AFRL-AFOSR-UK-TR-2023

Novel Computational Methods for Predicting Transitions in Spatiotemporal Neurodynamics between Attention and Mind-wandering

Nehaniv, Chrystopher
UNIVERSITY OF HERTFORDSHIRE
COLLEGE LANE
HATFIELD, HERTFORDSHIRE, ,
GB

02/16/2023
Final Technical Report

Air Force Research Laboratory
Air Force Office of Scientific Research
European Office of Aerospace Research and Development
Unit 4515 Box 14, APO AE 09421

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**	
20230216	Final	**START DATE** 20190801	**END DATE** 20220731

4. TITLE AND SUBTITLE
Novel Computational Methods for Predicting Transitions in Spatiotemporal Neurodynamics between Attention and Mind-wandering

5a. CONTRACT NUMBER	**5b. GRANT NUMBER**	**5c. PROGRAM ELEMENT NUMBER**
	FA9550-19-1-7034	
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Chrystopher Nehaniv

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
UNIVERSITY OF HERTFORDSHIRE COLLEGE LANE HATFIELD, HERTFORDSHIRE GB	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
EOARD UNIT 4515 APO AE 09421-4515	AFRL/AFOSR IOE	AFRL-AFOSR-UK-TR-2023-0038

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This report documents progress and results of the project “Novel Computational Methods for Predicting Transitions in Spatiotemporal Neurodynamics between Attention and Mind-wandering”. We aim to build an exploratory and predictive model of the brain that is sensitive to the transitions between sustained attention and mind-wandering behaviors. Such a predictive model potentially has applications in tracking attention during critical tasks as well as being of medical and diagnostic relevance. Towards this goal, we developed novel methods for characterizing and predicting the spatio-temporal dynamics of the brain at two complementary levels with differing types of spatial and temporal resolution:
Level 1. Electroencephalography (EEG) microstates, which are short quasi-stable topographies of brain electrical activity as measured at the scalp, on the order of 80-120 milliseconds.
Level 2. Functional Magnetic Resonance Imaging (fMRI) functional connectivity maps, which reveal networks of blood-oxygen-level-dependent (BOLD) activation in distributed brain areas at a slower time scale, on the order of seconds, with high spatial resolution.
This research examines and relates the regularities in patterns and sequences of EEG microstates and functional connectivity maps, with the intention of predicting transitions between states in humans of attention and mind-wandering.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	SAR	151

19a. NAME OF RESPONSIBLE PERSON	**19b. PHONE NUMBER** *(Include area code)*
NANDINI IYER	314-235-6161

Novel Computational Methods for Predicting Transitions in Spatiotemporal Neurodynamics between Attention and Mind-wandering
Project Final Report

Chrystopher L. Nehaniv[1,2], David G. Haydock[2], Hanna Derets[1], Sebastián Dohnány[1], Zixuan Gao[1], Reinoud Maex[2], Shabnam Kadir[2], Robert Leech[3], and Elena Antonova[4]

[1] Waterloo Algebraic Intelligence and Computation Laboratory
Dept. of Systems Design Engineering
Dept. of Electrical and Computer Engineering
Dept. of Applied Mathematics
Centre for Theoretical Neuroscience
University of Waterloo
Waterloo, Ontario, Canada

[2]Royal Society Wolfson Biocomputation Research Laboratory
School of Physics, Engineering and Computer Science
University of Hertfordshire
Hatfield, United Kingdom

[3]Centre for Neuroimaging Science
King's College London
Denmark Hill
London, United Kingdom

[4]Lab of Experiential Neuroscience and Analysis
Division of Psychology, Department of Life Sciences,
College of Health and Life Sciences & Centre for Cognitive Neuroscience
Brunel University London
Uxbridge, United Kingdom

This material is based upon work supported by the United States Air Force Office of Scientific Research (AFOSR) under award number FA9550-19-1-7034 to the first and last named authors.

Presented at Cognitive and Computational Neuroscience Review November 5, 2021
Project End: July 31, 2022 Submission Date: February 6, 2023

Contents

1 Project Objectives

This report documents progress and results of the project "Novel Computational Methods for Predicting Transitions in Spatiotemporal Neurodynamics between Attention and Mind-wandering".

We aim to build an exploratory and predictive model of the brain that is sensitive to the transitions between sustained attention and mind-wandering behaviors. Such a predictive model potentially has applications in tracking attention during critical tasks as well as being of medical and diagnostic relevance. Towards this goal, we developed novel methods for characterizing and predicting the spatio-temporal dynamics of the brain at two complementary levels with differing types of spatial and temporal resolution:

Level 1. Electroencephalography (EEG) microstates, which are short quasi-stable topographies of brain electrical activity as measured at the scalp, on the order of 80-120 milliseconds.

Level 2. Functional Magnetic Resonance Imaging (fMRI) functional connectivity maps, which reveal networks of blood-oxygen-level-dependent (BOLD) activation in distributed brain areas at a slower time scale, on the order of seconds, with high spatial resolution.

This research examines and relates the regularities in patterns and sequences of EEG microstates and functional connectivity maps, with the intention of predicting transitions between states in humans of attention and mind-wandering.

Scientific Objectives. The present project set out to address the following novel methodological objectives:

(1) derive a discrete 'alphabet' for fMRI functional connectivity maps analogous to the 'alphabet' of EEG microstates, quasi-stable patterns of electrical activity that cluster into small finite number of discrete types,
(2) characterize sequential dynamics of such simultaneous EEG and fMRI 'microstate' by building optimal generative automata models using the epsilon-machine approach previously applied by us to the EEG microstate sequences,
(3) identify rigorous mathematical measures for characterizing properties of generative automata models at subject- and group-levels, and
(4) apply the developed methods to differentiate between attention and mind-wandering with the view of predicting mind-wandering episodes during attention demanding conditions.

In the course of literature review, we assessed multiple approaches to classify fMRI states into types or discrete classes, referred to as fMRI dynamic functional connectivity (dFC) maps. In particular, the type of dFC maps referred to as *co-activation patterns* are derived in manner completely analogous to EEG microstates but at a time scale on the order of 1 to 2 seconds, and were seen as most relevant for our goals (see Section 2 for details). Therefore, objective (1) could be reformulated as:
(1) derive a small finite set of temporally discrete fMRI Co-Activation Patterns (CAPs), which are occurring simultaneously and in-tandem with, EEG microstates.

Simultaneous EEG-fMRI Dataset

The **target dataset for attention and mind-wandering** comprised simultaneous EEG and fMRI data acquired from 20 healthy right-handed individuals (age range 20-41 yo) by Prof. Robert Leech and Dr. Peter Hellyer as part of a larger MRC-funded study based at Imperial College London (PI, Prof. David Sharp). All participants were scanned during three runs: 1) eyes-open resting state lasting 8 minutes, where mind-wandering is expected; 2) a block-design focused attention paradigm using Choice Reaction Time (CRT) task blocks interleaved with eyes-open rest blocks as a baseline condition with 49s duration of each block; and 3) the same CRT task only continuous, without the blocked design. For more details on the sample, the paradigm, and scanning parameters see (Fagerholm et al., 2015). This dataset has prior ethical approval including secondary analysis for 20 healthy participants for both resting state and the attention task runs. The EEG data were captured at 5kHz with a 30-channel array, placed according to the extended international 10-20 system (Klem et al., 1999). The fMRI was recorded with the voxel size 2.00 x 2.00 x 2.00 mm, with a /echo time (TR/TE) ratio of 2000/30ms, using 35 ascending slices with 3.00mm thickness. The offset between the beginning of each participant's fMRI and EEG recordings was recorded. Three of the 18 participants were removed from analysis during processing - two for errors during data capture and one for showing characteristics of an outlier during microstate analysis.

Using a **complementary eyes-closed EEG dataset** (described in Section 6.1), our previous report (Nehaniv et al., 2021) focused on syntactic and discrete dynamical characterizations of EEG during mind-wandering and in this report Sections 5, 6 and 7 focus on assessing the use of such characterizations to distinguish mind-wandering from attention requiring conditions of visualization and verbalization.

With Prof. Leech joining the project in its second year, we extended the methods developed for EEG using the eyes-closed dataset to EEG and fMRI in the target dataset to complete objectives 1, 2, and 3. Integrating the methods for both EEG and fMRI in analysis of the target dataset in ongoing work focuses on the signature of mind-wandering vs. attention to complete objective 4 with partial results reported here. Due to the disruption of the COVID-19 pandemic, the integrated models for EEG-fMRI will be reported on in subsequent publications.

2 Background on Discrete Sequence Analysis for EEG and fMRI

Here we give context to the reader with a concise review of relevant literature on EEG microstates, fMRI functional connectivity dynamics, and sequence analysis for discretized neurodynamics, and describe the outstanding questions to be addressed.

Literature Review: Overview

Electroencephalography (EEG) has been used to characterize microstates: sets of transient, quasi-stable topographical maps of global EEG signal distributed across the scalp with an average duration of 80-120 milliseconds. In recent years, sequence analysis of EEG microstates has become of increased interest. Patterns in EEG microstate sequences have shown meaningful relations to mental states, psychopathologies, neurodevelopmental stages, and neurological diseases. Whereas EEG microstate syntax captures neural spatiotemporal dynamics at the scalp, understanding its relationship to large neural networks and their dynamics, as well as associated cognitive processes/mental states, has continued to be an elusive goal. Functional magnetic resonance imaging (fMRI) studies attempted to identify the relationship between EEG microstates and the brain's intrinsic functional networks. However, no previous study has attempted to simultaneously apply sequence analysis to characterize EEG microstate and fMRI functional connectivity (FC) network dynamics. Here, we suggest application of discrete sequence analysis to simultaneously acquired EEG and fMRI data by generating EEG microstates along with a set of dynamic FC (dFC) fMRI states referred to as *Co-activation Patterns (CAPs)*, doing so in both the resting state and in choice reaction time (CRT) tasks. This paves the way for use of EEG microstate and fMRI CAP sequences in both resting state and attention-requiring tasks in constructing predictive and generative epsilon-automata models of observed discrete sequences associated to these cognitive processes.

2.1 EEG Microstate Definition

Brain neural dynamics associated with cognitive functions, particularly unconstrained mental activities like the resting and mind-wandering states, are heavily studied topics. EEG has been used as a robust neuroimaging approach for decades (Tudor et al., 2005), with the dominating methodology being frequency analysis (Kubicki et al., 1979). EEG activity shows a multitude of superimposed frequencies which can be investigated separately in defined ranges referred to as "bands". EEG activity within the various bands has been associated with different mental states (Aftanas & Golocheikine, 2002; Borbély et al., 1981) and has different characteristics in those with disorders (Newson & Thiagarajan, 2019).

In the "alpha" band (8-12Hz) it has been shown that the brain's electrical activity is somewhat discontinuous. It is characterized by rapid and dramatic changes in topographic distribution, followed by periods of quasi-stable spatial distributions. These distributions are referred to as *EEG microstates* (Lehmann, 1987). Microstates maintain maximum and minimum potential points on the scalp for a short period of time, which then swap locations, with the maximum point taking the place of the minimum point, and vice versa. Formal definition of a microstate ignores this switch, only considering the pole locations. A single microstate is generally 80-120ms long, following the dominant alpha band frequency range (Teplan, 2002).

It has been shown that of the topographies observed across time points, a select few configurations are most dominant. EEG microstates cluster into an 'alphabet' of discrete classes, similar to the letters of

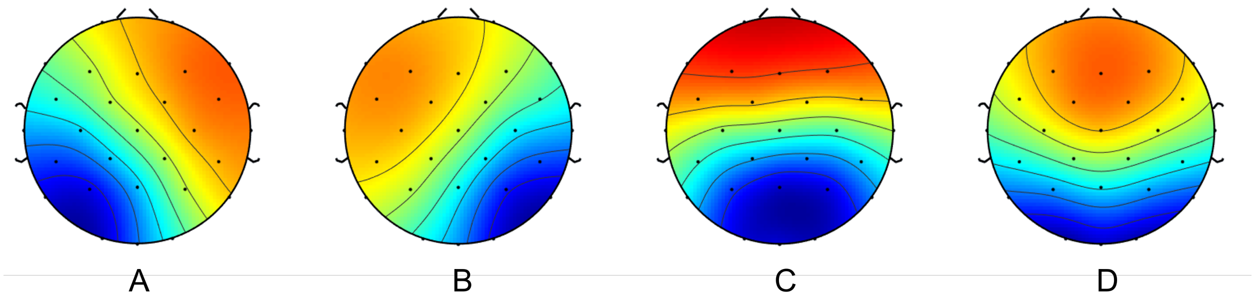

Figure 1: The canonical set of EEG microstates, labelled A, B, C and D left to right. Classes of microstate adapted from (Milz et al., 2016). Colours range from blue to red, with the most red indicating one pole, and the most blue indicating the other pole. Microstate A has poles in a left occipital to right frontal orientation, whereas microstate B is from right occipital to left frontal orientation. Class C has a symmetrical frontal to occipital orientation, and class D has a similar symmetrical orientation, but has a more frontocentral to occipital axis.

DNA code, with four topographic types referred to as microstate classes A, B, C, and D. These classes have been shown to 'explain' around 80% of variance in eyes-closed resting state EEG data, and are referred to as the 'canonical' EEG microstates (Koenig et al., 2002). The states have been found to be highly replicable in different populations, both healthy and clinical, in eyes-closed resting state EEG data (Michel & Koenig, 2018), and have demonstrated stability across methodological approaches (Khanna et al., 2014). Figure 1 shows the canonical classes of microstates. Microstate A has poles in a left occipital to right frontal orientation, whereas the axis through microstate B's poles has a right occipital to left frontal orientation. Class C has a symmetrical frontal to occipital orientation, and class D has a similar symmetrical orientation, but has a more frontocentral to occipital axis.

While the canonical set of four EEG microstates has been shown to be stable across studies (Michel & Koenig, 2018), data-driven approaches have been implemented to understand whether four is the optimal number of classes (Custo et al., 2014; Musso et al., 2010; Yuan et al., 2012). Data-driven approaches consider the amount of variance in topographies accounted by a candidate set of microstates, as well as cognitive functions that can be associated to each microstate class. For example, one study used a criterion algorithm to define the number of microstates to use and found their optimal number to be seven (Custo et al., 2017). Some classes defined in this seven appear to be more fundamental components of the canonical set; microstate F, named C′ in other studies (Michel & Koenig, 2018), has a spatially similar orientation to microstate C, but has a more occipitocentral pole, as opposed to microstate C′s occipital pole. This demonstrates the stability of the canonical set, yet raises questions as to whether clustering into four microstate classes is ideal, especially with many studies having used data-driven approaches successfully (Custo et al., 2014; Musso et al., 2010; Yuan et al., 2012). It is suggested that the best way to choose the number of EEG microstate classes might be a data-driven approach that also compares to the canonical set for association to functional significance.

2.2 Microstate parameters and sequences

While the topographies of the canonical microstate classes remain stable between studies, parameters of each individual class change based on the group and also cognitive state. The most commonly measured parameters are *duration* (average amount of time spent in a single microstate), *occurrence* (mean number of times a single microstate's class happens within a 1s window) and *coverage* (percentage of time covered by a single microstate class). These parameters, as well as pairwise transitions between microstate classes, have been shown to differ between mental states such as sleep (Brodbeck et al., 2012), hypnosis (Katayama et al., 2007) and meditation (Faber et al., 2005; Faber et al., 2015). Neurological disorders such as Alzheimer's (Strik et al., 1997) and Parkinson's disease (Pal et al., 2021) also show differences in these features. Psychopathologies like schizophrenia have also demonstrated altered parameters versus controls (Khanna et al., 2015; Koenig et al., 2002; Lehmann et al., 2005).

Short sequences of EEG microstates have already been investigated without relation to brain networks. Sequences of four microstates have been shown to differ between groups of paranormal believers and non-believers, with believers exhibiting the sequence $A \rightarrow B \rightarrow C \rightarrow A$ more predominantly, and non-believers exhibiting the reverse $A \rightarrow C \rightarrow B \rightarrow A$ more predominantly (Schlegel et al., 2012). Similarly, in a study conducted by Lehmann et al. (2005), length four sequences of microstates were shown to differ between schizophrenic patients and controls in the canonical set. The specific sequence $A \rightarrow C \rightarrow D \rightarrow A$ was common in controls, but its reverse, $A \rightarrow D \rightarrow C \rightarrow A$, was found as more common in the patients. It was shown by Nehaniv and Antonova (2017) that transition frequency differences found by Lehmann et al. (2005) do not explain the complexity of microstate syntax sufficiently, and that longer sequences are needed. Transition probabilities between microstates are often non-stationary across the time series (von Wegner et al., 2017), and hence a richer model may be required to capture the dynamics of EEG microstates. Van De Ville et al. (2010) also presented evidence of higher level structure regularities in microstate sequences by revealing scale-free dynamics in microstate syntax, concluding that "modelling microstate syntax needs to go beyond short-range interactions such as those modelled by n-step Markov chains".

2.3 Functional significance of EEG microstates

Although the topographies are replicable from study to study, the functional significance of EEG microstates is yet to be understood. Use of cognitive manipulations has been applied in attempts to relate each microstate with specific cognitive processes. Milz et al. (2016) reported microstate A as having a higher average duration during visual tasks, suggesting an association with visual processing. It has been suggested however that this increased duration during visual tasks may instead be due to microstate A being associated inhibition of language processing instead (Antonova et al., 2022). Microstate B has been associated with verbal processing (Milz et al., 2016) as well as visual processing when comparing occurrence and coverage during eyes-closed and eyes-open states (Seitzman et al., 2017).

Microstate C has a reported decrease in occurrence during a task relative to resting state (Milz et al., 2016; Seitzman et al., 2017), but specific differences between task conditions are unclear. Microstates A, B and C have also shown no meaningful differences in parameters between tasks in a separate study (Antonova et al., 2022).

Microstate D has been associated with attention orientation in different ways. Some suggest microstate D is associated with the resting state, reporting an increase in duration (Antonova et al., 2022) and occurrence (Milz et al., 2016) during resting state versus visualization and verbalization tasks, while others suggest an association with task conditions, reporting increase in duration and occurrence during a serial subtraction task versus resting state (Seitzman et al., 2017).

The lack of understanding of functional significance makes it clear that relating EEG microstates to brain networks may be a useful way of deriving their function. This may be done through source localization, or relation to fMRI data. Many studies have attempted to relate EEG microstates to functional networks in the brain using source localization (Milz et al., 2016; Milz et al., 2017; R. D. Pascual-Marqui et al., 2014). Source localization attempts to infer the position of the current sources in the brain from the electrode potentials at the scalp. The issue with this approach is that many different current density distributions in a 3D volume can produce the same potential distribution on the surface (Helmholtz, 1853). In the study of EEG, this is called the "inverse problem" (Grech et al., 2008). The problem states that one cannot discern the source of the neural signals in the brain based on activity at the scalp alone. Remedies to the issue are actively being investigated (Castaño-Candamil et al., 2015; Lopez Rincon & Shimoda, 2016), but have been a subject of investigation for decades (R. D. Pascual-Marqui, 1999; Vega-Hernández et al., 2008). The utility of source localization should not be discounted, but assumptions including the number of sources that give raise to a EEG microstate, result in an uncertainty in the true nature of the underlying generators. Hence, it may be more suitable to use EEG in-tandem with another form of measurement, using a "best of both worlds" approach (Manganas & Bourbakis, 2017). Different topographies at the scalp cannot be used to unambiguously discern the location of generators in the brain. However, differences in scalp topography imply the existence of different distributions of generators, even if those generators are unknown (Michel & Koenig, 2018).

2.4 Functional Magnetic Resonance Imaging as a comparative measure of neural activity

Functional magnetic resonance imaging (fMRI) is another form of non-invasive measurement of brain activity. It uses the blood-oxygen-level-dependency (BOLD) signal - a measure of blood flow in the brain - to investigate generators of observed neural activities. The neural basis of this blood flow measurement is rooted in the metabolic load change apparent in active regions of the brain. This change is called the hemodynamic response (HR), and lags behind neural activity by a few seconds (Buckner, 1998). The BOLD signal has proven to be a strong correlate of neural activity (Abreu et al., 2021; Britz et al., 2010; Brookes et al., 2009; Marino et al., 2019; Musso et al., 2010; Yuan et al., 2012), but it lacks direct association to an electrical signal. Using fMRI and EEG simultaneously allows the spatiotemporal dynamics of the brain to be observed at two complementary levels, i.e. at high spatial vs. temporal resolution, respectively (Manganas & Bourbakis, 2017). and gives fMRI a neuro-electrical association.

Attempts have been made to associate EEG microstates and fMRI using the hemodynamic response function (HRF), convolving it with the EEG signal to bring it into the same temporal domain of fMRI (Buckner et al., 2008). This approach has been incorporated to associate the functional significance of fMRI states (Damoiseaux et al., 2006; Karapanagiotidis et al., 2020; Yeo et al., 2011) to EEG microstates; such studies have generally associated microstate A with verbal processing, B with visual processing, C with subjective processing, and D with attention orientation (Britz et al., 2010; Milz et al., 2016; Yuan et al., 2012). While it is clear that some progress has been made using the HRF, there is contention on the significance of the approach (R. D. Pascual-Marqui et al., 2014). It may therefore be necessary to incorporate a different method that retains the temporal resolution of EEG and spatial resolution of fMRI data, instead of reducing one to meet the other. Investigation into candidate fMRI states that are analogous to an EEG microstates is a useful first step in this endeavour.

Similarly to EEG microstates and their characteristic sequences, canonical sets of functional maps have been studied in fMRI, referred to generally as functional connectivity (FC) maps (Biswal et al., 1995). There are many distinct types of these maps, each of which has its own prior assumptions and methodologies of generation.

2.4.1 Functional Connectivity Maps and Resting State Networks

In its most basic form, functional connectivity (FC) is the relationship between neural activation patterns of anatomically separated brain regions. Regions of the brain are associated by their activation in relation to the function of activation, rather than actual anatomical connectivity (Biswal et al., 1995). Correlation between the time series of different recorded voxels (a voxel is a unit of measurement in fMRI, a cubic volume of specified size within the 3D image) are what determine FC networks. There are many approaches developed for FC investigation, but when considering whole-brain connectivity patterns, the best approach is to use model-free methods. These methodologies are designed to find general patterns of connectivity across brain regions, rather than focusing on a single region. Such methods include principal component analysis (PCA), independent component analysis (ICA), and various types of clustering (Preti et al., 2017). The patterns of connectivity produced by these approaches aim to be maximally independent of one another, aiming to find the underlying sources of the resting state activity.

Resting state networks (RSNs) are identified components of fMRI activity. These networks are generated from resting state fMRI, commonly using ICA, and have been shown to be associated with different cognitive functions (Buckner et al., 2008; Damoiseaux et al., 2006; Fox et al., 2006). One such set of RSNs that are commonly used as a comparative set, generated a set of seven spatial functional networks in the cortex (Yeo et al., 2011). Figure 2 shows this set of RSNs. Each RSN is indicated by a different colour. Note that the RSNs are spatial parcellations and not whole-brain activity patterns, due to their nature as equivalence classes in a spatial partition (in that a single brain region cannot be a part of more than one network).

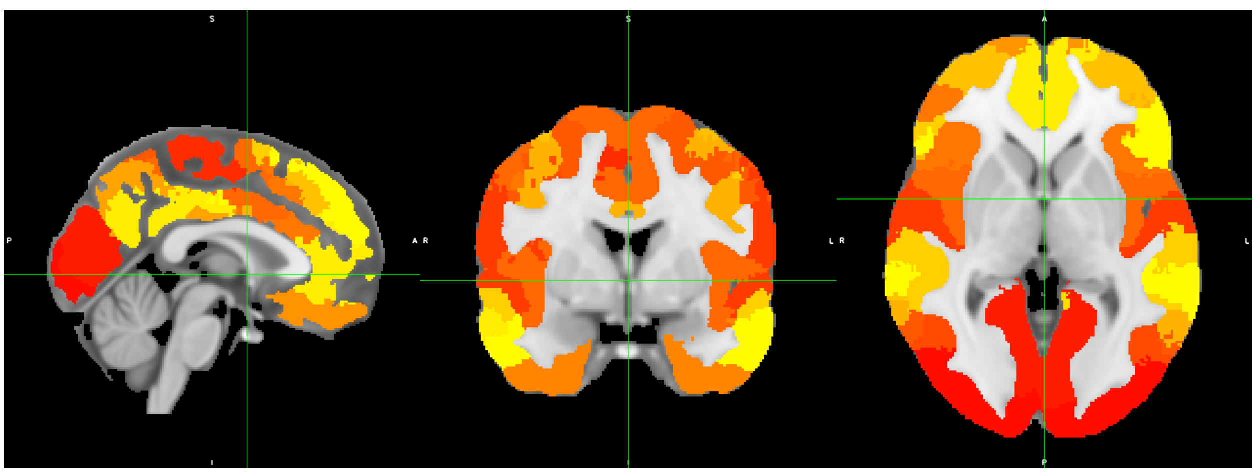

Figure 2: Set of spatial set of seven resting state networks, adapted from (Yeo et al., 2011). Each RSN is indicated by a different shade of red-yellow. Image was generated using FSLeyes. Cursor is located at MNI152 zero point.

The issue with the methodologies identified above is that they merely find correlated spatial sources. When the chosen method of analysis is applied, the output maps are a functional correlation of regions of the brain over the course of the whole time series, not capturing the temporal dynamics of the large-scale brain networks. This means that the FC networks are “temporally static", and therefore they are difficult to effectively relate directly to the continuously changing nature of spatiotemporal neural dynamics, e.g., as captured by the discrete sequence of EEG microstates over time.

2.4.2 Dynamic functional connectivity maps

Dynamic functional connectivity (dFC) approaches aim to capture the fluctuating nature of neural networks. Many methodologies have been developed to investigate dFC (Calhoun et al., 2014). All share the objective of analysing FC across a time scale, but do so in different ways. The most common approach is the sliding window, where connectivity between brain regions is computed as Pearson correlation between pairs of BOLD time courses over a temporal window (Hutchison et al., 2013). The computation is repeated iteratively, with the window moving along the whole time course to generate a connectivity time course. Performing this for all connections yields a dynamic characterization of connectivity of the whole brain. Many different window types have been proposed to improve upon this core concept (see Preti et al., 2017, Supp. Fig. 1), but all are limited by the necessity of dimensionality reduction. Since the analysis can only be applied when comparing specific regions of the brain, it is limited to correlations between the regions that were predefined. Importantly, the sliding window approach also works under the assumption that the brain’s activity is characterized by slow temporal dynamics (Hutchison et al., 2013), contrary to the fast temporal dynamics reflected in EEG microstates.

Frame-wise analysis was proposed as an alternative to the sliding-window approach for dFC analysis to do away with the assumption of slow dynamics. Frame-wise approaches fall into two categories: BOLD phase coherence and BOLD co-activation. Firstly, phase coherence studies were initially introduced to use phase synchronization between voxels as a measure of dFC, to improve temporal resolution versus the existing sliding window approach (Glerean et al., 2012). The approach has yielded studies which apply leading eigenvector dynamic analysis (LEiDA), which captures phase-locking states: states of phase coherence in activity between regions of the brain. Such states have been shown to overlap with functional subsystems in a meaningful way (Cabral et al., 2017; Vohryzek et al., 2020), and the parameters of these states have shown evidence of relation to cognitive performance (Cabral et al., 2017), also showing an ability to differentiate participants with major depressive disorder from healthy controls (Figueroa et al., 2019). However, these dFC maps are generally generated out of dimensionally reduced datasets, using atlases that reduce the brain into parcellated sub-regions. The phase correlation is hence between predefined regions of the brain. This reduction into atlases is not useful when the intention is to maintain the spatial resolution of the fMRI data for comparison with the relatively low spatial resolution of EEG.

BOLD co-activation approaches are the other form of frame-wise analysis, first being introduced with the method of point process analysis (PPA) (Tagliazucchi et al., 2012; Tagliazucchi et al., 2010). PPA only considers "relevant" time points, defining relevant as time points in the BOLD signal where a predefined threshold of activity is crossed. The peaks of activity in the BOLD time course are due to neuronal avalanching (Tagliazucchi et al., 2012), a phenomena observed in EEG studies (Fagerholm et al., 2015; Van De Ville et al., 2010). These peak time points are then subject to ICA, generating PPA-RSNs.

The BOLD co-activation approach was improved upon by applying the same process to a time course of fMRI activity, while retaining the original volumes of each time point, rather than subjecting them to ICA (X. Liu et al., 2013). A threshold is defined based on the overall time series, and all suprathreshold events in the fMRI signal are taken as relevant input volumes. Clustering analysis is then applied to each of these volumes to generate a set of states. These states are referred to as *co-activation patterns* (CAPs). In this way, CAPs have been shown to be temporal sub-component's of the known RSNs generated by means of FC approaches (X. Liu et al., 2018). The FC maps can, therefore, be thought of as the temporal averages of CAPs. This PPA-CAP method can be applied to a specific user defined brain region, or across the whole brain (X. Liu et al., 2013). It is commonplace in such an approach to measure equivalents to the EEG microstates parameters of occurrence, coverage, and duration (Chen et al., 2015; Koenig et al., 2002; Milz et al., 2016; Milz, 2015). In fact, the methodological steps of obtaining EEG microstates and CAPs are analogous. Furthermore, CAP approaches have already shown ability to differentiate across conscious states (Amico et al., 2014), and parameters of CAPs have been shown to differ between "rest" and working memory (Chen et al., 2015).

While PPA has the advantage of being an atlas-free analysis, avoiding a reduction of dimensionality due to parcellation, it should be noted that it has shortcomings in this particular case. Due to the selection of relevant time points of fMRI data, the number of time points used in analysis is reduced. Attempting to associate the higher temporal resolution of EEG with the lower temporal resolution of fMRI means that the number of time points in fMRI must be maximized to increase likelihood of identifying correspondences between the two types of neuroimaging data. There are however variants of the CAP approach which exist (X. Liu et al., 2018), and application of CAP analysis without use of a threshold is tractable.

2.5 FMRI sequence analysis

Investigation into transitions between the dFC states is part and parcel of dFC study in general (Preti et al., 2017). Many studies generate probabilistic state transition matrices to understand pairwise relationships between states. Any studies which have generated a discretized sequence of fMRI states have only done so for investigation of dynamics between pairs of states however, rather than the dynamics of longer-length sequences (Karapanagiotidis et al., 2020; Vohryzek et al., 2020). Hence application of state sequence analysis methods used in EEG microstate studies (Nehaniv & Antonova, 2017; Schlegel et al., 2012; Van De Ville et al., 2010) to fMRI CAPs may yield recurring sequences that can be associated with mental states, just as the states themselves have been. Even short sequences of fMRI states can last multiple seconds, allowing for an investigation of spatiotemporal patterns of activity that utilize a longer window of time vs. traditional sliding window approaches, while still retaining high spatial resolution.

By relating the two in complementary fashion, one may better understand the functional significance of EEG microstates, and characterize the spatiotemporal dynamics associated with different mental states/cognitive process at both high spatial (fMRI 'microstate' sequences) and temporal (EEG microstate sequences) resolution.

2.6 Micro-states and Macro-States - Utilizing Simultaneous EEG-fMRI

Britz et al. (2010) reported microstate A as associated with a *negative* blood-oxygen-level-dependency (BOLD) signal in the phonological network, and microstate B as associated with *negative* BOLD signal in the visual network (Britz et al., 2010). Custo et al. (2017) have also found similar associations. This appears to be in contradiction to the findings of Milz et al. (2016), where function of microstates A and B was indicated as the other way around: A in visual processing, and B in verbal processing. As pointed out by Antonova et al.

(2022) however, these two findings can be reconciled. EEG microstates are predominantly driven by sources in the alpha band frequency (Milz et al., 2017). Alpha is known to have inhibitory effects on modality-specific processing (O'Gorman et al., 2013), and negative BOLD activity may be related also to inhibitory neuronal activity (Sten et al., 2017). Hence, the observation that microstate A is predominant during visualization tasks (Milz et al., 2016), taken together with the suggestion that the generator of microstate A is a negative BOLD activity in the phonological network (Britz et al., 2010), may reflect an inhibiting action of the language processing areas during visualization tasks. Similarly, microstate B was implicated as predominant during verbalization tasks (Milz et al., 2016) due to its generator being the inhibition of visuo-spatial processing areas (Britz et al., 2010).

Microstate C has been associated with a *positive* BOLD signal in the posterior of the anterior cingulate cortex (ACC) along with other regions, with function being suggested as a subjective representation of the participant's own body (Britz et al., 2010). This was also found by Custo et al. (2017) in source localization methods, but in the non-canonical microstate C' (or F). These findings are in contradiction to the cognitive analyses described above however, which suggested an association with the default mode network (Seitzman et al., 2017).

The suggested generator of microstate D has been shown to be highly concordant in two fMRI studies (Britz et al., 2010; Custo et al., 2017) as a negative BOLD activity in the dorsal attention network. This is in line with the finding by Milz et al. (2016) that microstate D is more active during rest, but does not reflect that reported by Seitzman et al. (2017).

Worth noting is the "non-canonical" microstate E. This state has thus far not been related to fMRI. Source localization studies give conflicting suggestions of its source, with some suggesting it is generated from the default mode network (Custo et al., 2017), and others from the right medial pre-frontal cortex (mPFC) (Bréchet et al., 2019).

Despite these studies elucidating some conflicts in the field, it is clear that the specific functional significance of EEG microstates remains unclear. It is important to highlight that all canonical microstates have been shown as active in all cognitive processing modes (Antonova et al., 2022; Milz et al., 2016), suggesting that despite parameter differences between cognitive processes, each microstate's functional significance is not so clear cut. A more recent study has attempted to use microstates to predict dynamic functional connectivity (dFC) states (Abreu et al., 2021), fMRI states which consider whole-brain dynamics, rather than the activity of a particular network. In the study, microstates outperformed EEG spectral analysis in their predictive power of dFC states (although microstates were generated using a unique, non-stationary method). Abreu et al. (2021) highlighted how brain networks are continuously changing, and that consequently it would be expected that EEG microstate dynamics would reflect this. Due to this, it was found that the relationship between EEG and fMRI states is too complex to reduce to a one-to-one characterization of fMRI state network and EEG microstate, as in previous studies (Britz et al., 2010). Hence, we suggest here that it may be of importance to investigate the syntax of microstate sequences to better understand their relation to fMRI dFC states, and hence the functional significance of EEG microstates themselves.

2.7 Epsilon-machines as Models for Multi-scale Spatiotemporal Neurodynamics

Methods using epsilon-machines have recently been developed to derive and describe the 'grammar' of EEG microstate sequences (Nehaniv & Antonova, 2017). The aim of such approaches is to characterize mental processes in terms of their patterning of microstate sequences. Epsilon-machines are generative automata that can be used to derive optimal predictive models of underlying spatiotemporal neurodynamics. It has been shown that the complexity of EEG microstate dynamics in individuals cannot be captured by simple pairwise transition probabilities between microstates, but requires the generation of complex discrete dynamical systems with many causal states (Nehaniv & Antonova, 2017). Computational tools for understanding such systems from Krohn-Rhodes algebraic automata theory can be applied to analyse the complexity of predictive automata models (Egri-Nagy et al., 2014; Nehaniv et al., 2015; Rhodes, 2010). Epsilon-machine automata are constructed as generative brain models from observed discrete event sequences in EEG and/or fMRI. These sequences recorded during different cognitive states yield different epsilon-machine brain models which are used in order (1) to capture syntactic properties of the sequences that occur in the course of neurodynamics of these cognitive states, and (2) to aid in recognition of these cognitive states by comparing new or unseen EEG or fMRI sequences of discrete letters, i.e., EEG microsates or fMRI co-activation pattterns, from participants to match them to the neurodynamics that can be generated by those various brain models for different cognitive states. Furthermore, we aim to create coupled multiscale models linking slower temporal transitions at the fMRI timescale (transitions between CAPs) with different, temporally fine-grained EEG microstate models, each running at a faster time scale in tandem and simultaneously with the fMRI models, and providing syntactic (formal langauge/generative automata) characterizations of neurodynamics in order to identify corresponding cognitive states as well as to identify transitions between cognitive states. Such models could be applied then to other datasets, or in real-time to new data, e.g., from the same individuals whose EEG/fMRI was used to construct the models in order to do such cognitive processing mode and transition identifications.

3 Methods

3.1 Initial Investigation into Discretization of Simultaneous EEG-fMRI Data

Here we report the results from initial exploration of EEG and fMRI pre-processing methods for the simulataneous EEG/fMRI dataset, as well as the initial investigations into deriving an alphabet of discrete EEG patterns (microstates) and an alphabet of discrete fMRI activity patterns (co-activation patterns) in pursuit of project objective 1 (see Section 1).

3.1.1 Preprocessing EEG

EEG data were processed using the EEGLAB toolbox in MATLAB (Delorme & Makeig, 2004). The standard approach in EEG microstate analysis is to cut out any periods of noise from the data and use short periods of time named epochs (approx. 2s) for analysis (Koenig et al., 2002; Michel & Koenig, 2018; Milz et al., 2016). Since a goal of the project is to relate EEG microstates and their temporal sequence to discrete fMRI states and sequences, it is necessary to retain as much temporally continguous data as possible in much longer epochs, lasting minutes not seconds. Hence, before removing large portions of data, different methods of noise removal were attempted that would retain as much of the signal as possible for analysis. Various attempts at noise reduction are detailed below.

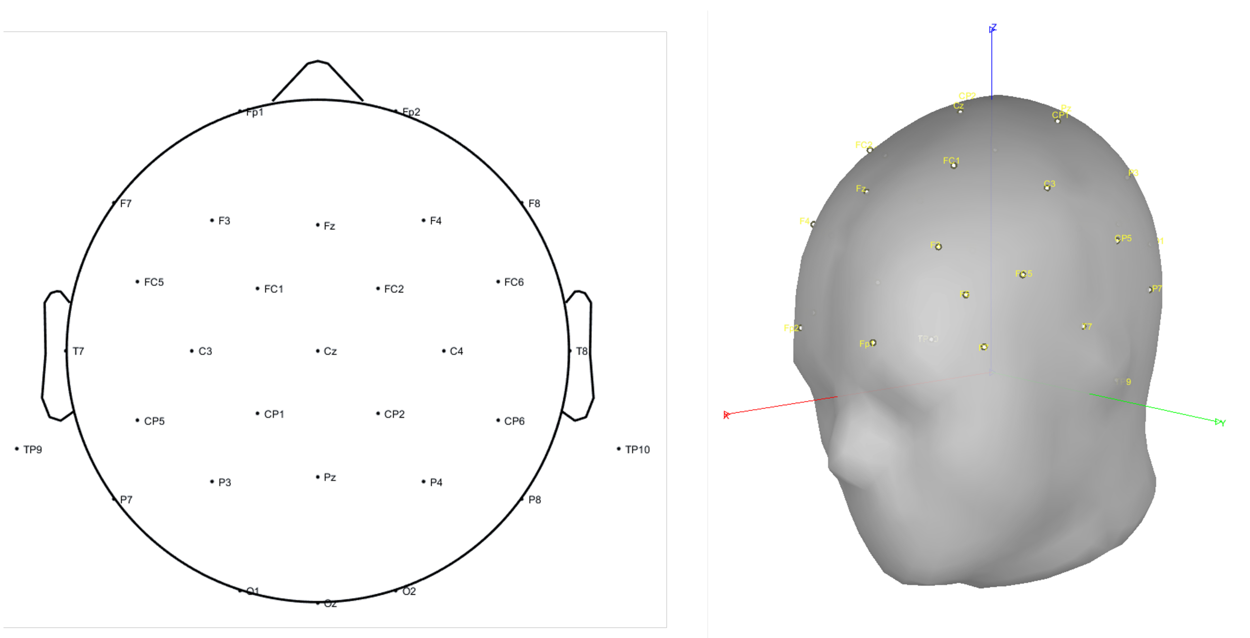

Figure 3: (Left) Top-down view of EEG channel locations of the 30 channel 10-20 system (Klem et al., 1999), on a two-dimensional cartoon head. Channels not on head are below neighbouring channels in 3D space. (Right) Three-dimensional representational image of electrodes on scalp. Image generated using Brainstorm (Tadel et al., 2011).

Basic preprocessing. First, since this data was recorded inside an fMRI scanner, artefacts due to interference of the scanner are apparent. The BrainVision Analyzer's magnetic resonance (MR) template was applied to all channels to remove artefacts associated the scanner (Brain Products, 2021). Next, each participant's EEG channels were referenced to the global average. In general, EEG recordings are the difference in electrical potential between two points expressed in micro-volts, which in this case is the difference between two recording sites. It is common practice to reference each channel to a ground electrode. Since there was no such electrode in this case, the average potential across all electrodes was calculated, and that output value was used as the reference. This was done for each participant individually (Figure 3).

The dataset recorded here contained an abnormal noise component. During the recording of the EEG and fMRI, the fMRI machine caused the channel leads to vibrate at 17Hz. Due to this, there were periods in the time course which contained bursts of high 17Hz activity. To counter this, on a participant-by-participant basis, a least squares linear regression was applied to the periods of high 17Hz activity with a 17Hz sinusoidal wave used as the regressor. In some participants, this 17Hz artifact was active throughout the time course but at a lower amplitude. In these cases, a windowed application of the same regression was applied throughout the time series. This technique of noise removal is a traditional approach in the field (Croft & Barry, 2000) and is used often to remove heart and eye associated artefacts. Figure 4 shows an example of a burst of the 17Hz artefact across recording channels (left), and the same period after regression (right).

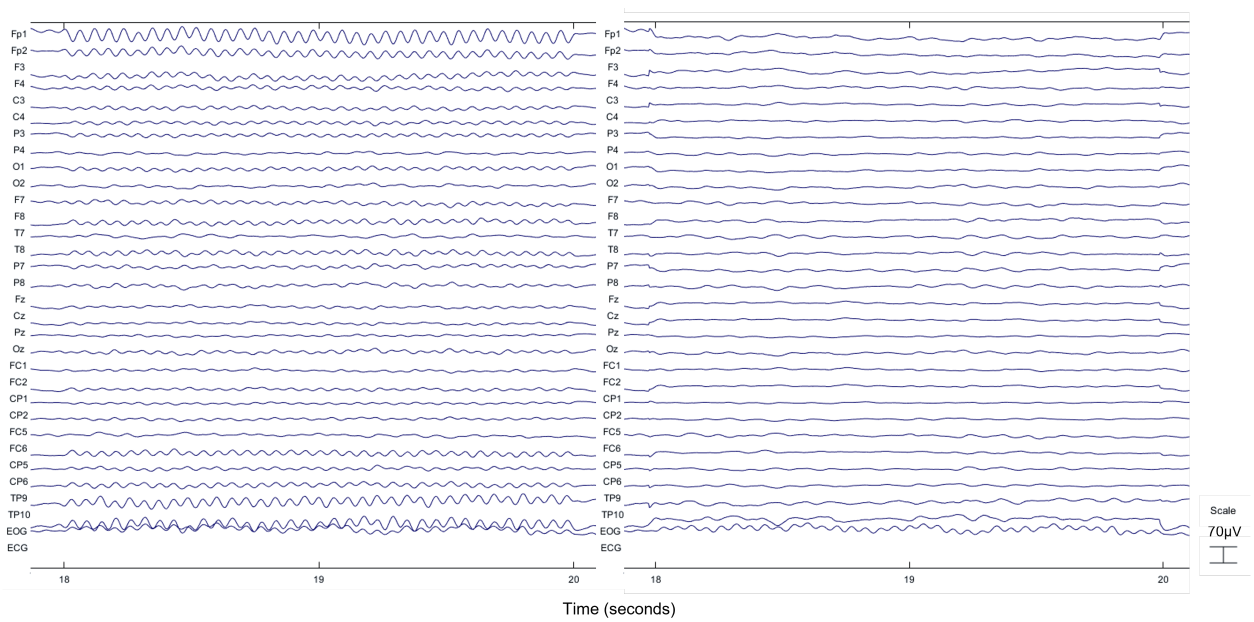

Figure 4: Comparison of a period of a single participants EEG time course, before (left) and after (right) application of a linear regression using a 17Hz signal as a regressor to remove the periods of noise and retain brain signal. X-axis indicates time point on the participant time course of recording. Y-axis shows label of each channel recording.

Following this, the data was down-sampled to 250Hz, from its collected frequency of 5kHz. The data was then band-pass filtered between 2 and 20Hz. This filter range is common practice in the field (Khanna et al., 2014; Milz et al., 2016; Milz et al., 2017), and retains alpha band activity.

Exploratory processing.

Along with the 30 EEG channels recorded across the scalp, one EOG channel and one ECG channel were recorded to measure the electrical activity of the eyes and the heart, respectively, potential sources of noise in the EEG signal. Eye movement, blinking and muscle movement around the eyes are especially apparent in the dataset due to the eyes-open state. In epochs of the time series that these types of artefacts were apparent, a linear regression was applied to each channel individually, using the EOG channel as a regressor. While this noise was mitigated in some cases, most applications were not successful in removing noise. Since the EOG channel itself contained noise, regression of the EOG channel with EEG only added noise. The same was the case for application of ECG as a regressor.

Other cases of noise had the potentiality of having both EOG and ECG noise present simultaneously. A multi-linear regression was applied to these cases (Croft & Barry, 2000), and again, in many cases, noise was introduced rather than mitigated. For this reason, any periods of noise that were apparent other than the 17Hz artefacts (which were successfully removed with regression), were excluded from further analysis. Lastly, some artefacts were associated with head movement. To confirm this, the points at which suspected head movement artefacts occurred on the time series were compared to the relative motion measured during fMRI registration (see section 3.1.3), as shown in figure 5. Such components cannot be removed via regression and were excluded from further analysis. The exclusion of these periods meant that the only noise remaining in the data was noise that was latent throughout the time series and could not be easily cut out.

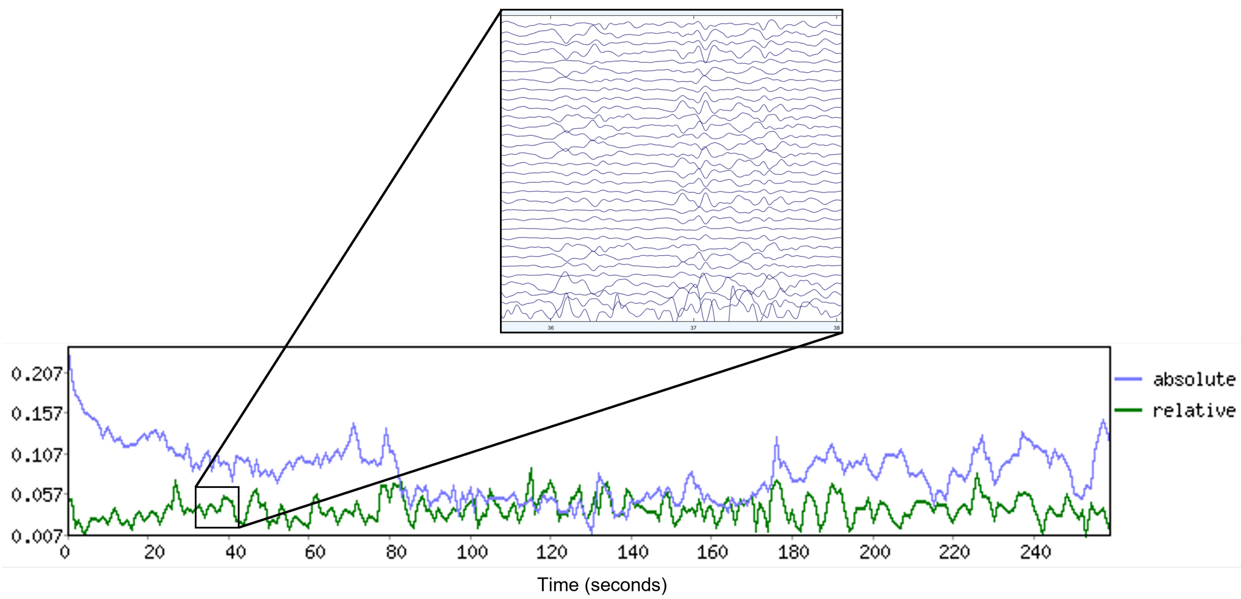

Figure 5: Estimated head movement generated by MCFLIRT program from registration and motion correction of fMRI data for single participant. Blue line shows absolute head movement from timepoint to timepoint, green line shows relative head movement from starting position at 0s. Y-axis given in mm. Inset shows the EEG time course of a three second time period. Noise that was not removed by noise corrections is apparent in the period, and relative head movement as well as absolute head movement are high during this period.

