, the object's shape is unchanged despite a large change in orientation (i.e., analogous to *NP−* as with *TapWait*, *SeqWait*, *TapMov*, *SeqMov*). The faulty inference can be attributed to the representation of the object (i.e., list of filled pixel locations) as its format is uninformed by the possible types of change. However, an ideal *change-informed* representation of object *X*'s shape would be invariant to irrelevant rotations as in Scenario *B* (analogous to cognitive state variation), while being
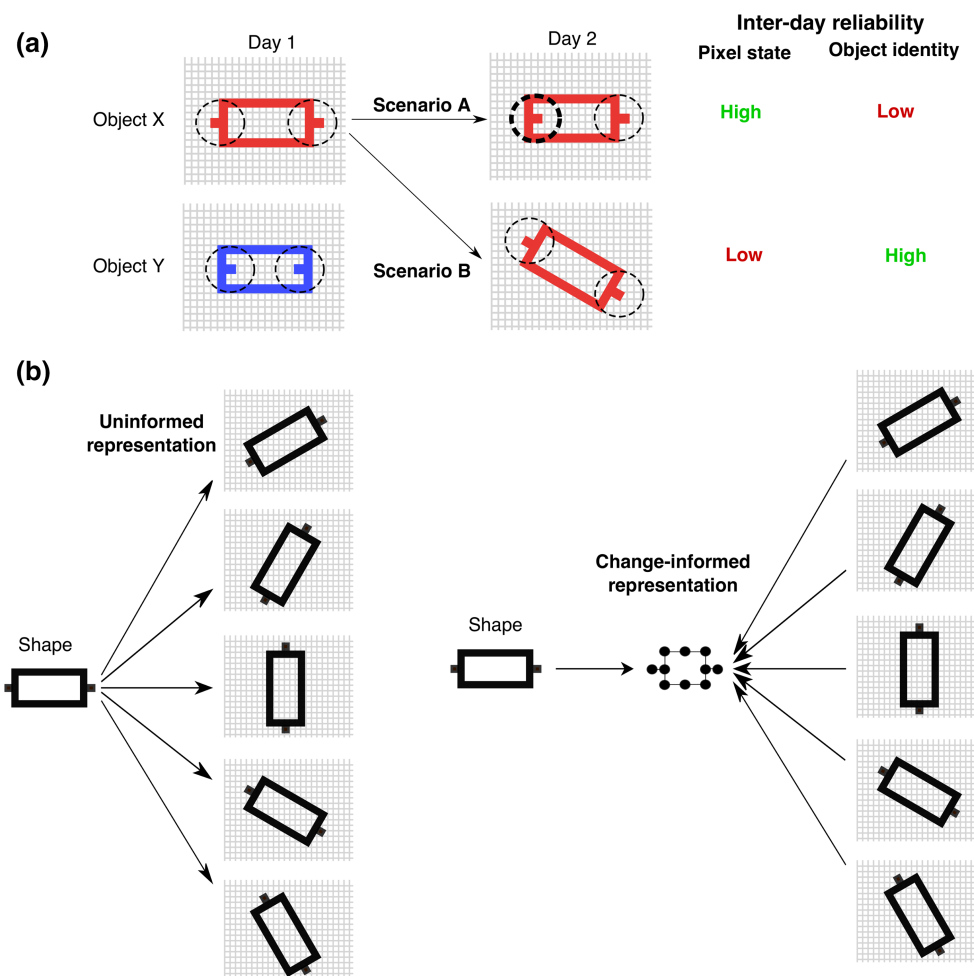
FIGURE 11   Test–retest reliability versus individual re-identification. (a) Object *X* (red) is defined by the configuration of filled and unfilled pixels. In Scenario *A*, most pixels of object *X* from Day 1 are filled on Day 2 except pixels within the left dotted circle. These few changed pixels prevent object *X* from being re-identified on Day 2 as its shape is confusable with object *Y*. In Scenario *B*, on Day 2, object *X* is rotated relative to its orientation on Day 1 thus filled pixels on Day 1 are not reliably filled on Day 2. However, accounting for the orientation differences allows object *X* to be distinguished in shape from *Y* on Day 2. Thus, high inter-day reliability in pixel state does not imply the same for object identity and vice versa. (b) Schematic of a *one-to-many* mapping between shape and representations that are uninformed about possible orientation changes (left panel); and a hypothetical *one-to-one* relationship associated with a change-informed representation, which is invariant to irrelevant orientation changes (right panel)

sensitive to true shape changes as in Scenario *A* (analogous to neurophysiological change) (Figure 11b).

Similarly, our approach was based on the view that RS change classification requires a change-informed representation of individual's RS activity for evaluation. Ideally, this representation would be sensitive to neurophysiological changes (*NP*+) but invariant to incidental cognitive variation (*NP*−). In analytical terms, we assume that there is a *one-to-many* mapping between an individual's neurophysiological phenotype and the multiple functional neural activity states (including at rest) that are uniquely configured by that phenotype (i.e., *NP* → [activity states]). Hypothetically, this *one-to-many* mapping (i.e., *NP* → [activity states]) could be analysed to extract a *one-to-one* mapping between an

individual's neurophysiology to a unique configuration of constraints shared by the many activity states (i.e., *NP* → [constrained configuration] → [activity states]) (Figure 11b). Such a constraint configuration could serve as a change-informed representation as it would be invariant to incidental cognitive variation (*NP*−) but sensitive to neurophysiological change (*NP*+).