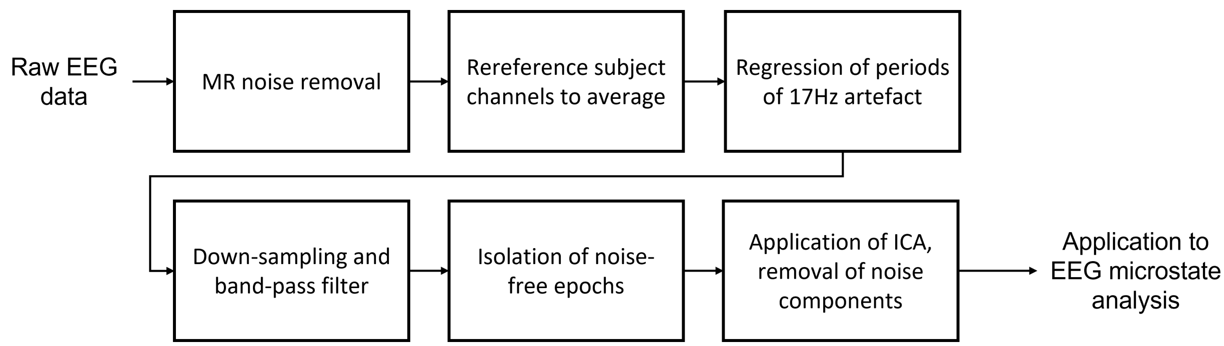

Figure 6: Finalized flowchart of EEG preprocessing steps. EEG channels are first rereferenced to the global average. Then, regression is applied to periods where 17Hz artefact is apparent. Data is then down-sampled to 250Hz, and band-pass filtered between 2 and 20Hz. Periods that have the minimum amount of noise are then isolated into epochs, and ICA is applied at an individual level. Components containing clear signs of noise are excluded from further analysis.

The final step of preprocessing was application of single-participant ICA. Components that were identified as noise were removed from the data. From the remaining signal, as many 2 second epochs as possible were taken from each participant that were the most noise-free, and these epochs were used to generate EEG microstate classes, which could be used then to derive EEG microstates sequences in over much longer epochs. Therefore, the approach used most prominently in the field of EEG microstate analysis was adopted. Figure 6 shows the finalized EEG preprocessing pipeline.

3.1.2 EEG Microstate Analysis

EEG microstates were generated using a data-driven technique with application of the KeyPy software package using Python (Milz et al., 2016; Milz, 2015). For each participant, the EEG topographies at time points of maximum global field power (GFP) (Skrandies, 1990) are collected and used as input to a specific k-means clustering algorithm (R. Pascual-Marqui et al., 1995). Clustering is a procedure of classifying a set of objects in different groups such that within group differences are smaller than across group differences. K-means clustering is applied here to classify EEG time points based on spatial similarity. The modified version of k-means applied here, differs in that polarity of channels is ignored (R. Pascual-Marqui et al., 1995).

If it is the case that four microstates are being sought after, the algorithm takes an extra labelling step, taking the existing canonical set of microstates and computing a similarity score between the data-driven set and the canonical sets from previous studies (Koenig et al., 2002; Milz et al., 2016). Once all similarities are computed, the data-driven set is labelled with the canonical set based on highest similarity. These classes are then used to compute mean classes across participants.

Figure 7 shows the four mean microstate classes across the 13 of the 18 participants (5 were excluded due to errors in processing). The mean microstates generated show clear similarities to those in figure 1, with similar orientations to those described in the canonical set, yet are not identical. The mean classes explain 81% of variance in the data, a similar percentage to past studies (Michel & Koenig, 2018).

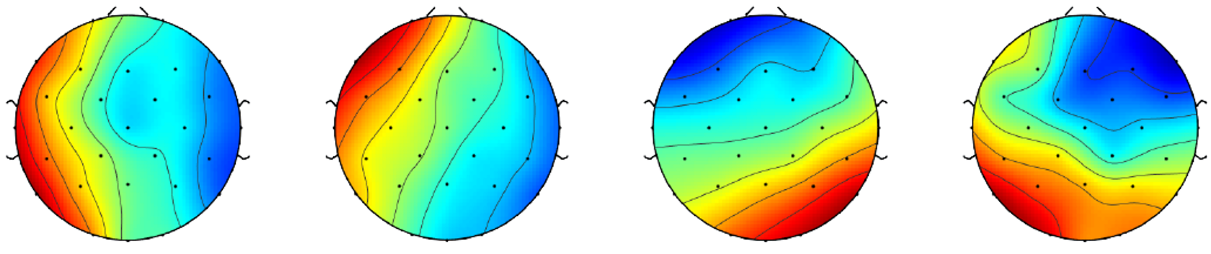

Figure 7: Data-driven EEG microstate maps generated from 13 of the 18 participants using KeyPy. Microstates are labelled A, B, C and D based on highest correlation with canonical maps. Note that polarities are not relevant when comparing to canonical set.

3.1.3 Preprocessing fMRI

Basic processing. FMRI data processing was carried out using FEAT version 6.00, part of the FSL software library (Jenkinson et al., 2002). First, non-brain removal was applied using the brain extraction tool (BET) (Smith, 2002). This extracts a basic "mask", which is used to differentiate the voxels associated with the brain, from those which are not. The MCFLIRT tool is next applied to the fMRI time series, a motion correction and registration tool(Jenkinson et al., 2002). Registration is the 3D alignment of two images. Alignment in this case implies that an anatomical location is the same voxel across participants. To ensure that this alignment takes place, all participants are registered against a standard image space. The functional data of a single participant is first aligned with its simultaneously captured structural data (T1 image). That alignment is then registered onto a standard space (MNI152 2mm), so that participants can be compared. This process is known as an inverse transform. In this case, the transform was an affine transform, which implies 12 degrees of freedom (DOF) for the registration, i.e., 12 degrees by which the images may be manipulated in 3D space to achieve alignment.

Next is spatial smoothing. This process assigns each voxel a weighted average of its neighbouring voxels, with each voxel being weighted based on its closeness to the voxel of interest. This is applied using a Gaussian kernel of full width at half maximum (FWHM) 5mm. The process reduces resolution of the data, but increases the signal to noise ratio, and a minimum smoothness is required for Gaussian random field theory to apply to analysis, a prior that must be adhered to for processing to be valid. A high-pass temporal filter is applied, using a weighted least squares regression line with sigma equal to 50s, to remove low frequency components that are commonly associated with noise in fMRI recordings.

Exploratory processing. The final step was application of ICA for noise removal. This was initially applied using the MELODIC tool, an ICA package available in the FSL software suite. ICA was applied across participants with 50 components as the output. Each of the ICs were manually inspected, and those which were clearly noise (e.g. a large portion of activity was outside the brain, clear movement artefacts) were excluded. The remaining components were then regressed back onto the time series of each individual participant, so that only those components were remaining. Noise in each of the components were still clearly apparent. This is likely due to the group approach of the ICA. Components which were removed from individual participants were deemed noise components across the group of participants, rather than at the individual level.

To alleviate these problems of persisting noise, the ICA package AROMA was applied to the dataset (Pruim et al., 2015). ICA-AROMA automatically detects noise components in ICA applications, and applies said ICA at the participant level, similar to approaches in EEG microstate processing. This approach removes participant specific noise components, instead of generalizing noise across the group. ICA-AROMA reduced noise drastically in the resulting dFC maps, and was hence used in subsequent analysis. The finalized pipeline of fMRI preprocessing is found in figure 8, using AROMA as an automatic noise removal tool.

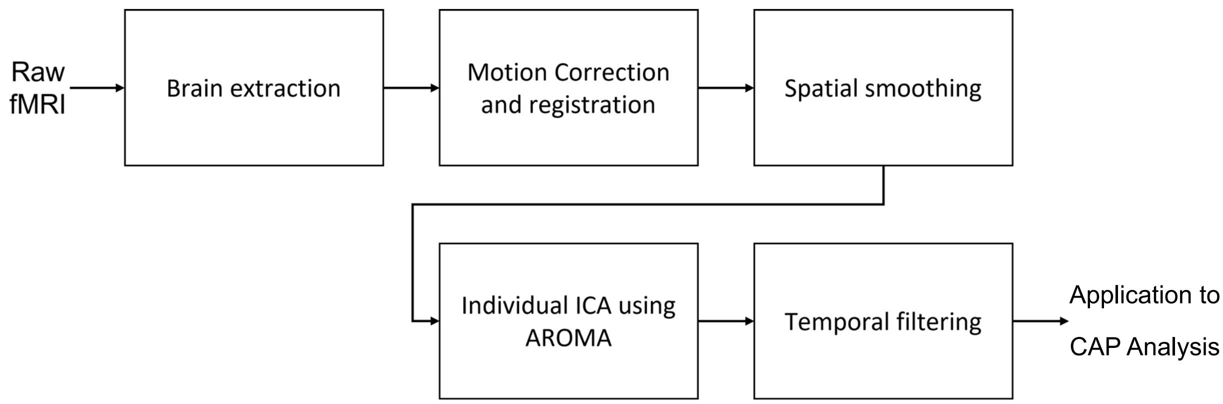

Figure 8: Finalized processing pipeline of fMRI data. Raw data is applied to brain extraction to isolate brain related voxels. Motion correction and registration are applied to the brain voxels to remove movement related noise, and to bring all participants to a common space. Spatial smoothing of voxels is then applied to increase SNR. Finally, participants are individually subjected to ICA using AROMA, and output has temporal filter applied.

3.1.4 Initial Investigation into CAP Analysis

To obtain the fMRI CAPs, a k-means clustering algorithm was applied to each of the preprocessing approaches outlined above. K-means was initially applied to all 8160 fMRI volumes (time frames) from 17 of the 18 participants (one excluded due to errors). The cluster centres were subjected to a z-statistic which normalized across the CAPs, so that the maps quantify the degree of significance to which the CAP values for each voxel deviate from zero. Figure 9 shows CAPs generated from the data processed using group-ICA, with 7 clusters. Upon visual inspection, it is clear that these CAPs contain excessive noise, and cannot be used to discern any meaningful associations.

Clustering was also applied to the ICA-AROMA outputted data, using 30 clusters. Ten of the 30 are shown in figure 10, labelled accordingly. These CAPs evidently have a much improved SNR to those found in figure 9. CAPs 7, 9, 12, 26 and 28 all have some coarse similarity to the default mode network (DMN). Furthermore, CAPs 6, 10 and 15 have a similar relationship to existing medial and occipital visual RSNs. CAP 2 shows coarse similarity to the salience network, while CAP 9 shows similar overlap with the sensorimotor network, and CAPs 16 and 26 have some clear frontal component (Damoiseaux et al., 2006; Fox et al., 2006; Heine et al., 2012). Each of these associations supports suggestions that CAPs are temporal components of RSNs (X. Liu et al., 2013; X. Liu et al., 2018).

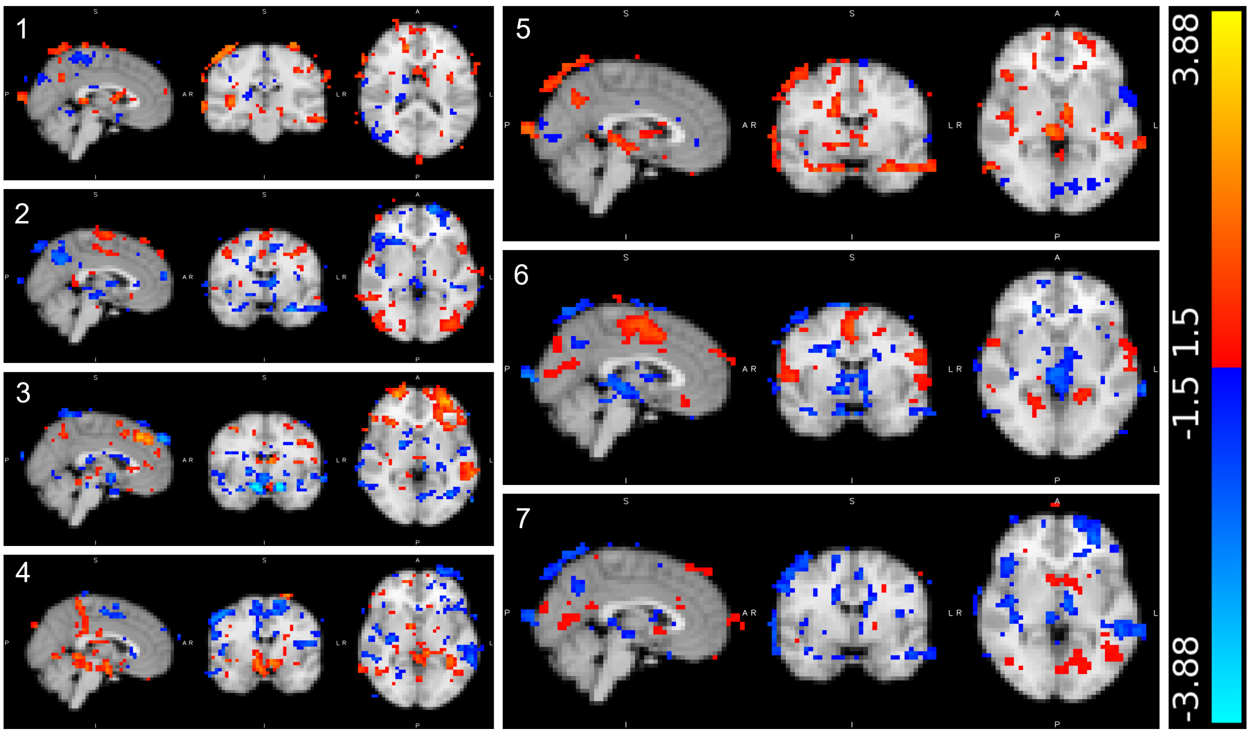

Figure 9: fMRI CAPs generated from clustering analysis of time point volumes using less robust noise removal techniques. Each map is viewed in three dimensions via its sagittal, coronal and transverse plane, left to right. Grey background image is the MNI152 standard. All states are shown at the zero point of standard space. Colours overlaying the standard brain indicate z-score of active regions of activity in each CAP. Colour bar shows the z-score range across the states.

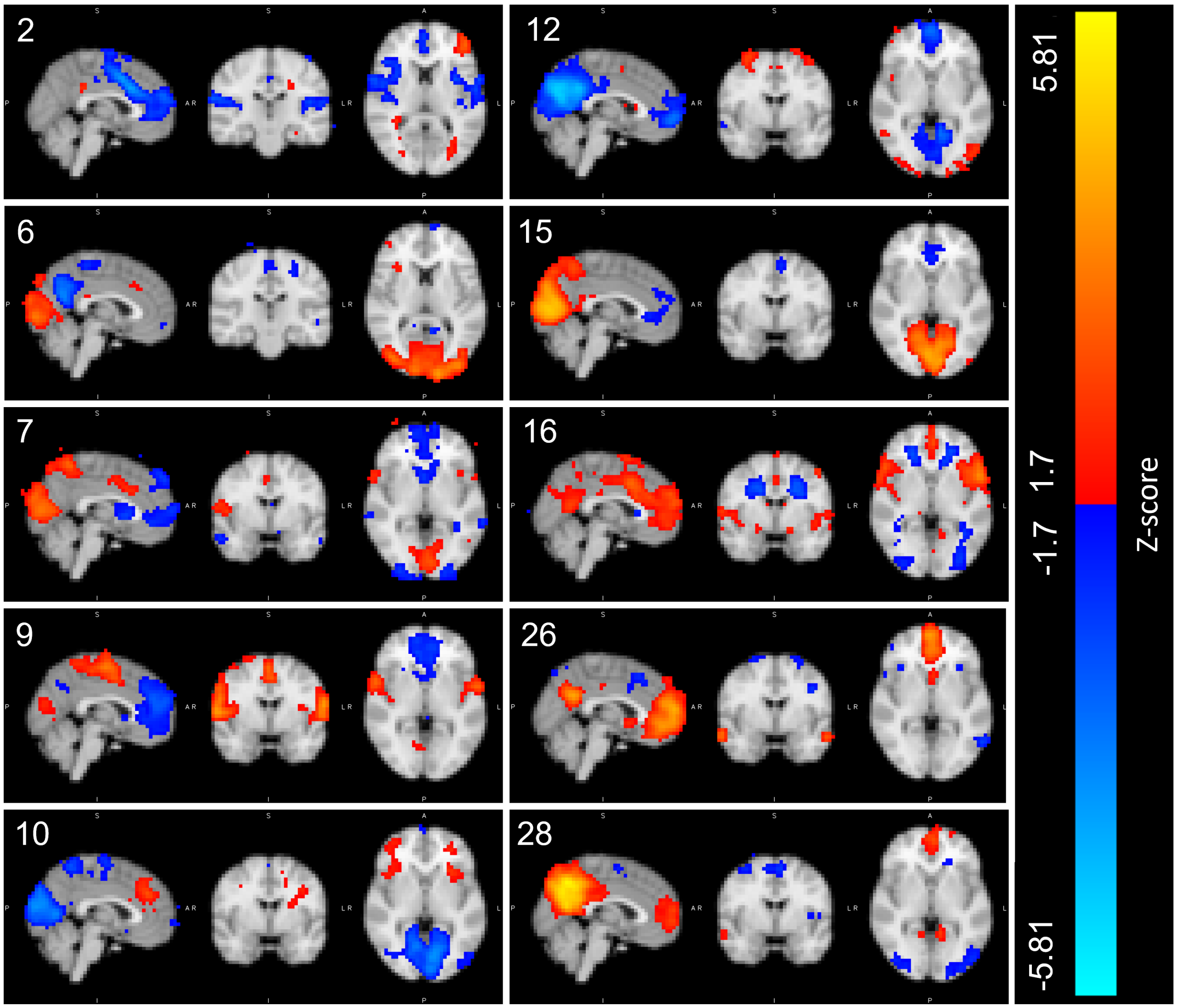

Figure 10: fMRI CAPs generated from clustering analysis of time point volumes using ICA-AROMA derived dataset. Grey background is the MNI152 2mm standard brain. Colours overlaying the standard brain indicate regions of activity in each CAP. Colour bar shows z-score. All CAPs shown at zero point of standard space.

3.1.5 Outstanding Questions and Follow-up Objectives following Initial Investigation

Here we record observations, questions and plans as we saw them immediately following the initial investigation just described: Each discrete set of states (microstates for EEG and CAPs for fMRI) must be subjected to stability checks, to confirm their validity. The number EEG microstates must be considered from a data-driven perspective. Using a criterion to determine the optimal number of states is an important step to take. Possible candidates are the minimum description length (MDL) (Yuan et al., 2012), and the more basic cross-validation criterion of R. Pascual-Marqui et al. (1995). Following a criterion check, the next steps are iterative application of microstate analysis, to ensure a convergence of centroids of the clustering algorithm. Parameters of microstates also need to be investigated further, as comparison to field consensus will validate findings, and to highlight individual differences and differences between conditions such as attention and mind-wandering.

Stability of fMRI Co-Activation Patterns (CAPs) is also essential, however, in contrast to EEG microstates which have been studied for decades, this is a new area of research. To first understand how the fMRI states generated would compare to those in the literature, a labelling procedure will be employed, similar to that employed on EEG data (Milz, 2015; R. Pascual-Marqui et al., 1995). A similarity measure will be used to compare the generated states to existing states in the field, so that a more robust verification of similarity can be confirmed, rather than a simple by-eye comparison. Investigation of clustering with different numbers of centroids is also necessary to validate the discovered CAPs. A basic iterative stability check can also employed for this purpose. CAPs are expected to show some differences between conditions and individuals.

Once both EEG microstates and fMRI CAP sets are confirmed as valid states, the next step will be the application of sequence analysis. The states generated in both domains will be initially remapped back onto their respective time series, and each sequence of discrete EEG microstates and fMRI co-activation patterns, an epsilon-machine model can be constructed (Nehaniv & Antonova, 2017). Application of sequence analysis to both the EEG microstates and fMRI CAPs can then be used to not only generate a simple transition matrix, but also identify common sequences of events (events being the occurrence of a given EEG microstate or fMRI CAP in the time series) in both domains, and to create generative models that could be used for prediction and identification of neurodynamics.

Overlap of the states, sequences, and grammars of each of EEG and fMRI states will next be analysed. Importantly, how EEG microstate sequences are associated to CAPs can then be analysed, and long sequences of fMRI states can be investigated. Both of these comparisons have not yet taken place in the field, and investigation will allow for identification of patterns of states in both domains that may be native to specific mental states. Generating a model of such mental states by analysing the sequential patterns at rest will then allow the mind-wandering states that are found at rest to be identified within attention tasks.

3.2 Follow-up Developments of Processing and Analysis Pipeline

Building on the experiences of our initial investigation of the target dataset described above, we refined our methods and tools. We describe here our developed methods and software processing software and their application to simultaneous EEG-fMRI datasets. These serve to generate and relate discrete alphabets and sequences of EEG microstates and fMRI Co-Activation Patterns. This will then serve as the basis for temporal analysis via studying short temporal syntactic information (*n*-grams) and the construction of generative and predictive neurodynamic models (epsilon-machines – see Section 5) in the sequel.

3.2.1 EEG Preprocessing

The EEG pre-processing pipeline was iterated upon to include better tools for cleaning, while still utilizing *EEGLAB*. The previously used BrainVision MR noise removal tool was replaced with the *FMRIB* suite for *EEGLAB* (Iannetti et al., 2005; Niazy et al., 2005), to streamline the pipeline. Magnetic resonance (MR) noise was removed from the EEG signal using the *FASTR* artefact slice removal template. The *FMRIB* suite also provided additional functions for removal of noise components that could not be controlled as in Section 3.1.1. Heartbeat detection was then used with the ECG channel as a reference with the *QRS* and *BCG* tools, which detect and remove artefacts caused by heartbeats. The vibration caused bursts of 17Hz signals in was regressed out for each individual with the *EEGLAB CleanLine* suite (Delorme & Makeig, 2004), which estimates and removes sinusoidal artifacts for each channel using frequency domain regression techniques (Mitra & Bokil, 2007). Artefact subspace reconstruction (ASR) was then used to clean the data of any remaining noise (Miyakoshi et al., 2020). Offset between the start of EEG and fMRI recordings was used to ensure that the recordings aligned temporally, along with an offset of 6 seconds in the fMRI due to the haemodynamic response function (HRF) (Buckner, 1998).

3.2.2 EEG Microstate Analysis

Application of a data-driven technique was employed over the previous year of investigation with application of the *EEGLAB* microstate plugin (Poulsen et al., 2018) rather than KeyPy. For each participant, the EEG topographies at time points of maximum global field power (GFP) (Skrandies, 1990) were collected as input into a k-means clustering algorithm. K-means clustering is applied here to classify EEG time points based on spatial similarity. Two thousand GFP peaks were taken from each of the fifteen participants (two were removed for issues during capture, one for clear outliers during microstate analysis), and variant *k*-means clustering was applied up to a maximum of 500 iterations, with 100 repetitions of each candidate number of clusters. The run with the highest explained variance was used as the result. The process was repeated with 3, 4, 5, 6, 7 and 8 clusters. Measures of fit were used to determine the best number of clusters to use, and the resulting cluster centroids were assessed as to whether they were physiologically feasible. The measures of fit were global explained variance (GEV) and the cross-validation (CV) criterion (R. Pascual-Marqui et al., 1995). Global explained variance measures how similar each EEG sample time point is to the cluster centre it has been assigned to (Murray et al., 2008). The CV criterion calculates an estimator of the variance of residual noise in the fit (R. Pascual-Marqui et al., 1995). Standard parameters were used to measure and compare the microstates. *Duration* is the average amount of time spent in a single microstate. *occurrence* is the mean number of instances of a single microstate's class per second. *Coverage* is the percentage of time covered by a single microstate class. Each of these parameters were calculated at the subject level, and compared across the group.

3.2.3 fMRI Preprocessing

Pre-processing of the fMRI pipeline was largely unchanged, but two new tools were used to refine the pipeline. A group masking was applied which removed both white matter regions, and voxels which were not apparent in all participants. This resulted in noise components that were active outside the brain being removed (note the voxels outside the brain in Figure 9). The global signal, i.e., the mean signal averaged over the whole brain, of each participant was regressed from the time series of each voxel so that the BOLD signal observed was the variance from the mean across the brain. This approach is a heavily debated one (Fox et al., 2009), but its use is retained here due to its utility in CAP analysis (T. T. Liu et al., 2017). Finally, a parcellation was applied to the data. Parcellation takes the average activity across a set of voxels that are labelled with the same region and assigns the mean value to the region rather than the individual voxel values. Existing parcellations were used here: the Schaeffer 1000 cortex (Schaefer et al., 2018), and the Tian S4 sub-cortex (Tian et al., 2020), totalling 1054 individual regions.

3.2.4 fMRI Co-activation Pattern Analysis

Co-activation patterns were generated by applying a k-means clustering algorithm to the time series of fMRI using Python 3.9 and the *scikit-learn* (Pedregosa et al., 2011) and *nilearn* (Abraham et al., 2014) packages. The individual participants were concatenated into a single time series, and the group mask was used to identify voxels of interest. The dataset then had its dimensionality transformed, resulting in a 2D matrix, the first axis being voxels and the second axis being time points. K-means clustering was applied along the temporal dimension, using a maximum of 1000 iterations for 100 repetitions. The best fit was identified for 6, 8 and 10 clusters. Only even numbers are considered due to the common activation/attenuation CAP pairs (X. Liu et al., 2018). Clusters at k=6, 8 and 10 were retained to compare their activation with microstate sequences. Upon generation of a CAP set, a spatial correlation was computed between each pair of CAPs in that set. To this matrix of spatial similarity between CAP volumes was then applied the Munkres algorithm. This combinatorial optimization algorithm was used to assign pairs of CAPs based on the highest level of dissimilarity (Munkres, 1957). The previous generation of CAP analysis as in section 3.1.4 did not attempt to identify pairs, nor did it use low numbers of CAPs. The low numbers of CAPs were chosen since the number of event types that can be used in the construction of dynamical sysetms models (epsilon-machines) must be kept low to avoid an increase in the number of possible sequences leading to a resulting paucity of data to be able capture transitions involving them (see Section 5 for an overview of epsilon-machines). Morevover, the initial "CAPs" generated and reported in Section 3.1.4 were generated from independent components instead of the data itself. These shortcomings were rectified, and the results with the correct implementation as just described are found in Section 4.1.2.

3.2.5 Additional Work: Initial Investigations into Sequence Analysis

It was necessary to review the occurrence of microstate and CAP n-grams. An *n-gram* is defined as a sequence of events of given length *n* (here the events in the temporal sequence are individual EEG microstates or fMRI CAPs, respectively) . The duration parameter, as in section 3.2.2 was also calculated for the n-grams of both microstates and CAPs. N-gram duration is defined as the total amount of time the individual n-gram lasted, simply the sum of each individual event's duration in the n-gram. However, coverage cannot be investigated in this way. Since n-grams greater than length one overlap (the sequence *ABCDE* in length 2-grams would give *AB*, *BC*, *CD*, *DE*), defining coverage as the percentage of the time series covered by the n-gram would not be sufficient. *N-gram frequency* is therefore a novel parameter that is defined here as the number of occurrences of an n-gram over the total possible number of occurrences of an n-gram. For example, if the full input sequence of observations is *ABCDE*, the possible number of occurrences of a length 3-gram would be : *ABC*, *BCD*, and *CDE*. Hence, *ABC*, *BCD* and *CDE* would have one occurrence over three possible occurrences, and hence a frequency of 33.3% each. This definition of n-gram frequency, therefore, accommodates the length of the n-gram considered.

Since CAPs and microstates are co-occurring, it is possible to review the microstate n-grams occurring during specific CAP occurrences. The alignment step outlined in section 3.2.1, combined with an offset of 6s due to the haemodynamic response function (HRF), allowed for an approximate temporal alignment of the microstate and CAP sequences arising in the course of the same underlying neural activity. Once aligned, the periods of activity where each CAP occurred were isolated in the microstate time series, generating a list of microstate sequences that were occurring during each CAP individually.

3.3 Predictive Generative Brain Models: Epsilon-Machines

Details of the development of epsilon-machine models of EEG and fMRI discrete sequences and associated methods, and attepmpts at their application, as achieved so far are reported in the remainder of this report.

4 Results: Progress toward Objectives for EEG/fMRI

Here were report the results achieved so far toward the project objectives focusing on the simultaneous EEG-fMRI dataset. Additional methods and results for epsilon-machines methods applicable to EEG or fMRI datasets for recognizing cognitive states are illustrated in the next chapters.

4.1 Objective 1: Deriving Sets of EEG Microstates and fMRI CAPs

4.1.1 Five EEG Microstates Optimally Explain Variance

Figure 11 shows the set of EEG microstates from 3-8 clusters, along with the global explaned variance (GEV) and cross-validation (CV) score of each number of clusters. While explained variance did increase as the number of microstates was increased, there is an apparent plateau as more are added. Since the CV criterion is an estimator of residual noise, a low score is desirable. As the number of clusters increases, the possibility of noise is amplified. An upward curve develops from five microstates onward. The combination of these two measures concluded that the optimal number of clusters to use was five, with a GEV of 69.5%.

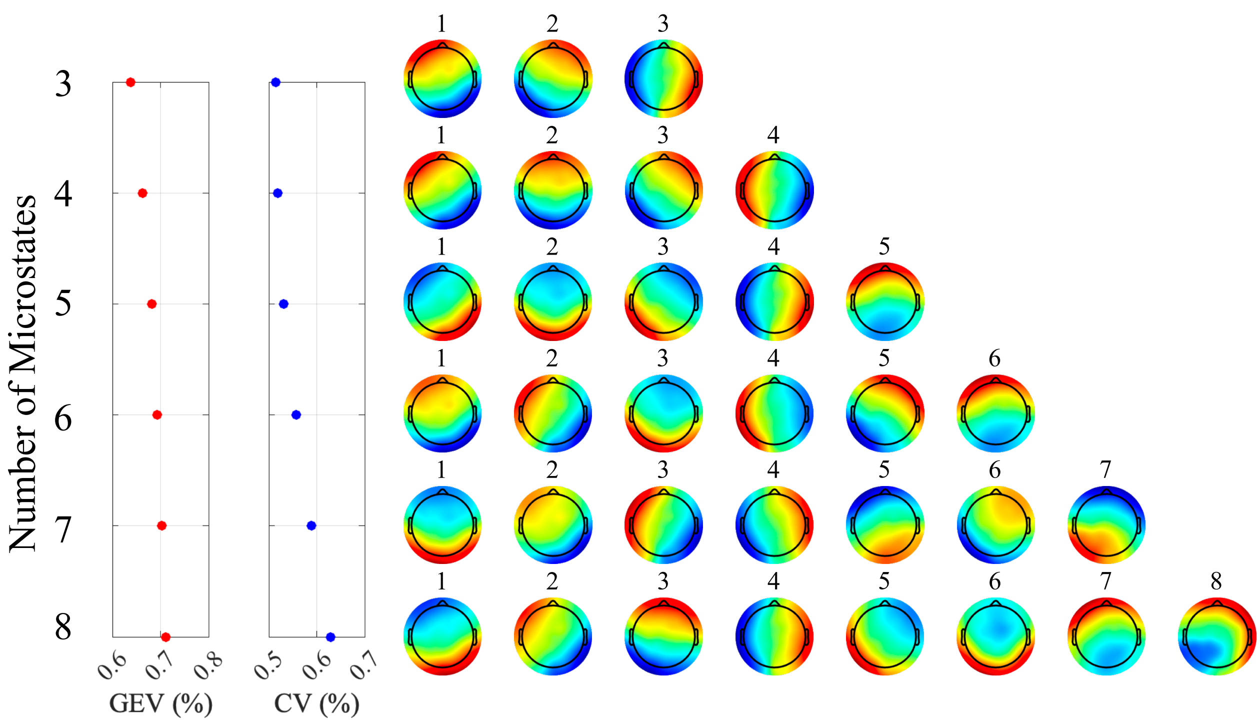

Figure 11: Overview of choosing optimal number of clusters for EEG microstate analysis. The two leftmost panels show the global explained variance (GEV) and cross-validation criterion (CV) score respectively, for each candidate number of clusters. Right panel shows the microstates for each number of clusters. Each row is ordered by most to least explained variance, left-to-right. Microstates are numbered for ease of reference, e.g., microstate 6-2 denotes microstate the in the row for six microstates in the second column .

Parameters were derived from the five chosen microstates through the back-fitting procedure for each participant as outlined in section 3.2.2. Whichever cluster centre each time point was most similar to was labelled with that cluster. This created a sequence of microstates which could then be used to identify the parameters of coverage and duration. Figure 12 shows these two parameters across participants for each of the five microstates. Note that the microstates have been reordered and labelled by the "canonical" set in the microstate literature.

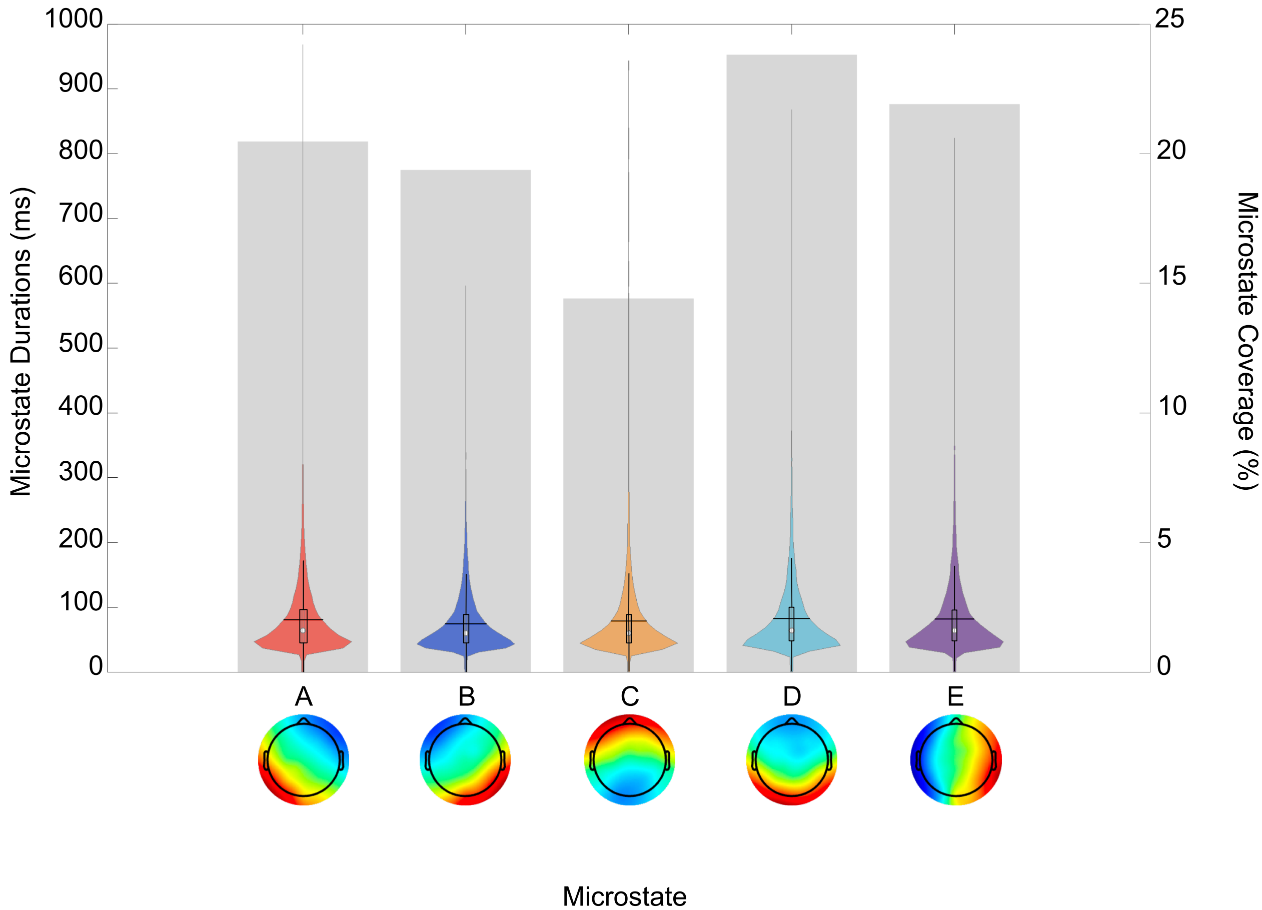

Figure 12: Plot of re-ordered five microstates. Violin plots show distribution of microstate durations in milliseconds, using the left side y-axis. Dash on each violin shows the average duration in each case, dot gives the median value. Colours are unique to each microstate. Width of the violin indicates the number of times the given microstate occurs for the given duration. Grey bar plot shows the coverage of the given microstate using the right-side y-axis.

In the resting state, the first four microstates in the five-cluster set show spatial similarity with the "canonical" set, which has been studied extensively in the existing literature (Koenig et al., 2002; Milz et al., 2016; Milz et al., 2017). Note the increased similarity to the canonical set in Figure 1 versus previous iterations (Figure 7). Microstate E has also been identified previously (Bréchet et al., 2019; Custo et al., 2017; Michel & Koenig, 2018). The average duration across all microstates was around 100ms (Figure 12), also agreeing with existing literature (Khanna et al., 2014; Koenig et al., 2002; Michel & Koenig, 2018; Milz et al., 2016; Milz et al., 2017).

4.1.2 Generating Co-Activation Patterns (CAPs): An Eight-letter Alphabet for fMRI

Figure 13 shows the CAPs using 8 clusters for the resting state. CAPs were also generated for 6 and 10 clusters, which can be found in Supplementary Figures 46 and 47 respectively. When increasing the number of clusters from 6 to 8, unique CAPs were generated which did not obviously belong to a larger cluster using 6 CAPs (see CAPs 1 and 2 in Figure 13). Ten CAPs were not used because pairings were not as obvious, and some pairs of CAPs appeared to be very similar states that likely belonged to the same larger cluster. Additionally, keeping the number of CAPs down is necessary in order to apply epsilon-machine analysis at the fMRI level as this requires more data with increasing alphabet size (i.e., distinct discrete observed event types; which here are the CAPs). The eight CAPs explained 36% of the variance.

4.1.3 Structure and Analysis of the CAPs for Resting State

The first two CAPs show a total attenuation and activation of the sub-cortex, respectively. CAP 1 is most generally similar to an activation pattern of the DMN (Leech et al., 2012). CAP 2, while conversely showing somewhat of an attenuation of the DMN, also demonstrates activation of the more posterior and parietal cortex. Activation of the sensorimotor cortex accompanies the attenuation of the thalamus in CAP 2, an activation pattern described in previous resting state CAP studies (X. Liu et al., 2018).

CAP 3 shows a clear attenuation of the visual network (Schaefer et al., 2018; van den Heuvel & Hulshoff Pol, 2010; Yeo et al., 2011). Interestingly, attenuation is apparent in frontal regions of the frontoparietal control network (FPN) (Menon, 2011) and regions indicative of the limbic system. The strongest regions of activation are in the dorsal and ventral attention networks.

In the subcortex, CAP 3 shows much of the thalamus, hippocampus and caudate nucleus as active, with the putamen and globus pallidus showing attenuation patterns. Conversely, CAP 4 shows an activation of the visual system, along with activation in similar regions of the FPN, but shows the same attenuation of activity as CAP 3 in OFC. Attention networks are also opposed to its pair here, with CAP 4 showing the most substantial attenuation in the same dorsal and ventral areas.

CAPs 5 shows an activation of the DMN and visual network. More lateral regions of the frontal lobe and precentral regions also display this activation pattern. The posterior regions of the dorsal attention network are attenuating here. The FPN and salience networks show attenuation also, as expected by the so-called triple-network model outlined by (Menon, 2011). CAP 5 shows mostly attenuation in the subcortex but interestingly has an asymmetry in activation/attenuation from left to right in the thalamus. CAP 6, the inverse of CAP 5, shows an attenuation of the DMN, with the FPN and salience network showing activation patterns. A clear demonstration of the utility of the CAP method is highlighted here. While the DMN has been derived statically in the past (Menon, 2011; Yeo et al., 2011), the DMN is active in CAPs 1 and 5, but both differ drastically across the rest of the brain.

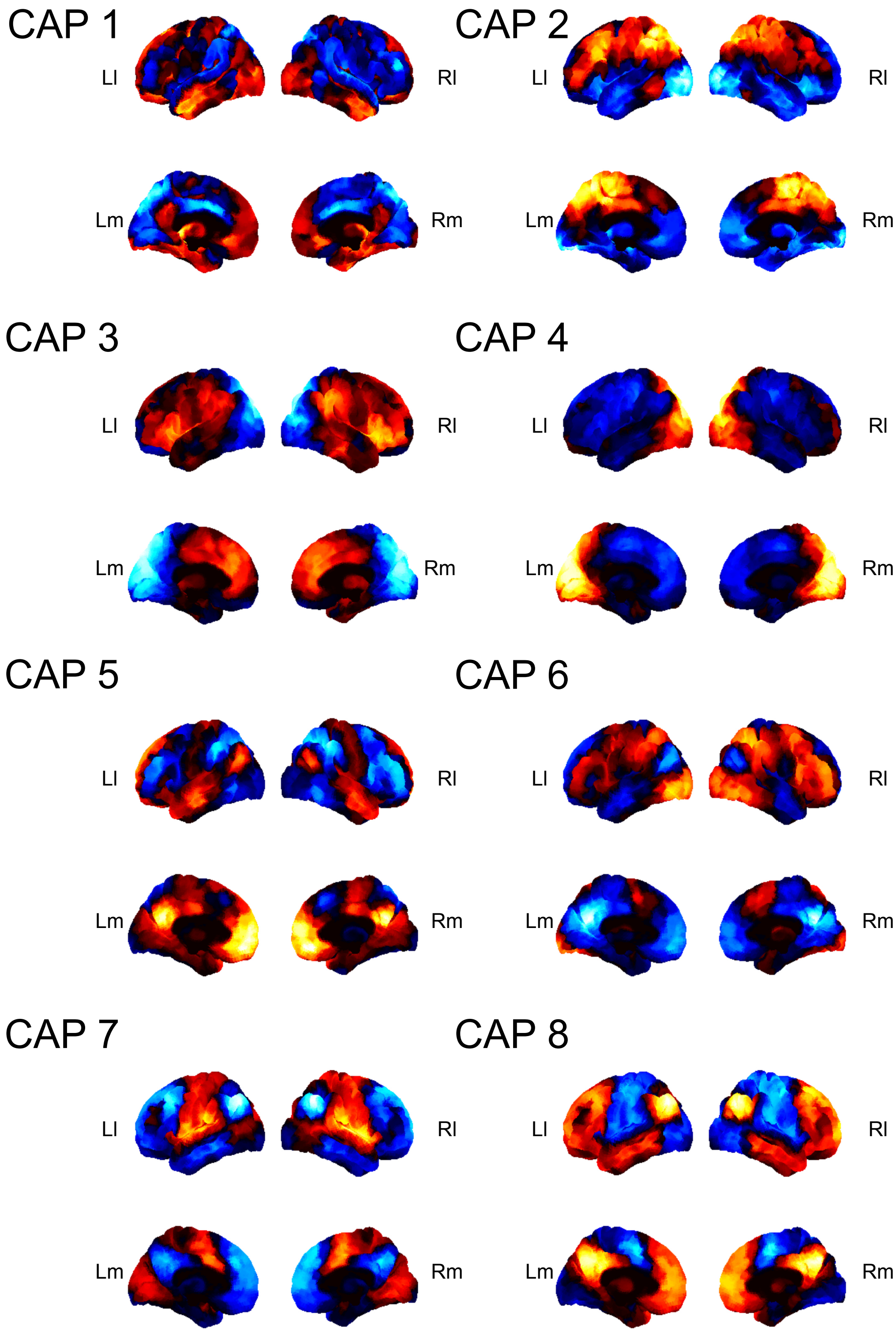

Figure 13: Eight Co-Activation Patterns of the resting-state eyes-open data across participants. CAPs are numbered in their pairs, with each row showing a pair. Brighter red-yellow values indicate higher activation, brighter blue values indicate higher deactivation. CAPs are represented on surfaces with a colour range of -1 to 1 for visualization purposes. L*l* and L*m* are left hemisphere lateral and medial views respectively. R*l* and R*m* are right hemisphere lateral and medial views respectively. The cortical surface is visualized on the *fsaverage5* surface.

CAPs 7 and 8 are representative of the heavily investigated gradient map referred to as the "sensory-association axis", or "principle gradient" (Haak & Beckmann, 2020; Hong et al., 2020; Margulies et al., 2016; Park et al., 2021). The axis shows a functional distinction in activity in unimodal and transmodal regions in the cortex. It distinguishes them as being on a gradient (or spectrum) rather than two distinct regions. The pattern outlined clearly has the same pattern as in CAPs 7 and 8; only in the case of the CAPs do we have an activation and attenuation pattern. It is worth noting here that the patterns shown in CAPs 3 and 4 resemble the secondary gradient of this analysis (Margulies et al., 2016). CAPs may represent the poles of the axes in this study. In the subcortex, CAP 7 has symmetrical activation patterns in the putamen, globus pallidus, amygdala, and ventroposterior thalamus, with the rest of the thalamus showing attenuation, along with the caudate, nucleus accumbens and hippocampus. CAP 8 has symmetrical activation patterns in much of the thalamus, caudate nucleus and the nucleus accumbens.

CAP Similarity, Transitions and Parameters. The spatial correlation matrix between the eight CAPs is shown in Figure 14 (left). The highest correlations are between CAPs 2 and 4, 3 and 6, and 4 and 7. Note the low correlations between the CAP pairs. CAPs were labelled as pairs based on their dissimilarity. CAPs 1 and 2 have the least dissimilarity of all pairs (Figure 14, left).

Additionally, the right panel of Figure 14 shows transition ratios between CAPs. Note the low transition ratios between paired CAPs, suggesting the need to transition into another CAP first to get from a CAP to its opposite. A simple and noteworthy pattern is the high ratio from 3 to 6 and 6 to 3 - suggesting a basic loop between two CAPs, which are relatively similar spatially.

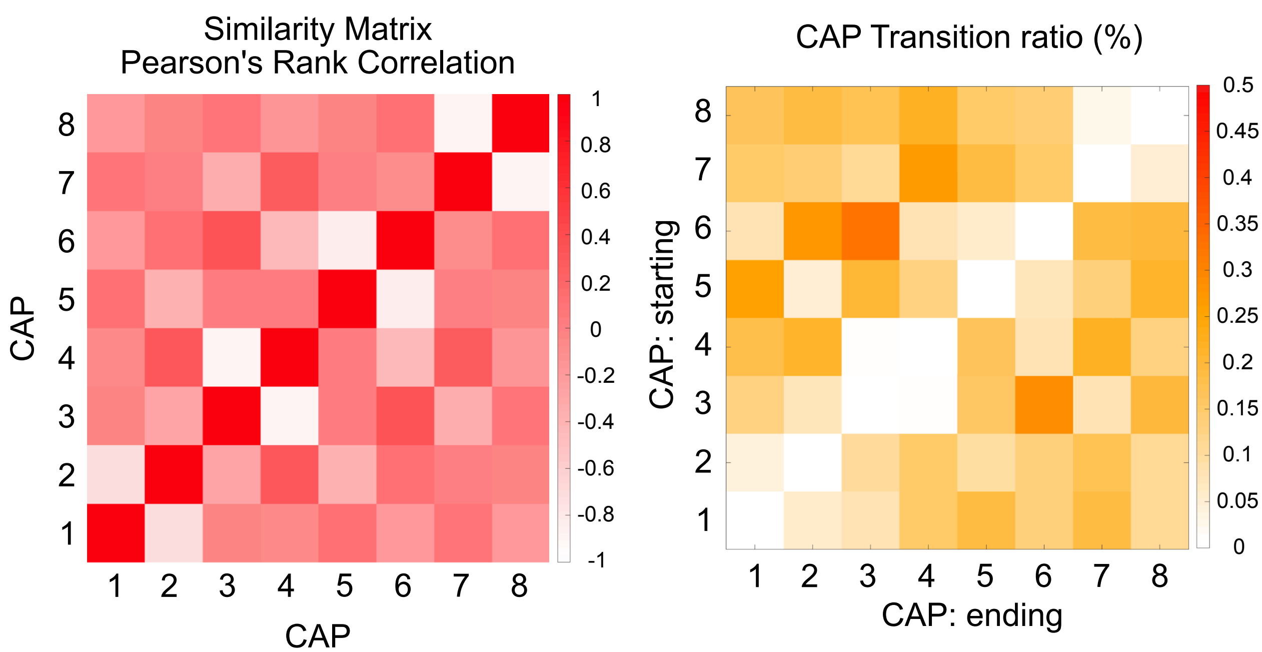

Figure 14: Similarity matrix between each of the eight data driven CAPs (left). Diagonal is comparison of a CAP to itself. Note white boxes between dissimilar pairs. Transition ratios between given CAPs (right) gives starting CAP on y-axis and ending CAP on x-axis. Note again low ratios of transition between paired CAPs.

Figure 15 shows the coverage (left) and average duration (right) of each CAP across participants, for the set of 8 CAPs. Colours indicate pairings of CAPs that were generated by the assignment algorithm outlined in section 3.2.4.

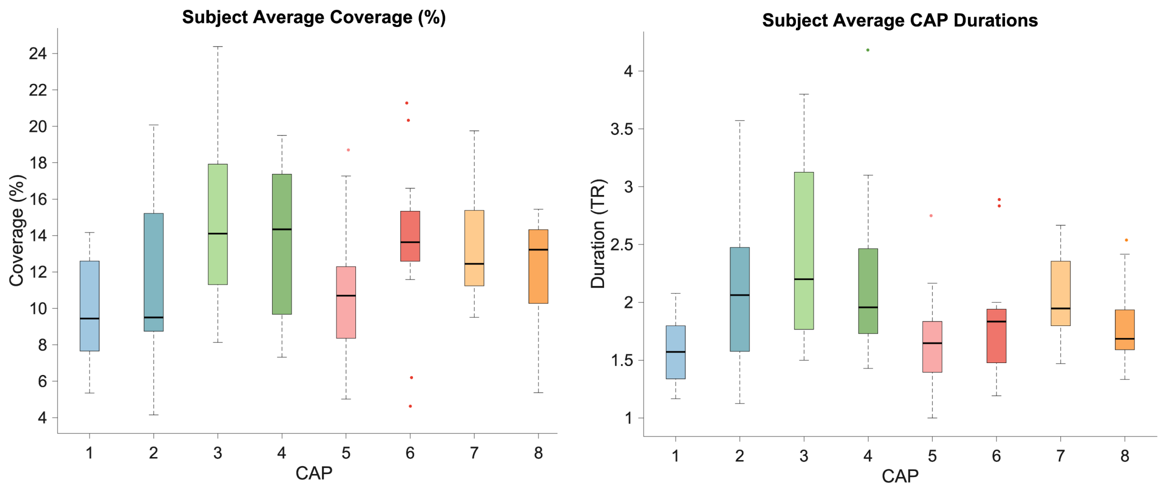

Figure 15: Left shows boxplots of coverage of CAPs across the time series across participants, given as a percentage of time points. Right shows boxplots of the average duration per participant, measured in number of TR's (units of fMRI recording time points, in this case each 2 seconds long). In both cases, mean line is therefore the average of the given CAP across participants for the given metric. Box indicates the interquartile range, stems show minimum and maximum values, dots show outliers, defined as a participant that has a metric more than 3 standard deviations from the mean. Colours are indicative of the pairs of opposite CAPs.

CAP duration and coverage have not yet been reported in the resting state alone, with accounts of duration and coverage being in relation to a patient group (Abreu et al., 2021). Statistical significance of parameters of CAPs could be established in the future.

Comparison with Resting State CAPs Generated from NKI Rockland Sample. Each set of CAPs was correlated against eight CAPs generated from the large NKI Rockland sample (Tobe et al., 2022). This set is shown in Figure 16. A correlation between the NKI set and the set of 8 CAPs was generated with the correlation coefficients given in Figure 17. Note the CAPs 1 and 4-8 all show a strong correlation/anti-correlation with at least one of the NKI CAPs. CAPs 2 and 3 do not show this strong correlation. CAP 3 does show a similar topography to NKI CAP 1 however, despite not showing strong correlations. Additionally, CAP 1 shows a slight correlation/anti-correlation with the NKI pair CAPs 1 and 2, but their topographies do. Note that the visualizations of the CAPs do not include the 54 sub-cortical regions, but the correlations do include them.

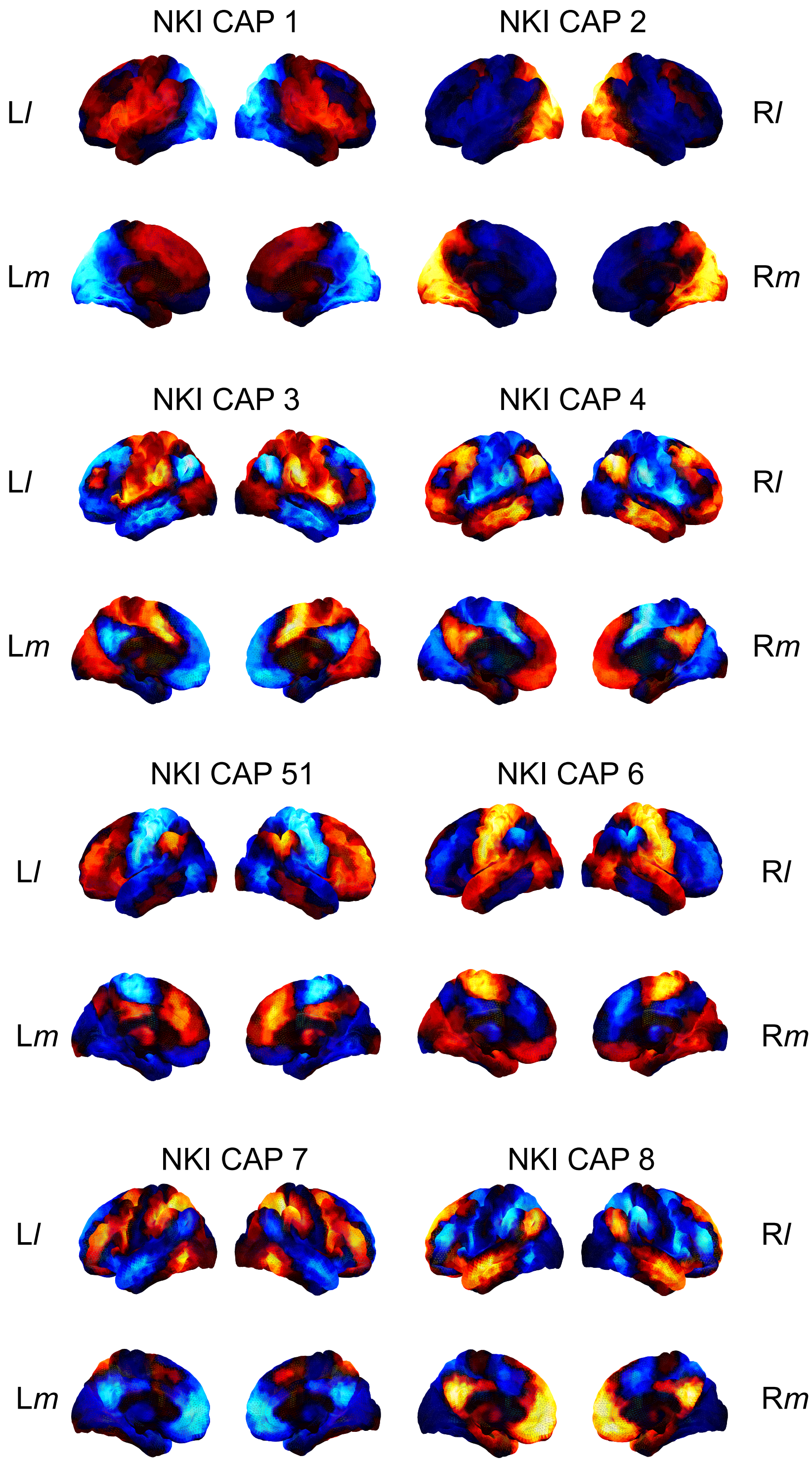

Figure 16: CAPs generated using eight clusters from the NKI Rockland sample, using 721 with each participant being recorded in resting state for approximately 5 minutes each. The same parcellation as in our approach was applied here to control the dimensionality of the dataset. L*l* and L*m* are left hemisphere lateral and medial views respectively. R*l* and R*m* are right hemisphere lateral and medial views respectively. The cortical surface is visualized on the *fsaverage5* surface.

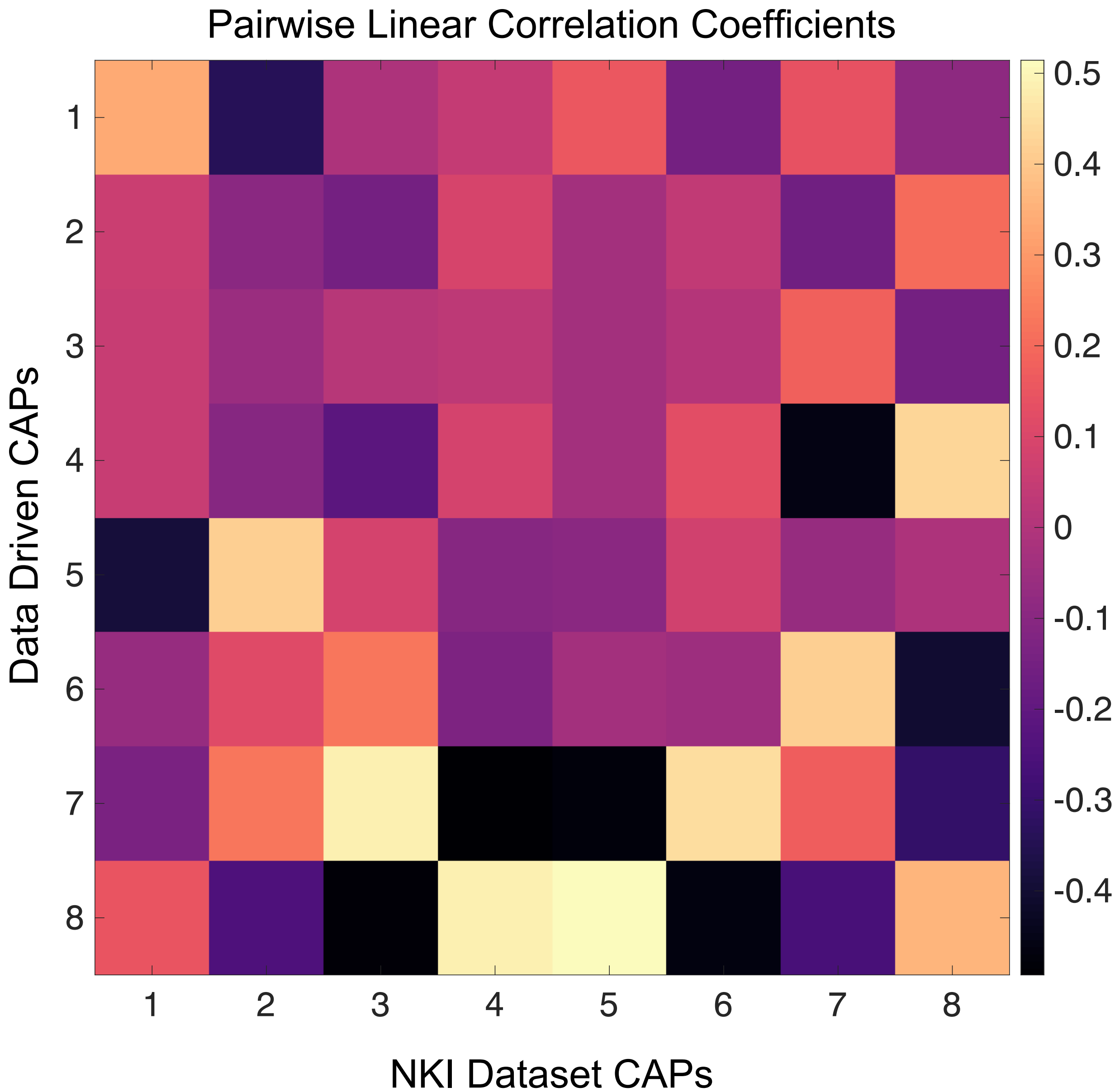

Figure 17: Pairwise linear correlation coefficients calculated between 8 data-driven CAPs in the present dataset and 8 CAPs generated in the large NKI sample dataset. Colour bar indicates the r value of each CAP comparison between datasets.

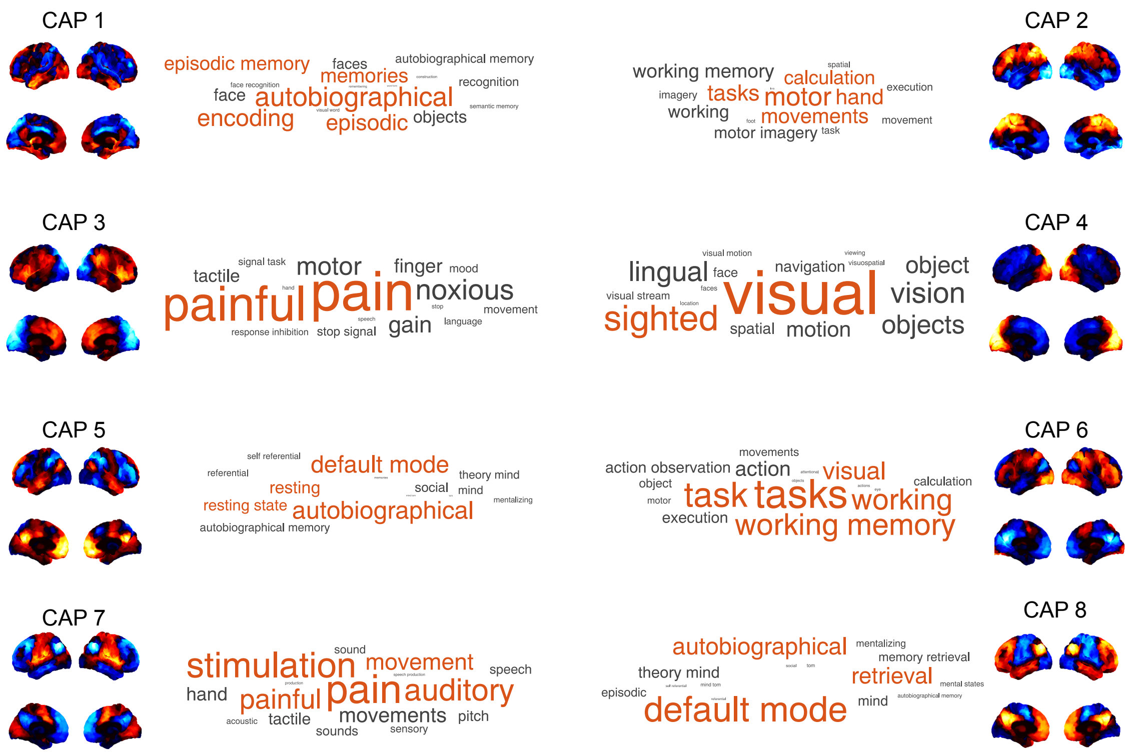

Figure 18: Word clouds indicating the top 15 most correlated topics to each CAP generated using NeuroSynth image decoder. Each CAP has its visualization shown next to its word cloud. Larger words indicate stronger correlations. Orange words indicate most strongly correlated topics for each CAP, where black text indicates less correlated topics. *TOM* is theory of mind.

Topics Correlated to the Resting State CAPs. For further clarification, the resting state CAPs were also assessed with *NeuroSynth* (Yarkoni et al., 2011) in order to associate their patterns with the wider literature in the field, as shown in Figure 18. CAP 1 was associated with various memory tasks and face recognition. CAP 2 on the other hand was associated with motor tasks and tasks in general. The lack of polarity between suggested functions of this pair may be due to their low dissimilarity (Figure 14 left), as well as their low coverage (Figure 15 left). CAPs 1 and 2 may also simply indicate more complex transitory states which are a combination of multiple cognitive activities. CAP 3 showed labelling with somatosensory and motor tasks, where CAP 4 showed visual tasks. Similarly, CAPs 5 and 6 show NeuroSynth labels for resting state, to task execution respectively. CAPs 7 and 8 show a general correlation with unimodal and transmodal activity labels respectively.

4.2 Objective 2: Characterizing the Sequential Dynamics of EEG Microstates and fMRI CAPs using Epsilon-Machines

4.2.1 Microstate N-Grams

Here, the calculation of two microstate n-gram parameters was derived, as outlined in section 3.2.5. The n-gram duration and frequencies of microstate n-grams of length 2 and 3 is shown in Figure 19. The variance of both parameters between n-grams is evident. Colours denote the starting microstate of each n-gram. The variance of parameters appears to increase as longer n-grams are considered. The duration of n-grams naturally gets longer with the length of the n-gram, and coverage gets smaller naturally, too, since the possibility of an n-gram occurring gets smaller in the pool of possible n-grams that could occur.

4.2.2 CAP Specific Microstate N-Gram Parameters

Parameters were calculated for the set of 5 microstates and their n-grams during each CAP from n=1-5. Figure 20 shows the n-gram parameters across each n-gram length.

Noteworthy is the uniformity of frequency and coverage between CAPs, but the greater variance of duration between CAPs as n is increased.

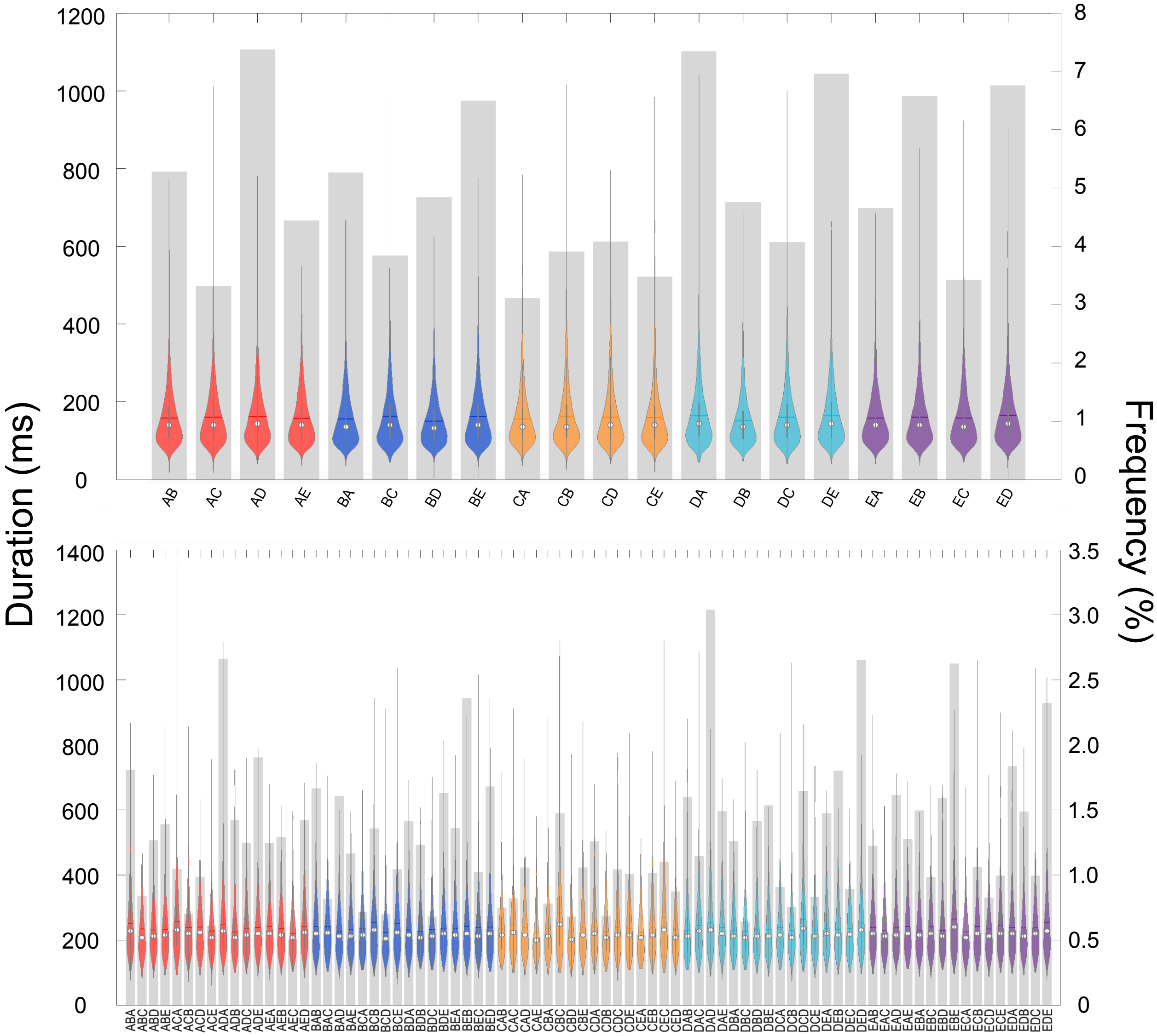

Figure 19: Violin plot of microstate n-gram durations, and bar plots of n-gram coverages of length 2 (top) and 3 (bottom). Colours indicate the starting microstate in the n-gram. Mean durations are more varied as the length of the n-gram increases.

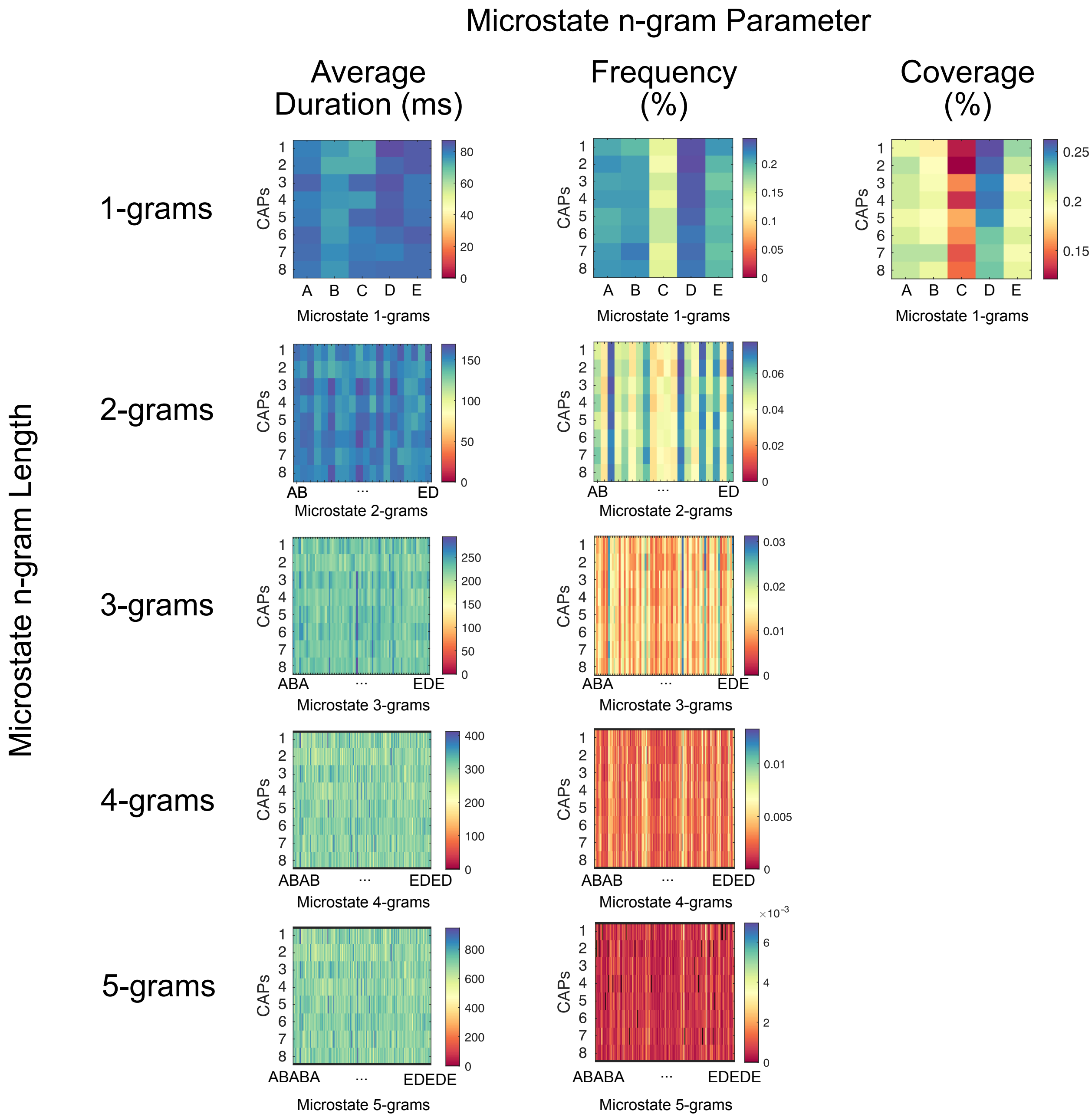

Figure 20: Microstate n-gram parameters during each CAP isolated period of the time series. The first column shows the average duration of each n-gram during each CAP in milliseconds (ms). The second column shows the frequency of each n-gram during each CAP. The third column shows the coverage for each of the microstates (1-grams) during each of the CAPs. Each row shows the parameters for a given n-gram length. Colour bars indicate the values of the given matrix. Note that none of the colour bars in each column or row are uniform, since each matrix is on a different scale. The x-axis labels of 2-grams and greater do not show all labels, but instead show the range alphabetically. There are 20 2–grams, 80 3-grams, 320 4-grams and 1280 5-grams. Black values on the 5-gram matrices indicate a zero - meaning the given n-gram did not occur for the given CAP.

4.2.3 Group-Level Epsilon-machines

This section uses the concept of epsilon-machines, which serve as predictive generative brain models constructed from observed discrete data sequences such as EEG microstate sequences or fMRI Co-activation Pattern (CAP) sequences. *For background and details on epsilon-machines see Section 5 or this project's first year's report* (Nehaniv et al., 2021).

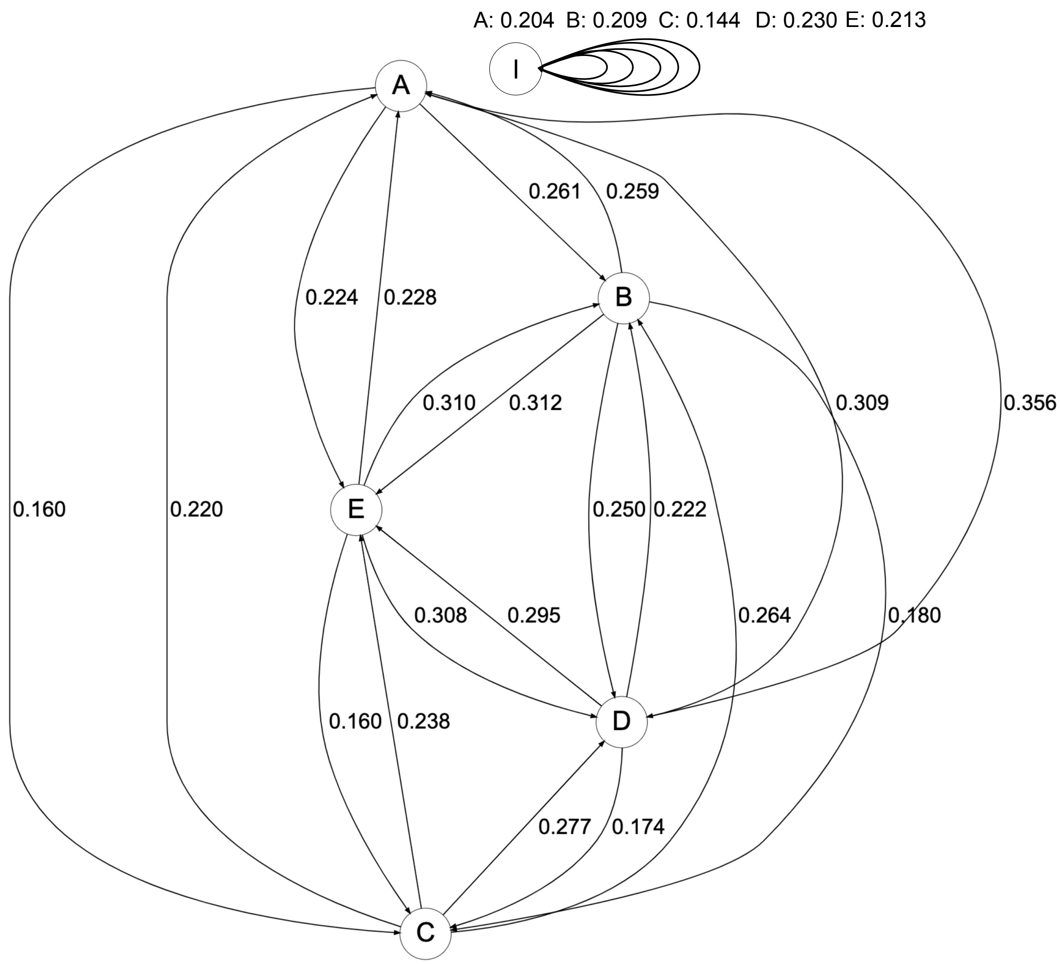

Figure 21: Epsilon automaton of five data driven EEG microstates using 1-grams in the resting state. Causal states are labelled with the microstates that they are associated with. Arrows denote a transition from one microstate to the next. Labels on each arrow are the probability of transition from the starting state to the ending state, given the starting state. All probabilities are rounded to three decimal places. State I denotes the impossible state. Labels next to the impossible state give the frequencies of occurrence of each microstate.

Preliminary investigations were carried out by building epsilon-machines using the five data-driven EEG microstates with 1- and 2-grams across all participants. Figure 24 shows the epsilon automaton for the 1-gram EEG microstates. In this case, minimization yielded no collapsing of states since that causal states correspond to the five microstates which have different probability distributions for the next observed microstate. The probability of transition between microstates is displayed on arrows between causal states. The epsilon-machine built and minimized using microstate 2-grams is given in Figure 22. The number of possible event-based 2-grams from 5 possible microstate observations is 20, which was reduced to 18 in the minimization. The 2-grams *BA* and *EA* were binned in the same causal state, and *BD* and *ED* were binned in the same causal state. The causal state groups and their transition probability distributions are outlined in Figure 23.

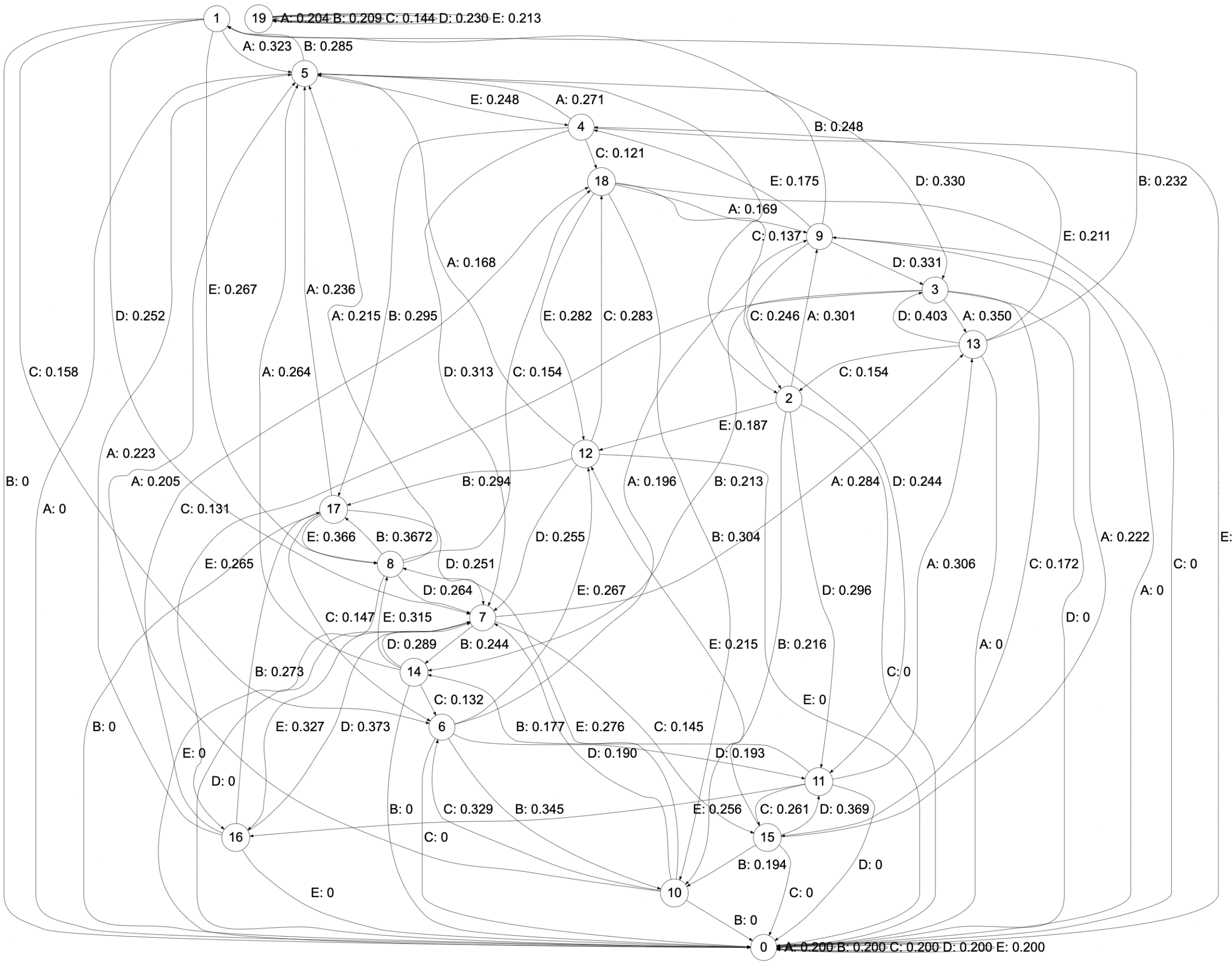

Figure 22: Data driven microstate 2-gram epsilon automaton in the resting state. Each node is a causal state, which contains within it a set of microstate 2-grams. Transitions between the casual states are indicated by arrows, and are labelled with the next observed microstate and the probability of occurrence of this microstate.

Causal State	Microstate 2-grams	
0	IMPOSSIBLE	
1	AB	
2	AC	
3	AD	
4	AE	
5	BA	EA
6	BC	
7	BD	ED
8	BE	
9	CA	
10	CB	
11	CD	
12	CE	
13	DA	
14	DB	
15	DC	
16	DE	
17	EB	
18	EC	
19	UNKNOWN	

Figure 23: Table of microstate 2-grams that correspond to the causal states in Figure 22 for resting state data. Impossible state and unknown or “dead" state are included.

Similarly, event-based 1- and 2- gram epsilon-machines were also constructed for the fMRI CAP sequences across participants. The 1-gram automaton is shown in Figure 24. Note the additional complexity of this representation over that given in Figure 21 due to the 8 CAPs used here versus the five microstates. Minimization of the 2-gram CAP epsilon-machine did not cause the binning of any CAPs, indicating uniqueness of transition probability distributions between all CAP 2-grams in the event-based approach.

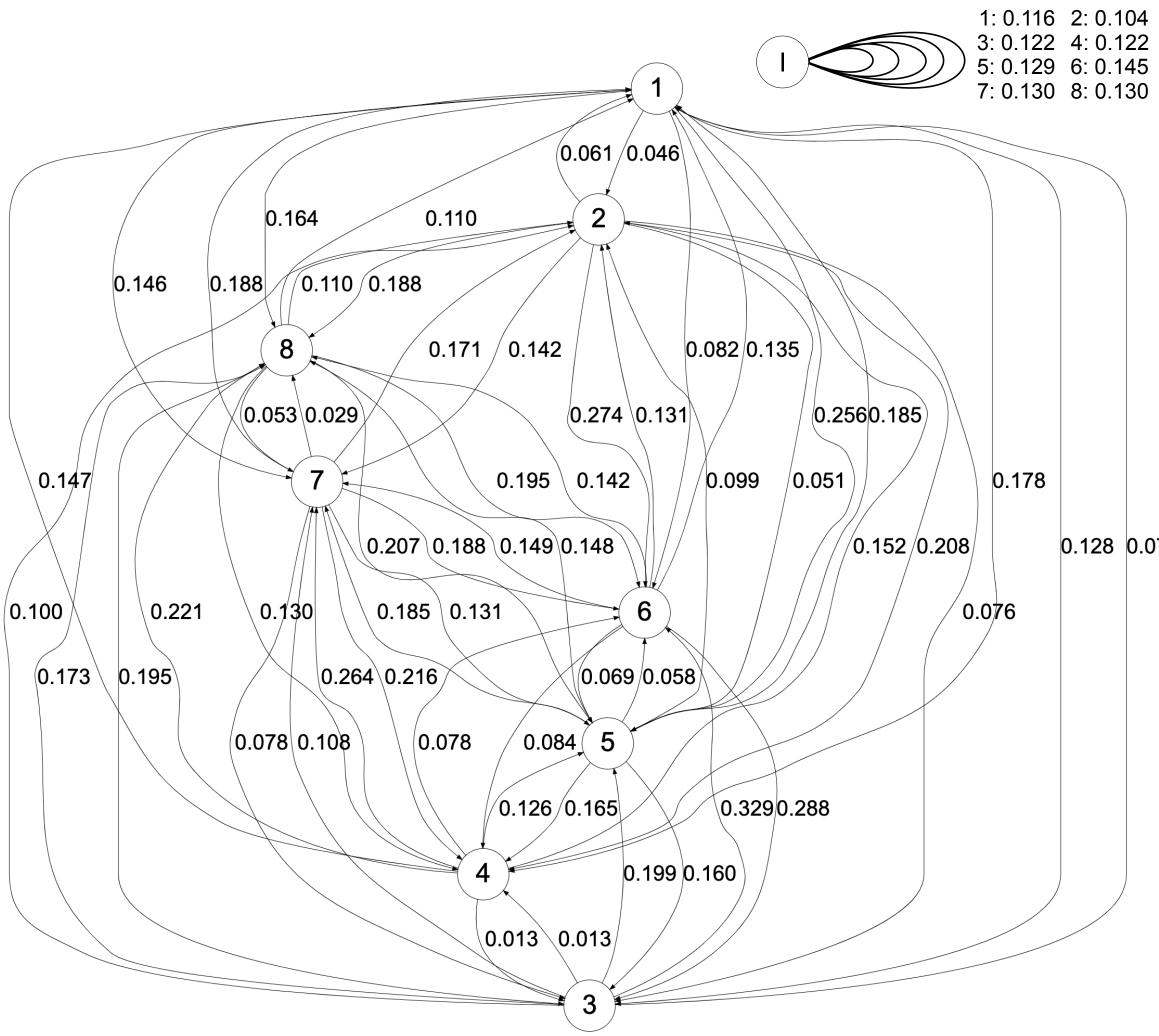

Figure 24: Epsilon automaton of eight data-driven CAPs using 1-grams. States are labelled with the CAPs that they are associated with. Arrows denote a transition from one CAP to the next. Labels on each arrow are the probability of transition from the starting state to the ending state, given the starting state. All probabilities are rounded to three decimal places. State I denotes the impossible state. Labels next to the impossible state are the frequency of occurrence of each CAP.

4.3 Objective 3: Identify Rigorous Mathematical Measures for Characterizing Properties of Generative Automata Models at Individual- and Group-Levels

Rigorous mathematical measures for epsilon-machines include their *statistical complexity*, the $\log_2$ of their number of causal states (Crutchfield, 1994); as well also algebraic characterizations of their dynamics such as the *semigroup of tranformations* the basic events generate on their state spaces; and their *natural subsystems* - maximal permutation groups occurring in these structure (Nehaniv et al., 2015); as well as their *Krohn-Rhodes complexity* - the necessary number of levels with group-computation needed to hierarchically construct the epsilon-machine Rhodes, 2010. An upper bound on complexity is given by the holonomy decomposition computed with the computer algebra package SGPDEC (Egri-Nagy et al., 2014), and if no permutation groups occur this gives instead a bound on the *aperiodic complexity* of the automaton Nehaniv, 1996.

In preliminary work on the simultaneous EEG/fMRI dataset, holonomy decompositions of EEG microstate event-based epsilon-machines constructed using the resting state condition in the target dataset, using length 1-, 2- and 3-grams, were carried out. The decompositions resulted in the lowest number of levels possible in every case: $n+1$, meaning also the lowest upper bound on the level of aperiodic complexity possible for a process that forgets its state after n events. No permutation group subsystems were found at any level.

The same was the case for the resting state CAP sequence data. At n-gram lengths 1-, 2-, and 3 in the event-based group epsilon-machine, the decompositions again had minimal complexity upper bounds in every case, with the number of hierarchy levels in each case being $n+1$. Once again, no permutation groups were found at any level in all cases.

These mathematical measures will be continue to be investigated more thoroughly in coming work (see Section 5 for information on epsilon-machines).

4.4 Objective 4: Apply the Developed Methods to Existing Datasets to Attempt to Differentiate Between Attention and Mind-Wandering to Predict Mind-Wandering Episodes During Attention-Demanding Conditions

Processing and analysing the attention task data was completed in both the blocked and continuous recordings. Five microstates were seen as optimal in both cases. A high similarity is seen across the task recordings topographically in the microstates generated. Figure 25 shows the microstates generated for each task, with Figure 26 showing the spatial similarity matrix between the resting state set and the two task sets of microstates. Note the high similarity score on the diagonal, indicating a consistency in the set of microstates across tasks. The standard parameters of the microstates were also calculated for both CRT tasks.

Eight CAPs were also generated in both the CRT task recordings. Figures 27 and 28 give the CAPs generated in each case. Here unlike in the EEG microstates, there is a lack of consistency between tasks. Spatial similarity matrices are given in Figure 29 to quantify the difference between the resting state CAPs and each of the two CRT tasks.

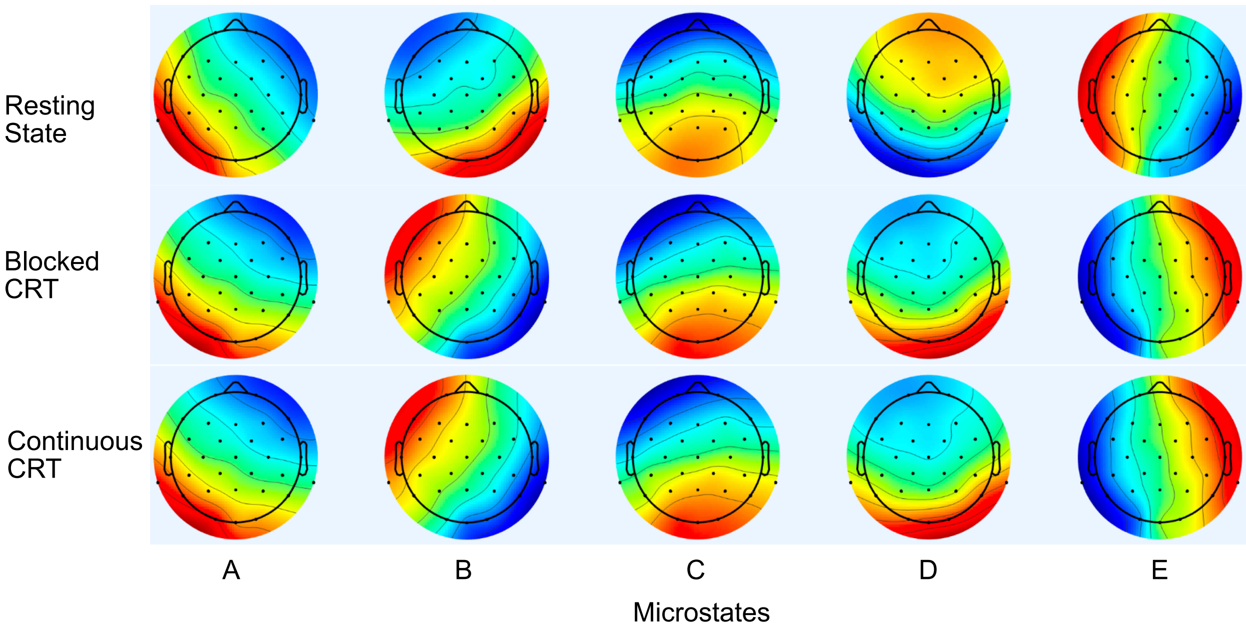

Figure 25: Five data-driven microstates for resting state (top) blocked CRT task (middle) and continuous CRT task (bottom). Each row label is indicative of the canonical state that has been reported in the literature. Note that the location poles of each microstate across tasks is relevant but the +/- locations are not, due to the oscillating nature of microstates.