Such a change-informed representation might not be known a priori. In such a scenario, training a decision rule for person identification (Figures 1c and 3a) could serve as a data-driven procedure to select such a change-informed representation. The term 'representation' is used here to refer to how information is carried by a configuration of features (e.g., by the specific assignment of weights to different features) rather than in the sense of

'feature selection' (Guyon et al., 2002; Guyon & Elisseeff, 2003), that is, to find a subset of features that carry relevant information.

4.2 | Person identification to select change-informed representations

Samples of same-day activity from (1) other individuals (not-S) and (2) person S serve as examples of the possible types of change, namely, $NP+$ and $NP-$, respectively. Therefore, training a decision rule on these examples (using machine learning) was a mechanism to select a putative change-informed representation.

As a simple analogy, in Figure 11a, object X's shape can be distinguished from that of object Y on Day 1 based on a few critical pixels (circled). In Scenario A, this pixel subset from Day 1 is unfilled on Day 2 implying that X's identity on Day 2 is now potentially confusable with object Y, which in turn suggests a possible cross-day change in shape ($NP+$). Thus, inter-individual comparisons can help to select critical feature relationships to represent an object X's identity, which is based on the object's shape. The critical pixels are unchanged across days in Scenario B suggesting an unchanged identity and hence shape ($NP-$). However, this representation suffers the limitation of being uninformed about rotations, as described above with the full list of filled pixels. Therefore, selecting a representation that was robust to incidental change was crucial.

Despite using an analogy of RS activity to a static object, training for person identification on same-day activity involved an assumption about dynamics and timescales. Each same-day measurement was segmented into 2-s non-overlapping activity samples. This inter-sample variability on short timescales (i.e., between the samples acquired within seconds/minutes of each other on the same day) was assumed to contain information about how RS activity could change in the absence of neurophysiological change (i.e., $NP-$). Therefore, successful cross-day identification was predicated on whether inter-sample differences on short timescales on the order of seconds were informative about inter-sample differences on long timescales (i.e., hours and days apart). Using moment-to-moment variability to 'account' for incidental RS differences was critical to bypass limits on available information. The cognitive state during rest measurements is related to experimental context and instructions (Duncan & Northoff, 2013; Kawagoe et al., 2018). However, beyond the assumption that participants were awake, we did not model the participant's cognitive state, for example, using participant's self-reported assessments of their cognitive state during the

RS measurement, sleepiness or coffee consumptions (Diaz et al., 2013; Guerra-Carrillo et al., 2014).

The individuality of RS activity has been studied with a variety of objectives, such as biometric identification (Campisi & Rocca, 2014; Gui et al., 2014; Valizadeh et al., 2019) and general questions related to the neural basis of individual differences and trait-identification (Demuru et al., 2017; Finn et al., 2017; Gratton et al., 2018; Smit et al., 2005; Smit et al., 2006). Consistent with these studies, our results also demonstrate the high distinctiveness of individual RS activity. Person identification was possible significantly above random chance from two-second snapshots of the power spectra at rest within the same day, as well as across days and tasks (Tables A.1, A.2 and A.4). However, in the current study, person identification served as a procedure to select change-informed representations. Hence, it was not sufficient to identify a person, for example, with a (dis)similarity-based measure that does not provide such a representation (e.g., Finn et al., 2015). Furthermore, the modulation of cross-day accuracy (i.e., accuracy loss) was a key indicator and classification above random chance was not itself informative about cross-day change. For these reasons, a machine learning approach was valuable as a principled method to obtain an explicit representation of the individual to be used for change classification.

4.3 | Trade-offs of data-driven selection of representations

Selecting representations based on a functional criterion (i.e., the ability to distinguish S from not-S) involved certain trade-offs.

There was no guarantee that the selected representations would be linked to individual neurophysiology rather than arbitrary day-specific properties. Even in the toy example (Figure 11a), the objects differ in colour (red = object X, blue = object Y), but using this feature to assess inter-day change would be uninformative about shape changes (as in Scenario A). Therefore, establishing the empirical soundness of this approach was critical. Because cognitive variability across days is difficult to establish, the high identification accuracy with $RS1$ could have been attributed to highly motivated and instruction-compliant participants rather than the neural characteristics of the rest state. However, the *Tap* and *Sequence* tasks provided verifiable within-subject examples of states that deviated from rest in order to assess the generality of RS-based inferences. Additional validity checks were provided by the battery of empirical tests, for example, the effects of aggregated training; assessing day versus measurement-specificity; and the weight distribution.

An important boundary condition is that high same-day accuracy for person identification is not sufficient to obtain a change-informed representation as shown by the poor cross-day accuracy with mono-band/mono-location feature subsets (Figures 3 and 6, Figure S4).

As an individual's identity is defined relative to other individuals in the studied group, an individual's representation would vary depending on the diversity of the group and properties of the most-similar individuals (as illustrated by the confusion matrix and inter-individual clustering in Figure 5). Furthermore, features *shared* by all individuals would be excluded. For example, in a study of the heritability of individual RS-connectivity properties with magnetoencephalography (MEG) (Demuru et al., 2017), the explicit removal of connectivity characteristics shared by all individuals in the group significantly improved individual identification. Thus, downweighting the role of shared features (explicitly or implicitly) would prevent changes to these shared features from being detected.

Furthermore, selecting representations based on a functional criterion enables considerable generality as the criterion does not specify the kind of information being represented. For instance, in our study, each activity sample was defined by the power in the canonical frequency bands. However, each sample could alternatively be defined by, for example, the dynamic connectivity estimated from the oscillatory phase (Bonkhoff et al., 2021; Calhoun et al., 2014; Rosjat et al., 2018; Rosjat et al., 2020). Exploring such extensions to other forms of information and information obtained from other imaging modalities is a key topic for future studies.

4.4 | Outlook

Our results support the technical feasibility and potential value of RS change classification to support the use of RS to track neuroplastic change. Notwithstanding its trade-offs, person identification suggests a powerful and convenient strategy to select appropriate change-informed representations to support change classification. We assumed that individuals in the studied group did not undergo extensive plastic changes. If individual identification was not possible with longitudinal RS even with such a group of healthy individuals over a period of 5 days, then the merits of using RS as a tracking indicator would seem to require critical re-evaluation especially for tracking over longer periods of time and with populations where such neuroplastic changes would be expected. Prior studies have found changes to the power spectrum with aging (Chiang et al., 2011; Knyazeva et al., 2018; van Albada et al., 2010; Voytek et al., 2015), for example,

age-related reductions in the frequencies of the alpha and beta band peaks. Voytek et al. (2015) suggest that such changes might indicate a change in the $1/f$ baseline possibly due to increased physiological noise with aging (also see Demuru & Fraschini, 2020). Furthermore, systematic longitudinal changes in the power spectrum have been observed following stroke (Giaquinto et al., 1994; Saes et al., 2020). Thus, implementing change classification to support longitudinal RS in clinical populations is an important future priority.

CONFLICT OF INTEREST
None.

AUTHOR CONTRIBUTION
Maximilian Hommelsen: Conceptualization, Methodology, Data acquisition, Software, Validation, Formal analysis, Writing: Original draft, Review & Editing, Visualization. **Shivakumar Viswanathan**: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing: Original Draft, Review & Editing, Visualization. **Silvia Daun:** Conceptualization, Writing: Review & Editing, Visualization, Supervision, Project administration, Funding acquisition.

PEER REVIEW
The peer review history for this article is available at https://publons.com/publon/10.1111/ejn.15673.

DATA AVAILABILITY STATEMENT
The data and code that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID
Maximilian Hommelsen https://orcid.org/0000-0003-1972-0465
Shivakumar Viswanathan https://orcid.org/0000-0002-7513-3778
Silvia Daun https://orcid.org/0000-0001-7342-1015

REFERENCES

Alegre, M., de Gurtubay, I. G., Labarga, A., Iriarte, J., Malanda, A., & Artieda, J. (2004). Alpha and beta oscillatory activity during a sequence of two movements. *Clinical Neurophysiology*, *115*(1), 124–130. https://doi.org/10.1016/S1388-2457(03)00311-0

Barry, R. J., Clarke, A. R., Johnstone, S. J., Magee, C. A., & Rushby, J. A. (2007). EEG differences between eyes-closed and eyes-open resting conditions. *Clinical Neurophysiology*, *118*(12), 2765–2773. https://doi.org/10.1016/j.clinph.2007.07.028

Benjamin, C., Lieberman, D. A., Changl, M., Ofen, N., Whitfield-Gabrieli, S., Gabrieli, J. D. E., & Gaab, N. (2010). The influence of rest period instructions on the default mode network. *Frontiers in Human Neuroscience*, *4*(December), 1–9. https://doi.org/10.3389/fnhum.2010.00218

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*(July), 57–69. https://doi.org/10.1016/j.neuroimage.2017.07.014

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer New York. https://link.springer.com/book/9780387310732

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, *34*(4), 537–541. https://doi.org/10.1002/mrm.1910340409

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Blum, A., Kalai, A. & Langford, J. (1999). Proceedings of the twelfth Annual ACM Conference on Computational Learning Theory

Boersma, M., Smit, D. J. A., de Bie, H. M. A., van Baal, G. C. M., Boomsma, D. I., de Geus, E. J. C., Delemarre-van de Waa, H. A., & Stam, C. J. (2011). Network analysis of resting state EEG in the developing young brain: Structure comes with maturation. *Human Brain Mapping*, *32*(3), 413–425. https://doi.org/10.1002/hbm.21030