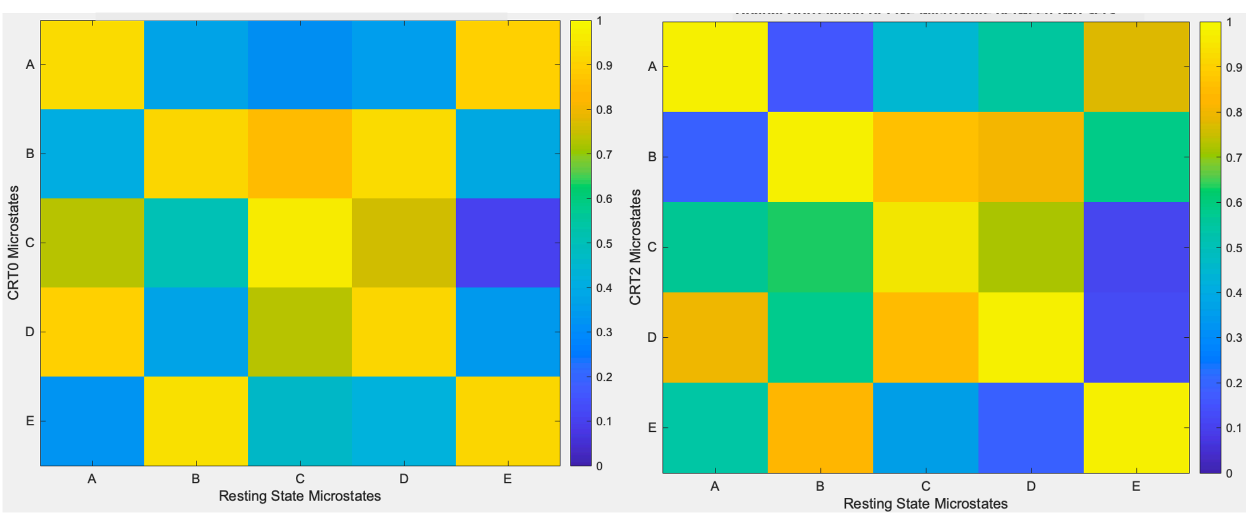

Figure 26: Spatial correlation of the five data-driven microstates for resting state versus the blocked CRT task (left) and resting state versus the continuous CRT task (right). Colour bar indicates correlation score, blue being low, yellow being high.

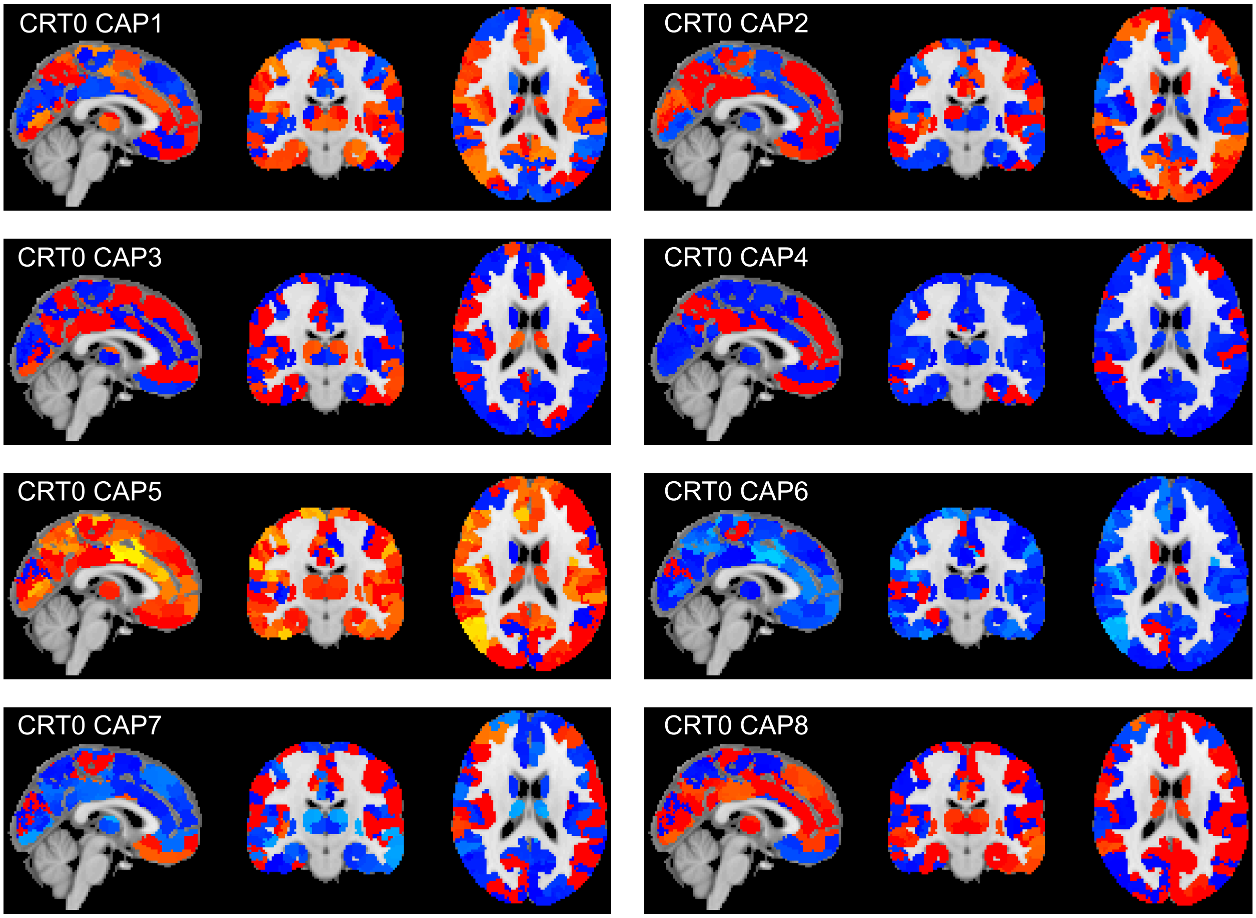

Figure 27: Eight CAPs generated using the blocked Choice Reaction Task (CRT) data. CAPs are in the same row as their respective activation-deactivation pair. CRT0 is used as a shorthand label to specify the blocked CRT task.

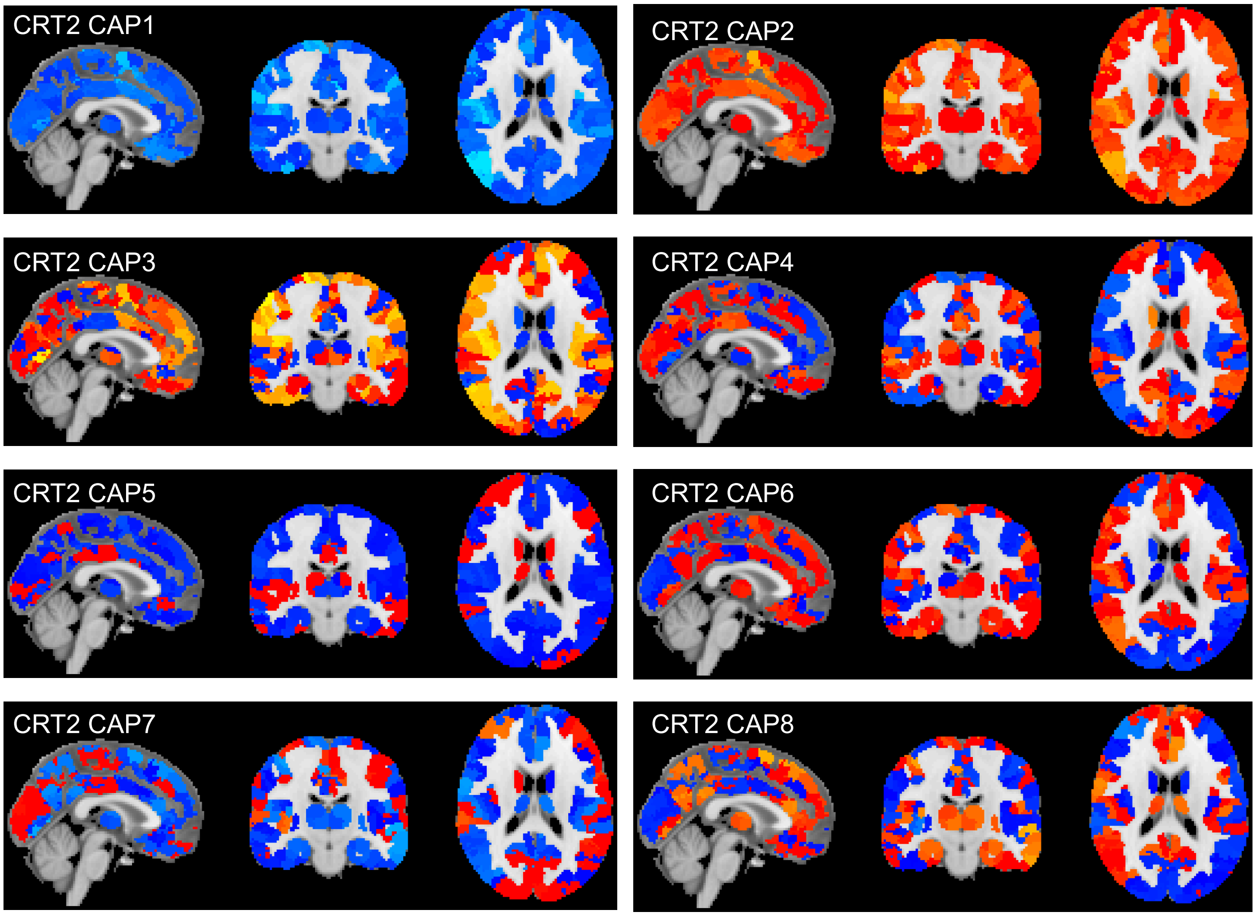

Figure 28: Eight CAPs generated using the continuous Choice Reaction Task (CRT) data. CAPs are in the same row as their respective activation-deactivation pair. CRT2 is used as a shorthand label to specify the continuous CRT task.

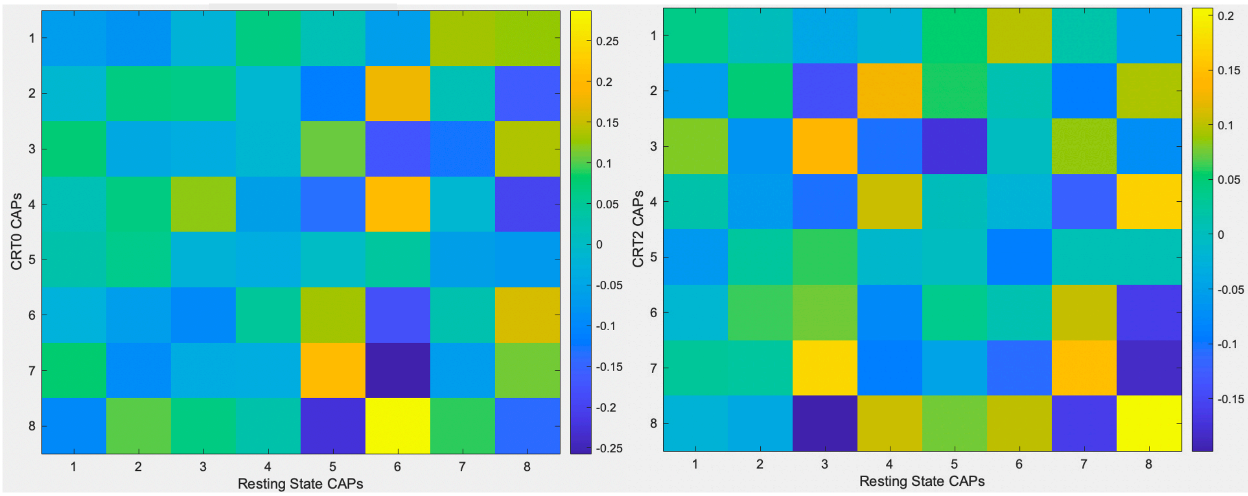

Figure 29: Spatial correlation of the eight data-driven parcellated CAPs for the resting state versus the blocked CRT task (left) and resting state versus the continuous CRT task (right). Resting state is on the x-axis in both cases. Colour bar indicates correlation score, blue being low, yellow being high. Each plot uses the corresponding colour bar.

In addition to these group-level structural differences between EEG microstates and fMRI CAPs in mind-wandering vs. attention-requiring tasks, there are dynamical differences at individual and group level. In Sections 6 and 7, we report on differentiating cognitive processing modes (mind-wandering, verbalization and visualization conditions) using epsilon-machines constructed at individual- and group-levels based on another, eyes-closed dataset described in section 6.1. This includes explorations of metric spaces of these brain models by studying the separation relations between epsilon-machines constructed for different cognitive processing modes, as well as tests of differentiating between individuals. Further, competitive log-likelihood assessments are introduced to detect an unknown discrete EEG sequence of microstates' best-matching brain models. The same techniques can all be applied to CAP sequences from fMRI data.

5 Epsilon-Machines: Optimal Generative Models from Finite Datasets

Following the viewpoint introduced by Nehaniv and Antonova (2017), we construct optimal discrete generative dynamical systems from the observed EEG microstate sequences or from fMRI co-activation patterns sequences from the different recordings to determine syntactic properties of mind-wandering or other cognitive processing modes.

Crutchfield and Young proposed the general method of studying information sources of discrete observations and constructing provably optimal dynamical recognizers/generators called ε-machines that optimally capture or approximate the behaviour of an observed process (Crutchfield & Young, 1989).[1] Originally, the construction supposed access to an infinite past history and a probability distribution over possible futures. Analogous to minimization processes for finite automata, infinite past histories leading to the present moment that all have the same probability distribution over all possible futures are identified as ***causal states*** of the ε-machines. In a given causal state, the different discrete observations ("letters") that are possible are described by their probability distribution of occurrence. A single observation leads from a causal state to the next according to the next observed letter. The state transition is deterministic, but which observation occurs is probabilistic. Causal states make the future conditionally independent from the past.

5.1 Constructing ε-machines from n-gram Transitions Observed in Data

Instead of infinite past histories and infinite possible futures, finite histories are used when working with real-world datasets. We call a sequence of n successive discrete observations of a process an ***n*-gram**, or ***history of length*** n. One partitions the histories a given length (n-grams) into classes with (statistically) indistinguishable futures, according to time series data. We can consider all histories of a given length n and of these only those that occur in the observations under consideration, e.g. within a single 3-minute epoch.

Given such an n-gram $X_1 \dots X_n$ of observations, the dataset may include $n+1$-grams of the form $X_1 \dots X_n Y$ for an observation Y, where Y and each X_i ($1 \leq i \leq n$) are in the discrete alphabet A of possible discrete observations. The frequency of Y given that the preceding n-gram is $X_1 \dots X_n$ defines a probability distribution on the next observation Y conditioned on the current n-gram. This observation occurs with probability $p(Y \mid X_1 \dots X_n)$, as estimated from the data, and one has a deterministic state-transition function between n-grams:

$$\delta(X_1 \dots X_n, Y) = X_2 \dots X_n Y,$$

when Y is observed following $X_1 \dots X_n$.

This construction specifies an automaton that is like an epsilon-machine, except that equivalent states are not merged. We call such an automata a ***pre-epsilon machine***. A pre-epsilon machine is a dynamical model of an observed process that can be constructed from the data before one performs any minimization to merge equivalent states. Minimization of the number of states is important in charaterizing dynamics and studying the complexity of the process under observation. One can identify two states if their probability distributions over all possible futures (sequences of observations over a temporal window after the current state) differ by

[1] Here ε is the small Greek letter epsilon (ε) and refers to a measure of the precision of observation. The precision of observation and more generally the fact that any 'objective' measures depend on apparatus, means, and precision of the observation process are explicitly acknowledged as part of the approach and reflected in the models that result. Indeed, ε-machines model and optimally predict sequences of discrete observations of a system.

no more than some small ***threshold*** $\delta \geq 0$. By iteratively merging states with such equivalent futures one obtains an epsilon-machine whose states are then referred to as causal states (following Crutchfield).

5.2 Three (or rather Five) Levels of Temporal Resolution: Clock, Peak and Event Mode

After processing and discretization in the alphabet of EEG microstates or fMRI Co-Activation Patterns (CAPs), each epoch or condition in an experiments yields an EEG microstate sequence or fMRI CAP sequence. Due to artifact removal, neurodynamic observations in EEG and fMRI broken in contiguous substrings whose syntax we study. We use three different scales of temporal analysis for EEG and two for fMRI yielding five different temporal scales of analysis in total at different orders of magnitude in time.

Here is how we use three time-scales for EEG and two for fMRI when constructing sequences from which to do *n*-gram analysis and epsilon-machine construction:

- Clock Time. Observations of EEG microstates are encoded into strings of the letters corresponding the observed microstate A, B, C, D or E (here at a rate 250 Hz, with each letter corresponding to 4 ms.) For fMRI Co-activation patterns, the timescale is 2000 ms (the scan time) for observation corresponding to the repetition time (TR) of fMRI scanning to record a CAP.

- Global Field Power Peak-based Time. The EEG microstate at each successive peak of the global field power in electroencephalographic recording gives a string of observed microstates at these peaks ($\sim$50 msec per observation or 'letter').

- Event-based Time. The sequence of *distinct* EEG microstates is reported – on average every $\sim$80-120 msec per *event* or 'new letter' recording the newly observed microstate, i.e., when there is a transition to an EEG microstate of different type. This sequence contains no repeated letter XX for any X. Similarly for fMRI, this is the temporal sequence of *distinct* fMRI CAPs. For EEG, we have on average one microstate event every 80 - 100 ms, and for fMRI, we have a new CAP approximately every 2-7 seconds.

5.3 Epsilon-Machine Construction

We give more detail on our epsilon-machine construction and minimization. Let there be a system that is in a continuous dynamic process of change. We may not have any accurate description of the working this system, we can only observe it, but we want to investigate it and build a discrete dynamic model that simulates its behavior.

We suppose the process of change of our model is discrete and our model of system can be in one of a finite set of possible states. There is also the set of possible observations of the system we can make, and we regard these as events that act to transition our model of system from one state to another. Each event occurs with a certain probability in our model given its state, and for each state we know the only state the system will go to after this event. Such a model is called a deterministic finite probabilistic automaton - ***epsilon machine*** $\mathscr{A} = (S, A, \delta, p)$, where S is a state space, A is an alphabet of possible discrete observations (and/or events), $\delta : S \times A \to S$ is a transition function and $p : S \times A \to [0,1]$ is probability function, with $\sum_{a_j \in A} p(s, a_j) = 1$ for each state $s \in S$.

We construct epsilon machines according to the sequences of either microstates of EEG recordings, which are the sets of finite words over the alphabet of EEG microstates $\{A, B, C, D\}$ (or possibly larger alphabets of microstates, or sequences of fMRI Co-activation Patterns - it does not matter - the formalism is the same, as long as the alphabet is a discrete finite set). Each letter of this alphabet is considered as an event, which occurs in our system and transfers it to the next state.

To define the state space of our model, we suppose that the occurrence of each event within the system depends on the ***history*** of a fixed length - the last L events that occurred within the system. So we assume that all possible histories of length L bring the system into different states and the state space S consists of $|A|^L$ possible words of the length L. Any future event from the current state occurs with different probabilities and puts the system into a certain state defined as follows.

Transition function $\delta : S \times A \to S$ from state $w \in S$ which is the word $w = X_1 ... X_L \in A^L$, with the input letter (or event) $Y \in A$ puts the system into a state

$$\delta(w, Y) = X_2 ... X_L Y \in A^L \tag{1}$$

Probabilities of each transition are approximately determined over the entire available set of observations as the ratio of the number of each transition after a specific word to the number of occurrences of this word with a letter after it in observations.

Let us denote for all $a_i \in A$ and $w \in A^L$,

- $num(w)$ - the number of occurrences of word w with a letter after it in data,
- $num(w, a_i)$ the number of occurrences of letter $a_i \in A$ after word w.

Then we have

$$p(w, a_i) = \frac{num(w, a_i)}{num(w)}. \tag{2}$$

There is a problem with this definition of probabilities when the set of sequences we are dealing with contains words that occur only at a sequence in the data with no successor letter since $num(w) = 0$, giving division by zero. Such states are called *dead states*, after them there is no information on the future and it is impossible to determine the probabilities of any event from them. There are several possibilities in the construction, including 1) one can terminate the run of system if a dead state is encountered, 2) one can estimate the probability of the next event from the frequency of occurrence of events in the data. Our model has several rules for handling such states, but these are rather technical and will be described elsewhere. Events of probability zero can be treated as transitioning to an *impossible state* from which there is no return.

As mentioned, probabilities are estimated from the frequency with which a given n-gram is followed by the particular letter as in equation (2), but that in case an n-gram does not occur with any successor letter in the data, then the ε-machine will terminate upon reaching this state. As a convention, the probability of seeing any observation in such a state is treated as zero (even though these sum to zero in a dead state – i.e., outgoing probabilities do not sum to 1).

Also, if an n-gram does not occur, its frequency is zero, and we also assign the value zero to each transition probability from the n-gram. That is, if $num(w) = 0$, we take $p(w, a_i) = 0$ as regards equation (2). *This allows us to regard the ε-machine as having all possible n-grams as states.*

One can then merge states whose probability distributions over possible futures are within a threshold $\delta \geq 0$ of each other as mentioned, and iterate the process to converge to a minimized epsilon-machine whose states are causal states. The result of minimization is provably well-defined, i.e., always yielding the same result. (Full details will be published elsewhere.) This result and the C++ software implementing epsilon-machine minimization are the work of Hanna Derets, and the access to the software is available on request to the University of Waterloo Algebraic Intelligence and Computation Laboratory. Figures 21 and 22 and Figure 24 in Section 4.2 show the models that result from epsilon-machine construction and minimization for EEG microstates and for fMRI CAP sequences, respectively, from the resting state condition of the target dataset.

5.4 Geometry of the Space of ε-Machines

To compare two epsilon machines $\mathscr{E}_1$ and $\mathscr{E}_2$, two types of distance metric between them are defined as follows:

Epsilon Distance Metric We define a distance metric on ε-machines constructed according to transition probabilities from n-grams as in Section 5.3. The *distance* between two such ε-machines $\mathscr{E}_1$ and $\mathscr{E}_2$ is defined as:

$$D(\mathscr{E}_1, \mathscr{E}_2) = \frac{1}{|A|^n} \sum_{\text{all } n\text{-grams } (a_1,..,a_n)} \sum_{x \in A} |p_1(x|a_1,...a_n) - p_2(x|a_1,a_2,..,a_n)|,$$

where $\mathscr{A}$ is the alphabet, and $p_i(x|a_1,...a_n)$ is the probability of letter x being observed after n-gram $a_1,\ldots,a_n$ according to ε-machine $\mathscr{E}_i$ for $i = 1$ or 2. Recall that if no letter has been observed after $a_1,...a_n$ we defined $p_i(x\,|\,a_1,...a_n) = 0$ by convention for all $x \in A$, and the ε-machine terminates upon reaching the state with this n-gram.

This is a distance metric on n-gram transition ε-machines.[2]

We normalize its values by dividing by the number of n-grams, which for our alphabet $\mathscr{A}$ is $|A|^n$. (Note for event-mode, $|A|^n$ needs be replaced by the number of event-mode n-grams, i.e., $|A|(|A|-1)^{n-1}$ to avoid repeated letters in the n-grams.) Normalization helps when we extend this to all ε-machines: We extend to distances to k-gram transition ε-machines as follows. Without loss of generality, suppose $k \leq n$, and let $\mathscr{E}_1$ be a k-gram ε-machine and $\mathscr{E}_2$ describe an n-gram ε-machine, then we define:

$$D(\mathscr{E}_1, \mathscr{E}_2) = \frac{1}{|A|^n} \sum_{\text{all } n\text{-grams } (a_1,..,a_n)} \sum_{x \in A} |p_1(x|a_{n-k+1},..,a_n) - p_2(x|a_1,a_2,..,a_n)|.$$

That is, only the last k letters of the n-gram are considered for epsilon-machine $\mathscr{E}_1$.[3] (As before, for event-mode, $|A|^n$ needs be replaced by $|A|(|A|-1)^{n-1}$.)

Now formally, if $\mathscr{E}_2$ is an ℓ-gram ε-machine (with $\ell \leq n$), we can complete the definition of the metric:

[2] This is essentially an L_1- or "Manhattan" metric, and was introduced in our previous report (Nehaniv et al., 2021).

[3] This metric is valid not just for tables describing ε-machines with n-grams, but any ε-machines whose next state depends only on the last n observations.

$$D(\mathscr{E}_1,\mathscr{E}_2) = \frac{1}{|A|^n} \sum_{\text{all } n\text{-grams } (a_1,..,a_n)} \sum_{x\in A} |p_1(x|a_{n-k+1},..,a_n) - p_2(x|a_{n-\ell+1},a_2,..,a_n)|.$$

To see that the triangle inequality holds, observe that this follows from the fact the first formula above is a metric (essentially the Manhattan distance on functions from the set of $n+1$-grams to non-negative real values in $[0,1]$) and that we are in effect treating the k-gram and ℓ-gram ε-machines as n-gram ε-machines sensitive only to the last k, resp. ℓ observations ($1 \leq k, \ell \leq n$). Symmetry $D(\mathscr{E}_1,\mathscr{E}_2) = D(\mathscr{E}_2,\mathscr{E}_1)$ holds trivially. However, the definiteness requirement for a metric $D(\mathscr{E}_1,\mathscr{E}_2) = 0$ if and only if $\mathscr{E}_1 = \mathscr{E}_2$, means that we must identify ε-machines with the same behaviour. So, technically speaking we have a metric space whose points are classes of ε-machines producing the same sequences with the same probabilities.

Remark: Note that the above metrics are well-defined also for ε-machines for which there exists an n such that a history of the last n observations determines the probability of the next observation (letter) – or, equivalently, whenever the n-gram of most recent observations determines the causal state. [4]

Jaccard Distance between ε-machines. Another distance metric on ε-machines that we define for further analysis is a *Jaccard distance* variation. Here we assume that in addition to the fact that epsilon-machines are constructed according to transition probabilities from n-grams in the sequences of data, we also know some additional information about n-grams in the data: the number of occurrences of the letter $a_i \in A$ after all of the n-grams, as well as the number of occurrence of all n-grams in data. Knowing this information for pre-epsilon-machines we can also maintain this information for minimized machines. The *Jaccard distance* between two such ε-machines $\mathscr{E}_1$ and $\mathscr{E}_2$ is defined as:

$$J(\mathscr{E}_1,\mathscr{E}_2) = \frac{1}{|A|^n} \sum_{w\in A^n} 1 - \frac{\sum_{a_i\in A} \min(n_1([w]_1,a_i), n_2([w]_2,a_i))}{\sum_{a_i\in A} \max(n_1([w]_1,a_i), n_2([w]_2,a_i))} \tag{3}$$

$n_x([w]_x, a_i)$ - the number of occurrences of the event a_i after all of the n-grams w in the state $[w]_x$ in the epsilon-machine $\mathscr{E}_x$, where $[w]_x$ is equivalence class of the word w in the x^{th} epsilon-machine.

Epsilon distance can be used if we want to compare two machines build on the different lengths of history (see (Nehaniv et al., 2021) for details), while Jaccard distance for this use is meaningless. However, if a transition is equally likely in both machines, but much rarer in one epsilon-machine than in the other, the Jaccard distance captures this difference, while the epsilon distance misses it.[5][6]

[4] In particular, we can use this metric to measure distance to a minimized ε-automata such as those produced by CSSR (Shalizi & Klinkner, 2004) and similar methods.

[5] Epsilon distance is the metric introduced in the project's first year report (Nehaniv et al., 2021). Jaccard distance for automata is new (due to authors H. Derets and C.L. Nehaniv), and is introduced here for the first time. Note that computing Jaccard distance requires knowledge not only of probabilities but of event counts.

[6] The original classical Jaccard distance between two non-empty *multi-sets* is defined as

$$d_J(A,B) = 1 - \frac{|A \bigcap B|}{|A \bigcup B|}$$

As this is a metric and we can denote the following:

$$A = \{(x_1,r_1),...,(x_n,r_n)\}, \;\; B = \{(x_1,k_1),...,(x_n,k_n)\},$$

where x_i - are the elements of the multi set $A \cup B$, $r_i \geq 0$, $k_i \geq 0$ - amount of their occurrences in the multi-set A and B respectively.

5.5 Difference between using epsilon and Jaccard distances

By definition of epsilon and Jaccard distances, we see that they consist of differences in the probabilities of the corresponding transitions (epsilon) and differences in the number of occurrences of these transitions (Jaccard) accumulated over all states and events. The probabilities of transitions in the constructed machines are the frequencies of these transitions. The Jaccard distance is introduced to add sensitivity to how typical transitions are to particular machines. For example, a certain state in one machine with alphabet $\{A, B\}$ appears 10 times and the transition A from it occurs in 5 cases, which means that the transition probability is 1/2. In the second epsilon machine, this state occurs 1000 times, and the transition A occurs 500 times, so the transition probability is also 1/2. However, this transition in the first machine occurs much less frequently, which will be taken into account in the Jaccard distance through the difference in the number of transitions (500 - 5 = 495) yielding as summand $(1-(5+5)/(500+500)) = 495/500$ instead of (1/2-1/2=0) for the epsilon distance type - see equation (3) for Jaccard distance.

Therefore, the difference or closeness between machines determined using these two types of distance only indicates the presence or absence of accumulated differences in the above properties.

Factors such as the number of causal states after minimization, the absence of some n-grams in the data of one of the sets, difference in the number of occurances of duration of the microstates in the number of occurrences or in duration data - all these factors increase the difference if they are different in the compared groups.

The classical Jaccard distance formula is equivalent to the following equation:

$$d_J(A,B) = 1 - \frac{\sum_{i=1}^{n} \min(r_i, m_i)}{\sum_{i=1}^{n} \max(r_i, m_i)}$$

Which implies that formally speaking the only difference between the Jaccard distance for epsilon-machines and Jaccard distance for non-empty multi-sets lies into normalization: we its values by dividing by the number of n-grams, which for our alphabet A is $|A|^n$. So, technically speaking in case of Jaccard distance we also have a metric space whose points are classes of ε-machines producing the same sequences with the same probabilities.

6 Neurodynamical Brain Models for Cognitive States

6.1 Another Dataset and Its Processing to Extract EEG Sequences

To identify signatures of mind-wandering and compare two other cognitive processing modalities (silent verbalization, and visualization), we employed another EEG dataset with 26 participants in eyes-closed resting state (unlike the fMRI-EEG eyes-open dataset we have discussed so far). Traditionally EEG and the derived microstates have been studied in eyes-closed conditions - the general trend is that alpha-power is reduced in most people when the eyes are open. This EEG eyes-closed dataset complements the simultaneous EEG-fMRI study and was also employed in the analyses of the project's first year report (Nehaniv et al., 2021), which focused on characterizing the mind-wandering condition. This dataset contains for each participant approximately two 3-minute epochs each of EEG data from interleaved conditions designed to elicit mind-wandering, verbalization and visualization; followed by a second run through the sequence in order to allow us to compare EEG syntactic properties in different recurring conditions.

6.1.1 Participants

A total of 26 participants were recruited via university circular emails and local on-line forums. The inclusion criteria were English fluency, right-handedness (as ascertained using the Edinburgh Handedness Inventory (Oldfield, 1971)), and age between 18 and 65 years. The exclusion criteria were a history of: i) mental health problems, ii) drug and/or alcohol abuse, and iii) neurodevelopmental and/or neurodegenerative disorders. One participant was excluded due to a pre-existing neurological disorder.

The methods development/main analysis sample consisted of 20 participants (Mean age: 40.2, SD=14.63, range: 21-62); Male/Female=12/8; Mean age in higher education: 4.11, SD=.96, range: 2-5).

The validation sample consisted of 5 male participants (Mean age: 21.2, SD=2.28, range: 19-24; Mean years in higher education: 2.2, SD=1.3, range: 1-4).

The study received ethics approval from King's College London Ethics Approval Board (Ref: HR-16/17-4092) and informed consent was gained from the participants prior to participation.

6.1.2 Experimental Paradigm

The experimental paradigm consisted of two identical runs of two repetitions of the same condition sequence: Runs 1 & 2: mind-wandering $\rightarrow$ verbalization $\rightarrow$ visualization (Repetition 1) $\rightarrow$ mind-wandering $\rightarrow$ verbalization $\rightarrow$ visualization (Repetition 2). For the mind-wandering condition, the participants were instructed to relax and allow their minds to wander naturally. For the verbalization and visualization conditions, the participants were asked to repeat silently the word 'square' or to visualize a square, respectively, at a self-paced rate of approximately 2 sec (without silent counting). Each condition was 3 min long, with a total duration of 18 min for each run. Participants had their eyes closed throughout the run. All participants were given a practice run of a full condition sequence lasting 1 min. (For the details of paradigm development and piloting, see (Antonova et al., 2022).)

Abbreviations. For brevity in presentation, we denote the two runs by R1 (Run 1) and R2 (Run 2). Within each run we denote the epochs in the condition sequence just mentioned as MW1, Ver1, Vis1, MW2, Ver2, and Vis2, respectively, in that order. Thus R1-Vis2 denotes second visualization epoch in Run 1.

6.1.3 Experimental Presentation and Procedure

The paradigm (written instructions and an example of a square outline for the visualization condition) was programmed in OpenSesame (Mathôt et al., 2012) and presented on a 24" Dell SE2416H6 LCD monitor. Auditory instructions before the start of each condition were relayed using a prerecorded voice presented through the PC speakers.

6.1.4 EEG Data Collection and Analysis

EEG data recording. EEG data were recorded using a 40-channel Neuroscan Quikcap system (Compumedics, USA). A total of 36 channels were recorded including 4 electrodes for the EOG signal and two references. The EEG electrodes were arrayed according to the standard 10-20 EEG setup. The VEOG electrodes were positioned above and below the left eye, the HEOG electrodes at the outer canthi. The online reference electrode was set to A2, and the ground electrode was at AFz. The data were low-pass filtered at 250 Hz and sampled at 1 kHz.

EEG data preprocessing. The EEG data were inspected in Brainstorm (Tadel et al., 2011), imported into EEGLAB (Delorme & Makeig, 2004), and segmented into twelve 3-min epochs (i.e. 2 runs x 2 repetitions x 3 conditions = 12 epochs). Each EEG epoch was re-referenced to the common average and band-pass filtered (1–20Hz, 4th-order Butterworth filter). Eye-movement artefacts were removed by regressing the EEG channels with the EOG (Croft & Barry, 2002). Obvious (mostly movement and swallowing) artifacts were cut out (ranging between 0 and 11 in number per epoch) and noisy electrodes rejected (0-3 per epoch) before the application of independent component analysis (ICA) on the remaining channels (27-30 per epoch) and time-frames (93,000 – 185,000 per epoch). For each ICA component, the time-course, 2D skull map of inverse weights, and power spectrum were inspected, and the percentage variance of the EEG signal explained by each component was compared to the explained variance of the EOG signal. Standard criteria were used for component rejection based on their spatial profile and temporal-frequency composition. The number of retained ICA components varied between recording epochs (range 7-26) but had little effect on the derived micro-state outcomes, as also observed by others (Dinov & Leech, 2017). Finally, the EEG data were down-sampled to 250 Hz.

EEG microstate computation. The EEG Analysis Toolbox KeyPy (Milz et al., 2016; Milz, 2015) assigns each EEG time-frame to a microstate class based on its 2D spatial organization. Microstate classes can be pre-defined, or calculated from the data using a clustering algorithm. For the present study, both strategies were used. First, as a quality test for our recordings, four data-driven maps were generated by running (200 repetitions of) KeyPy's modified k-means clustering algorithm, and assigning them the labels A, B, C and D based on their correlation with the canonical class maps of Milz et al., 2016. Only the time-frames at the local global-field-power (GFP) peaks were used to minimize the effect of noise, with low-power intermediate time-frames being interpolated by KeyPy. As most of the GFP peaks occur at the crests and troughs of the alpha waves, the average interval between peaks was about 50ms (50.2 ± 3.4 over all epochs for the main group). In a second stage, KeyPy was rerun skipping the clustering phase and assigning directly the peaks of the GFP function to one of the four canonical maps of (Milz et al., 2016), as with the data-driven maps, the

Milz et al. maps were first assigned to the GFP peaks and interpolated in between.

6.1.5 Canonical vs. Data-driven EEG Microstates

Microstate classes can be pre-defined, or calculated from the data using a clustering algorithm. For the present study, both strategies were used in order to be able compare their performance in the analysis of spatiotemporal neurodynamics: **canonical** EEG microstates from the KeyPy library, as well as **data-driven** EEG topographic maps generated for each participant and epoch by running KeyPy's clustering algorithm and assigning them the labels A, B, C and D based on their correlation with the canonical class maps of (Milz et al., 2016).

Canonical EEG microstates and a population average of data-driven EEG microstates from the dataset are shown here in Figure 30.

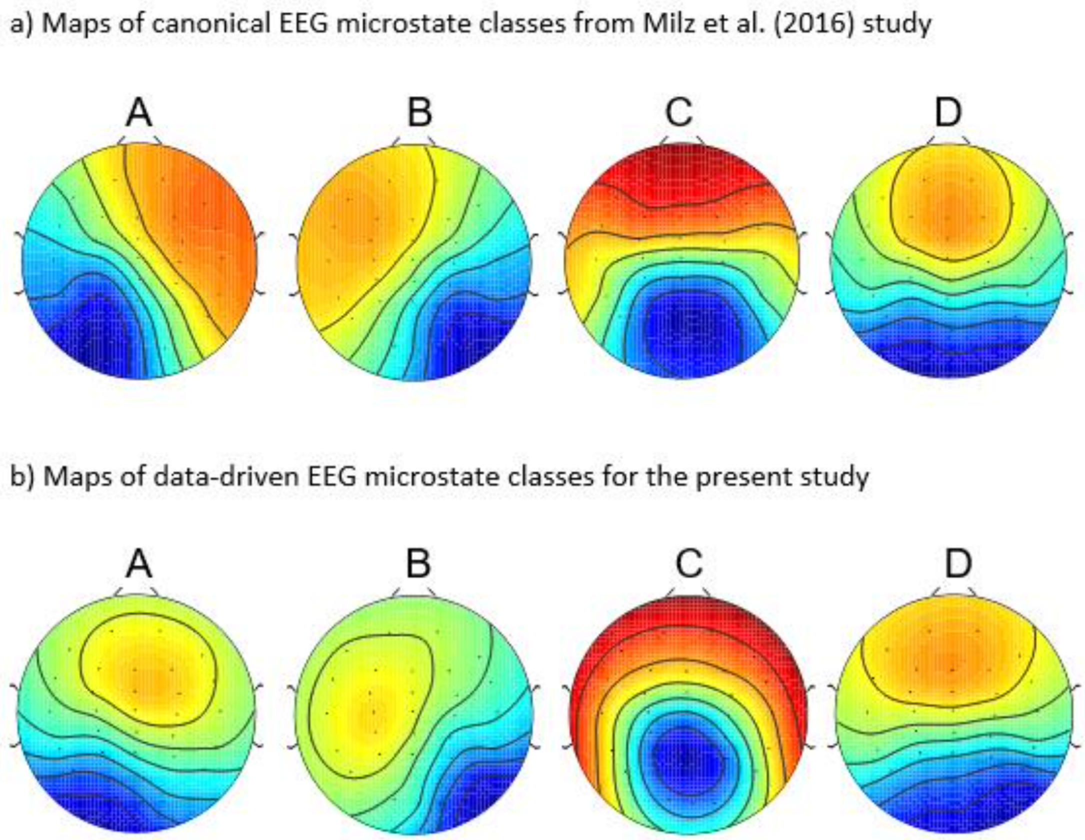

Figure 30: The topographical maps for (a) canonical EEG microstate classes A, B, C, and D of (Milz et al., 2016) from the KeyPy template library (reduced from 64 to 30 channels and rendered with the EEGLAB topoplot function), and b) data-driven EEG microstate maps derived from the present dataset's recordings computed in KeyPy (Milz, 2015) using a variant *k*-means clustering (average from 20 participants across epochs (Antonova et al., 2022)).

The EEG microstates sequences for each epoch were studied to extract syntactic structure at three scales of temporal resolution explained in Section 5.2. In this section and the next we limit the discussion to reporting results using the four canonical EEG microstate.

6.1.6 Epoch, Individual & Group Level Models of Mind-Wandering, Verbalization & Visualization

After processing and discretization of EEG recodings in the alphabet of EEG microstates, each 3-minute session yields an EEG microstate sequence (possibly interrupted by the removal of several artifacts) of approximately 45,000 EEG microstates (180 seconds at 250 Hz). Due to artifact removal, this is broken into contiguous substrings whose syntax we study. We use three different scales of temporal analysis for epsilon-machine construction (namely, clock, peak and event temporal scales as explained in Section 5.2).

Using the method of epsilon-machine construction minimization described in section 5.3 we build generative models of neurodynamics **for each temporal mode** at epoch, individual and group level :

1. For each epoch in our dataset we construct such an ε-machine.
2. Combining the data from a given participant's mind-wandering epochs we can construct an brain model of mind-wandering for that individual. Similary, for verbalization and visualization.
3. For all sequences in Run 1 (respectively Run 2), we can construct a single epsilon-machine from the two visualization epochs (Ver1 and Ver2) combining the data from a group of individuals. Similarly, we can construct two mind-wandering epsilon-machine from the two mind-wandering epochs (MW1 and MW2) of the run from the group; and the this is also done for the visualization epoches (Ver1, Ver2).

6.2 Analysing the Geometric Space of Brain Models

From the set of all files with sequences of microstates of 3-minute EEG recordings of individuals, in this work we consider only the data of the R1 run, which correspond to the 20 participants coded as C01-C10, C12, C14, C17-C18, C20, C22-C26. For R1 of these individuals we have three cognitive processes and 2 repetitions of each. In the work reported in this section, we did not study the individual files separately, but the concatenation of all individual files into one for each of the 6 cognitive process conditions R1-MW1, R1-MW2, R1-Ver1, R1-Ver2, R1-Vis1, R1-Vis2. Using the algorithm described earlier, we build epsilon automata from the concatenated data, and for the comparison of the epsilon machines of different cognitive process conditions we use the 6 by 6 matrices of pairwise epsilon and Jaccard distances. The automata have parameters for building and minimizing, which in the range of all considered values gives us: 3 time modes, 7 history lengths, 21 minimization delta values (starting from $\delta = 0$ all values with a step of 0.05 up to $\delta = 1$) and two types of distance - a total of 882 parameter configurations for minimized machines and 42 pre-epsilon automata, so together we obtain 924 pairwise distance matrices.

6.2.1 Embedding the Metric Space in Euclidean Space

When considering a set of epsilon machines, we pass to a matrix of pairwise distances, however, having only this matrix, it is difficult to say anything about the difference between automata within the set. Therefore, we associate a set of points on a two-dimensional plane with a set of machines, preserving the pairwise distances between them as much as possible. The set of points is found according to the following algorithm, described in detail in the article (Young & Householder, 1938).

1. Let D be a distance matrix $n \times n$, where D_{ij} is the distance between epsilon machines $\mathscr{E}_i$ and $\mathscr{E}_j$.
2. Define matrix M_{ij} as follows
$$M_{ij} = \frac{D_{1j}^2 + D_{i1}^2 - D_{ij}^2}{2}$$
3. Find the coordinates of the points by eigenvalue decomposition of matrix M_{ij}.

   S - diagonal matrix of eigenvalues, U - normalized matrix of eigenvectors, $X = U\sqrt{S}$ - matrix of points corresponding to machines.

   If there exists a set of points in a k-dimensional space $k < n$ that exactly preserves the pairwise distances between machines of the set, then only k columns of X will be non-zero (corresponding to k non-zero eigenvalues of M).
4. For the two-dimensional embedding use coordinates in columns corresponding to the two biggest eigenvalues.

To assess the quality of displaying real distances on a two-dimensional plane, the following numerical characteristics are calculated:

Maximum Local Distortion - maximum relative error of the approximation of the distance between 2D points representing machines.

$$d_L = \max_{i,j} \frac{|D_{ij} - D'_{ij}|}{D_{ij}}$$

D_{ij} - distance between epsilon machines $\mathscr{E}_i$ and $\mathscr{E}_j$.
D'_{ij} - distance between corresponding two-dimensional points, that represent machines $\mathscr{E}_i$ and $\mathscr{E}_j$.

Maximum Global Distortion - maximum absolute error of the approximation of the distance between 2D points representing machines divided by the maximum distance between the machines of this set.

$$d_G = \max_{i,j} \frac{|D_{ij} - D'_{ij}|}{\max_{i,j} D_{ij}}$$

Variance of the selected display directions - what fraction of the original distance falls on the selected projections.

$$V_{all} = \frac{|\lambda_x| + |\lambda_y|}{\sum_{\lambda \in Spec} |\lambda|} \qquad V_+ = \frac{|\lambda_x| + |\lambda_y|}{\sum_{\lambda \in Spec^+} |\lambda|}$$

λ_x, λ_y - two biggest the eigenvalues of matrix M which correspond to the selected columns of coordinates.
$Spec$ - the spectrum of the matrix M, i.e., the set of eigenvalues of matrix M (with multiplicities, if any)
$Spec^+$ - the positive eigenvalues in the spectrum of M only.

Obtaining negative eigenvalues means the impossibility of Euclidean embedding in n-dimensional space, which indicates the loss of information on n-dimensional points, however, in our case, we are still considering a two-dimensional projection, so it may be useful to see what proportion of the total value of non-negative eigenvalues are the two chosen directions.

Maximum Euclidean Embedding Error - maximum absolute error of the approximation of the distance between n-dimensional points representing machines.

$$e = \max_{i,j} \frac{|D_{ij} - D^*_{ij}|}{\max_{i,j} D_{ij}}$$

D_{ij} - distance between epsilon machines $\mathscr{E}_i$ and $\mathscr{E}_j$.
D^*_{ij} - distance between corresponding n-dimensional points that represent epsilon machines $\mathscr{E}_i$ and $\mathscr{E}_j$.

6.2.2 Results: 2D projection of Euclidean embedding for pairwise distance matrices

Having a matrix of pairwise distances between 6 group-level epsilon machines of cognitive process conditions in R1: MW1, MW2, Ver1, Ver2, Vis1, Vis2, we can calculate its Euclidean embedding in 6-dimensional space, since both types of distances are metrics, and then consider how the machines corresponding to the cognitive processes are located on the 2D projection. Such an embedding, however, is not always possible, and sometimes it is possible into a space of smaller dimensions. Considering the success of the Euclidean embedding, we obtain 33 out of 42 successful embeddings in 6-dimensional space for pre-epsilon machines, and 459 out of 882 for minimized machines. For each particular distance matrix we can consider the maximum Euclidean embedding error, as the maximum absolute difference between epsilon machines and corresponding 6-dimensional points, i.e., in the case where we have Euclidean embedding this error equals zero. Calculating the maximum Euclidean embedding error averaged among all failed embeddings we obtain: for pre-epsilon machines it equals $e = 0.0022$ with the standard deviation $\sigma_e = 0.0025$ and for minimized machines - $e = 0.0053$ with the standard deviation $\sigma_e = 0.0065$. For the comparison of the epsilon machines, we do not look at the entire embedding (6 dimensional space), but only at the two columns corresponding to the largest eigenvalues—that is, the 2D projection of the points, as it can be considered even if we do not have Euclidean embedding with the zero error.

As an illustrative example, Fig. 31 shows 2D projections of the Euclidean embedding for distance matrices between group-level pre-epsilon machines for history length $L = 3$ according to epsilon and Jaccard distance types as an example of visually visible the cognitive process condition ‘clustering’ result, while all of the 42 figures are provided in the supplementary materials appendix B.2.

Linear separation or clustering of cognitive processes in the 2D projections has not been verified with an automated computer check in these 2D visualization, but different quantitive notions of separation in the metric spaces (as it is, without projection) are reported on in section 6.2.3. After plotting the points corresponding to the pre-epsilon machines on the plane, we see that the group-level pre-epsilon machines corresponding to 3 cognitive process conditions are not explicitly clustered into 3 well-separated clusters, but visually we can see that on many graphs the visualization machines are located close to each other, verbalization is also sometimes visually separable.

As we consider only two dimensional projections we can check what is V_{all} in distance between ε-machines accounted for by the distances in selected 2D projection averaged among all configurations of the parameters, as well as the biggest difference between pairwise distance matrices of 2D points and epsilon machines i.e., *maximum global distortion* d_G and *maximum local* d_L *distortion* averaged for all distance matrices.

Characteristics of 2D projections are listed in the Table 1, which shows that for minimized machines the quality of displaying information is better on the average sense. In general, the percentage of the information displayed on the plots is quite high, although the spread in the variance of displayed directions for all parameters is also large. Local and global distortions have a small spread for both pre-epsilon machines and minimized machines, and the global distortion of the matrices of pairwise distances between 2D points from the distances between epsilon machines is quite small.

Table 1: Quality measures of displaying information on 2D projections of Euclidean embedding for 6 group-level epsilon machines of participants, corresponding to three cognitive processes of mind-wandering, verbalization and visualization. (See text for definition of measures used.)

Minimization	V_{all}	$\sigma_{V_{all}}$	V_{all}^{min}	V_{all}^{max}	d_L	σ_{d_L}	d_G	σ_{d_G}
no minimization	85.91%	9.54	69.70%	99.94%	0.7797	0.2064	0.5144	0.2812
minimized	91.67%	6.49	68.38%	99.93%	0.6929	0.1580	0.3389	0.2016

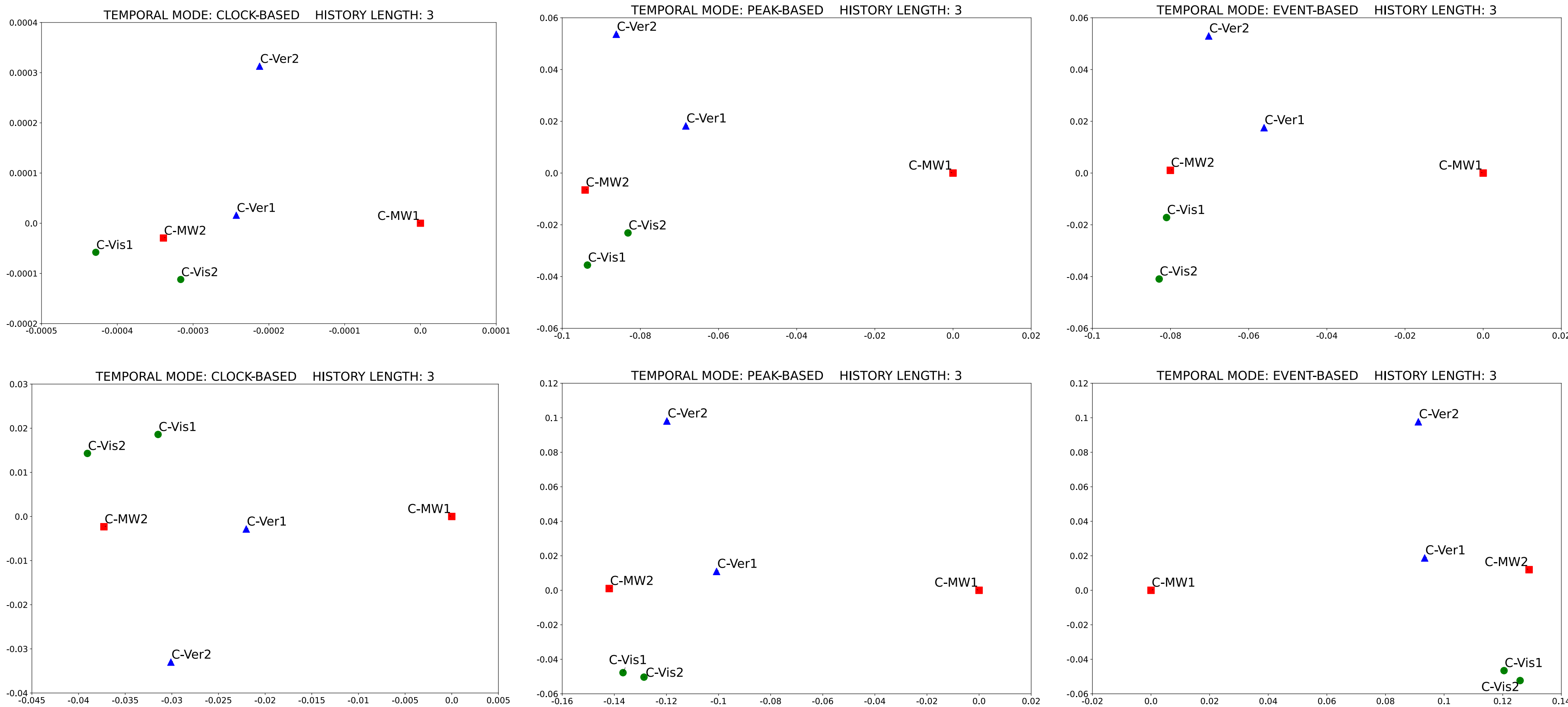

Figure 31: 2D projections of Euclidean embedding for pairwise distance matrices for group-level pre-epsilon machines for mind-wandering (square shaped points), verbalization (triangle shaped points) and visualization (circle shaped points) conditions. Three columns correspond to the three time modes: clock, peak and event respectively. The first line corresponds to epsilon distance type, second to Jaccard, history length is $L = 3$.

6.2.3 Separation of epsilon machines using original matrices of pairwise distances

The transition from matrices of pairwise distances between machines to pairwise distances between points in the 2D plane allows us to consider the visualization of the differences and grouping of automata, however, such a transition entails the loss of some information. To cut off the influence of data loss on the comparison results, we introduce such a concept as the separation between groups of machines, determined only by the original matrices of pairwise distances.

A group of machines A ***is separable from*** group B if the distance between any two pairs of machines $\mathscr{A}$, $\mathscr{A}'$ from group A is less than the distance from any machine $\mathscr{A}$ from group A to any machine $\mathscr{B}$ from group B:

$$\max_{\mathscr{A},\mathscr{A}'}(D(\mathscr{A},\mathscr{A}')) < \min_{\mathscr{A},\mathscr{B}}(D(\mathscr{A},\mathscr{B})),\ \ \mathscr{A},\mathscr{A}' \in A,\ \ \mathscr{B} \in B$$

where $D(\mathscr{A},\mathscr{A}')$ denotes the distance between two epsilon machines $\mathscr{A}$ and $\mathscr{A}'$, which can be both epsilon or Jaccard, depending on the pairwise distance matrix we use.

When one group is separable from another we have ***one-sided separation***, but if the separability condition is met simultaneously for two groups, then we call it ***two-sided separation***. In the presence of such a separation, a numerical measure of it can also be introduced - the separation ratio.

The ***separation ratio*** $R_S(A,B)$ for two groups of machines A and B is defined as the ratio of the minimum distance between machines from different groups and the diameter of a set of machines from two groups:

$$R_S(A,B) = \frac{\min_{\mathscr{A},\mathscr{B}}(D(\mathscr{A},\mathscr{B}))}{\min_{\mathscr{C},\mathscr{C}'}(D(\mathscr{C},\mathscr{C}'))},\ \ \mathscr{A} \in A,\ \ \mathscr{B} \in B,\ \ \mathscr{C},\mathscr{C}' \in C = A \cup B$$

Suppose we are comparing two groups of machines: A, which has n machines, and B, which has m machines. Then, to test the criterion for separability of group A from group B, we perform $(\frac{n(n-1)}{2} \cdot nm)$ comparisons between pairwise distances, i.e., check that the distance between each pair of machines in group A is less than each distance between all pairs of machines, such that one is from group A and the other is from group B. Similarly, to check the separability of group B from group A, we do $(\frac{m(m-1)}{2} \cdot nm)$ comparisons. If all comparisons gave us a positive result, then we have an absolute two-sided group separation, however, if there are pairwise distance pairs that fail the test, then we can see what percentage of them passed the test. We will consider the percentage of pairwise distance pairs that passed the test as a ***percentage of separation*** between groups, which can also be one-sided or two-sided depending on comparisons of pairwise distances that we make.

6.2.4 Results: Separation for cognitive process conditions of mind-wandering, verbalization and visualization

Checking if we can separate the machines of cognitive processes from the rest of the machines in the group we do not need two-sided separation, because we are not interested in the machines of different cognitive process conditions to be as close to each other as two machines of the same cognitive process condition are. In this section we only check one-sided separation of machines of the same cognitive process condition (e.g., mind-wandering machines R1-MW1 and R1-MW2) from other four machines for the other conditions (e.g, from the verbalization and visualization machines: R1-Ver1, R1-Ver2, R1-Vis1, R1-Vis2).

In Table 2 we can see the successful (one-sided) separation cases for mind-wandering and verbalization cognitive process conditions of group-level machines. The biggest separation ratio corresponds to verbalization cognitive process condition and the following configuration parameters: event time mode, history length $L = 7$ and minimization threshold $\delta = 0.2$ when using Jaccard distance type, where the separation ratio equals $d = 0.981078$. Visualization cognitive process condition again shows the best result for separation from the other cognitive procsssing conditions, as it is found in 195 cases and that is why all of them are not included in the table, only parameters for the biggest separation ratio. The biggest ratio for this cognitive process condition is $d = 0.985358$ and it corresponds to the peak time mode, history length $L = 7$, $\delta = 0.150$, Jaccard distance type.

Taking the average of separation percentages for all parameters configurations yields $P_{MW} = 35.03\%$, $P_{Ver} = 56.62\%$, $P_{Vis} = 77.30\%$. As we can see separation percentage for visualization cognitive process condition is higher than for other cognitive process conditions.

In addition to cases where a separation criteria using pairwise distance matrices is met, it is also possible to calculate the percentage ofseparation (defined above) in cases where we do not have separation, to see how it changes with different parameters. As an illustrative example of the impact of choices of temporal mode, distance metric, minimization threshold δ and history length, Fig. 32 shows that the increase of history length very likely will lead to obtaining a bigger percentage of separation for peak time mode. In contrast, the increase of minimization delta will lead to the decrease of percentage of separation eventually, after a possible increase in the first several steps, and drops drammatically for $\delta > 0.4$. For the event time mode, a similar trend is observed for all three cognitive process conditions - the corresponding graphs are presented in the supplementary materials Appendix B.3. At the same time for the clock mode the situation is somewhat different: the percentage of separation is generally poor and is not so sensitive to the length of the history nor the minimization delta threshold when changing parameters, it does not display a clear change trend but, in the case of Jaccard distance, is slightly better for $\delta = 0$ and history lengths $L \geq 2$.

Table 2: Instances of successful one-sided separation for the (two) machines of the same cognitive process condition from the (four) other machines in the set of group-level machines. Each row for mind-wandering (MW) and Verbalization (Ver) conditions describes parameters of one successful separation case, while the row for Visualization (Vis) lists the parameters for the maximum separation ratio among 195 cases of successful separation. Separation ratio is fraction of diameter of the space separating the machines from the other others. L denotes the history length and δ the threshold for epsilon-machine minimization.

cognitive process condition	№	time	L	δ	dist type	separation ratio
MW	1	event	2	0.050	Jaccard	0.537823
Ver	2	clock	6	0.900	epsilon	0.149514
Ver	3	peak	3	0.000	epsilon	0.731105
Ver	4	peak	3	0.050	epsilon	0.770639
Ver	5	peak	3	0.100	epsilon	0.425755
Ver	6	peak	4	0.100	epsilon	0.379429
Ver	7	event	2	0.050	epsilon	0.508708
Ver	8	event	3	0.000	epsilon	0.654215
Ver	9	event	3	0.050	epsilon	0.716513
Ver	10	event	4	0.000	epsilon	0.812468
Ver	11	event	4	0.050	epsilon	0.787321
Ver	12	event	5	0.150	epsilon	0.560754
Ver	13	event	5	0.200	epsilon	0.569803
Ver	14	event	6	0.250	epsilon	0.610883
Ver	15	event	6	0.350	epsilon	0.0876152
Ver	16	peak	2	0.050	Jaccard	0.711039
Ver	17	peak	3	0.100	Jaccard	0.337218
Ver	18	event	2	0.050	Jaccard	0.328331
Ver	19	event	5	0.200	Jaccard	0.651849
Ver	20	event	6	0.250	Jaccard	0.465677
Ver	21	event	7	0.200	Jaccard	0.981078
Ver	22	event	7	0.250	Jaccard	0.943592
Vis	**23-217**	peak	7	0.150	Jaccard	$\max R_S = 0.985358$

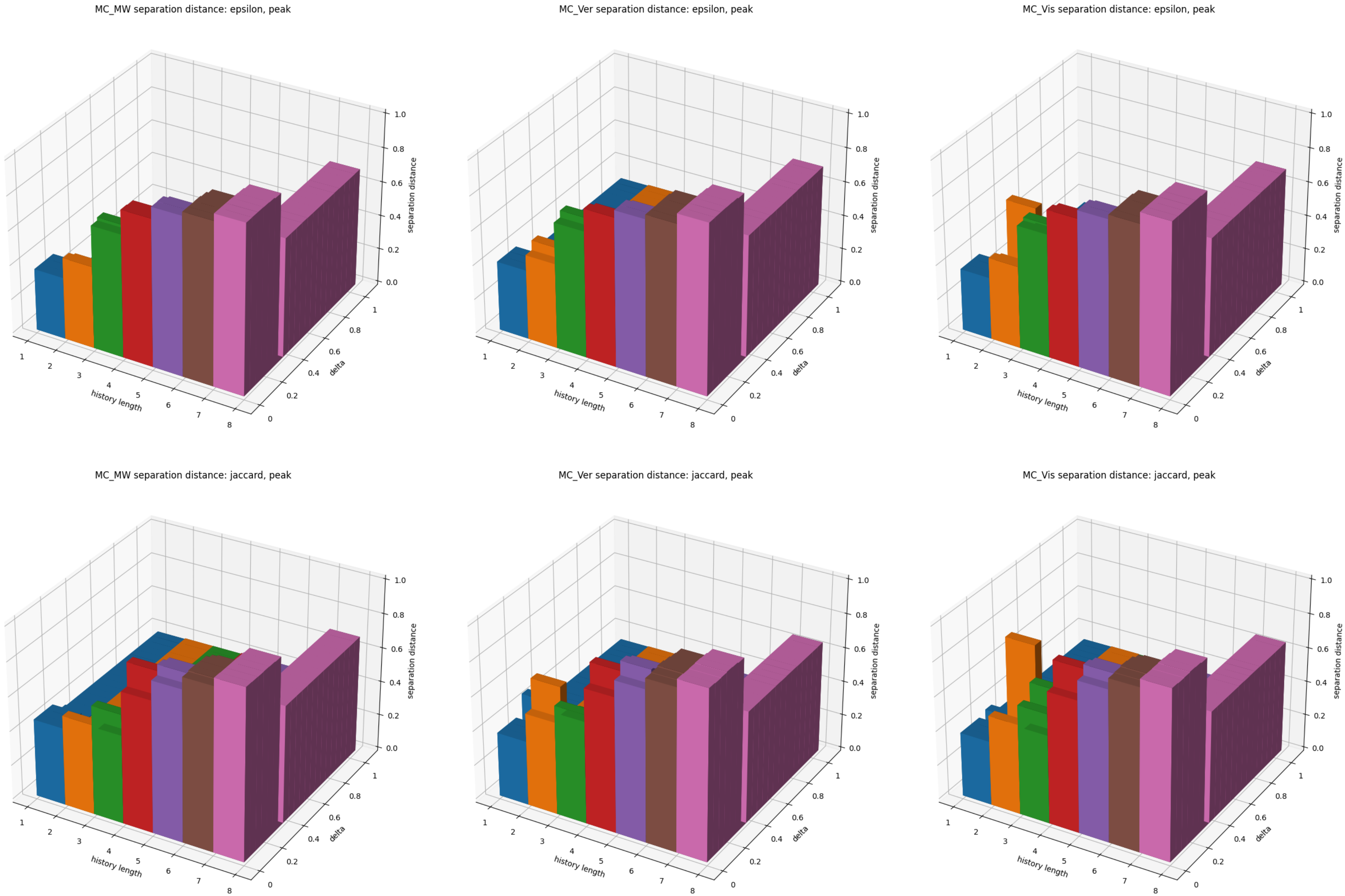

Figure 32: Percentage of separation for the minimized cognitive process conditions machines, corresponding to different history lengths (from $L = 1$ to $L = 7$) and minimization deltas (from $\delta = 0$ to $\delta = 1$ with the step size 0.05). First line corresponds to the epsilon distance type and second line corresponds to Jaccard distance. Three cognitive process conditions: mind wandering, verbalization and visualization correspond to the three columns respectively. Time mode for all represented pictures is peak.

Separation between machines after permuted label assignments.

Also, to check whether the fact that 2 machines of cognitive processes have separated from 4 other machines in the group is not a random phenomenon, we check whether there is a separation between all possible pairs of machines and the remaining 4 machines in the group, i.e., between misassigned pairs of cognitive process machines. In total, for each of 882 parameter configurations, there are 15 ways to select a pair of machines out of 6, three of which correctly correspond to pairs of machines of one process (MW pair , Ver pair, Vis pair). The result shows that absolute one-sided separation was still found in some incorrect machine relabelings, however this phenomenon is not too frequent: in no more than 1-2 cases out of 12 incorrect partitions of a set of machines into groups, for each of 882 parameter configurations.

More precisely, we have $882 \cdot 12 = 10584$ incorrect ways to select two cognitive process machines taking all parameters configurations, out of which 838 gave a separation (7.91%). The average separation percentage for all permutations is $P_p = 48.42\%$, which is quite high compared to the separation percentage of real cognitive process condition machines. To validate the capacity for separation of group machines for different cognitive modes, repeating this experiment with more than two machines per mode would allow for permutation testing potentially to establish the statistical significance of observed one-sided separations.

6.3 Summary

One-sided separation for cognitive processing modes at group level was achieved for all cognitive processing conditions for some parameter values, and was best using peak and event temporal modes, while clock mode generally showed poor performance with little sensitivity to parameters but achieved one-sided . In contrast peak and event epsilon-machine showed increasing separation ratios for all cognitive processing modes as history length increased. However, for minimization thresholds $\delta \geq 0.4$ separation performance was generally but not always negatively impacted. The tentative recommendation is to use peak and event temporal modes, with as large a history length as possible and minimization thresholds $\delta < 0.4$. Mind-wandering was most difficult to achieve one-sided separation for in group machines. This may be since it tends occurs in all conditions. Verbalization could be separated for various parameter settings in peak and event mode using either epsilon or Jaccard distance. Visualization was particularly robustly separated for a large range of parameters, with the best separation ratio achieved using the Jaccard distance metric in peak mode.

7 Recognizing Cognitive Modes and Individuals using Epsilon-machines Generated from EEG Microstate Sequences

7.1 Likelihood Analysis: cognitive process condition Recognition using Epsilon-Machines

In order to verify that the epsilon machines capture in their construction the microstate dynamics characterizing cognitive process conditions: mind-wandering (MW), verbalization (Ver) and visualization (Vis), we can, as one other possible approach, directly execute participants' EEG microsate sequences on the machines and evaluate and compare the runs.

7.1.1 Running an EEG microstate sequence on an Epsilon-Machine

Each state of the machine corresponds to a set of words of particular length L, therefore we can determine the initial state of the computation by finding the state that corresponds to the first L microstates of the sequence. The computation then follows along reading the input sequence letter-by-letter to determine the epsilon-machine transitions, where each has a specified probability.

7.1.2 Log-likelihood of a Sequence

We define a measure of how well an epsilon machine $\mathscr{E}$ matches a sequence $S = s_1 s_2 ... s_n$ as the logarithm of the probability of the sequence had it been produced by the machine $\mathscr{E}$, also called *log-likelihood* ℓ. This probability is the product of the probability of the starting state $s_1 ... s_L$ in the epsilon machine $\mathscr{E}$, built with the history length L, and probabilities of transitions t along the computation, where the whole computation transitions T times. Therefore,

$$\ell = \log_2 p(s_1 ... s_L) + \log_2(\prod_{t=1}^{T} p(t)) = \log_2 p(s_1 ... s_L) + \sum_{t=1}^{T} \log_2 p(t), \tag{4}$$

where $p(s_1 ... s_L)$ is the probability of the L-gram $s_1 ... s_L$ occurring in the data, according to which particular epsilon machine was built. Here we assume that in addition to the fact that epsilon machines are constructed according to transition probabilities from L-grams in the sequences of data, we also know some additional information about L-grams in the data: the number of occurrences for each of the n-gram in data.

7.1.3 Error Conditions and Penalties: Restarts

Sometimes the initial state of the computation cannot be found, because the corresponding state does not exist in the machine. Also, some transitions specified by a microstate in a sequence might have $probability = 0.0$ in the machine. In these cases, we need to move one letter further down the sequence and attempt to restart to a new state, in a similar fashion as for the initial state, and add a penalty to current cumulative log-likelihood. Additionally, there might be a situation where there is a gap in the measured EEG sequence, in which case we also need to restart, however without incurring a penalty. These cases can be summarized as follows:

1. Restart caused by the absence of a state in the epsilon machine. (penalized)
2. Restart caused by an impossible, or zero probability transition. (penalized)
3. Restart caused by a break in the sequence of observations. (not penalized)

L (length)	Clock mode	Peak mode	Event mode
1	0.0102599746210696	0.1274556400506971	0.2062156118885511
2	0.0106796052063905	0.1020918599363347	0.2062156118885511
3	0.0111350251558834	0.0785025666708401	0.1902689873417721
4	0.0116310170091877	0.0630252100840336	0.1531728665207877
5	0.0000154621640845	0.0196078431372549	0.1122448979591837
6	0.0000154621640845	0.0142857142857143	0.0344827586206897
7	0.0000154621640845	0.0117647058823529	0.0192307692307692

Table 3: Minimum observed transition probability values in each length and mode

The penalty is defined as the logarithm of the smallest possible non-zero transition probability in all the machines for a particular word length L and timing mode divided by the base of the logarithm used (2). The values are listed in table 3. The resulting value can also be interpreted as twice as 'bad' as the least likely existing transition.

7.1.4 Ranking and Evaluation

To evaluate how well the machines in a given set of epsilon-machines match a candidate sequence, we need to compare the performance of multiple machines against a single sequence, as ℓ is just a relative measure. This will create an ordered list of machines according to how well they matched the sequence. The values of ℓ in this list can be normalized (denoted as ℓ_{norm}) to the worst performing machine, so that its $\ell_{\text{norm}} = 0$ and other machines have $\ell_{\text{norm}} = \ell - \ell(\text{worst performing machine})$.

7.1.5 Determining the Best Matching Epsilon-Machines from a Set for a Given EEG Microstate Sequence

As we can construct many different machines, it will be significant what particular set of machines we choose to run the sequence on. Epsilon-machines may combine data from many epochs, conditions and/or participants. The sets of machines we consider are constructed from the EEG canonical microstate sequence data from the 3 conditions (cognitive process conditions: mind-wandering, verbalization, visualization), 12 epochs (2 runs x 2 repetitions x 3 cognitive process conditions), and 240 individual epoch recordings (20 participants x 12 epochs) :

- **M3**: One machine per each cognitive process condition, where each combines all runs and repetitions of the particular cognitive process condition for all 20 main participants.
- **M12**: One machine for each cognitive process condition in a particular run and repetition, combining all 20 main participants.
- **M240**: One machine for each cognitive process condition in a particular run and repetition and for a particular participant. That is, no sequences are combined in its construction.

Since we are looking to determine whether we can predict the cognitive process condition of the sequence with the machines and as some sets of machines, such as M12 and M240 in this study, may contain more than one which matches the cognitive process condition, we are interested in more than the first machine in the list. The measures we use are:

- **Top 1**: 100 if the cognitive process condition of machine in first position matches the sequence cognitive process condition, 0 otherwise.
- **Top 3**: Machines in first 3 positions are considered, where each position is weighted (match in first: 100, second: 50, third: 33), the corresponding values are then sum-ed up for each match. The whole value is normalized by $\frac{100}{183}$ so that the maximum attainable value is 100.
- **Rank**: This measure considers matches at any rank, where the final score is calculated from the following values: the sum of ranks of all the machines that match the sequence cognitive process condition and the maximum and minimum attainable scores i.e., if n is the number of matching machines, it would be the sum of the first n ranks and last n ranks. The final score is

$$\text{rank} = 100\frac{\text{score}_{\text{max}} - \text{score}}{\text{score}_{\text{max}} - \text{score}_{\text{min}}}.$$

7.1.6 EEG Microstate Sequence Test Sets

To assess the quality of recognition of candidate EEG microstate sequences using a collections of epsilon machines, we test a set of sequences against the currently selected set of machines, evaluating each sequence according to the measures and then summarizing the results. Two test subsets are following:

- **N5**: The EEG sequences of 5 participants whose data that have not been part of the epsilon-machine construction. These participants coded are as C11, C13, C15, C16 and C19.
- **N20**: The EEG sequences of the 20 participants that have been used to construct the epsilon-machines. These participants are coded as C01 through C26, excluding the above five and excluding C21.

7.1.7 Examples

Figure 33 shows an example where we execute the second run and second repetition of a mind-wandering sequence of a validation (**N5**) participant C11 against the set of three machines that combine the data from all main participants in a cognitive process condition (**M3**). This particular example is in peak mode. We plot the normalized ℓ_{norm} of the sequence on each of the machines, which are ordered along the horizontal from the best matching machine to worst. Red dots signify a match in cognitive process condition. All scores are evaluated and can be seen in the legend of the plot. We note that we attained a match in the first position (Top 1 = 100), which is also captured by the rank score. (Top 3 score is not appropriate in this situation.)

Moving on, the figure 34 shows the same against 12 machines. Note that without combining all the data for a single cognitive process condition into single machine, we have now lost the match in the first position in this example, however the overall result is still very good as the rank score shows.

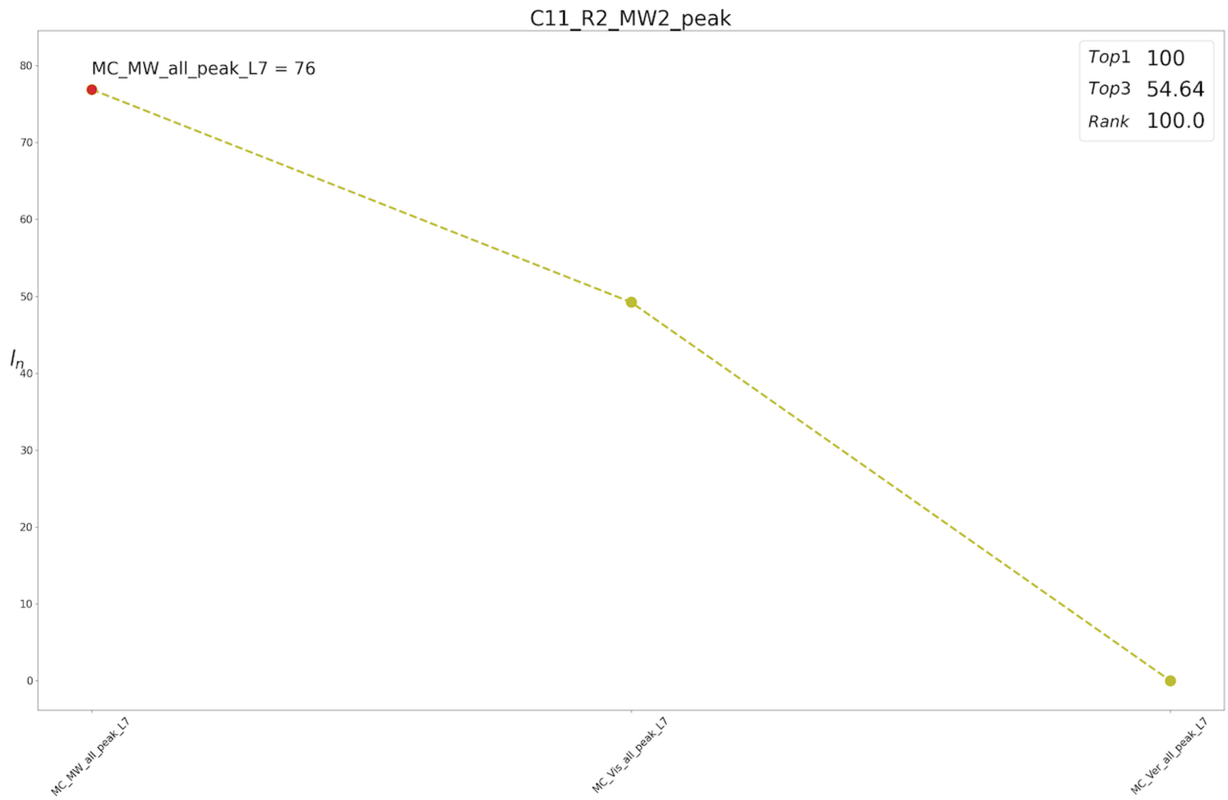

Figure 33: Validation participant C11's EEG microstate sequence from run 2 and repetition 2 of MW task, peak timing mode, run against M3 (epsilon-machines from the MW, Ver, and Vis data of 20 other participants), is recognized as mind-wandering.

Figure 35 shows the same as figure 33, however this time for an a participant whose data was included in the construction of the epsilon-machines. Note the much stronger distinction between the machines: the log-likelihood difference between the first and second best matching epsilon-machines for the validation participant was 30 and for the main participant it is 350. (Scales of the plots are not the same).

Figures 36 and 37 show results in the same repetition, run and cognitive process condition for 12 machines in event and clock modes respectively for validation participant C11. In this case the results are quite similar, compared to peak mode the scales are smaller, i.e., the distinctions less strong. Rank and Top 3 score point us to conclude that the cognitive process conditions were best recognized in event mode.

Finally, figure 38 shows results for M240 and a matched participant C01. As this participant has a corresponding machines in the pool of 240, it clearly stands out in the results.

To conclude, these examples show just a few trends and observations that can be made looking at particular sequences run against different sets of machines. In the next section, we do similar runs for the groups of participants whose data were included (N20) and not included (N5) in the epsilon-machine constructions, summarizing the results for many sequences to obtain generalized results.

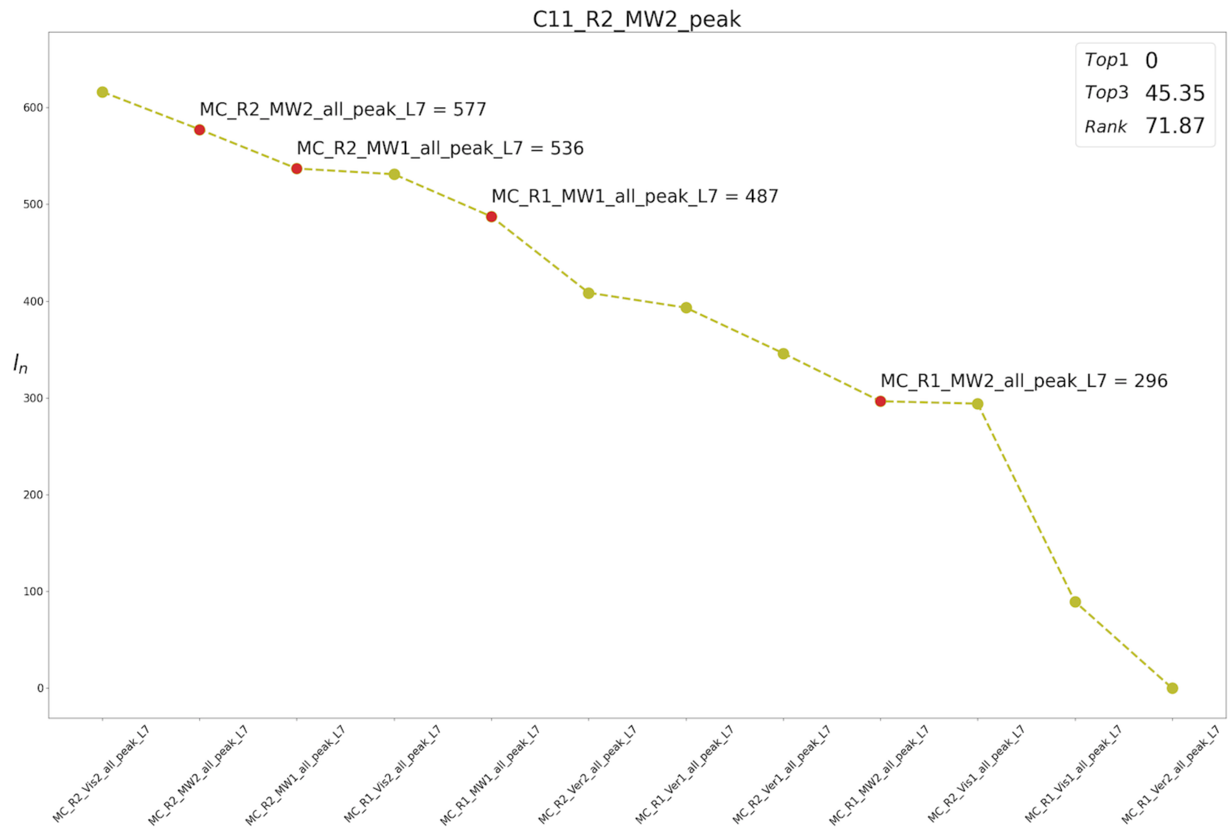

Figure 34: Validation oarticipant C11's EEG microstate sequence from run 2 and repetition 2 of MW task, peak timing mode, run against M12 (12 epsilon-machines from the MW, Ver, and Vis epochs of 20 other participants), shows rank score of 71.87 for mind-wandering.

7.2 Summary Results

The following results bar charts are obtained by running multiple sequences against differing subsets of machines in the different timing-modes and for different word lengths. The resulting scores are averaged for each category and the average is showed in the bar charts.

The figures 39 and 40 show the results when we limit ourselves to Top 1 score. Figure 39 shows the results, when the sequence has been included in the machine construction while figure 40 uses the set of sequences that have not been included in machine construction. It can be easily noticed that by including the sequence in machine construction, the results are much improved. Especially when we consider M240, where we do not combine sequences to construct machines, whatever the length of the words or timing mode, the best ranking machine is always the one that corresponds to the sequence. Once we combine some sequences to create M12 and M3, the results are almost unmistakably good for lengths of word larger than 6, except for clock mode. When we look at sequences that have not been part of the machines (figure 40), the results are more varied. Especially exceptional are M12 machines at recognizing Verbalization in clock and peak modes.

Let us also look at histogram which uses other than Top 1 measure. Figure 41 shows the performance of

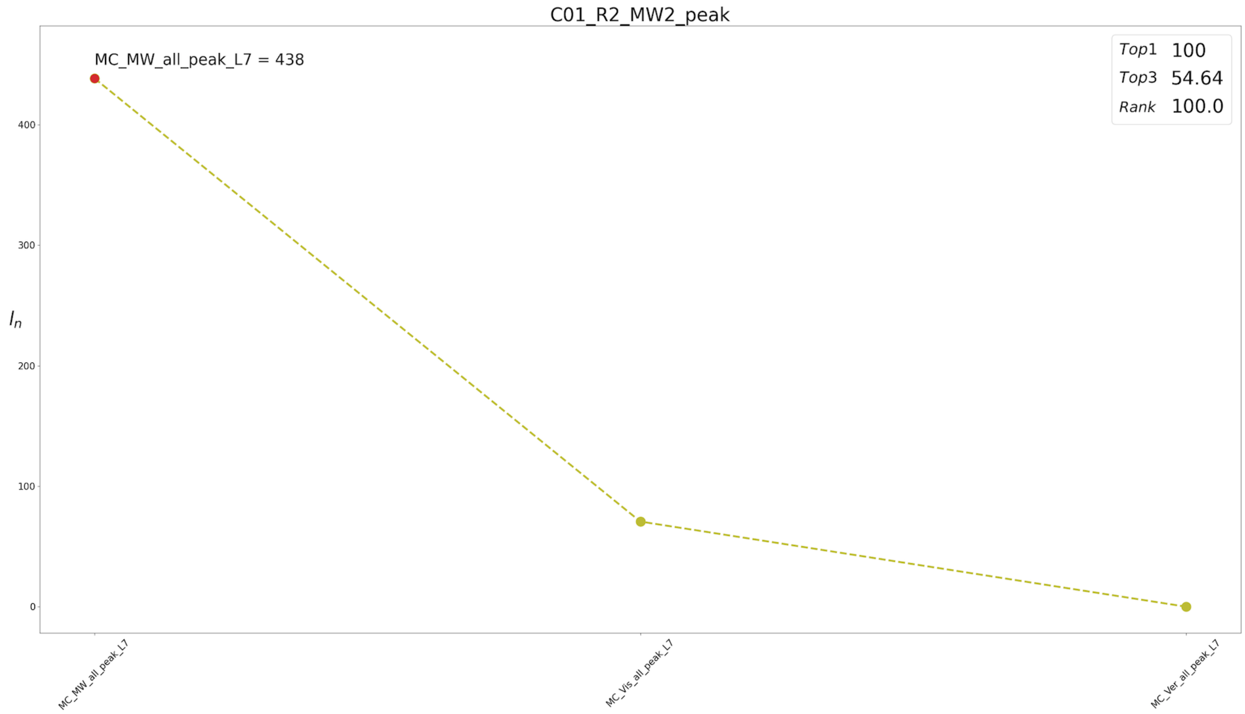

Figure 35: Main participant 1's EEG microstate sequence from run 2 and repetition 2 of MW task, peak timing mode, run against M3

M240 machines evaluated by Top 1, Top 3 and Rank scores. We can see than while Top 1 score or Top 3 score point to some differences in performance in different timing modes, cognitive process conditions and for different word lengths, actually the distribution of matching machines among the 240 are very similar in each case as shown by Rank score.

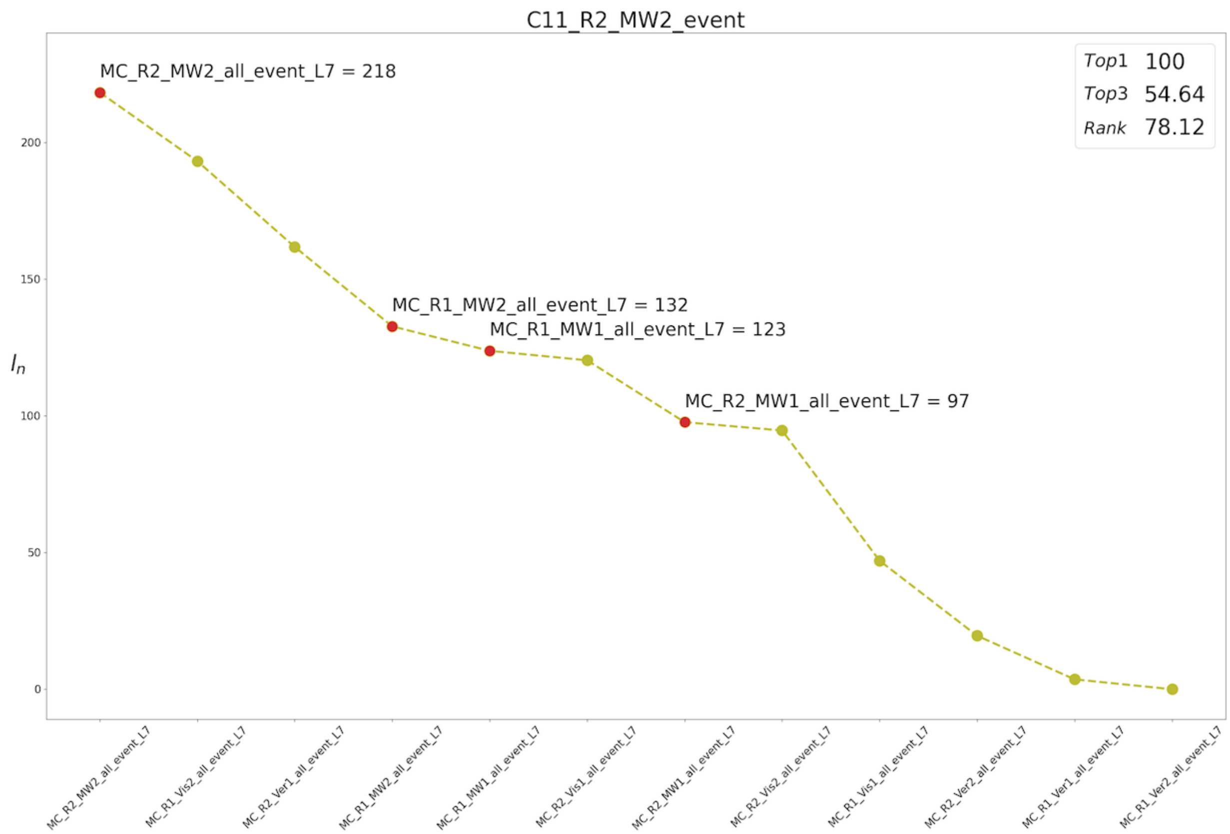

Figure 36: Validation participant C11; run 2 and repetition 2 of MW task; event timing mode; run against M12

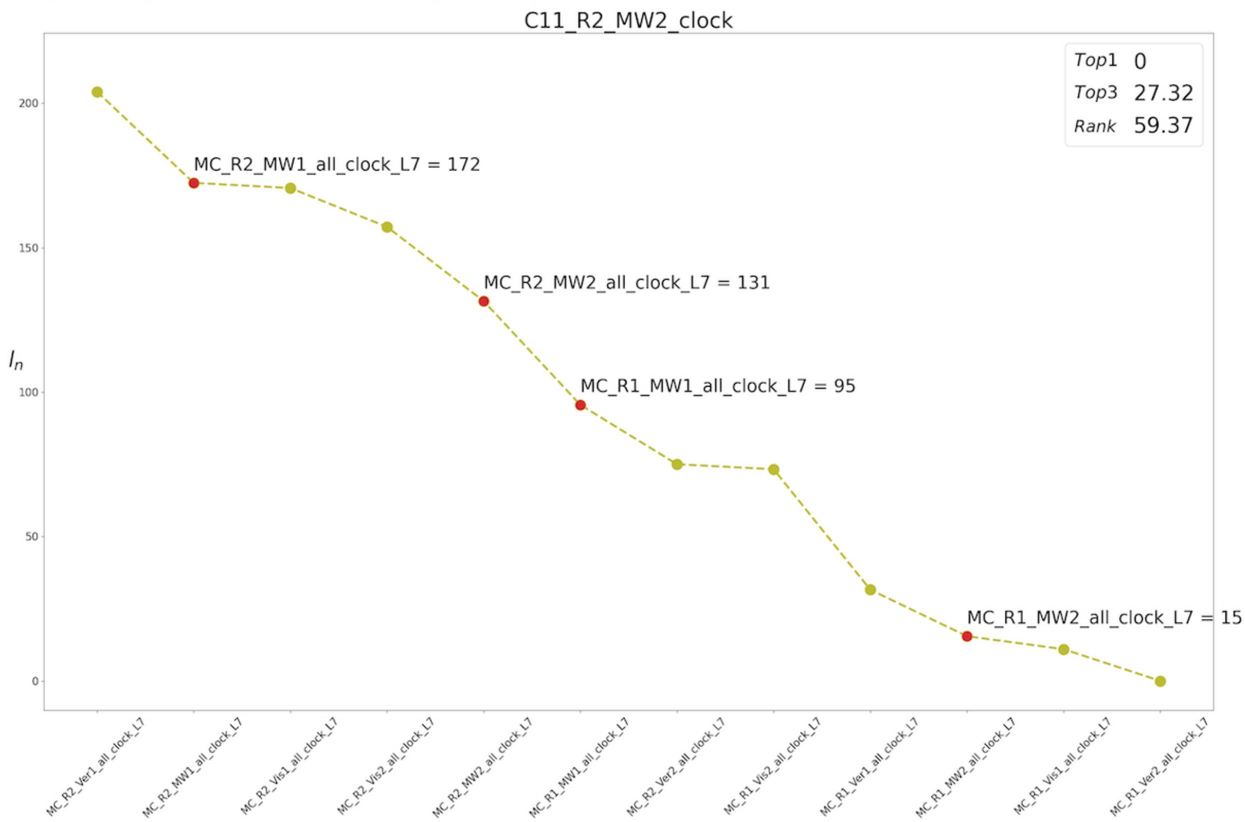

Figure 37: Validation participant C11; run 2 and repetition 2 of MW task; clock timing mode; run against M12

Figure 38: Main participant 1; run 2 and repetition 2 of MW task; peak timing mode; run against M240

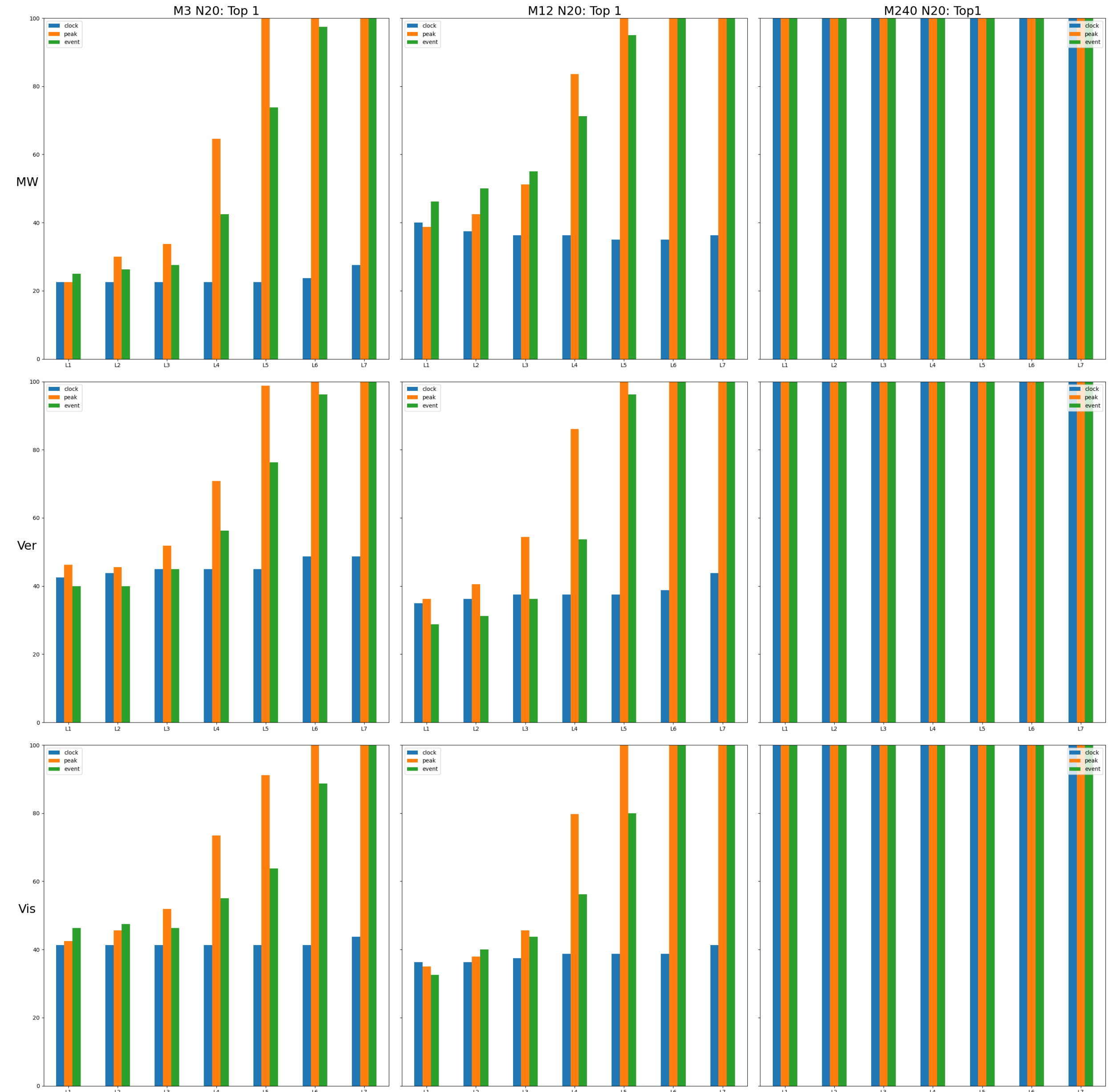

Figure 39: Cognitive Processing Condition Recognition Results for the N20 EEG microstate sequences for M3, M12 and M240 machines evaluated by the Top 1 score. Each row corresponds to a cognitive process condition (mind-wandering, verbalization, visualization). Columns give the results for the M3, M12, M240 sets of epsilon-machines. In each panel, mean Top 1 scores are plotted for clock, peak and event mode (blue, orange, green) for epsilon-machines constructed with history lengths from 1 to 7. Recognition of cognitive process conditions is 100% against M240 for temporal scales and at all history lengths. For the sets of group machines M3 and M12, recognition performance of cognitive process condition rises monotonically to 100% for peak and event time-modes as history length increases, but not for clock time-mode whose performance is poor.