Bonkhoff, A. K., Hope, T., Bzdok, D., Guggisberg, A. G., Hawe, R. L., Dukelow, S. P., Rehme, A. K., Fink, G. R., Grefkes, C., & Bowman, H. (2020). Bringing proportional recovery into proportion: Bayesian modelling of post-stroke motor impairment. *Brain*, *143*(7), 1–18. https://doi.org/10.1093/brain/awaa146

Bonkhoff, A. K., Schirmer, M. D., Bretzner, M., Etherton, M., Donahue, K., Tuozzo, C., Nardin, M., Giese, A. K., Wu, O., Calhoun, D. V., & Grefkes, C. (2021). Abnormal dynamic functional connectivity is linked to recovery after acute ischemic stroke. *Human Brain Mapping*, *42*(January), 1–14. https://doi.org/10.1002/hbm.25366

Boser, B. E., Guyon, I. M. & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory, pp. 144-152. https://doi.org/10.1145/130385.130401

Box, G. E. P., & Andersen, S. L. (1955). Permutation theory in the derivation of robust criteria and the study of departures from assumption. *Journal of the Royal Statistical Society: Series B (Methodological)*, *17*(1), 1–26. https://doi.org/10.1111/j.2517-6161.1955.tb00176.x

Brandmaier, A. M., Wenger, E., Bodammer, N. C., Kühn, S., Raz, N., & Lindenberger, U. (2018). Assessing reliability in neuroimaging research through intra-class effect decomposition (ICED). *eLife*, *7*, 1–19. https://doi.org/10.7554/elife.35718

Buckner, R. L., & DiNicola, L. M. (2019). The brains default network: Updated anatomy, physiology and evolving insights. *Nature Reviews Neuroscience*, *20*(10), 593–608. https://doi.org/10.1038/s41583-019-0212-7

Cabeza, R., Albert, M., Belleville, S., Craik, F. I. M., Duarte, A., Grady, C. L., Lindenberger, U., Nyberg, L., Park, D. C., Reuter-Lorenz, P. A., Rugg, M. D., Steffener, J., & Rajah, M. N. (2018). Maintenance, reserve and compensation: The cognitive neuroscience of healthy ageing. *Nature Reviews Neuroscience*, *19*(11), 701–710. https://doi.org/10.1038/s41583-018-0068-2

Calhoun, V. D., Miller, R., Pearlson, G., & Adali, T. (2014). The chronnectome: Time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron*, *84*(2), 262–274. https://doi.org/10.1016/j.neuron.2014.10.015

Campisi, P., & Rocca, D. L. (2014). Brain waves for automatic biometric-based user recognition. *IEEE Transactions on Information Forensics and Security*, *9*(5), 782–800. https://doi.org/10.1109/TIFS.2014.2308640

Carino-Escobar, R. I., Carrillo-Mora, P., Valdés-Cristerna, R., Rodriguez-Barragan, M. A., Hernandez-Arenas, C., Quinzaños-Fresnedo, J., Galicia-Alvarado, M. A., & Cantillo-Negrete, J. (2019). Longitudinal analysis of stroke patients brain rhythms during an intervention with a brain-computer interface. *Neural Plasticity*, *2019*, 1–11. https://doi.org/10.1155/2019/7084618

Cassani, R., Estarellas, M., San-Martin, R., Fraga, F. J., & Falk, T. H. (2018). Systematic review on resting-state EEG for Alzheimers disease diagnosis and progression assessment. *Disease Markers*, *2018*, 1–26. https://doi.org/10.1155/2018/5174815

Cassim, F., Szurhaj, W., Sediri, H., Devos, D., Bourriez, J. L., Poirot, I., Derambure, P., Defebvre, L., & Guieu, J. D. (2000). Brief and sustained movements: Differences in event-related (de)synchronization (ERD/ERS) patterns. *Clinical Neurophysiology*, *111*(11), 2032–2039. https://doi.org/10.1016/S1388-2457(00)00455-7

Chiang, A. K. I., Rennie, C. J., Robinson, P. A., van Albada, S. J., & Kerr, C. C. (2011). Age trends and sex differences of alpha rhythms including split alpha peaks. *Clinical Neurophysiology*, *122*(8), 1505–1517. https://doi.org/10.1016/j.clinph.2011.01.040

Cox, R., Schapiro, A. C., & Stickgold, R. (2018). Variability and stability of large-scale cortical oscillation patterns. *Network Neuroscience*, *2*(4), 481–512. https://doi.org/10.1162/netn_a_00046

Damoiseaux, J. S., & Greicius, Æ. M. D. (2009). Greater than the sum of its parts: A review of studies combining structural connectivity and resting-state functional connectivity. *Brain Structure and Funtion*, *213*, 525–533. https://doi.org/10.1007/s00429-009-0208-6

Davis, J. & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. ICML '06: Proceedings of the 23rd International Conference on Machine Learning, 233–240. https://doi.org/10.1145/1143844.1143874

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including

independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Demuru, M., & Fraschini, M. (2020). EEG fingerprinting: Subject-specific signature based on the aperiodic component of power spectrum. *Computers in Biology and Medicine*, *120*(April), 103748. https://doi.org/10.1016/j.compbiomed.2020.103748