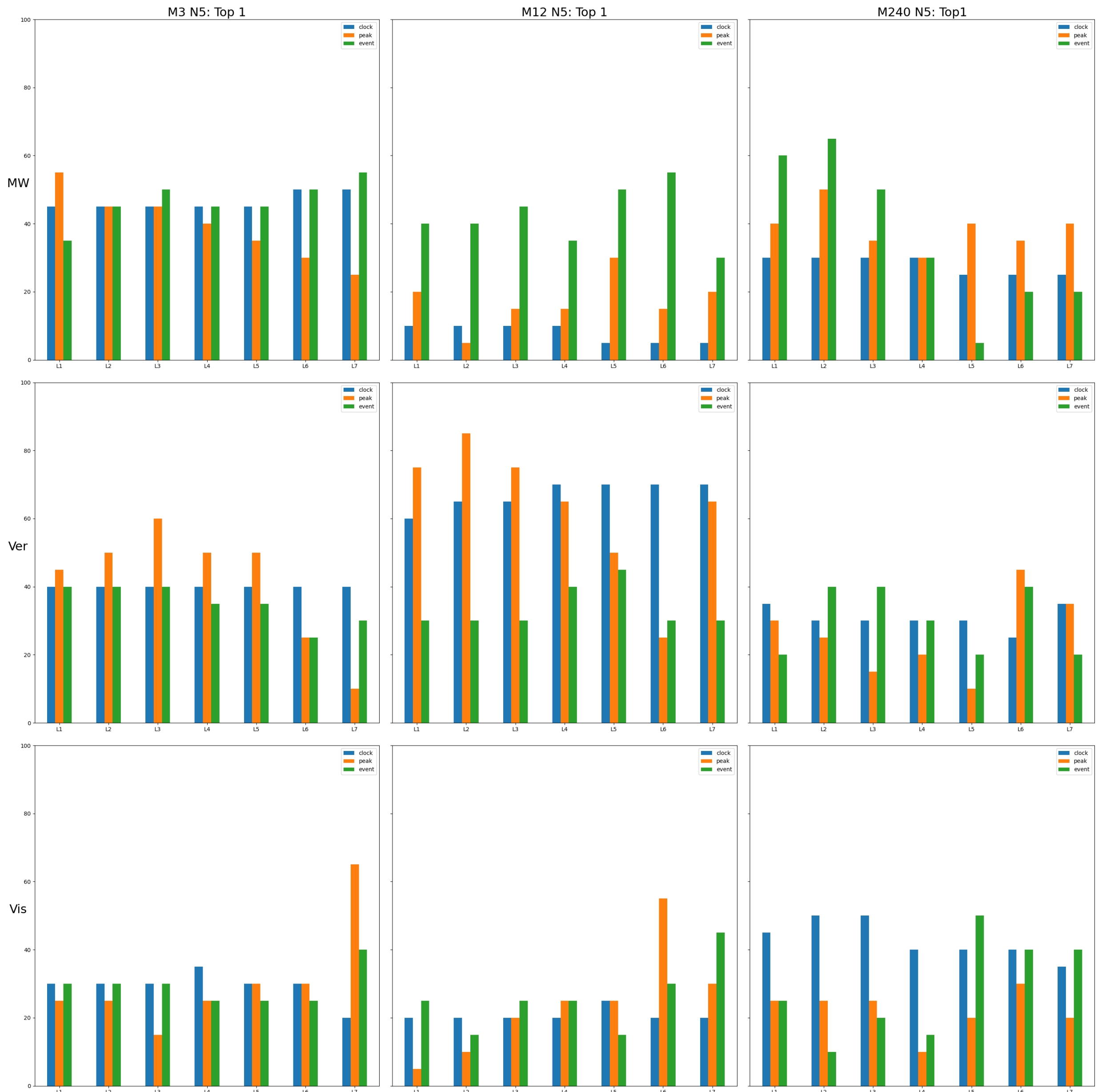

Figure 40: Results for validation (N5) sequences, for M3, M12 and M240 machines evaluated by the Top 1 score.

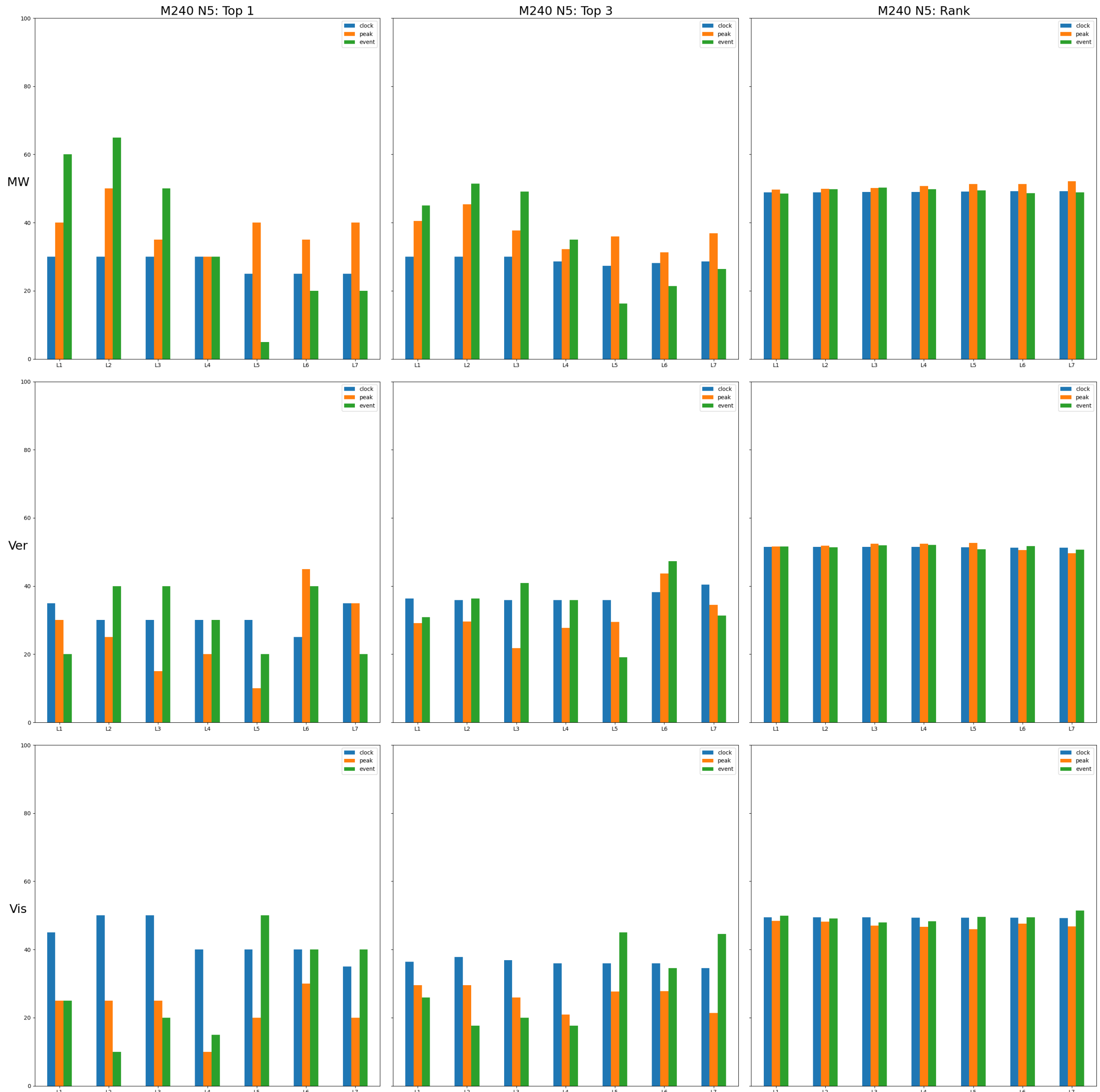

Figure 41: Cognitive Process condition Recognition Results for the validation (N5) EEG microstate sequences, for the M240 epsilon-machines. Each row corresponds to a cognitive process condition (mind-wandering, verbalization, visualization). Columns of panels show evaluation by Top 1, Top 3 and Rank scores, respectively. Within each panel, scores are plotted for clock, peak and event temporal modes (blue, orange, green) for epsilon-machines constructed with history lengths from 1 to 7.

7.3 Testing Run 2 sequences on Run 1 machines

The data being tested are not included in the data that the epsilon automata are built from, but these data are the second repetition EEG recordings from the same participants. The sequences for checking the likelihood of matching are EEG microstate sequences of 3-minute individual recordings.

7.3.1 Individual sequences on 3 group-level machines: cognitive process condition matching

In the first step, we only distinguish between data corresponding to different cognitive process conditions, and we can check how often we can match a machine with the correct cognitive process condition to the sequences using the log-likelihood.

The sequences for constructing epsilon machines are a combination of 3-minute EEG recordings of 25 participants C01 - C20, C22-C26, combined into 3 files according to the cognitive process conditions: mind wandering (MW), verbalization (Ver), visualization (Vis).

time	mind mode	L1	L2	L3	L4	L5	L6	L7
clock	MW	28.0 ± 6.0	28.0 ± 6.0	28.0 ± 6.0	28.0 ± 6.0	28.0 ± 6.0	30.0 ± 6.2	32.0 ± 6.3
	Ver	32.0 ± 6.3	30.0 ± 6.2	30.0 ± 6.2	32.0 ± 6.3	32.0 ± 6.3	32.0 ± 6.3	32.0 ± 6.3
	Vis	38.0 ± 6.5	38.0 ± 6.5	38.0 ± 6.5	38.0 ± 6.5	38.0 ± 6.5	36.0 ± 6.4	34.0 ± 6.4
peak	MW	34.0 ± 6.4	34.0 ± 6.4	40.0 ± 6.6	42.0 ± 6.6	38.0 ± 6.5	46.0 ± 6.7	40.0 ± 6.6
	Ver	32.0 ± 6.3	30.0 ± 6.2	30.0 ± 6.2	28.0 ± 6.0	40.0 ± 6.6	34.0 ± 6.4	52.0 ± 6.7
	Vis	34.0 ± 6.4	36.0 ± 6.4	32.0 ± 6.3	34.0 ± 6.4	38.0 ± 6.5	20.0 ± 5.4	2.0 ± 1.9
event	MW	22.0 ± 5.6	26.0 ± 5.9	30.0 ± 6.2	34.0 ± 6.4	34.0 ± 6.4	32.0 ± 6.3	36.0 ± 6.4
	Ver	34.0 ± 6.4	34.0 ± 6.4	34.0 ± 6.4	38.0 ± 6.5	34.0 ± 6.4	34.0 ± 6.4	30.0 ± 6.2
	Vis	40.0 ± 6.6	40.0 ± 6.6	38.0 ± 6.5	36.0 ± 6.4	36.0 ± 6.4	44.0 ± 6.7	44.0 ± 6.7

Table 4: Average Top1 score (with 95% Confidence Interval) for matching mind mode of 150 files with **sequences of R2** data to the three group-level epsilon **machines built from R1** data.

The best performance here with just 3 machines to match against for each cognitive process condition occurs for peak mode with history length 5, then 4 or 3, for event mode with length 6 and 7, which appear to be performing above chance level, while clock mode performs around chance level or worse.

7.3.2 Individual sequences on 6 group-level machines: cognitive process condition matching

In the second step, we distinguish between data corresponding to different cognitive process conditions and different repetitions with six group-level machines from Run 1, and again we check how often we can match a machine with the correct cognitive process condition to the sequences using the log-likelihood.

time	mind mode	L1	L2	L3	L4	L5	L6	L7
clock	MW	42.0 ± 6.6	40.0 ± 6.6	40.0 ± 6.6	40.0 ± 6.6	40.0 ± 6.6	40.0 ± 6.6	40.0 ± 6.6
	Ver	24.0 ± 5.7	22.0 ± 5.6	22.0 ± 5.6	22.0 ± 5.6	22.0 ± 5.6	22.0 ± 5.6	24.0 ± 5.7
	Vis	36.0 ± 6.4	36.0 ± 6.4	36.0 ± 6.4	36.0 ± 6.4	36.0 ± 6.4	34.0 ± 6.4	34.0 ± 6.4
peak	MW	50.0 ± 6.7	54.0 ± 6.7	52.0 ± 6.7	46.0 ± 6.7	24.0 ± 5.7	38.0 ± 6.5	44.0 ± 6.7
	Ver	22.0 ± 5.6	28.0 ± 6.0	26.0 ± 5.9	34.0 ± 6.4	44.0 ± 6.7	52.0 ± 6.7	54.0 ± 6.7
	Vis	34.0 ± 6.4	26.0 ± 5.9	30.0 ± 6.2	34.0 ± 6.4	36.0 ± 6.4	0.0 ± 0.0	8.0 ± 3.6
event	MW	44.0 ± 6.7	44.0 ± 6.7	46.0 ± 6.7	38.0 ± 6.5	36.0 ± 6.4	32.0 ± 6.3	56.0 ± 6.7
	Ver	24.0 ± 5.7	26.0 ± 5.9	28.0 ± 6.0	30.0 ± 6.2	34.0 ± 6.4	34.0 ± 6.4	42.0 ± 6.6
	Vis	36.0 ± 6.4	34.0 ± 6.4	32.0 ± 6.3	34.0 ± 6.4	32.0 ± 6.3	36.0 ± 6.4	2.0 ± 1.9

Table 5: Average Top1 score (with 95% Confidence Interval) for matching mind mode of 150 files with **sequences of R2** data to the six group-level epsilon **machines built from R1** data.

Here with six group-level machines from Run 1, the best performance is in peak mode at a number of history lengths 1 to 5, especially at history length 4, while in event mode for length 3 performs better than chance at distinguishing cognitive conditions. The best setting for clock mode is length 1 (marginally above chance overall).

7.3.3 Individual sequences on 150 individual-level machines: cognitive process condition matching

In the third step, we distinguish between data corresponding to different cognitive process conditions and different repetitions (e.g., MW1 and MW2) and different individuals as well, and again we check how often we can match a machine with the correct cognitive process condition to the sequences using the log-likelihood.

time	mind mode	L1	L2	L3	L4	L5	L6	L7
clock	MW	30.0 ± 6.2	30.0 ± 6.2	30.0 ± 6.2	30.0 ± 6.2	30.0 ± 6.2	30.0 ± 6.2	34.0 ± 6.4
	Ver	40.0 ± 6.6	40.0 ± 6.6	42.0 ± 6.6	46.0 ± 6.7	46.0 ± 6.7	44.0 ± 6.7	44.0 ± 6.7
	Vis	46.0 ± 6.7	46.0 ± 6.7	44.0 ± 6.7	42.0 ± 6.6	42.0 ± 6.6	40.0 ± 6.6	48.0 ± 6.7
peak	MW	38.0 ± 6.5	38.0 ± 6.5	68.0 ± 6.3	68.0 ± 6.3	32.0 ± 6.3	4.0 ± 2.6	2.0 ± 1.9
	Ver	48.0 ± 6.7	38.0 ± 6.5	30.0 ± 6.2	22.0 ± 5.6	18.0 ± 5.2	24.0 ± 5.7	10.0 ± 4.0
	Vis	42.0 ± 6.6	42.0 ± 6.6	10.0 ± 4.0	12.0 ± 4.4	46.0 ± 6.7	68.0 ± 6.3	82.0 ± 5.2
event	MW	40.0 ± 6.6	40.0 ± 6.6	100.0 ± 0.0	0.0 ± 0.0	10.0 ± 4.0	40.0 ± 6.6	54.0 ± 6.7
	Ver	36.0 ± 6.4	32.0 ± 6.3	2.0 ± 1.9	68.0 ± 6.3	38.0 ± 6.5	12.0 ± 4.4	8.0 ± 3.6
	Vis	40.0 ± 6.6	28.0 ± 6.0	0.0 ± 0.0	30.0 ± 6.2	56.0 ± 6.7	48.0 ± 6.7	46.0 ± 6.7

Table 6: Average Top1 score (with 95% Confidence Interval) for matching mind mode of 150 files with **sequences of R2** data to the 150 individual epsilon **machines built from R1** data.

Here with comparison over 150 individual epoch machines from Run 1, performance improves for clock mode with 150 machines, performing better than chance with better performances generally as history length increases. Even better is peak mode wth lengths 1 and then 2, whereas event mode is similar to clock for length 1 only, and above chance for length 2.

7.3.4 Individual sequences on 150 individual-level machines: person matching

Here we check whether or not the best matching machine for a sequence from Run 2 from the set of 150 machines epochs from Run 1 is from the same participant.

time	person	L1	L2	L3	L4	L5	L6	L7
clock	C01	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0
	C02	66.7 ± 18.3	66.7 ± 18.3	50.0 ± 19.4	66.7 ± 18.3	66.7 ± 18.3	83.3 ± 14.5	66.7 ± 18.3
	C03	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	33.3 ± 18.3	0.0 ± 0.0	0.0 ± 0.0
	C04	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	50.0 ± 19.4	16.7 ± 14.5	0.0 ± 0.0
	C05	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5
	C06	33.3 ± 18.3	33.3 ± 18.3	33.3 ± 18.3	33.3 ± 18.3	50.0 ± 19.4	33.3 ± 18.3	0.0 ± 0.0
	C07	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C08	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5
	C09	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0
	C10	0.0 ± 0.0	16.7 ± 14.5	50.0 ± 19.4	66.7 ± 18.3	33.3 ± 18.3	16.7 ± 14.5	33.3 ± 18.3
	C11	16.7 ± 14.5	0.0 ± 0.0	16.7 ± 14.5	33.3 ± 18.3	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0
	C12	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5
	C13	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C14	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C15	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C16	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C17	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0
	C18	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C19	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C20	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C22	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0
	C23	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5
	C24	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5
	C25	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C26	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	16.7 ± 14.5
	All	10.7 ± 5.7	8.7 ± 5.8	11.3 ± 5.7	12.0 ± 7.4	12.7 ± 7.4	9.3 ± 6.9	8.0 ± 5.8
peak	C01	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C02	50.0 ± 19.4	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C03	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C04	33.3 ± 18.3	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C05	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C06	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C07	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C08	16.7 ± 14.5	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5
	C09	16.7 ± 14.5	0.0 ± 0.0	33.3 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C10	16.7 ± 14.5	50.0 ± 19.4	33.3 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C11	0.0 ± 0.0	0.0 ± 0.0	33.3 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0

	C12	0.0 ± 0.0	16.7 ± 14.5	33.3 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C13	33.3 ± 18.3	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C14	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C15	16.7 ± 14.5	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C16	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C17	0.0 ± 0.0	0.0 ± 0.0	50.0 ± 19.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C18	16.7 ± 14.5	0.0 ± 0.0	33.3 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0
	C19	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C20	0.0 ± 0.0	16.7 ± 14.5	33.3 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C22	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5
	C23	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	33.3 ± 18.3
	C24	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	33.3 ± 18.3	0.0 ± 0.0
	C25	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C26	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	33.3 ± 18.3	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0
	All	10.0 ± 5.2	10.0 ± 4.5	13.3 ± 6.1	1.3 ± 2.5	0.7 ± 1.3	2.0 ± 2.8	2.7 ± 3.0
event	C01	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C02	33.3 ± 18.3	33.3 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C03	33.3 ± 18.3	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C04	16.7 ± 14.5	0.0 ± 0.0	50.0 ± 19.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C05	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C06	33.3 ± 18.3	16.7 ± 14.5	33.3 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C07	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C08	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C09	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C10	16.7 ± 14.5	16.7 ± 14.5	50.0 ± 19.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C11	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C12	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5
	C13	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C14	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5
	C15	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C16	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C17	50.0 ± 19.4	16.7 ± 14.5	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C18	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C19	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C20	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C22	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5
	C23	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C24	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C25	0.0 ± 0.0	0.0 ± 0.0	33.3 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5
	C26	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	All	10.7 ± 5.4	6.7 ± 3.7	11.3 ± 6.0	2.0 ± 2.1	0.0 ± 0.0	0.0 ± 0.0	2.7 ± 2.4

Table 7: Average Top1 score (with 95% Confidence Interval) for matching person of 150 files with **sequences of R2** data to the 150 individual epsilon **machines built from R1** data.

Performance overall appears to be above chance, but better for some particiapnts than others (e.g., participant C02 is matched most of the time for clock at different history lengths; in contrast C13 is never matched). Performance at person matching is for history length 4 in clock mode, length 3 in peak mode, and length 1 or 3 in event mode, though shows high variability. For some settings, performance is abysmal, e.g., event mode machines rarely matches the person for history lengths 3 and above, never matching the person for lengths 5 and 6.

7.3.5 Individual sequences on 150 individual-level machines: cognitive process condition matching for each person individually

Checking per participant whether their EEG sequences from Run 2 match the cognitive process condition when run against the 150 machines from Run 1 epochs is done in this section.

time mode	mind mode	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	50	50	50	50	50	50	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	50	0	0	0	0	0
	Vis	0	0	50	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	50	0	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 8: Average Top1 score (with 95% Confidence Interval) for matching cognitive process conditions of the person C01 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

The tables corresponding to the rest of the participants, i.e. C03 - C20, C22- C26 are placed in the corresponding section of the supplementary materials section A.2 as well as all of the tests repeated using Run 1 sequences on Run 2 machines in section A.3.

Performance is generally very poor, with certain individuals having more than 50% success in identifying the mode for history length 3 in peak more, but generally most participants' conditions not identified even at chance level.

7.4 Discussion

The log-likelihood method for recognizing cognitive modes did not perform as well as the geometric distance metrics studied in the previous section. This may perhaps be due to certain epsilon-machines attracting many matches disproportionately. For sequences used in the construction of the epsilon-machines though that

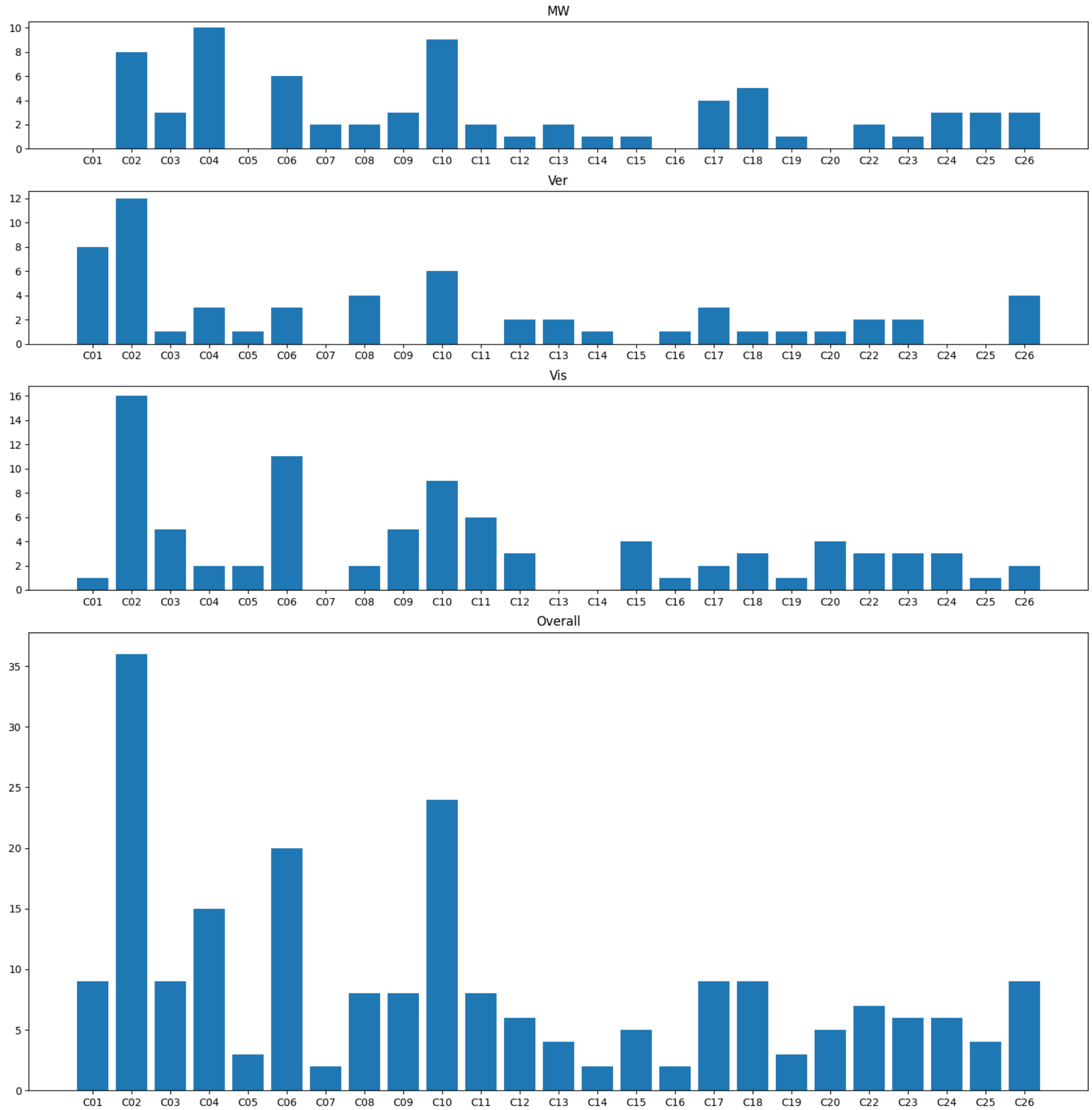

Figure 42: The number of correct matching of cognitive process conditions across individuals, i.e., cases where the machine with the highest likelihood correspond to the same cognitive process condition as the sequences. The graph displays the result of testing 150 individual **R2 sequences on 150 R1 epsilon automata**, for three cognitive mode conditions in all three time modes (clock, peak, event) and 7 history lengths (from $L = 1$ to $L = 7$). Since two files with sequences correspond to each person, in all respects we have 42 tests for each person. The last table shows the sum of the result for three cognitive process conditions, displaying the overall process matching success across people.

time mode	mind mode	L1	L2	L3	L4	L5	L6	L7
clock	MW	50	50	50	50	50	50	0
	Ver	50	50	50	50	50	100	100
	Vis	100	100	50	100	100	100	100
peak	MW	50	0	0	0	0	0	0
	Ver	50	0	0	0	0	0	0
	Vis	50	50	0	0	0	0	0
event	MW	0	50	0	0	0	0	0
	Ver	50	50	0	0	0	0	0
	Vis	50	0	0	0	0	0	0

Table 9: Average Top1 score (with 95% Confidence Interval) for matching cognitive process conditions of the person C02 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

method showed very specific capacity to match the sequence over a broad range of temporal modes and history lengths, especially for machines constructed from a single epoch's data. There appears to be significant variabily between individauls in how well their sequences are identified both for cognitive processing conditon and for detecting the person who was the source of the EEG microsate sequence. Appendix A.3 contains similar analyses but using Run 1 data on Run 2 epsilon-machines.

8 Status and Further Planned Methodological Developments

8.1 Objective 1: EEG Microstates and Co-activation PatternsCAPs derived in Resting State and Attenton Tasks

The EEG microstates we derived for the simultaneous EEG/fMRI target dataset have an explicit derivation justifying the choice of five types that is supported by the explained variance and cross-validation criteria, and are very similar topographically to the canonical set, as well as the previously investigated microstate E in the literature. It is remarkable that this is the case for this eyes-open dataset as most microstate studies have used eyes-closed data only. As expected from the literature, the discrete alphabet of eight Co-activation Patterns (CAPs) derived for the target dataset has clear pairs that show functional correspondence (X. Liu et al., 2018). Some of these CAP pairs exhibit commonly found "gradients", while other pairs appear to include commonly known networks co-activating and attenuating. Objective 1 has been achieved. Next steps beyond the project will investigate a robust criterion for defining the optimal number of CAPs to use in conjunction with the EEG, to confirm the robustness of the eight generated CAPs for resting state and attention-demanding tasks (such as the CRT task). Unlike what we found for EEG where very similar microstates arise for both resting state and attention tasks, the eight CAPs generated for attention tasks differ more substantially from those for the resting state. This raises the possibility of using multiple sets of eight CAPs simultaneously in applications to detecting cognitive mode.

8.2 Objective 2: EEG Microstate and CAP n-grams and Basic Epsilon-machines

Our studies of epsilon-machines and words (*n*-grams) occurring and not occurring showed differences between cognitive modes in the arrangement of these machines in metric spaces. Also, log-likelihood techniques for classifying EEG microstate state sequences according to cognitive mode and person-detection were tested. These led to recognition of substantial between-subject variability in these success rates and performed above chance accuracy even when averaging over parameter spaces. Based on the exploration so far, one can identify the most promising ranges of parameters (temporal mode, history lengths, and minimization thresholds) to concentrate on in order to refine and validate the methods on other datasets. In ongoing work (not yet reported) we find large group differences in the metric spaces for epsilon-machine brain models. The work so far mainly used eyes-closed EEG and the four canonical microstates. As a complement to this, it would be desirable to explore how variability between participants as well as how particular characteristics of individuals might affect performance of the methods.

For the EEG/fMRI target dataset, we are in a position to investigate these methods using the five microstates and the CAP sets obtained. So far investigation of epsilon-machines, with minimization of 2-grams linked *BA* and *EA*, and *BD* and *ED*, respectively, in the resting state EEG microstate sequence. Both causal states here indicate a similarity between microstates *B* and *E*, which may be due to spatial similarity. Figure 43 shows the spatial similarity matrix between resting state microstates. *B* and *D* are two of the most similar microstates spatially, and the similarity of transition from these microstates to others may be due to this fact. Further investigations are required to confirm this, however. First investigations into CAP epsilon-machines in resting state have shown no significant merging of states minimization. Dataset size limitations and the slow rate of fMRI scanning currently limit CAP epsilon-machines to history lengths for CAP sequences to just 1-grams. Nevertheless, we are finding that the the EEG characteristics of microstate *n*-gram coverage, frequency and durations differ between the CAPs (Haydock et al., in prep.). This gives us the insight to associate to each CAP a characteristic EEG sequence epsilon-machine to capture the multiscale neurodynamics at group- and individual-level in different cognitive conditions.

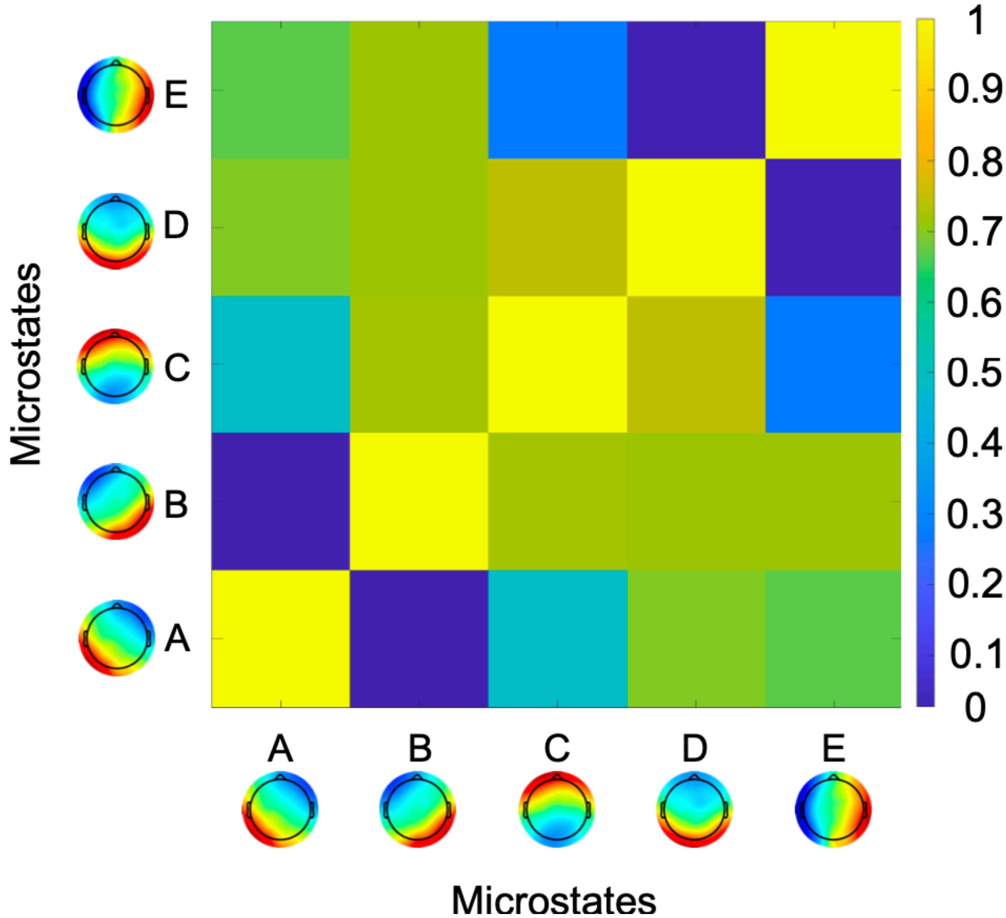

Figure 43: Spatial similarity matrix between each pair of the five microstates. Colour indicates correlation coefficient in each case. Diagonal is self comparison with score of one.

Further study using different length n-grams with event, peak and clock mode during CAPs in resting state and during CAPs in attention-demanding tasks will be required to clarify this relationship further. Epsilon-machines that serve as brain models for EEG microstate generation and prediction during each CAP will clarify further any simple connections between EEG and fMRI by co-occurrence. Furthermore, if an epsilon-machine is generated for the microstates that are occurring during each CAP are generated, and at the CAP level, an epsilon-machine is also generated, a combined model of nested automata can be used to model a task condition. Figure 44 gives a schematic of this nested model. Similarly, a complementary multi-level model with resting state CAP transitions at the high level and EEG microstate epsilon-machine models nested within each CAP state can be constructed combining data from the fMRI and EEG domains can be constructed and evaluated. Such models could use clock or event model at the CAP level and any of clock, peak and event modes at the EEG level. We are now in a position to construct such models.

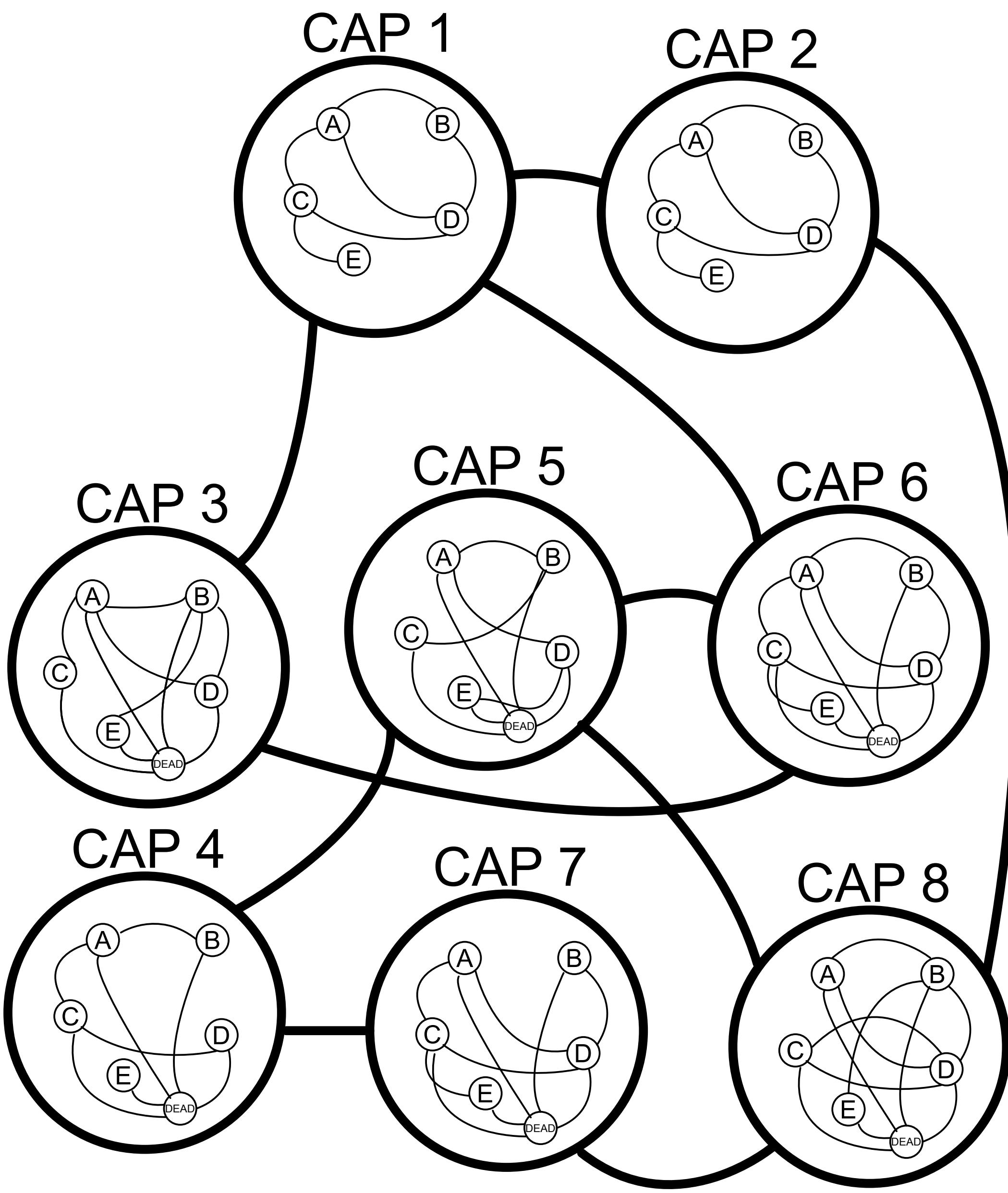

Figure 44: Basic schematic of a potential model to bridge the gap between EEG and fMRI. An epsilon-machine can be generated using only the EEG time series which is occurring during a given CAP (inside a single CAP circle). This may be applied to the EEG time series for all CAPs. An epsilon-machine can then be generated at the fMRI level, creating an automaton that transitions between CAPs. The result is an fMRI epsilon-machine which contains within it nested machines for the EEG microstates. Significant differences between microstate dynamics during different CAP must first be confirmed – see forthcoming work (Haydock et al., in prep.).

8.3 Objective 3: Investigation into Characterization of Epsilon-machine Models at Individual and Group-Level

Geometric and log-likelihood investigation of epsilon-machines has been explored in this report, and must be further validated on additional datasets. Techniques of Krohn-Rhodes theory and rigorous measures of epsilon-machine methods are applied according to our methodology (Nehaniv & Antonova, 2017). Preliminary investigations into the holonomy decomposition of the resting state event-based EEG microstate and fMRI CAP epsilon-machines from our target dataset found no high levels of complexity and no natural group subsystems. This may be due to the particular epsilon-machine construction method that was used, in contrast to NehanivAntonova2017. Next steps will apply other epsilon-machine construction methods in an attempt to identify natural subsystems and complexity. Immediate work to investigate these findings will run the same analyses on peak- and clock-based epsilon-machines to check for complexity of these models. There will also be an investigation into the epsilon-machines of EEG microstates occurring during each CAP to see whether there is a connection between microstates and CAPs in this way, as there may be natural subsystems in the EEG microstate behaviour that are only apparent during specific CAPs. Finally, it is natural to apply and study the other measures already developed in this report for EEG (Sections 6, 7, and Appendix C) to CAP and EEG epsilon-machines of the target dataset, such as the geometric distance, log-likelihood matching and shortest zero occurrence *n*-grams, and also to apply these methods between subject-level epsilon-machines, as well as to study distance geometry between the EEG microstate epsilon-machines occurring during each CAPs. By linking CAPs and their transitions to EEG microstate syntax in this way, functional significance of microstate syntax might be better elucidated.

8.4 Objective 4: Attention Task Data and Investigation of Subject Task Performance

The EEG microstates we derived are consistent in the resting state, the simple blocked choice reaction time (CRT) task, and the attention-demanding, continuous CRT task. Interestingly, the CAPs generated significantly differed between tasks, with spatial similarities showing few consistent co-activation patterns (Figure 29). This finding may indicate that different generators in each task recording are yielding similar electrical activity patterns on the scalp. Further investigation into this claim will be conducted in the future. Methods of epsilon-distance and Jaccard distance comparison between reference epsilon-machines and epsilon-machines constructed on the fly (using a sliding window - it is fast to construct a single epsilon machine) from neurophysiological data will be assessed for detecting attention and mind-wandering. This will be compared to log-likehood matching of reference epsilon-machines for different cognitive modes (at group- and individual-level) for efficacy vis à vis the geometric methods – a comparison suggested by the results of this project.

While CAPs and microstates have been generated in all task recordings, future work will focus on measures of the epsilon-machines in the resting state, comparing with other task conditons and cognitive modes. The models generated from the resting state data, as per objective 3, will be characterized for both EEG microstates and fMRI CAPs, and the measures will be employed to out seek out when same characteristics, dynamics and presumably mechanisms arise in the CRT task data to identify periods of mind-wandering. CRT task performance data will be investigated to achieve this. Periods where subjects make errors or react more slowly to the task will be compared to the resting state models, and capacity for prediction of mind-wandering using the models on the performance data (errors and delays) in the EEG/fMRI dataset will be assessed.

Acknowledgments

The research described here is based upon work supported by the United States Air Force Office of Scientific Research (AFOSR) under award number FA9550-19-1-7034. R. Maex was supported by a University of Hertfordshire Vice Chancellor's Post-Doctoral Fellowship in Biocomputation for Mental and Well-being. David G. Haydock was supported by a PhD studentship from the Royal Society / Wolfson Biocomputation Research Laboratory at the University of Hertfordshire. Hanna Derets and Sebastián worked on the project as MITACS Interns, and Thomas George contributed while supported by a University of Waterloo President's Award Undergraduate Research Student Assistantship. We thank Ting Xu (Child Mind Institute, USA) and Jonathan Smallwood (Queen's University, Canada) for discussions on fMRI CAPs, and are especially grateful to Ting Xu for access to her relevant software. Both Hanna Derets and ZiXuan Thomas Gao of the University of Waterloo Algebraic Intelligence and Computation Laboratory assisted with algorithm design, programming and data analysis as graduate students funded under a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (funding reference number RGPIN-2019-04669) to CLN. Cette recherche a été financée en partie par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), numéro de référence RGPIN-2019-04669.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*.

Abreu, R., Jorge, J., Leal, A., Koenig, T., & Figueiredo, P. (2021). EEG Microstates Predict Concurrent fMRI Dynamic Functional Connectivity States. *Brain Topography*, *34*(1), 41–55.

Aftanas, L. I., & Golocheikine, S. A. (2002). Non-linear dynamic complexity of the human EEG during meditation. *Neuroscience Letters*, *330*(2), 143–146.

Amico, E., Gomez, F., Di Perri, C., Vanhaudenhuyse, A., Lesenfants, D., Boveroux, P., Bonhomme, V., Brichant, J.-F., Marinazzo, D., & Laureys, S. (2014). Posterior cingulate cortex-related co-activation patterns: A resting state FMRI study in propofol-induced loss of consciousness. *PloS One*, *9*(6), e100012.

Antonova, E., Holding, M., Suen, H. C., Sumich, A., Maex, R., & Nehaniv, C. (2022). EEG microstates: Functional significance and short-term test-retest reliability. *Neuroimage: Reports*, *2*(2), 100089.

Biswal, B., Yetkin, F. Z., 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.

Borbély, A. A., Baumann, F., Brandeis, D., Strauch, I., & Lehmann, D. (1981). Sleep deprivation: Effect on sleep stages and EEG power density in man. *Electroencephalography and Clinical Neurophysiology*, *51*(5), 483–493.

Brain Products, B. P. (2021). *BrainVision analyzer* (Version 2.2.0). Brain Products GmbH, Gilching, Germany.

Bréchet, L., Brunet, D., Birot, G., Gruetter, R., Michel, C. M., & Jorge, J. (2019). Capturing the spatiotemporal dynamics of self-generated, task-initiated thoughts with EEG and fMRI. *NeuroImage*, *194*, 82–92.

Britz, J., Van De Ville, D., & Michel, C. M. (2010). BOLD correlates of EEG topography reveal rapid resting-state network dynamics. *NeuroImage*, *52*(4), 1162–1170.

Brodbeck, V., Kuhn, A., von Wegner, F., Morzelewski, A., Tagliazucchi, E., Borisov, S., Michel, C. M., & Laufs, H. (2012). EEG microstates of wakefulness and NREM sleep. *NeuroImage*, *62*(3), 2129–2139.

Brookes, M. J., Vrba, J., Mullinger, K. J., Geirsdóttir, G. B., Yan, W. X., Stevenson, C. M., Bowtell, R., & Morris, P. G. (2009). Source localisation in concurrent EEG/fMRI: Applications at 7t. *NeuroImage*, *45*(2), 440–452.

Buckner, R. L. (1998). Event-related fMRI and the hemodynamic response. *Human Brain Mapping*, *6*(5), 373–377.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *The year in cognitive neuroscience 2008* (pp. 1–38). Blackwell Publishing.

Cabral, J., Vidaurre, D., Marques, P., Magalhães, R., Silva Moreira, P., Miguel Soares, J., Deco, G., Sousa, N., & Kringelbach, M. L. (2017). Cognitive performance in healthy older adults relates to spontaneous switching between states of functional connectivity during rest. *Scientific Reports*, *7*(1), 5135.

Calhoun, V. D., Miller, R., Pearlson, G., & Adalı, T. (2014). The chronnectome: Time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron*, *84*(2), 262–274.