Demuru, M., Gouw, A. A., Hillebrand, A., Stam, C. J., van Dijk, B. W., Scheltens, P., Tijms, B. M., Konijnenberg, E., ten Kate, M., den Braber, A., Smit, D. J. A., Boomsma, D. I., & Visser, P. J. (2017). Functional and effective whole brain connectivity using magnetoencephalography to identify monozygotic twin pairs. *Scientific Reports*, *7*(1), 1–11. https://doi.org/10.1038/s41598-017-10235-y

Diaz, B. A., van der Sluis, S., Moens, S., Benjamins, J. S., Migliorati, F., Stoffers, D., Den Braber, A., Poil, S.-S., Hardstone, R., Van't Ent, D., Boomsma, D. I., de Geus, E., Mansvelder, H. D., van Someren, E. J. W., & Linkenkaer-Hansen, K. (2013). The Amsterdam resting-state questionnaire reveals multiple phenotypes of resting-state cognition. *Frontiers in Human Neuroscience*, *7*(JUL, 1–15. https://doi.org/10.3389/fnhum.2013.00446

Diedrichsen, J., & Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLoS Computational Biology*, *13*(4), e1005508. https://doi.org/10.1371/journal.pcbi.1005508

Dimsdale-Zucker, H. R., & Ranganath, C. (2019). Representational similarity analyses: A practical guide for functional MRI applications. *Handbook of Behavioral Neuroscience*, *28*, 509–525. https://doi.org/10.1016/B978-0-12-812028-6.00027-6

Duncan, N. W., & Northoff, G. (2013). Overview of potential procedural and participant-related confounds for neuroimaging of the resting state. *Journal of Psychiatry and Neuroscience*, *38*(2), 84–96. https://doi.org/10.1503/jpn.120059

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 Research*, *1169*(2004), 44–56. https://doi.org/10.1016/j.brainres.2007.07.014

Finn, E. S., Scheinost, D., Finn, D. M., Shen, X., Papademetris, X., & Constable, R. T. (2017). Can brain state be manipulated to emphasize individual differences in functional connectivity? *NeuroImage*, *160*(March), 140–151. https://doi.org/10.1016/j.neuroimage.2017.03.064

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., & Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, *18*(11), 1664–1671. https://doi.org/10.1038/nn.4135

Giaquinto, S., Cobianchi, A., Macera, F., & Nolfe, G. (1994). EEG recordings in the course of recovery from stroke. *Stroke*, *25*(11), 2204–2209. https://doi.org/10.1161/01.STR.25.11.2204

Gonzalez-Castillo, J., Kam, J. W. Y., Colin, H. W., & Bandettini, P. A. (2021). How to interpret resting-state fMRI. *Ask your Participants*, *41*, Box 1(6), 1–19. https://doi.org/10.1523/JNEUROSCI.1786-20.2020

Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., Ortega, M., Hoyt-Drazen, C.,

Gratton, C., Sun, H., Hampton, J. M., Coalson, R., Nguyen, A. L., McDermott, K., Shimony, J., Snyder, A., Schlaggar, B., Petersen, S., Nelson, S. M., & Dosenbach, N. U. F. (2017). Precision functional mapping of individual human brains. *Neuron*, *95*(4), 791–807e7. https://doi.org/10.1016/j.neuron.2017.07.011

Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., Nelson, S. M., Coalson, R. S., Snyder, A. Z., Schlaggar, B. L., Dosenbach, N. U. F., & Petersen, S. E. (2018). Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron*, *98*, 439–452. https://doi.org/10.1016/j.neuron.2018.03.035 e5

Grefkes, C., & Fink, G. R. (2020). Recovery from stroke: Current concepts and future perspectives. *Neurological Research and Practice*, *2*(1), 1–10. https://doi.org/10.1186/s42466-020-00060-6

Guerra-Carrillo, B., MacKey, A. P., & Bunge, S. A. (2014). Resting-state fMRI: A window into human brain plasticity. *The Neuroscientist*, *20*(5), 522–533. https://doi.org/10.1177/1073858414524442

Gui, Q., Jin, Z., & Xu, W. (2014). Exploring EEG-based biometrics for user identification and authentication. *2014 IEEE Signal Processing in Medicine and Biology Symposium, IEEE SPMB 2014 - Proceedings*, 1–6. https://doi.org/10.1109/SPMB.2014.7002950

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *The Journal of Machine Learning Research*, *3*, 1157–1182. Retrieved from http://dl.acm.org/citation.cfm?id=944968

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1), 389–422. https://doi.org/10.1023/A:1012487302797

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J. D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067

Haynes J. D. (2015). A primer on pattern-based approaches to fMRI: Principles, pitfalls, and perspectives. *Neuron*, *87*, (2), 257. –270. https://doi.org/10.1016/j.neuron.2015.05.025

Hermundstad, A. M., Bassett, D. S., Brown, K. S., Aminoff, E. M., Clewett, D., Freeman, S., Frithsen, A., Johnson, A., Tipper, C. M., Miller, M. B., Grafton, S. T., & Carlson, J. M. (2013). Structural foundations of resting-state and task-based functional connectivity in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 6169–6174. https://doi.org/10.1073/pnas.1219562110

Hoenig, M. C., Bischof, G. N., Seemiller, J., Hammes, J., Kukolja, J., Onur, Ö. A., Jessen, F., Fliessbach, K., Neumaier, B., Fink, G. R., van Eimeren, T., & Drzezga, A. (2018). Networks of tau distribution in Alzheimers disease. *Brain*, *141*(January), 568–581. https://doi.org/10.1093/brain/awx353

Hohenfeld, C., Werner, C. J., & Reetz, K. (2018). Resting-state connectivity in neurodegenerative disorders: Is there potential for an imaging biomarker? *NeuroImage: Clinical*, *18*(November 2017), 849–870. https://doi.org/10.1016/j.nicl.2018.03.013

Huang, X., Altahat, S., Tran, D., & Sharma, D. (2012). Human identification with electroencephalogram (EEG) signal processing.

2012 International Symposium on Communications and Information Technologies, ISCIT 2012, 1021–1026. https://doi.org/10.1109/ISCIT.2012.6380841

Huber, P. J. (1981). Robust statistics. In Advances in mathematics (Vol. 60). John Wiley & Sons, Inc. https://doi.org/10.1002/0471725250

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., Penna, S. D., Duyn, J. H., Glover, G. H., Gonzalez-Castillo, J., Handwerker, D. A., Keilholz, S., Kiviniemi, V., Leopold, D. A., de Pasquale, F., Sporns, O., Walter, M., & Chang, C. (2013). Dynamic functional connectivity: Promise, issues, and interpretations. NeuroImage, 80, 360–378. https://doi.org/10.1016/j.neuroimage.2013.05.079

Kawagoe, T., Onoda, K., & Yamaguchi, S. (2018). Different pre-scanning instructions induce distinct psychological and resting brain states during functional magnetic resonance imaging. European Journal of Neuroscience, 47(1), 77–82. https://doi.org/10.1111/ejn.13787

Kehagias, A. (2021). Community Detection Toolbox, (https://www.mathworks.com/matlabcentral/fileexchange/45867-community-detection-toolbox), MATLAB Central File Exchange.

Knyazeva, M. G., Barzegaran, E., Vildavski, V. Y., & Demonet, J.-F. (2018). Aging of human alpha rhythm. Neurobiology of Aging, 69, 261–273. https://doi.org/10.1016/j.neurobiolaging.2018.05.018

Laumann, T. O., Gordon, E. M., Adeyemo, B., Snyder, A. Z., Joo, S. J., Chen, M. Y., Gilmore, A. W., McDermott, K. B., Nelson, S. M., Dosenbach, N. U. F., Schlaggar, B. L., Mumford, J. A., Poldrack, R. A., & Petersen, S. E. (2015). Functional system and areal organization of a highly sampled individual human brain. Neuron, 87(3), 657–670. https://doi.org/10.1016/j.neuron.2015.06.037

Lim, J., Wu, W. c., Wang, J., Detre, J. A., Dinges, D. F., & Rao, H. (2010). Imaging brain fatigue from sustained mental workload: An ASL perfusion study of the time-on-task effect. NeuroImage, 49(4), 3426–3435. https://doi.org/10.1016/j.neuroimage.2009.11.020

Mišic, B., Betzel, R. F., de Reus, M. A., van den Heuvel, M. P., Berman, M. G., McIntosh, A. R., & Sporns, O. (2016). Network-level structure-function relationships in human neocortex. Cerebral Cortex, 26(7), 3285–3296. https://doi.org/10.1093/cercor/bhw089

Newbold, D. J., Laumann, T. O., Hoyt, C. R., Hampton, J. M., Montez, D. F., Raut, R. V., Ortega, M., Mitra, A., Nielsen, A. N., Miller, D. B., Adeyemo, B., Nguyen, A. L., Scheidter, K. M., Tanenbaum, A. B., Van, A. N., Marek, S., Schlaggar, B. L., Carter, A. R., Greene, D. J., … Dosenbach, N. U. F. (2020). Plasticity and spontaneous activity pulses in disused human brain circuits. Neuron, 0(0), 1–10. https://doi.org/10.1016/j.neuron.2020.05.007

Noble, S., Scheinost, D., & Constable, R. T. (2019). A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. NeuroImage, 203(December 2018), 116157. https://doi.org/10.1016/j.neuroimage.2019.116157

O'Brien, F., & Cousineau, D. (2014). Representing error bars in within-subject designs in typical software packages. The Quantitative Methods for Psychology, 10(1), 56–67.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia, 9(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Pani, S. M., Ciuffi, M., Demuru, M., la Cava, S. M., Bazzano, G., DAloja, E., & Fraschini, M. (2020). Subject, session and task effects on power, connectivity and network centrality: A source-based EEG study. Biomedical Signal Processing and Control, 59, 101891. https://doi.org/10.1016/j.bspc.2020.101891

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. Journal of Machine Learning Research, 12, 2825–2830. Retrieved from http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