Castaño-Candamil, S., Höhne, J., Martínez-Vargas, J.-D., An, X.-W., Castellanos-Domínguez, G., & Haufe, S. (2015). Solving the EEG inverse problem based on space–time–frequency structured sparsity constraints. *NeuroImage*, *118*, 598–612.

Chen, J. E., Chang, C., Greicius, M. D., & Glover, G. H. (2015). Introducing co-activation pattern metrics to quantify spontaneous brain network dynamics. *NeuroImage*, *111*, 476–488.

Croft, R. J., & Barry, R. J. (2000). Removal of ocular artifact from the EEG: A review. *Neurophysiologie Clinique/Clinical Neurophysiology*, *30*(1), 5–19.

Croft, R. J., & Barry, R. J. (2002). Issues relating to the subtraction phase in EOG artefact correction of the EEG. *International Journal of Psychophysiology*, *44*(3), 187–195.

Crutchfield, J. P., & Young, K. (1989). Inferring statistical complexity. *Physical Review Letters*, *62*(2), 105–108.

Crutchfield, J. P. (1994). Observing Complexity and the Complexity of Observation. In H. Haken, H. Atmanspacher, & G. J. Dalenoort (Eds.), *Inside Versus Outside* (pp. 235–272). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-48647-0_14

Custo, A., Van De Ville, D., Wells, W. M., Tomescu, M. I., Brunet, D., & Michel, C. M. (2017). Electroencephalographic Resting-State Networks: Source Localization of Microstates. *Brain Connectivity*, *7*(10), 671–682.

Custo, A., Vulliemoz, S., Grouiller, F., Van De Ville, D., & Michel, C. (2014). EEG source imaging of brain states using spatiotemporal regression. *NeuroImage*, *96*, 106–116.

Damoiseaux, J. S., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., & Beckermann, C. F. (2006). Consistent resting-state networks across healthy subjects. *PNAS*, *103*(37), 6.

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.

Dinov, M., & Leech, R. (2017). Modeling uncertainties in EEG microstates: Analysis of real and imagined motor movements using probabilistic clustering-driven training of probabilistic neural networks. *Frontiers in Human Neuroscience*, *0*.

Egri-Nagy, A., Mitchell, J. D., & Nehaniv, C. L. (2014). SgpDec: Cascade (De)Compositions of Finite Transformation Semigroups and Permutation Groups. In H. Hong & C. Yap (Eds.), *Mathematical Software – ICMS 2014* (pp. 75–82). Springer. https://doi.org/10.1007/978-3-662-44199-2_13

Faber, P. L., Lehmann, D., Barendregt, H., Kaelin, M., & Gianotti, L. R. R. (2005). Increased duration of EEG microstates during meditation. The KEY-Institute for Brain-Mind Research.

Faber, P. L., Lehmann, D., Gianotti, L. R. R., Milz, P., Pascual-Marqui, R. D., Held, M., & Kochi, K. (2015). Zazen meditation and no-task resting EEG compared with LORETA intracortical source localization. *Cognitive Processing*, *16*(1), 87–96.

Fagerholm, E. D., Lorenz, R., Scott, G., Dinov, M., Hellyer, P. J., Mirzaei, N., Leeson, C., Carmichael, D. W., Sharp, D. J., Shew, W. L., & Leech, R. (2015). Cascades and Cognitive State: Focused Attention Incurs Subcritical Dynamics. *Journal of Neuroscience*, *35*(11), 4626–4634.

Figueroa, C. A., Cabral, J., Mocking, R. J. T., Rapuano, K. M., Hartevelt, T. J. v., Deco, G., Expert, P., Schene, A. H., Kringelbach, M. L., & Ruhé, H. G. (2019). Altered ability to access a clinically relevant control network in patients remitted from major depressive disorder. *Human Brain Mapping*, *40*(9), 2771–2786.

Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2006). Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of the National Academy of Sciences*, *103*(26), 10046–10051.

Fox, M. D., Zhang, D., Snyder, A. Z., & Raichle, M. E. (2009). The Global Signal and Observed Anticorrelated Resting State Brain Networks. *Journal of Neurophysiology*, *101*(6), 3270–3283.

Glerean, E., Salmi, J., Lahnakoski, J. M., Jääskeläinen, I. P., & Sams, M. (2012). Functional magnetic resonance imaging phase synchronization as a measure of dynamic functional connectivity. *Brain Connectivity*, *2*(2), 91–101.

Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., Xanthopoulos, P., Sakkalis, V., & Vanrumste, B. (2008). Review on solving the inverse problem in EEG source analysis. *Journal of NeuroEngineering and Rehabilitation*, *5*(1), 25.

Haak, K. V., & Beckmann, C. F. (2020). Understanding brain organisation in the face of functional heterogeneity and functional multiplicity. *NeuroImage*, *220*, 117061.

Haydock, D. G., Leech, R., Kadir, S., Xu, T., Smallwood, J., Nehaniv, C. L., & Antonova, E. (in prep.). Short EEG microstate sequences show significant differences in duration during different fMRI Co-Activation Patterns *[title tentative]. journal preprint*.

Heine, L., Soddu, A., Gomez, F., Vanhaudenhuyse, A., Tshibanda, L., Thonnard, M., Charland-Verville, V., Kirsch, M., Laureys, S., & Demertzi, A. (2012). Resting state networks and consciousness. *Frontiers in Psychology*, *3*, 295.

Helmholtz, H. (1853). Ueber einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern, mit Anwendung auf die thierisch-elektrischen Versuche (Schluss.) *Annalen der Physik*, *165*(7), 353–377 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.18531650702.

Hong, S.-J., Xu, T., Nikolaidis, A., Smallwood, J., Margulies, D. S., Bernhardt, B., Vogelstein, J., & Milham, M. P. (2020). Toward a connectivity gradient-based framework for reproducible biomarker discovery. *NeuroImage*, *223*, 117322.

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*, 10.1016/j.neuroimage.2013.05.079.

Iannetti, G. D., Niazy, R. K., Wise, R. G., Jezzard, P., Brooks, J. C. W., Zambreanu, L., Vennart, W., Matthews, P. M., & Tracey, I. (2005). Simultaneous recording of laser-evoked brain potentials and continuous, high-field functional magnetic resonance imaging in humans. *NeuroImage*, *28*(3), 708–719.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841.

Karapanagiotidis, T., Vidaurre, D., Quinn, A. J., Vatansever, D., Poerio, G. L., Turnbull, A., Ho, N. S. P., Leech, R., Bernhardt, B. C., Jefferies, E., Margulies, D. S., Nichols, T. E., Woolrich, M. W., & Smallwood, J. (2020). The psychological correlates of distinct neural states occurring during wakeful rest. *Scientific Reports*, *10*(1), 21121.

Katayama, H., Gianotti, L. R. R., Isotani, T., Faber, P. L., Sasada, K., Kinoshita, T., & Lehmann, D. (2007). Classes of Multichannel EEG Microstates in Light and Deep Hypnotic Conditions. *Brain Topography*, *20*(1), 7–14.

Khanna, A., Pascual-Leone, A., Michel, C. M., & Farzan, F. (2015). Microstates in resting-state EEG: Current status and future directions. *Neuroscience and Biobehavioral Reviews*, *49*, 105–113.

Khanna, A., Pascual-Leone, A., & Farzan, F. (2014). Reliability of Resting-State Microstate Features in Electroencephalography (T. Koenig, Ed.). *PLoS ONE*, *9*(12), e114163.

Klem, G. H., Lüders, H. O., Jasper, H. H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalography and Clinical Neurophysiology. Supplement*, *52*, 3–6.

Koenig, T., Prichep, L., Lehmann, D., Sosa, P. V., Braeker, E., Kleinlogel, H., Isenhart, R., & John, E. (2002). Millisecond by Millisecond, Year by Year: Normative EEG Microstates and Developmental Stages. *NeuroImage*, *16*(1), 41–48.

Kubicki, S., Herrmann, W. M., Fichte, K., & Freund, G. (1979). Reflections on the topics: EEG frequency bands and regulation of vigilance. *Pharmacopsychiatry*, *12*(2), 237–245.

Leech, R., Braga, R., & Sharp, D. J. (2012). Echoes of the Brain within the Posterior Cingulate Cortex. *Journal of Neuroscience*, *32*(1), 215–222.

Lehmann, D. (1987). Principles of spatial analysis. *Handbook of electroencephalography and clinical neurophysiology : Methods of analysis of brain electrical and magnetic signals*, *1*, 309–354.

Lehmann, D., Faber, P. L., Galderisi, S., Herrmann, W. M., Kinoshita, T., Koukkou, M., Mucci, A., Pascual-Marqui, R. D., Saito, N., Wackermann, J., Winterer, G., & Koenig, T. (2005). EEG microstate duration and syntax in acute, medication-naïve, first-episode schizophrenia: A multi-center study. *Psychiatry Research: Neuroimaging*, *138*(2), 141–156.

Liu, T. T., Nalci, A., & Falahpour, M. (2017). The global signal in fMRI: Nuisance or Information? *NeuroImage*, *150*, 213–229.

Liu, X., Chang, C., & Duyn, J. (2013). Decomposition of Spontaneous Brain Activity into Distinct fMRI Co-activation Patterns. *Frontiers in Systems Neuroscience*, *7*.

Liu, X., Zhang, N., Chang, C., & Duyn, J. H. (2018). Co-activation patterns in resting-state fMRI signals. *NeuroImage*, *180*, 485–494.

Lopez Rincon, A., & Shimoda, S. (2016). The inverse problem in electroencephalography using the bidomain model of electrical activity. *Journal of Neuroscience Methods*, *274*, 94–105.

Manganas, S., & Bourbakis, N. (2017). A Comparative Survey on Simultaneous EEG-fMRI Methodologies. *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*, 1–8.

Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., Bezgin, G., Eickhoff, S. B., Castellanos, F. X., Petrides, M., Jefferies, E., & Smallwood, J. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences*, *113*(44), 12574–12579.

Marino, M., Arcara, G., Porcaro, C., & Mantini, D. (2019). Hemodynamic Correlates of Electrophysiological Activity in the Default Mode Network. *Frontiers in Neuroscience*, *13*, 1060.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324.

Menon, V. (2011). Large-scale brain networks and psychopathology: A unifying triple network model. *Trends in Cognitive Sciences*, *15*(10), 483–506.

Michel, C. M., & Koenig, T. (2018). EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review. *NeuroImage*, *180*, 577–593.

Milz, P., Faber, P., Lehmann, D., Koenig, T., Kochi, K., & Pascual-Marqui, R. (2016). The functional significance of EEG microstates—Associations with modalities of thinking. *NeuroImage*, *125*, 643–656.

Milz, P., Pascual-Marqui, R., Achermann, P., Kochi, K., & Faber, P. (2017). The EEG microstate topography is predominantly determined by intracortical sources in the alpha band. *NeuroImage*, *162*, 353–361.

Milz, P. (2015). KeyPy: The KEY EEG analysis toolbox. Zenodo. doi: 10.5281.

Mitra, P., & Bokil, H. (2007). *Observed Brain Dynamics*. Oxford University Press.

Miyakoshi, M., Jurgiel, J., Dillon, A., Chang, S., Piacentini, J., Makeig, S., & Loo, S. K. (2020). Modulation of Frontal Oscillatory Power during Blink Suppression in Children: Effects of Premonitory Urge and Reward. *Cerebral Cortex Communications*, *1*(1), tgaa046.

Munkres, J. (1957). Algorithms for the Assignment and Transportation Problems. *Journal of the Society for Industrial and Applied Mathematics*, *5*(1), 32–38.

Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP Analyses: A Step-by-Step Tutorial Review. *Brain Topography*, *20*(4), 249–264.

Musso, F., Brinkmeyer, J., Mobascher, A., Warbrick, T., & Winterer, G. (2010). Spontaneous brain activity and EEG microstates. A novel EEG/fMRI analysis approach to explore resting-state networks. *NeuroImage*, *52*(4), 1149–1161.

Nehaniv, C. L. (1996). Complexity of finite aperiodic semigroups and star-free languages. In J. Almeida, G. Gomes, & P. Silva (Eds.), *Semigroups, automata, languages* (pp. 195–209). World Scientific.

Nehaniv, C. L., & Antonova, E. (2017). Simulating and reconstructing neurodynamics with epsilon-automata applied to electroencephalography (EEG) microstate sequences. *IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (IEEE CCMB'17), at IEEE Symposium Series on Computational Intelligence, 27 November - 1 December 2017, Honolulu, Hawaii, U.S.A.*, 1753–1761.

Nehaniv, C. L., Maex, R., Gao, Z. X., Chen, A. C., Kadir, S., Leech, R., & Antonova, E. (2021). From letters to syntax in electroencephalography: Grammatical models of the human brain during resting-state mind-wandering - Year 1 Project Report, Award Number FA9550-19-1-7034, United States Air Force Office of Scientific Research (AFOSR), Cognitive and Computational Neuroscience Program Review, presented 23 October 2020, report date: 27 January 2021.

Nehaniv, C. L., Rhodes, J., Egri-Nagy, A., Dini, P., Morris, E. R., Horváth, G., Karimi, F., Schreckling, D., & Schilstra, M. J. (2015). Symmetry structure in discrete models of biochemical systems: Natural subsystems and the weak control hierarchy in a new model of computation driven by interactions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*, 20140223.

Newson, J. J., & Thiagarajan, T. C. (2019). EEG frequency bands in psychiatric disorders: A review of resting state studies. *Frontiers in Human Neuroscience*, *12*, 521.

Niazy, R. K., Beckmann, C. F., Iannetti, G. D., Brady, J. M., & Smith, S. M. (2005). Removal of FMRI environment artifacts from EEG data using optimal basis sets. *NeuroImage*, *28*(3), 720–737.

O'Gorman, R. L., Poil, S.-S., Brandeis, D., Klaver, P., Bollmann, S., Ghisleni, C., Lüchinger, R., Martin, E., Shankaranarayanan, A., Alsop, D. C., & Michels, L. (2013). Coupling Between Resting Cerebral Perfusion and EEG. *Brain Topography*, *26*(3), 442–457.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Pal, A., Behari, M., Goyal, V., & Sharma, R. (2021). Study of EEG microstates in Parkinson's disease: A potential biomarker? *Cognitive Neurodynamics*, *15*(3), 463–471.

Park, B.-y., Vos de Wael, R., Paquola, C., Larivière, S., Benkarim, O., Royer, J., Tavakol, S., Cruces, R. R., Li, Q., Valk, S. L., Margulies, D. S., Mišić, B., Bzdok, D., Smallwood, J., & Bernhardt, B. C. (2021). Signal diffusion along connectome gradients and inter-hub routing differentially contribute to dynamic human brain function. *NeuroImage*, *224*, 117429.

Pascual-Marqui, R. D., Lehmann, D., Faber, P., Milz, P., Kochi, K., Yoshimura, M., Nishida, K., Isotani, T., & Kinoshita, T. (2014). The resting microstate networks (RMN): Cortical distributions, dynamics, and frequency specific information flow. *arXiv:1411.1949 [q-bio]*.

Pascual-Marqui, R., Michel, C., & Lehmann, D. (1995). Segmentation of brain electrical activity into microstates: Model estimation and validation. *IEEE Transactions on Biomedical Engineering*, *42*(7), 658–665.

Pascual-Marqui, R. D. (1999). Review of Methods for Solving the EEG Inverse Problem. *International Journal of Bioelectromagnetism*, *1*(1), 77–90.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

Poulsen, A. T., Pedroni, A., Langer, N., & Hansen, L. K. (2018). Microstate EEGlab toolbox: An introductory guide. *bioRxiv*.

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

Pruim, R. H. R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J. K., & Beckmann, C. F. (2015). ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage*, *112*, 267–277.

Rhodes, J. (2010). *Applications of Automata Theory and Algebra: Via the Mathematical Theory of Complexity to Biology, Physics, Psychology, Philosophy, and Games,* edited by C. L. Nehaniv, foreword by M. W. Hirsch, World Scientific Publishing Company.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex (New York, NY)*, *28*(9), 3095–3114.

Schlegel, F., Lehmann, D., Faber, P. L., Milz, P., & Gianotti, L. R. R. (2012). EEG Microstates During Resting Represent Personality Differences. *Brain Topography*, *25*(1), 20–26.

Seitzman, B. A., Abell, M., Bartley, S. C., Erickson, M. A., Bolbecker, A. R., & Hetrick, W. P. (2017). Cognitive manipulation of brain electric microstates. *NeuroImage*, *146*, 533–543.

Shalizi, C. R., & Klinkner, K. L. (2004). Blind construction of optimal nonlinear recursive predictors for discrete sequences. *Uncertainty in ai: Proc. 20th conf.* (pp. 504–511). AUAI Press.

Skrandies, W. (1990). Global field power and topographic similarity. *Brain Topography*, *3*(1), 137–141.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155.

Sten, S., Lundengård, K., Witt, S. T., Cedersund, G., Elinder, F., & Engström, M. (2017). Neural inhibition can explain negative BOLD responses: A mechanistic modelling and fMRI study. *NeuroImage*, *158*, 219–231.

Strik, W. K., Chiaramonti, R., Muscas, G. C., Paganini, M., Mueller, T. J., Fallgatter, A. J., Versari, A., & Zappoli, R. (1997). Decreased EEG microstate duration and anteriorisation of the brain electrical fields in mild and moderate dementia of the Alzheimer type. *Psychiatry Research: Neuroimaging*, *75*(3), 183–191.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, *2011*, e879716.

Tagliazucchi, E., Balenzuela, P., Fraiman, D., & Chialvo, D. (2012). Criticality in large-scale brain fMRI dynamics unveiled by a novel point process analysis. *Frontiers in Physiology*, *3*, 15.

Tagliazucchi, E., Balenzuela, P., Fraiman, D., & Chialvo, D. R. (2010). Brain resting state is disrupted in chronic back pain patients. *Neuroscience Letters*, *485*(1), 26–31.

Teplan, M. (2002). Fundamentals of EEG Measurement. *Measurement Science Review*, *2,2*.

Tian, Y., Margulies, D. S., Breakspear, M., & Zalesky, A. (2020). Topographic organization of the human subcortex unveiled with functional connectivity gradients. *Nature Neuroscience*, *23*(11), 1421–1432.

Tobe, R. H., MacKay-Brandt, A., Lim, R., Kramer, M., Breland, M. M., Tu, L., Tian, Y., Trautman, K. D., Hu, C., Sangoi, R., Alexander, L., Gabbay, V., Castellanos, F. X., Leventhal, B. L., Craddock, R. C., Colcombe, S. J., Franco, A. R., & Milham, M. P. (2022). A longitudinal resource for studying connectome development and its psychiatric associations during childhood. *Scientific Data*, *9*(1), 300.

Tudor, M., Tudor, L., & Tudor, K. I. (2005). [hans berger (1873-1941)–the history of electroencephalography]. *Acta Medica Croatica: Casopis Hravatske Akademije Medicinskih Znanosti*, *59*(4), 307–313.

Van De Ville, D., Britz, J., & Michel, C. M. (2010). EEG microstate sequences in healthy humans at rest reveal scale-free dynamics. *Proceedings of the National Academy of Sciences*, *107*(42), 18179–18184.

van den Heuvel, M. P., & Hulshoff Pol, H. E. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, *20*(8), 519–534.

Vega-Hernández, M., Martínez-Montes, E., Sanchez-Bornot, J., Castellanos, A., & Valdés-Sosa, P. (2008). Penalized Least Squares Methods For Solving The Eeg Inverse Problem. *Statistica Sinica*, *18*, 1535–1551.

Vohryzek, J., Deco, G., Cessac, B., Kringelbach, M. L., & Cabral, J. (2020). Ghost attractors in spontaneous brain activity: Recurrent excursions into functionally-relevant BOLD phase-locking states. *Frontiers in Systems Neuroscience*, *14*.

von Wegner, F., Tagliazucchi, E., & Laufs, H. (2017). Information-theoretical analysis of resting state EEG microstate sequences - non-Markovianity, non-stationarity and periodicities. *NeuroImage*, *158*, 99–111.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Essen, D. C. V., & Wager, T. D. (2011). NeuroSynth: A new platform for large-scale automated synthesis of human functional neuroimaging data. *Frontiers in Neuroinformatics*, *5*.

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165.

Young, G., & Householder, A. (1938). Discussion of a set of points in terms of their mutual distances. *3*, 19–22.

Yuan, H., Zotev, V., Phillips, R., Drevets, W. C., & Bodurka, J. (2012). Spatiotemporal dynamics of the brain at rest — Exploring EEG microstates as electrophysiological signatures of BOLD resting state networks. *NeuroImage*, *60*(4), 2062–2072.

A Appendices

A.1 List of Software Pacakages Developed

Significant software development toward the project goals has been achieved since the previous report under the direction of C.L. Nehaniv.

A.1.1 Porting KeyPy to Python 3.7

The EEG processing software **KeyPy** developed by Patricia Milz (KEY Institute for Brain-Mind Research, Switzerland) (Milz, 2015) has been ported from Python 2.7 to Python 3.7 and the repository made publicly available on github `https://github.com/uwaicl/keypy`. Also the calculation of variance explained by a set of EEG microstates has been corrected.
Project Members: Thomas George, Reinoud Maex, David G. Haydock

A.1.2 Epsilon-machine construction and minimization

epsilon-machine-minimization is written for constructing and minimization of epsilon automata, according to the data - set of strings (formal alphabet words), for example, sequences of microstates of EEG observations.

Additional functions are also possible after the construction of the automata, such as the calculation of the distance between the machines or the likelihood of the input sequence to be accepted or generated by the automaton.

Developed by project member Hanna Derets (MITACS intern in Summer 2021) under supervision of C. L. Nehaniv.

A.1.3 Log-likelihood Matching of ε-machines and Visualizations

Sebastián Dohnány and Hanna Derets (MITACS interns in Summer 2021) developed software **epsilon-evaluate** and others to compute log-likehood of data sequences on epsilon-machines; also various software for **Euclidean embedding, visualization in 2D and 3D** for sets of points with pairwise distances given by any distance metric.

A.1.4 Automata Animation

Sebastián Dohnány developed **epsilon-animate**: Animation of probabilisticly generated or pre-defined sequences executing on finite automata and epsilon-machines.

To request access to the software contact Prof. C. L. Nehaniv, University of Waterloo Algebraic Intelligence and Computation Lab at `chrystopher.nehaniv@uwaterloo.ca` .

A.2 Recognizing cognitive process condition using top 1 likelihood matching for Run 2 sequences tested on individual epsilon machines based on Run 1 data, for 23 individuals.

Individual sequences on 150 individual-level machines: cognitive process condition matching for the persons 3 - 20, 22 - 26 individually

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	50	0	0
	Vis	0	50	50	50	50	0	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	50	50	50	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	0	0	0	0	0	0

Table 10: Average Top1 score for matching cognitive process conditions of the person C03 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	50	50	50	50	100	50	0
	Ver	0	0	0	0	50	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	50	50	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	0	0	0	0	0	0
event	MW	50	0	0	0	0	0	0
	Ver	0	0	100	0	0	0	0
	Vis	0	0	50	0	0	0	0

Table 11: Average Top1 score for matching cognitive process conditions of the person C04 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	50
	Vis	0	0	0	0	0	0	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	50	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	50	0	0	0	0	0

Table 12: Average Top1 score for matching cognitive process conditions of the person C05 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	50	0	50	50	50	0	0
	Ver	0	50	50	0	0	0	0
	Vis	50	50	0	50	100	100	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	50	0	0	0	0	0
event	MW	0	50	50	0	0	0	0
	Ver	0	0	50	0	0	0	0
	Vis	100	0	0	0	0	0	0

Table 13: Average Top1 score for matching cognitive process conditions of the person C06 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	50	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	50	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 14: Average Top1 score for matching cognitive process conditions of the person C07 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	50
	Ver	0	0	50	50	50	50	0
	Vis	0	0	0	0	0	0	0
peak	MW	50	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	50	0	0	0	50
event	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 15: Average Top1 score for matching cognitive process conditions of the person C08 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	50	50	0	0	50	0
peak	MW	50	0	50	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	50	0	0	0	0
event	MW	0	0	50	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 16: Average Top1 score for matching cognitive process conditions of the person C09 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	50	50	50	50	50	50
	Ver	0	0	50	50	0	0	0
	Vis	0	0	50	100	50	0	50
peak	MW	0	50	50	0	0	0	0
	Ver	50	50	0	0	0	0	0
	Vis	0	50	50	0	0	0	0
event	MW	0	0	50	0	0	0	0
	Ver	50	0	50	0	0	0	0
	Vis	0	50	50	0	0	0	0

Table 17: Average Top1 score for matching cognitive process conditions of the person C10 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	50	50	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	0	0	50	50	0	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	100	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	50	0	0	0	0

Table 18: Average Top1 score for matching cognitive process conditions of the person C11 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	0	0	0	0	0	50
peak	MW	0	0	50	0	0	0	0
	Ver	0	0	50	0	0	0	0
	Vis	0	50	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	50
	Vis	0	0	0	0	0	0	0

Table 19: Average Top1 score for matching cognitive process conditions of the person C12 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	50	0	50	0	0	0	0
	Ver	50	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	0	50	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 20: Average Top1 score for matching cognitive process conditions of the person C13 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	50	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	50
	Vis	0	0	0	0	0	0	0

Table 21: Average Top1 score for matching cognitive process conditions of the person C14 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	50	0	0	0	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	0	50	0	0	0	0
event	MW	0	50	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	0	0	0	0	0	0

Table 22: Average Top1 score for matching cognitive process conditions of the person C15 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	50	0	0	0	0	0	0
	Vis	0	0	0	50	0	0	0

Table 23: Average Top1 score for matching cognitive process conditions of the person C16 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	50	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	100	0	0	0	0
	Vis	0	0	50	0	0	0	0
event	MW	100	0	0	50	0	0	0
	Ver	50	0	0	0	0	0	0
	Vis	0	50	0	0	0	0	0

Table 24: Average Top1 score for matching cognitive process conditions of the person C17 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	50	50	50	50	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	0	50	0	0	0	0
	Ver	0	0	50	0	0	0	0
	Vis	50	0	0	0	0	50	0
event	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	0	0	0	0	0	0

Table 25: Average Top1 score for matching cognitive process conditions of the person C18 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	50	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	0	0	50	0	0	0
	Vis	0	50	0	0	0	0	0

Table 26: Average Top1 score for matching cognitive process conditions of the person C19 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	0	0	50	0	0	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	50	0	0	0	0
	Vis	0	50	50	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 27: Average Top1 score for matching cognitive process conditions of the person C20 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	50	50	0
peak	MW	0	50	0	0	0	0	0
	Ver	0	0	0	0	0	0	50
	Vis	0	0	50	0	0	0	0
event	MW	0	0	0	0	0	0	50
	Ver	0	0	50	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 28: Average Top1 score for matching cognitive process conditions of the person C22 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	50	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	50
peak	MW	0	0	0	0	0	0	50
	Ver	0	0	0	0	0	0	0
	Vis	50	0	0	0	0	0	50
event	MW	0	0	0	0	0	0	0
	Ver	0	0	50	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 29: Average Top1 score for matching cognitive process conditions of the person C23 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	50	0	0	0	0	0	50
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	0	0	0	0	100	0
event	MW	50	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 30: Average Top1 score for matching cognitive process conditions of the person C24 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	50	0	0	0	0	0
event	MW	0	0	100	0	0	0	50
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 31: Average Top1 score for matching cognitive process conditions of the person C25 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	50
	Ver	50	50	0	0	0	0	0
	Vis	0	0	0	0	50	0	0
peak	MW	0	0	0	50	0	0	0
	Ver	0	50	0	0	50	0	0
	Vis	0	0	0	50	0	0	0
event	MW	0	0	50	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 32: Average Top1 score for matching cognitive process conditions of the person C26 for 6 files with **sequences of R2** data to the 150 individual-level epsilon **machines built from R1** data.

A.3 Recognizing cognitive process condition and person using top 1 likelihood matching for Run 1 sequences tested on individual and group-level epsilon machines based on Run 2 data.

Repeating the same test but swapping data for testing sequences and for building machines, i.e. using individual R1 files to check their likelihood matching with machines built from R2 data, we get the following results.

A.3.1 Individual sequences on 3 group-level machines: cognitive process condition matching

time	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	32.0 ± 6.3	32.0 ± 6.3	30.0 ± 6.2	26.0 ± 5.9	24.0 ± 5.7	28.0 ± 6.0	34.0 ± 6.4
	Ver	50.0 ± 6.7	50.0 ± 6.7	52.0 ± 6.7	52.0 ± 6.7	54.0 ± 6.7	54.0 ± 6.7	52.0 ± 6.7
	Vis	38.0 ± 6.5	38.0 ± 6.5	36.0 ± 6.4	42.0 ± 6.6	38.0 ± 6.5	34.0 ± 6.4	34.0 ± 6.4
peak	MW	24.0 ± 5.7	20.0 ± 5.4	14.0 ± 4.7	22.0 ± 5.6	34.0 ± 6.4	28.0 ± 6.0	36.0 ± 6.4
	Ver	40.0 ± 6.6	42.0 ± 6.6	44.0 ± 6.7	36.0 ± 6.4	38.0 ± 6.5	60.0 ± 6.6	56.0 ± 6.7
	Vis	46.0 ± 6.7	54.0 ± 6.7	52.0 ± 6.7	60.0 ± 6.6	50.0 ± 6.7	28.0 ± 6.0	16.0 ± 4.9
event	MW	24.0 ± 5.7	20.0 ± 5.4	24.0 ± 5.7	32.0 ± 6.3	32.0 ± 6.3	30.0 ± 6.2	34.0 ± 6.4
	Ver	44.0 ± 6.7	40.0 ± 6.6	32.0 ± 6.3	38.0 ± 6.5	36.0 ± 6.4	28.0 ± 6.0	38.0 ± 6.5
	Vis	44.0 ± 6.7	46.0 ± 6.7	48.0 ± 6.7	44.0 ± 6.7	44.0 ± 6.7	54.0 ± 6.7	48.0 ± 6.7

Table 33: Average Top1 score (with 95% Confidence Interval) for matching condition of 150 files with **sequences of R1** data to the three group-level epsilon **machines built from R2** data.

A.3.2 Individual sequences on 6 group-level machines: cognitive process condition matching

time	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	48.0 ± 6.7	46.0 ± 6.7	46.0 ± 6.7	46.0 ± 6.7	48.0 ± 6.7	48.0 ± 6.7	36.0 ± 6.4
	Ver	44.0 ± 6.7	48.0 ± 6.7	48.0 ± 6.7	50.0 ± 6.7	50.0 ± 6.7	52.0 ± 6.7	58.0 ± 6.6
	Vis	4.0 ± 2.6	2.0 ± 1.9	2.0 ± 1.9	2.0 ± 1.9	2.0 ± 1.9	0.0 ± 0.0	2.0 ± 1.9
peak	MW	48.0 ± 6.7	38.0 ± 6.5	34.0 ± 6.4	32.0 ± 6.3	36.0 ± 6.4	28.0 ± 6.0	34.0 ± 6.4
	Ver	48.0 ± 6.7	54.0 ± 6.7	58.0 ± 6.6	42.0 ± 6.6	38.0 ± 6.5	78.0 ± 5.6	30.0 ± 6.2
	Vis	4.0 ± 2.6	16.0 ± 4.9	22.0 ± 5.6	32.0 ± 6.3	44.0 ± 6.7	6.0 ± 3.2	32.0 ± 6.3
event	MW	50.0 ± 6.7	50.0 ± 6.7	48.0 ± 6.7	50.0 ± 6.7	46.0 ± 6.7	40.0 ± 6.6	34.0 ± 6.4
	Ver	26.0 ± 5.9	26.0 ± 5.9	30.0 ± 6.2	30.0 ± 6.2	32.0 ± 6.3	26.0 ± 5.9	50.0 ± 6.7
	Vis	24.0 ± 5.7	28.0 ± 6.0	28.0 ± 6.0	34.0 ± 6.4	42.0 ± 6.6	46.0 ± 6.7	8.0 ± 3.6

Table 34: Average Top1 score (with 95% Confidence Interval) for matching condition of 150 files with **sequences of R1** data to the six group-level epsilon **machines built from R2** data.

A.3.3 Individual sequences on 150 individual-level machines: cognitive process condition matching

time	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	40.0 ± 6.6	42.0 ± 6.6	40.0 ± 6.6	38.0 ± 6.5	38.0 ± 6.5	38.0 ± 6.5	24.0 ± 5.7
	Ver	40.0 ± 6.6	38.0 ± 6.5	38.0 ± 6.5	36.0 ± 6.4	34.0 ± 6.4	34.0 ± 6.4	38.0 ± 6.5
	Vis	50.0 ± 6.7	50.0 ± 6.7	50.0 ± 6.7	46.0 ± 6.7	46.0 ± 6.7	48.0 ± 6.7	44.0 ± 6.7
peak	MW	38.0 ± 6.5	28.0 ± 6.0	28.0 ± 6.0	32.0 ± 6.3	68.0 ± 6.3	58.0 ± 6.6	46.0 ± 6.7
	Ver	42.0 ± 6.6	50.0 ± 6.7	52.0 ± 6.7	52.0 ± 6.7	22.0 ± 5.6	18.0 ± 5.2	44.0 ± 6.7
	Vis	44.0 ± 6.7	42.0 ± 6.6	28.0 ± 6.0	20.0 ± 5.4	24.0 ± 5.7	24.0 ± 5.7	20.0 ± 5.4
event	MW	44.0 ± 6.7	32.0 ± 6.3	28.0 ± 6.0	4.0 ± 2.6	46.0 ± 6.7	44.0 ± 6.7	56.0 ± 6.7
	Ver	36.0 ± 6.4	46.0 ± 6.7	68.0 ± 6.3	58.0 ± 6.6	66.0 ± 6.4	42.0 ± 6.6	48.0 ± 6.7
	Vis	44.0 ± 6.7	32.0 ± 6.3	12.0 ± 4.4	34.0 ± 6.4	2.0 ± 1.9	2.0 ± 1.9	8.0 ± 3.6

Table 35: Average Top1 score (with 95% Confidence Interval) for matching condition of 150 files with **sequences of R1** data to the 150 individual epsilon **machines built from R2** data.

A.3.4 Individual sequences on 150 individual-level machines: person matching

time	person	L1	L2	L3	L4	L5	L6	L7
clock	C01	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C02	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	33.3 ± 18.3	0.0 ± 0.0	33.3 ± 18.3	16.7 ± 14.5
	C03	50.0 ± 19.4	33.3 ± 18.3	16.7 ± 14.5	50.0 ± 19.4	50.0 ± 19.4	33.3 ± 18.3	33.3 ± 18.3
	C04	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	33.3 ± 18.3	33.3 ± 18.3	33.3 ± 18.3
	C05	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C06	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0
	C07	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C08	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	33.3 ± 18.3	16.7 ± 14.5	0.0 ± 0.0	16.7 ± 14.5
	C09	33.3 ± 18.3	33.3 ± 18.3	16.7 ± 14.5	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0
	C10	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5
	C11	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5
	C12	33.3 ± 18.3	33.3 ± 18.3	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C13	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C14	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	33.3 ± 18.3	16.7 ± 14.5
	C15	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C16	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C17	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0
	C18	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C19	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C20	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	33.3 ± 18.3	16.7 ± 14.5
	C22	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	33.3 ± 18.3	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0
	C23	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	16.7 ± 14.5

	C24	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5
	C25	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5
	C26	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5
	All	10.0 ± 5.2	10.0 ± 4.5	8.7 ± 3.2	11.3 ± 5.4	8.0 ± 4.9	10.7 ± 5.1	9.3 ± 4.1
peak	C01	0.0 ± 0.0	0.0 ± 0.0	66.7 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C02	33.3 ± 18.3	50.0 ± 19.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C03	16.7 ± 14.5	33.3 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C04	16.7 ± 14.5	33.3 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C05	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C06	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C07	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C08	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C09	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C10	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C11	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C12	0.0 ± 0.0	0.0 ± 0.0	33.3 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C13	0.0 ± 0.0	0.0 ± 0.0	33.3 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C14	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C15	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C16	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C17	0.0 ± 0.0	0.0 ± 0.0	66.7 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C18	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	33.3 ± 18.3	0.0 ± 0.0
	C19	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C20	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C22	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C23	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5
	C24	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C25	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C26	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	All	5.3 ± 3.5	7.3 ± 5.2	11.3 ± 7.5	2.0 ± 2.1	0.0 ± 0.0	1.3 ± 2.5	0.7 ± 1.3
event	C01	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C02	33.3 ± 18.3	50.0 ± 19.4	66.7 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C03	16.7 ± 14.5	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C04	16.7 ± 14.5	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C05	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C06	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C07	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C08	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C09	16.7 ± 14.5	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C10	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C11	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C12	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0

	C13	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C14	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C15	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	16.7 ± 14.5	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0
	C16	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C17	0.0 ± 0.0	0.0 ± 0.0	33.3 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C18	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	16.7 ± 14.5
	C19	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C20	33.3 ± 18.3	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C22	33.3 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C23	0.0 ± 0.0	0.0 ± 0.0	33.3 ± 18.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5
	C24	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C25	0.0 ± 0.0	33.3 ± 18.3	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	C26	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	16.7 ± 14.5	0.0 ± 0.0	0.0 ± 0.0
	All	10.0 ± 4.5	8.7 ± 4.9	10.0 ± 6.1	2.7 ± 2.4	2.0 ± 2.1	0.0 ± 0.0	1.3 ± 1.8

Table 36: Average Top1 score (with 95% Confidence Interval) for matching person of 150 files with **sequences of R1** data to the 150 individual epsilon **machines built from R2** data.

A.3.5 Individual sequences on 150 individual-level machines: cognitive process condition matching for each person individually

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	50	50	50	50	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	0	100	0	0	0	0
	Ver	0	0	50	0	0	0	0
	Vis	0	0	50	0	0	0	0
event	MW	0	50	0	0	0	0	0
	Ver	50	0	0	0	0	0	0
	Vis	0	0	0	50	0	0	0

Table 37: Average Top1 score for matching cognitive process conditions of the person C01 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	50	50	50	50	0	50	0
	Ver	0	0	0	50	0	0	0
	Vis	0	0	0	0	0	50	50
peak	MW	50	100	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	50	0	0	0	0	0
event	MW	50	50	0	0	0	0	0
	Ver	0	50	100	0	0	0	0
	Vis	50	50	100	0	0	0	0

Table 38: Average Top1 score for matching cognitive process conditions of the person C02 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	100	50	50	100	100	0	50
	Vis	50	50	0	50	50	100	50
peak	MW	0	0	0	0	0	0	0
	Ver	0	50	0	0	0	0	0
	Vis	50	50	0	0	0	0	0
event	MW	0	0	50	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	50	0	0	0	0	0

Table 39: Average Top1 score for matching cognitive process conditions of the person C03 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	50	50	50
	Ver	0	0	0	0	0	0	0
	Vis	0	0	50	50	50	50	50
peak	MW	0	0	0	0	0	0	0
	Ver	0	100	0	0	0	0	0
	Vis	50	0	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	0	50	0	0	0	0

Table 40: Average Top1 score for matching cognitive process conditions of the person C04 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	50	50	50	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	0	0	0	0	0	0
event	MW	0	0	50	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 41: Average Top1 score for matching cognitive process conditions of the person C05 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	50	0
	Vis	50	50	0	0	0	0	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	0	0	50	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	50	0	0	0	0	0

Table 42: Average Top1 score for matching cognitive process conditions of the person C06 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	50	0	0	0	0
peak	MW	0	50	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	0	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	50	0	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 43: Average Top1 score for matching cognitive process conditions of the person C07 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	50	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	50	50	50	50	0	50
peak	MW	50	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	0	0	50	0	0	0

Table 44: Average Top1 score for matching cognitive process conditions of the person C08 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	50	50	50	0	50	0	0
	Vis	50	50	0	0	0	50	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	50	0	50	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 45: Average Top1 score for matching cognitive process conditions of the person C09 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	50	0	0	0	0	50	50
	Ver	0	0	0	0	50	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	50	0	0	0	0	0
	Ver	0	0	50	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 46: Average Top1 score for matching cognitive process conditions of the person C10 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	50	50
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	50	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 47: Average Top1 score for matching cognitive process conditions of the person C11 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	100	100	50	50	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	50	0	0	0	0
	Vis	0	0	50	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	50	0	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 48: Average Top1 score for matching cognitive process conditions of the person C12 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	0	50	0	0	0	0
	Ver	0	0	50	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	50	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	50	0	0	0	0	0

Table 49: Average Top1 score for matching cognitive process conditions of the person C13 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	50	50	50
	Vis	0	0	50	50	0	50	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	0	0	0	0	0	0

Table 50: Average Top1 score for matching cognitive process conditions of the person C14 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	50	50	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	0	50	50	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	0	50	0	0	50	0	0
	Ver	0	0	0	50	0	0	0
	Vis	0	0	0	0	0	0	0

Table 51: Average Top1 score for matching cognitive process conditions of the person C15 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	0	50	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	0	0	50	0	0	0
	Vis	0	0	0	0	0	0	0

Table 52: Average Top1 score for matching cognitive process conditions of the person C16 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	50	50	0	0	50	0	0
	Vis	0	0	0	0	0	50	0
peak	MW	0	0	100	0	0	0	0
	Ver	0	0	50	0	0	0	0
	Vis	0	0	50	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	0	100	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 53: Average Top1 score for matching cognitive process conditions of the person C17 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	0	0	0	0	50	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	50	0
event	MW	50	0	0	0	50	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	50

Table 54: Average Top1 score for matching cognitive process conditions of the person C18 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	50	0	0	0
	Vis	0	0	0	0	0	0	0
peak	MW	0	0	0	50	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 55: Average Top1 score for matching cognitive process conditions of the person C19 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	100	0
	Vis	0	0	0	0	0	0	50
peak	MW	50	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	50	0	50	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	50	0	0	0	0	0	0

Table 56: Average Top1 score for matching cognitive process conditions of the person C20 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	50	50	50	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	50	0	0	0
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	50	0	0	0	0	0	0
	Vis	50	0	0	0	0	0	0

Table 57: Average Top1 score for matching cognitive process conditions of the person C22 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	50	50	50	50	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	50
peak	MW	0	0	0	0	0	0	50
	Ver	0	0	50	0	0	0	0
	Vis	0	50	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	100	0	0	0	50

Table 58: Average Top1 score for matching cognitive process conditions of the person C23 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	50	50	0	0	50
	Vis	0	0	0	0	0	0	0
peak	MW	0	50	0	0	0	0	0
	Ver	0	0	50	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	0	50	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 59: Average Top1 score for matching cognitive process conditions of the person C24 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	50	0	0	0
	Ver	50	50	0	0	0	0	0
	Vis	0	0	0	0	0	0	50
peak	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	0	0	0	0	0	0	0
	Ver	0	50	50	0	0	0	0
	Vis	0	50	0	0	0	0	0

Table 60: Average Top1 score for matching cognitive process conditions of the person C25 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

time mode	condition	L1	L2	L3	L4	L5	L6	L7
clock	MW	0	0	0	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	50	50
peak	MW	0	0	50	0	0	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0
event	MW	0	0	0	0	50	0	0
	Ver	0	0	0	0	0	0	0
	Vis	0	0	0	0	0	0	0

Table 61: Average Top1 score for matching cognitive process conditions of the person C26 for 6 files with **sequences of R1** data to the 150 individual-level epsilon **machines built from R2** data.

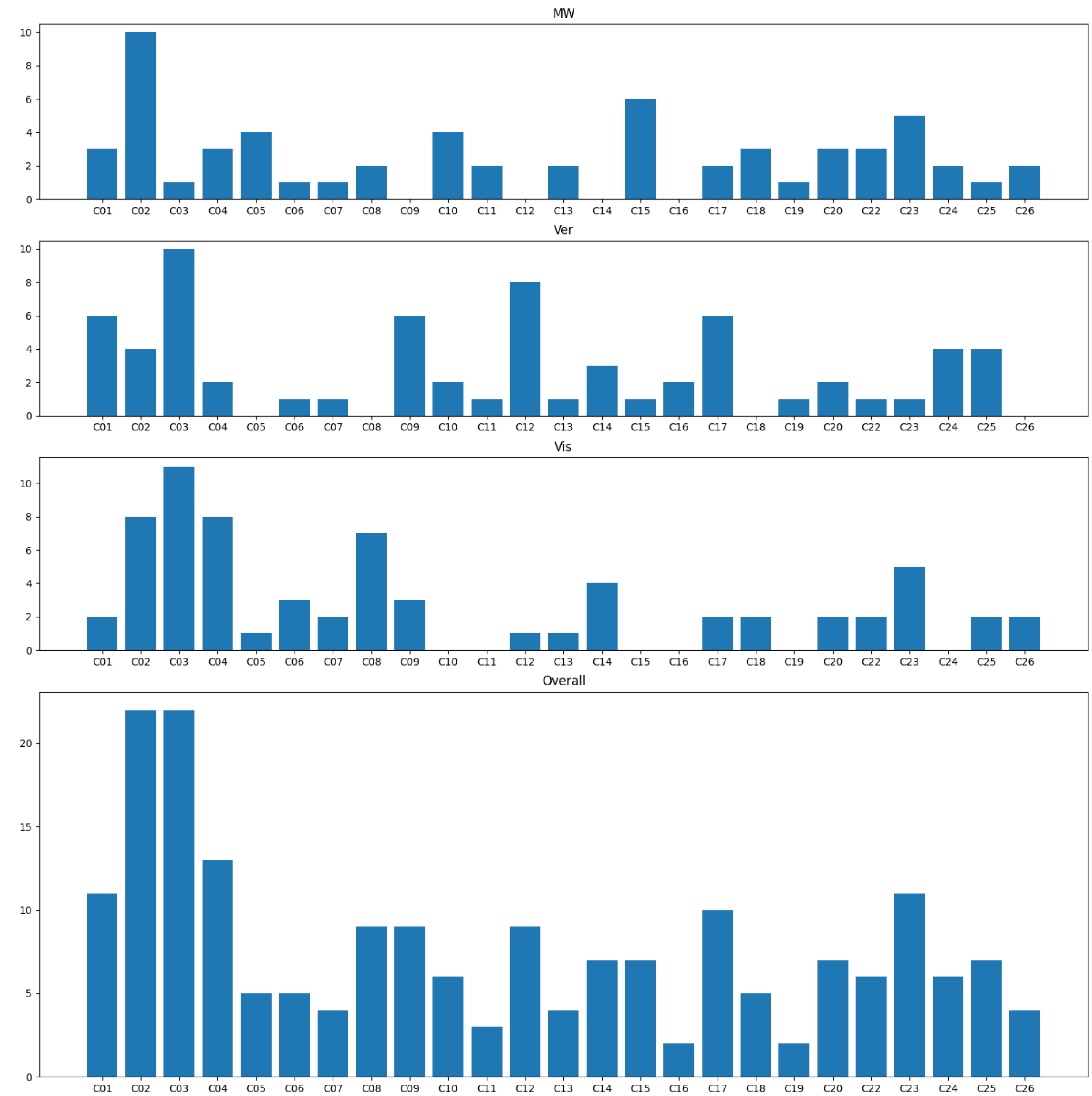

Figure 45: The number of correct matching of cognitive process conditions across individuals, i.e., cases where the machine with the highest likelihood corresponds to the same cognitive process condition as the sequence. The graph displays the result of testing 150 individual **R1 sequences on 150 R2 epsilon automata**, for three cognitive mode conditions in all three time modes (clock, peak, event) and 7 history lengths (from $L = 1$ to $L = 7$). Since two files with sequences correspond to each person, in all respects we have 42 tests for each person. The last table shows the sum of the result for three cognitive process conditions, displaying the overall process matching success across people.