Pfurtscheller, G., & da Silva, F. H. L. (1999). Event-related EEG/MEG synchronization and desynchronization: Basic principles. Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 110(11), 1842–1857Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10576479. https://doi.org/10.1016/S1388-2457(99)00141-8

Poldrack, R. A. (2017). Precision neuroscience: Dense sampling of individual brains. Neuron, 95(4), 727–729. https://doi.org/10.1016/j.neuron.2017.08.002

Postema, M. C., de Marco, M., Colato, E., & Venneri, A. (2019). A study of within-subject reliability of the brains default-mode network. Magnetic Resonance Materials in Physics, Biology and Medicine, 32(3), 391–405. https://doi.org/10.1007/s10334-018-00732-0

Pritschet, L., Santander, T., Taylor, C. M., Layher, E., Yu, S., Miller, M. B., Grafton, S., & Jacobs, E. G. (2020). Functional reorganization of brain networks across the human menstrual cycle. NeuroImage, 220(December 2019), 117091. https://doi.org/10.1016/j.neuroimage.2020.117091

Rashid, B., & Calhoun, V. (2020). Towards a brain-based predictome of mental illness. Human Brain Mapping, 41(12), 3468–3535. https://doi.org/10.1002/hbm.25013

Rehme, A. K., Fink, G. R., Cramon, D. Y. V., & Grefkes, C. (2011). The role of the contralesional motor cortex for motor recovery in the early days after stroke assessed with longitudinal fMRI. Cerebral Cortex, 21(April, 4), 756–768. https://doi.org/10.1093/cercor/bhq140

Rifkin, R., & Klautau, A. (2004). In defense of one-vs-all classification. Journal of Machine Learning Research, 5, 2–6.

Rosjat, N., Liu, L., Wang, B. A., Popovych, S., Tóth, T., Viswanathan, S., Grefkes, C., Fink, G. R., & Daun, S. (2018). Aging-associated changes of movement-related functional connectivity in the human brain. Neuropsychologia, 117, 520–529. https://doi.org/10.1016/j.neuropsychologia.2018.07.006

Rosjat, N., Wang, B. A., Liu, L., Fink, G. R., & Daun, S. (2020). Stimulus transformation into motor action: Dynamic graph analysis reveals a posterior-to-anterior shift in brain network communication of older subjects. Human Brain Mapping, 42(May), 1–17. https://doi.org/10.1002/hbm.25313

Saes, M., Zandvliet, S. B., Andringa, A. S., Daffertshofer, A., Twisk, J. W. R., Meskers, C. G. M., van Wegen, E. E. H., & Kwakkel, G. (2020). Is resting-state EEG longitudinally associated with recovery of clinical neurological impairments early

poststroke? A prospective cohort study. *Neurorehabilitation and Neural Repair*, *34*(5), 389–402. https://doi.org/10.1177/1545968320905797

Schrouff, J., & Mourao-Miranda, J. (2018). Interpreting weight maps in terms of cognitive or clinical neuroscience: Nonsense? *2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, 1–4. https://doi.org/10.1109/PRNI.2018.8423944

Smit, C. M., Wright, M. J., Hansell, N. K., Geffen, G. M., & Martin, N. G. (2006). Genetic variation of individual alpha frequency (IAF) and alpha power in a large adolescent twin sample. *International Journal of Psychophysiology*, *61*(2), 235–243. https://doi.org/10.1016/j.ijpsycho.2005.10.004

Smit, D. J. A., Posthuma, D., Boomsma, D. I., & de Geus, E. J. C. (2005). Heritability of background EEG across the power spectrum. *Psychophysiology*, *42*(6), 691–697. https://doi.org/10.1111/j.1469-8986.2005.00352.x

Valizadeh, S. A., 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*(April), 470–481. https://doi.org/10.1016/j.neuroimage.2019.04.005

Vallat, R. (2018). Pingouin: Statistics in Python. *Journal of Open Source Software*, *3*(31), 1026. https://doi.org/10.21105/joss.01026

van Albada, S. J., Kerr, C. C., Chiang, A. K. I., Rennie, C. J., & Robinson, P. A. (2010). Neurophysiological changes with age probed by inverse modeling of EEG spectra. *Clinical Neurophysiology*, *121*(1), 21–38. https://doi.org/10.1016/j.clinph.2009.09.021

van den Heuvel, M. P., Mandl, R. C. W., Kahn, R. S., & Hulshoff Pol, H. E. (2009). Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Human Brain Mapping*, *30*(10), 3127–3141. https://doi.org/10.1002/hbm.20737

van der Vliet, R., Selles, R. W., Andrinopoulou, E. R., Nijland, R., Ribbers, G. M., Frens, M. A., Meskers, C., & Kwakkel, G. (2020). Predicting upper limb motor impairment recovery after stroke: A mixture model. *Annals of Neurology*, *87*(3), 383–393. https://doi.org/10.1002/ana.25679

Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage*, *145*(August 2015), 166–179. https://doi.org/10.1016/j.neuroimage.2016.10.038

Vecchio, F., Babiloni, C., Lizio, R., de Vico Fallani, F., Blinowska, K., Verrienti, G., Frisoni, G., & Rossini, P. M. (2013). Resting state cortical EEG rhythms in Alzheimers disease: Toward EEG markers for clinical applications: A review.

In *Supplements to clinical neurophysiology* (1st ed., Vol. 62). Elsevier B.V. https://doi.org/10.1016/B978-0-7020-5307-8.00015-6

Voytek, B., Kramer, M. A., Case, J., Lepage, K. Q., Tempesta, Z. R., Knight, R. T., & Gazzaley, A. (2015). Age-related changes in 1/f neural electrophysiological noise. *Journal of Neuroscience*, *35*(38), 13257–13265. https://doi.org/10.1523/jneurosci.2332-14.2015

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Winkler, I., Debener, S., Muller, K. R., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2015*(Novem), 4101–4105. https://doi.org/10.1109/EMBC.2015.7319296

Woo, C.-W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better biomarkers: Brain models in translational neuroimaging. *Nature Neuroscience Reviews*, *20*(3), 365–377. https://doi.org/10.1038/nn.4478

Wu, J., Quinlan, E. B., Dodakian, L., McKenzie, A., Kathuria, N., Zhou, R. J., Augsburger, R., See, J., Le, V. H., Srinivasan, R., & Cramer, S. C. (2015). Connectivity measures are robust biomarkers of cortical function and plasticity after stroke. *Brain*, *138*(8), 1–11. https://doi.org/10.1093/brain/awv156

Wu, J., Srinivasan, R., Quinlan, E. B., Solodkin, A., Small, S. L., & Cramer, S. C. (2016). Utility of EEG measures of brain function in patients with acute stroke. *Journal of Neurophysiology*, *115*(5), 2399–2405. https://doi.org/10.1152/jn.00978.2015

Xu, H., & Mannor, S. (2012). Robustness and generalization. *Machine Learning*, *86*(3), 391–423. https://doi.org/10.1007/s10994-011-5268-1

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

APPENDIX A

TABLE A1 Identification accuracies in different experimental states reported as mean % (SD)

	Type			
States	**CV**	**1-day**	**2-day**	**3-day**
*RS*1 ($N = 24$)	99.98 (0.04)	92.10 (6.84)	95.93 (3.63)	97.39 (2.65)
*RS*1 ($N = 18$)	99.98 (0.06)	92.79 (6.76)	96.61 (3.30)	97.53 (2.51)
*RS*2 ($N = 24$)	99.99 (0.04)	91.58 (7.49)	95.86 (4.18)	96.99 (3.50)
TapWait ($N = 18$)	99.99 (0.02)	92.58 (6.39)	96.36 (3.42)	97.60 (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 ($N = 18$)	100.00 (0.00)	93.47 (8.41)	96.67 (4.64)	97.95 (2.99)

Note: All values were significantly above random chance (50%) (see Table S1).

TABLE A2 Identification accuracies for *RS*1 with mono-band and mono-location feature subsets reported as mean % (SD)

	Type			
Subset ($N = 24$)	**CV**	**1-day**	**2-day**	**3-day**
B_δ	96.10 (2.54)	64.66 (7.92)	67.87 (8.12)	70.12 (8.01)
B_θ	97.63 (1.52)	76.99 (7.69)	81.76 (7.11)	83.70 (6.94)
B_α	98.51 (1.17)	84.20 (7.74)	88.38 (6.34)	89.59 (5.67)
$B_{\beta 1}$	99.65 (0.57)	81.41 (10.44)	87.03 (9.03)	88.92 (8.14)
$B_{\beta 2}$	99.74 (0.30)	76.37 (9.98)	83.22 (9.00)	86.66 (7.96)
L_F	98.01 (1.80)	82.68 (8.89)	87.30 (7.45)	88.87 (6.56)
L_{FC}	98.54 (1.55)	86.93 (9.38)	90.39 (7.45)	91.76 (6.12)
L_{CP}	97.94 (1.78)	85.28 (8.57)	89.43 (7.30)	90.37 (6.52)
L_{PO}	97.96 (2.02)	81.02 (8.32)	86.47 (7.35)	87.97 (7.06)

Note: All values were significantly above random chance (50%) (see Table S2).

TABLE A3 Classification accuracy of *RS*1 versus task state (binary, within-subject) reported as mean % (SD)

	Type			
***RS*1 vs. ($N = 18$)**	**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	95.12 (3.74)	67.79 (8.53)	73.05 (11.01)	74.86 (11.37)
TapMov	93.56 (4.12)	79.04 (7.17)	82.75 (5.99)	84.18 (5.92)
SeqMov	97.81 (1.76)	88.77 (5.21)	92.81 (3.43)	93.32 (3.77)

Note: All values were significantly above random chance (50%) (see Table S3).

T A B L E A 4 Accuracy of cross-task $^{RS1}\mathbf{I}_p \rightarrow {}^{X}\mathbf{I}_q$ identification reported as mean % (SD)

Test states (N = 18)	Type		
	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)

Note: All values were significantly above random chance (50%) (see Table S4).