B Supplementary Figures

B.1 Clustering with 6 and 10 fMRI Coactivation Patterns (CAPs)

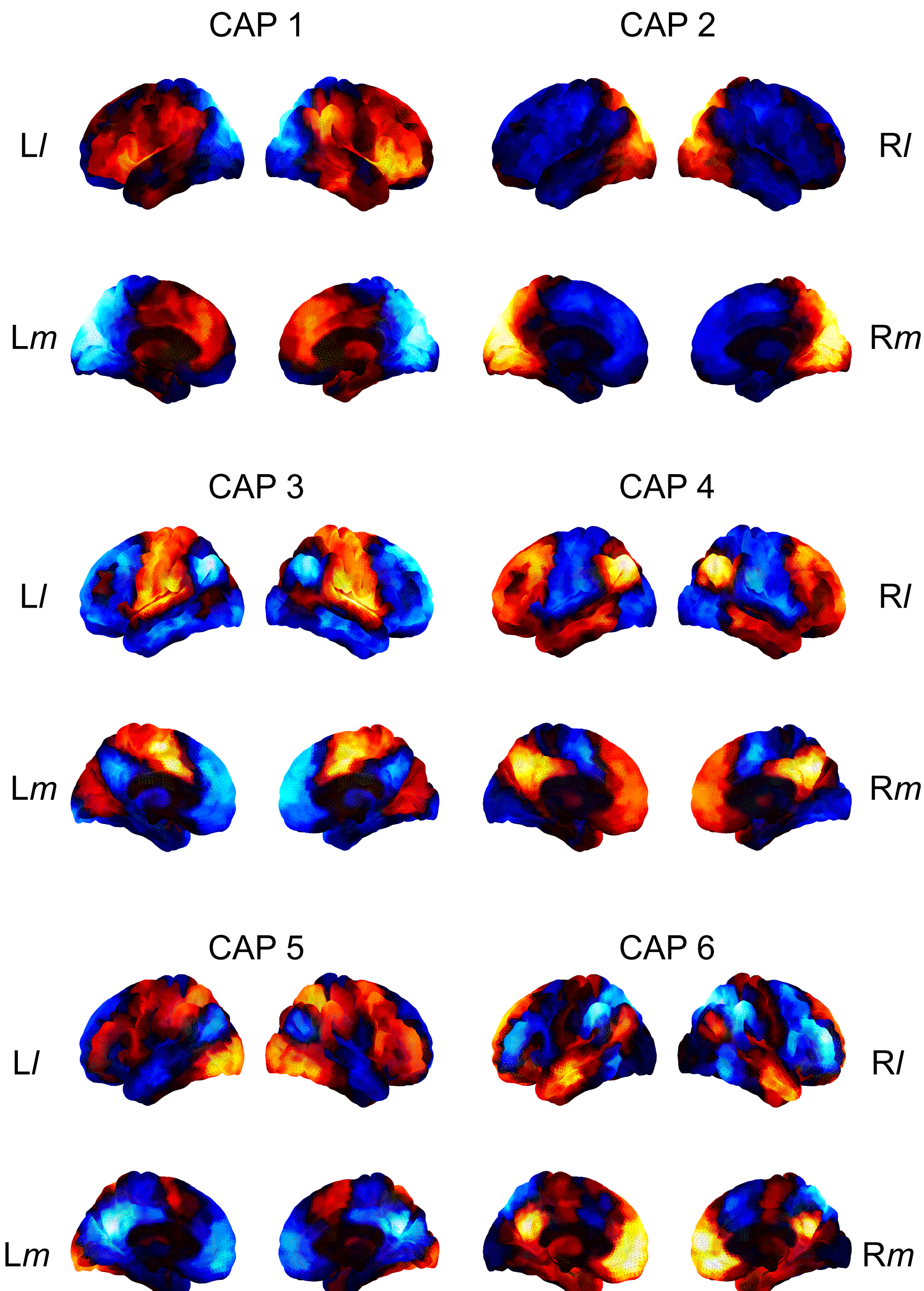

Figure 46: CAPs generated using six clusters. CAPs are paired based on Munkres assignment algorithm applied to their spatial correlations.

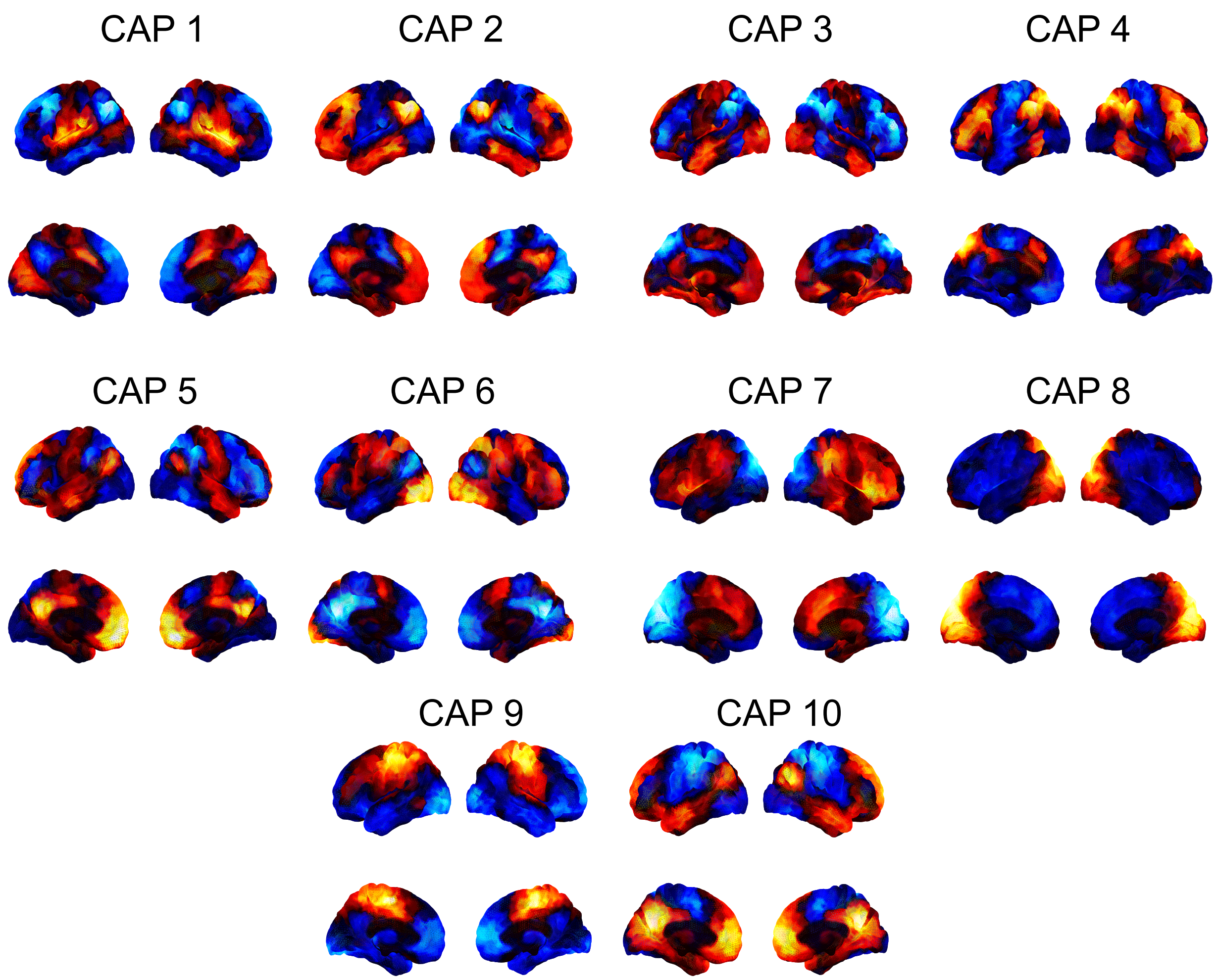

Figure 47: CAPs generated using ten clusters. CAPs are paired based on Munkres assignment algorithm applied to their spatial correlations.

,

B.2 Two-Dimensional Projections for Metric Geometries on Epsilon-Machines for Mind-Wandering, Visualization and Verbalization Conditions

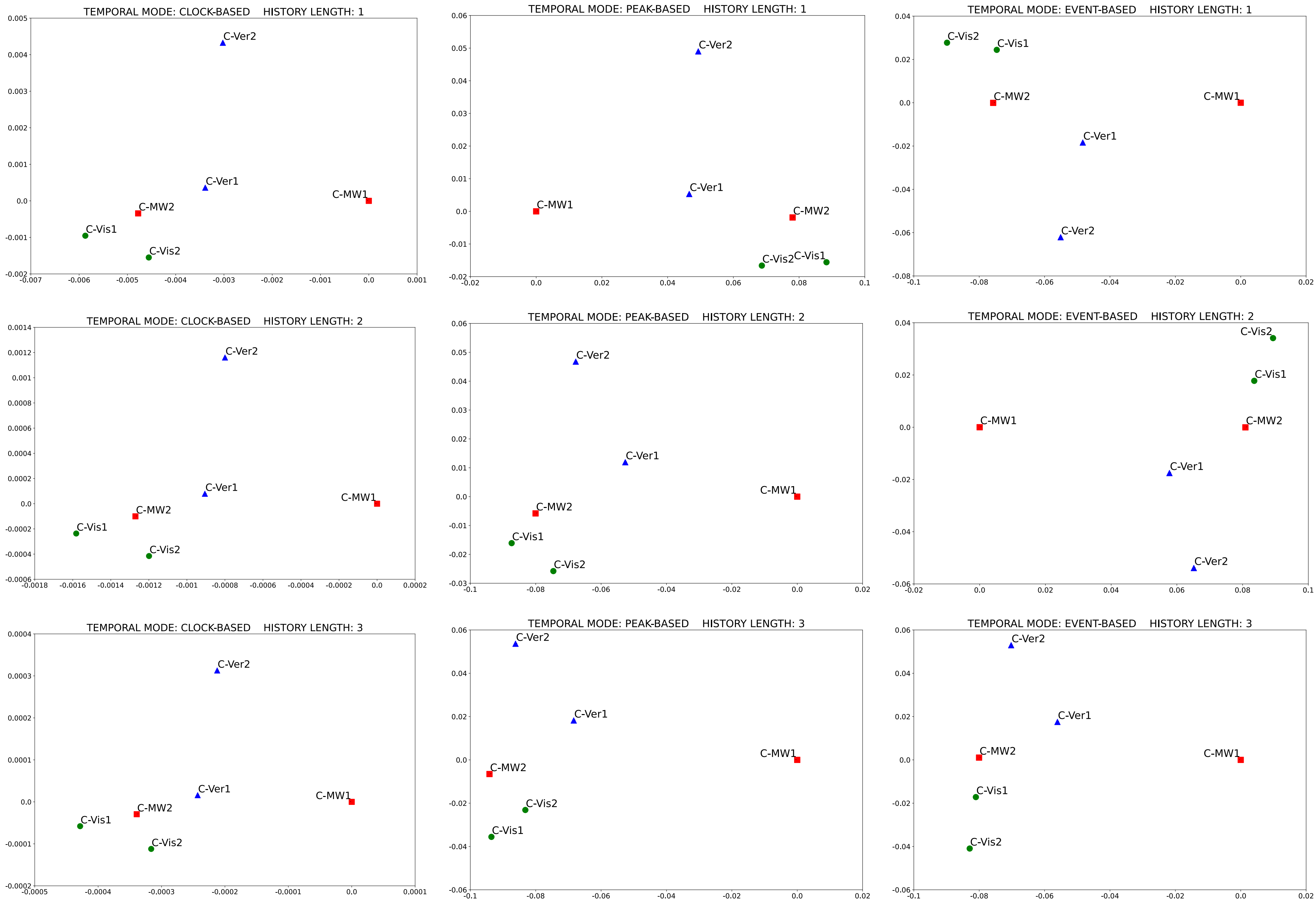

Figure 48: 2D projections of Euclidean embedding of metric spaces for group-level pre-epsilon-machines for mind-wandering (square shaped points), verbalization (triangle shaped points) and visualization (circle shaped points) conditions. Three columns correspond to the three time modes: clock, peak and event respectively. All figures correspond to the **epsilon distance** type. The three rows correspond to three history lengths $L = 1$, $L = 2$, $L = 3$ respectively.

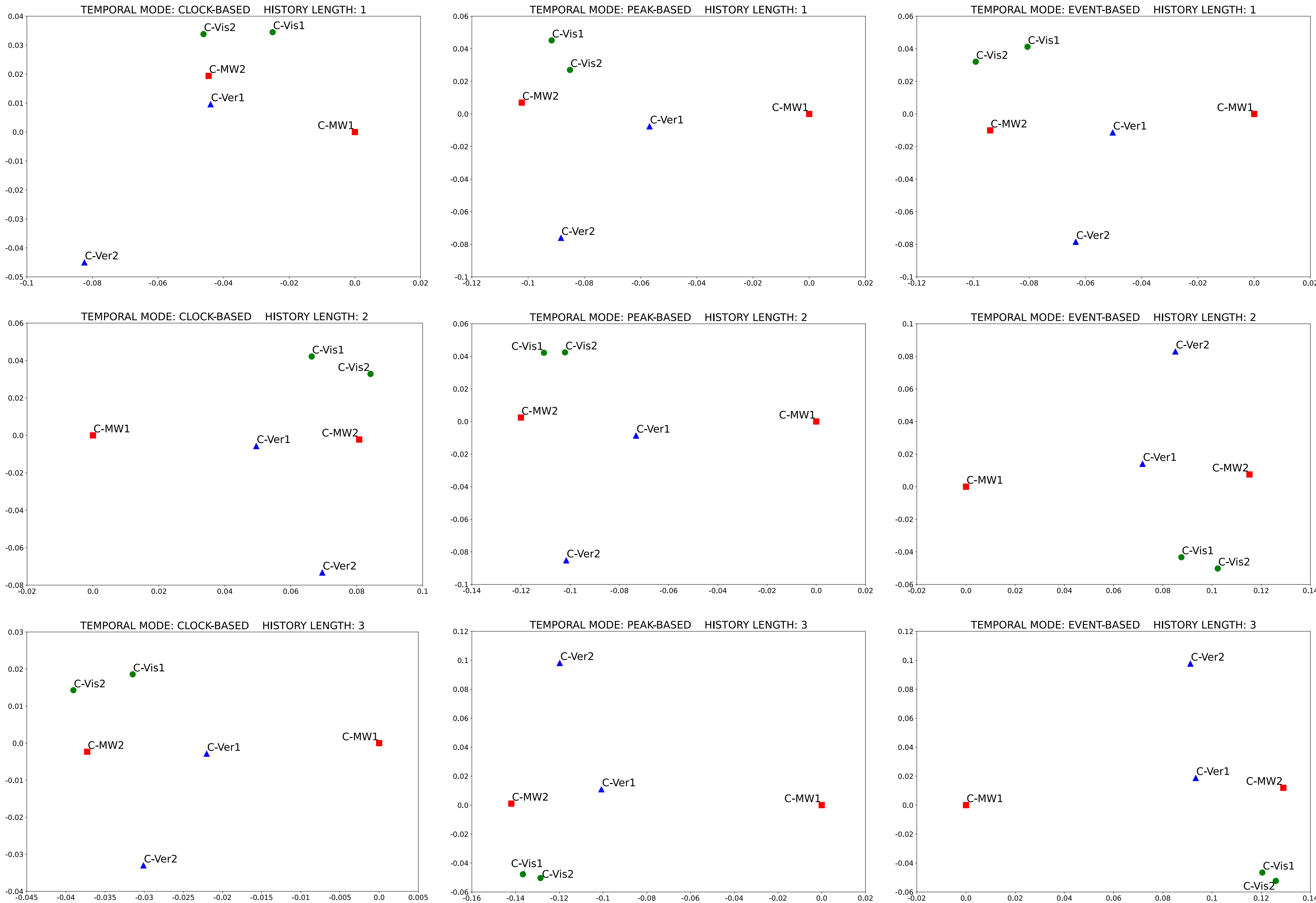

Figure 49: 2D projections of Euclidean embedding of the metrics spaces for group-level pre-epsilon-machines for mind-wandering (square shaped points), verbalization (triangle shaped points) and visualization (circle shaped points) conditions. Three columns correspond to the three time modes: clock, peak and event respectively. All figures correspond to the **Jaccard distance** type. The three rows correspond to three history lengths $L = 1$, $L = 2$, $L = 3$ respectively.

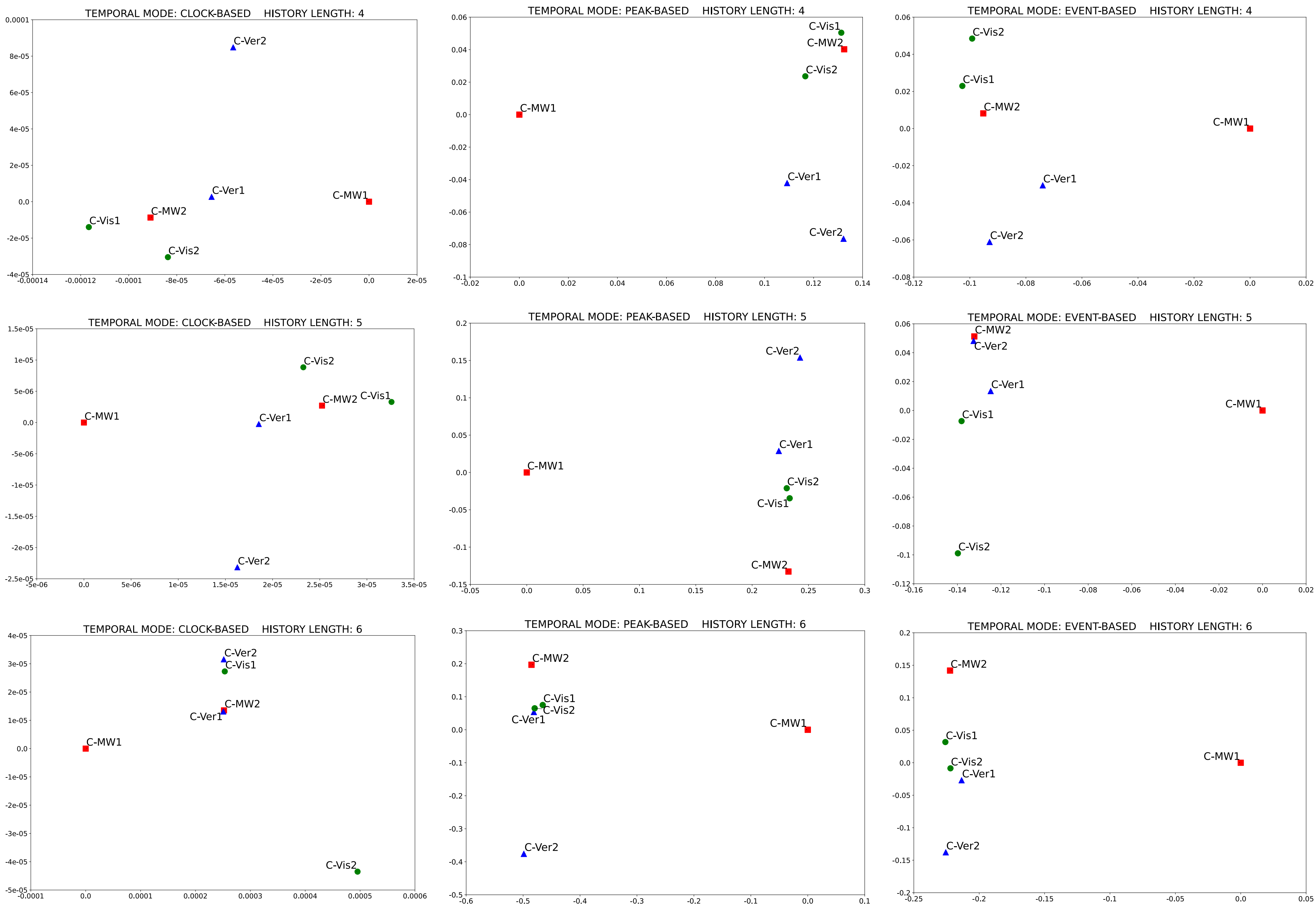

Figure 50: 2D projections of Euclidean embedding of the metrics spaces for group-level pre-epsilon-machines for mind-wandering (square shaped points), verbalization (triangle shaped points) and visualization (circle shaped points) conditions. Three columns correspond to the three time modes: clock, peak and event respectively. All figures correspond to the **epsilon distance** type. The three rows correspond to three history lengths $L = 4$, $L = 5$, $L = 6$ respectively.

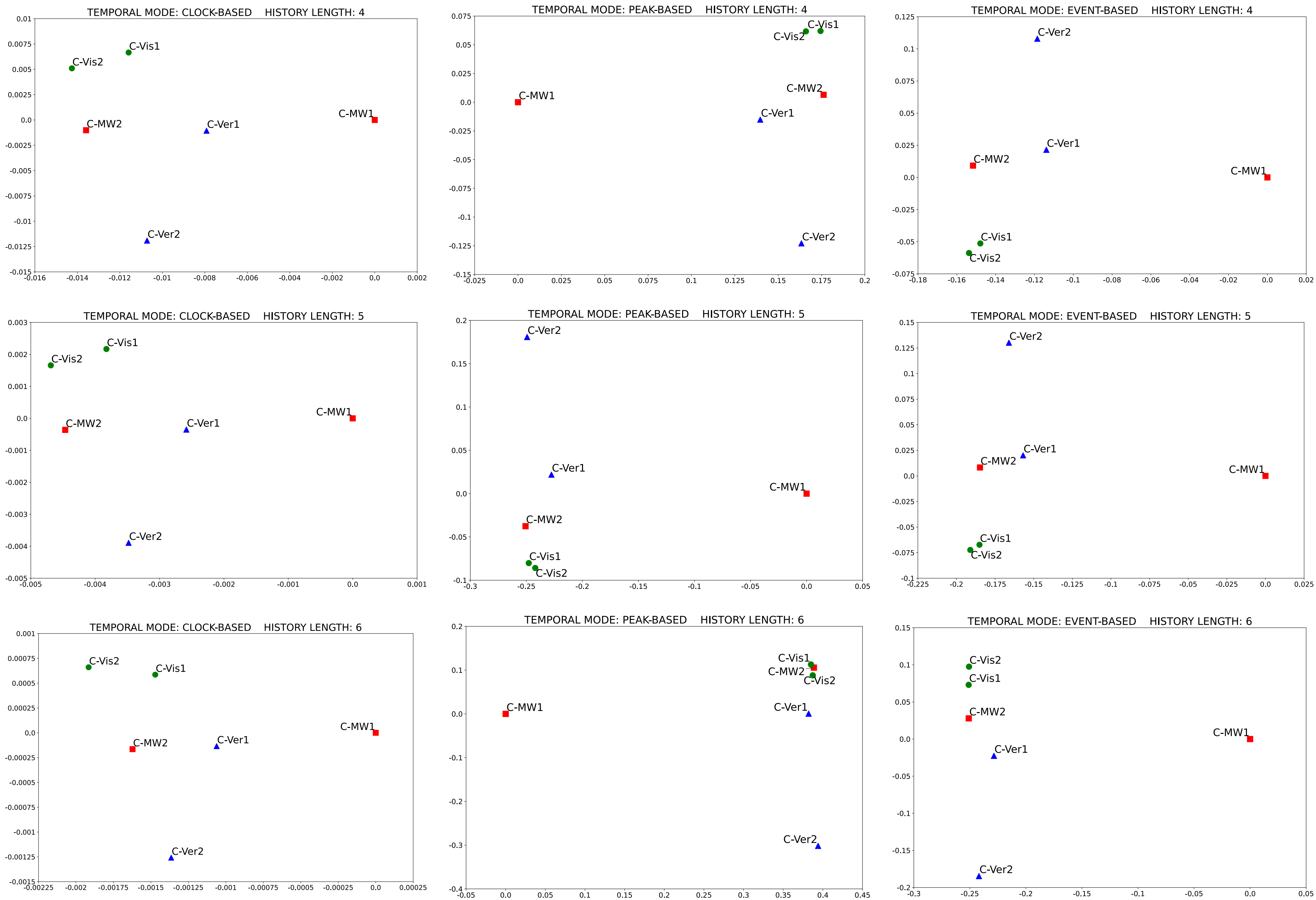

Figure 51: 2D projections of Euclidean embedding of the metrics spaces for group-level pre-epsilon-machines for mind-wandering (square shaped points), verbalization (triangle shaped points) and visualization (circle shaped points) conditions. Three columns correspond to the three time modes: clock, peak and event respectively. All figures correspond to the **Jaccard distance** type. The three rows correspond to three history lengths $L = 4$, $L = 5$, $L = 6$ respectively.

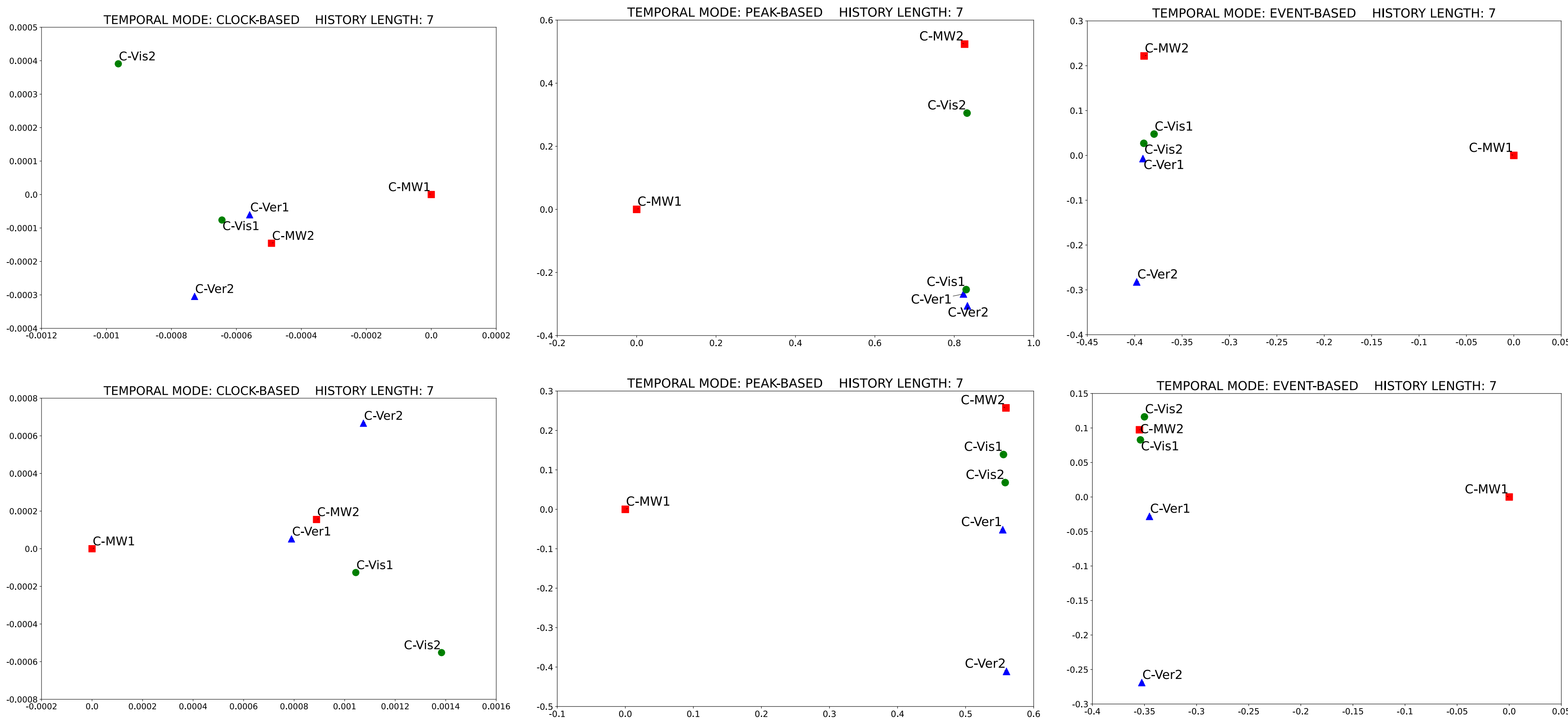

Figure 52: 2D projections of Euclidean embedding of the metrics spaces for group-level pre-epsilon-machines for mind-wandering (square shaped points), verbalization (triangle shaped points) and visualization (circle shaped points) conditions. Three columns correspond to the three time modes: clock, peak and event respectively. The first row corresponds to **epsilon distance** type, second to **Jaccard distance**, history length is $L = 7$.

B.3 Metric Space Percentage of Separation for Cognitive Processing Modes using Epsilon-Machines with Different Temporal Scales, Different History Lengths and Differing Thresholds for Causal State Identification

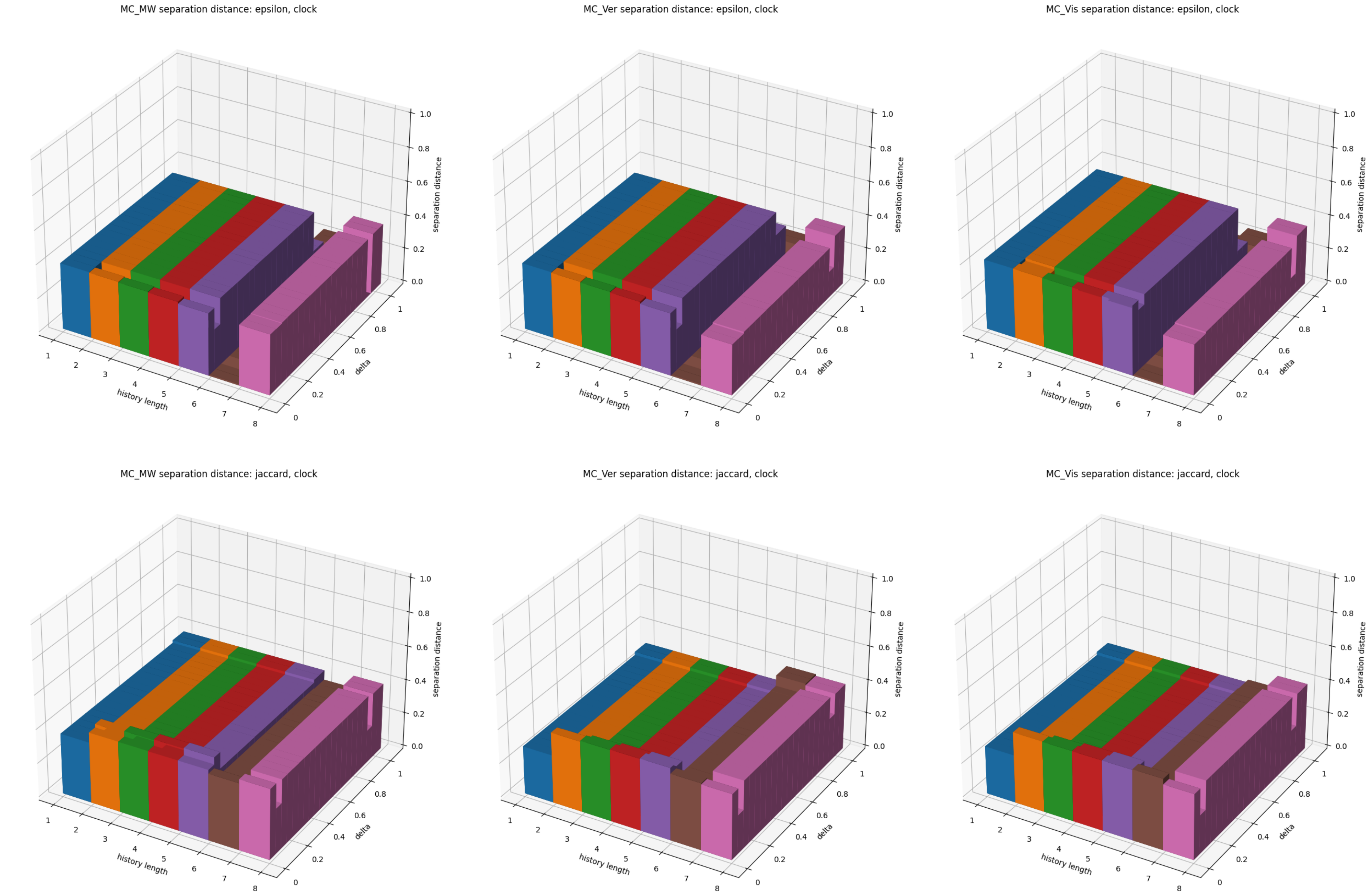

Figure 53: Percentage of separation for the minimized cognitive process condition machines, corresponding to different history lengths (from $L = 1$ to $L = 7$) and minimization deltas (from $\delta = 0$ to $\delta = 1$ with the step size 0.05). First row corresponds to the **epsilon distance** type and second row corresponds to **Jaccard distance**. Three conditions: mind-wandering, verbalization and visualization correspond to the three columns respectively. Time mode for all represented pictures is ***clock***.

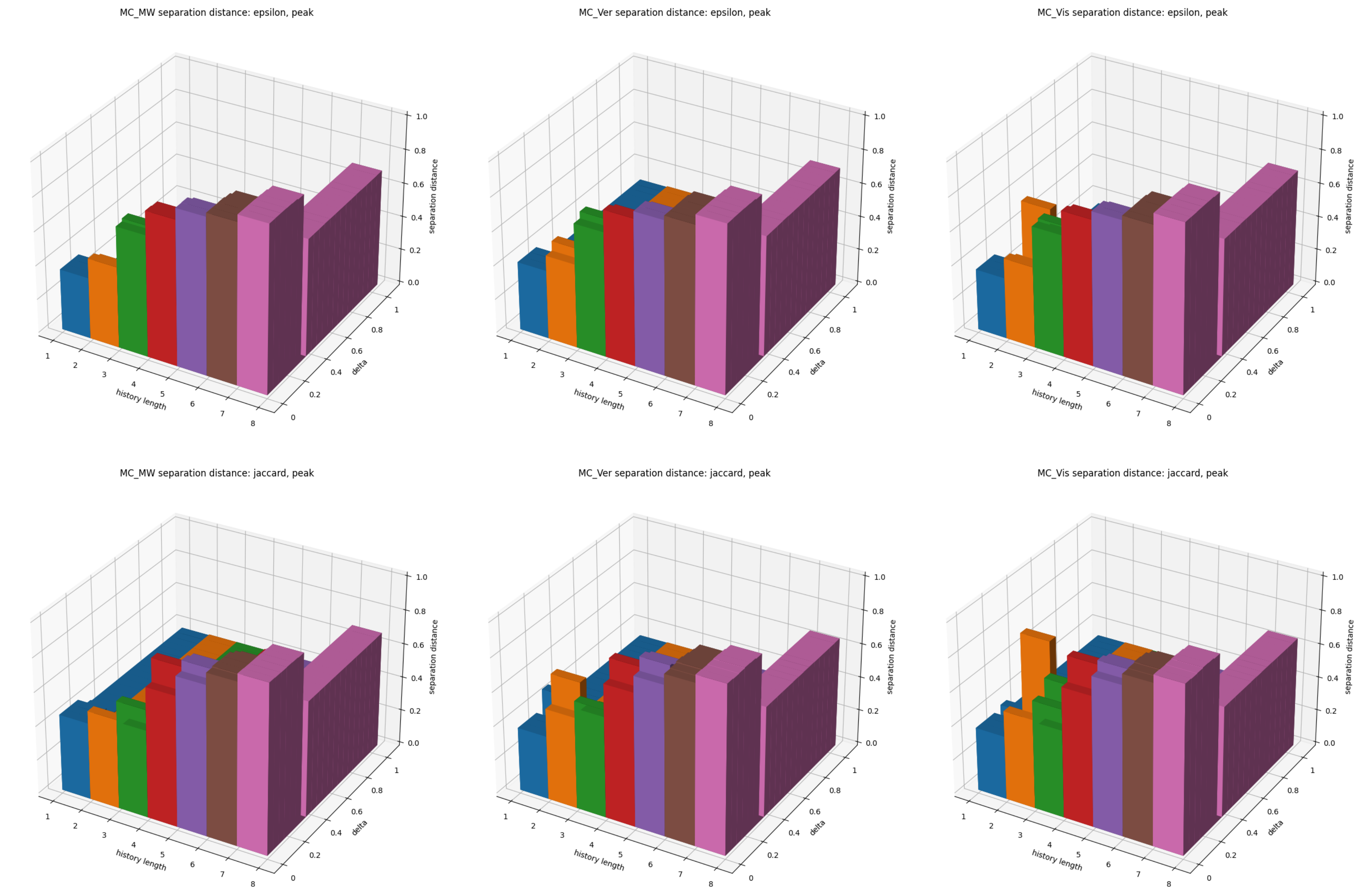

Figure 54: Percentage of separation for the minimized cognitive process condition machines, corresponding to different history lengths (from $L = 1$ to $L = 7$) and minimization deltas (from $\delta = 0$ to $\delta = 1$ with the step size 0.05). First row corresponds to the **epsilon distance** type and second row corresponds to **Jaccard distance**. Three conditions: mind-wandering, verbalization and visualization correspond to the three columns respectively. Time mode for all represented pictures is ***peak***.

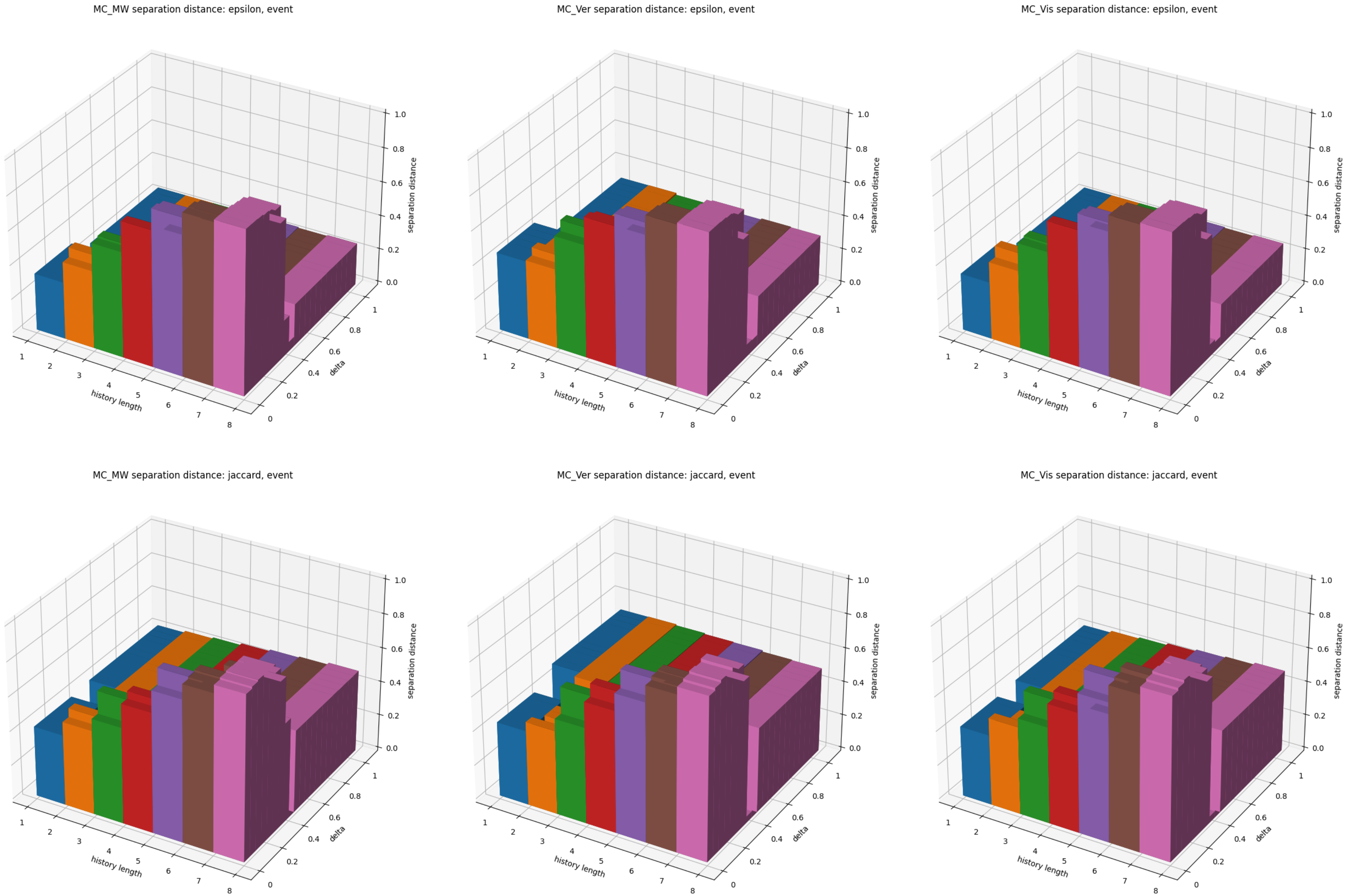

Figure 55: Percentage of separation for the minimized cognitive process condition machines, corresponding to different history lengths (from $L = 1$ to $L = 7$) and minimization deltas (from $\delta = 0$ to $\delta = 1$ with the step size 0.05). First row corresponds to the **epsilon distance** type and second row corresponds to **Jaccard distance**. Three conditions: mind-wandering, verbalization and visualization correspond to the three columns respectively. Time mode for all represented pictures is ***event***.

C Appendix: Shortest Zero Occurrence Words in EEG

C.1 Shortest zero occurrence n-grams

In this section we are aiming to distinguish cognitive processing conditions: mind-wandering (MW), verbalization (Ver) and visualization (Vis) by looking at words of microstates that do not occur within the measured participant sequences. For brevity, we use the acronym *SZO* to stand for *shortest zero-occurrence n*-gram. Note that if z does not occur then any word containing z also does not occur. Therefore, if z is an SZO, then for all (possibly empty) prefixes w_1 and suffixes w_2, $w = w_1 z w_2$ implies w does not occur. Thus to know which words do not occur, it suffices to find the shortest ones that do not occur.

We did the analysis for both *canonical* and *data-driven* microstates and we also distinguish three timing modes: *clock*, *peak* and *event*. Then we distinguish between *individual* and *global* data, where the former contains sets of SZO's for participants individually and the latter stands for intersection of sets of SZO's corresponding to individual participants within a particular timing mode, mind-mode and word length.

In the following sub-sections we both list the full sets of SZO's, but as it is sometimes hard to determine whether there are any differences between sets of non-occurring words, we also calculate set distances using the Jaccard distance metric on sets, which we recall is defined for two sets S_1 and S_2 as

$$d_{\text{Jaccard}}(S_1, S_2) = 1 - \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} \tag{5}$$

where bars denote cardinality (the number of elements in a set).

C.2 Canonical Global Data

In the following sub-section, non-occurring n-grams are listed for canonical microstates where individual participants' data are intersected to provide non-occurring n-grams globally in all matched participants. There are never any globally non-occurring n-grams in peak mode, and the SZOs in event mode are identical for all cognitive processing conditions and are as shown in this list of SZOs:

n=1 (none)

n=2 : AA, BB, CC, DD

n=3, n=4, n=5, n=6, n=7, n=8, n=9: (none)

SZOs List 1: Global MW, Ver and Vis in event mode: Double letters cannot occur by defintion of event mode

SZOs by Cognitive Processing Condition. There are greater differences when we come to clock mode, below are the tables for all of the three cognitive processing conditions separately.

n=1, n=2: (none)

n=3: ABA, ABC, ABD, ACA, ACB, ACD, ADA, ADB, ADC, BAB, BAC, BAD, BCA, BCB, BCD, BDA, BDB, BDC, CAB, CAC, CAD, CBA, CBC, CBD, CDA, CDB, CDC, DAB, DAC, DAD, DBA, DBC, DBD, DCA, DCB, DCD

n=4: ABBA, ABBC, ABBD, ACCA, ACCB, ACCD, ADDA, ADDB, ADDC, BAAB, BAAC, BAAD, BCCA, BCCB, BCCD, BDDA, BDDB, BDDC, CAAB, CAAC, CAAD, CBBA, CBBC, CBBD, CDDA, CDDB, CDDC, DAAB, DAAC, DAAD, DBBA, DBBC, DBBD, DCCA, DCCB, DCCD

n=5: ABBBA, ABBBC, ABBBD, ACCCA, ACCCB, ACCCD, ADDDA, ADDDB, ADDDC, BAAAB, BAAAC, BAAAD, BCCCA, BCCCB, BCCCD, BDDDA, BDDDB, BDDDC, CAAAB, CAAAC, CAAAD, CBBBA, CBBBC, CBBBD, CDDDA, CDDDB, CDDDC, DAAAB, DAAAC, DAAAD, DBBBA, DBBBC, DBBBD, DCCCA, DCCCB, DCCCD n=6: ABBBBA, ABBBBC, ABBBBD, ACCCCB, ACCCCD, ADDDDA, ADDDDB, ADDDDC, BAAAAB, BAAAAC, BAAAAD, BCCCCA, BCCCCB, BCCCCD, BDDDDA, BDDDDB, BDDDDC, CAAAAB, CAAAAC, CAAAAD, CBBBBA, CBBBBC, CBBBBD, CDDDDA, CDDDDB, CDDDDC, DAAAAB, DAAAAC, DAAAAD, DBBBBA, DBBBBC, DBBBBD, DCC-CCA, DCCCCB, DCCCCD

n=7: ABBBBBA, ABBBBBC, ABBBBBD, ACCCCCB, ACCCCCD, ADDDDDB, BAAAAAD, BCCCCCA, BCCCCCB, BCCCCCD, BDDDDDA, BDDDDDC, CAAAAAB, CAAAAAC, CAAAAAD, CBBBBBA, CBBBBBC, CDDDDDA, CDDDDDB, DAAAAAB, DAAAAAC, DBBBBBC, DBBBBBD, DCCCCCB

n=8, n=9: (none)

SZOs List 2: Global Mind-wandering SZOs in clock mode (canonical EEG Micorstates)

n=1, n=2: (none)

n=3: ABA, ABC, ABD, ACA, ACB, ACD, ADA, ADB, ADC, BAB, BAC, BAD, BCA, BCB, BCD, BDA, BDB, BDC, CAB, CAC, CAD, CBA, CBC, CBD, CDA, CDB, CDC, DAB, DAC, DAD, DBA, DBC, DBD, DCA, DCB, DCD

n=4: ABBA, ABBC, ABBD, ACCA, ACCB, ACCD, ADDA, ADDB, ADDC, BAAB, BAAC, BAAD, BCCA, BCCB, BCCD, BDDA, BDDB, BDDC, CAAB, CAAC, CAAD, CBBA, CBBC, CBBD, CDDA, CDDB, CDDC, DAAB, DAAC, DAAD, DBBA, DBBC, DBBD, DCCA, DCCB, DCCD

n=5: ABBBA, ABBBC, ABBBD, ACCCA, ACCCB, ACCCD, ADDDA, ADDDB, ADDDC, BAAAB, BAAAC, BAAAD, BCCCA, BCCCB, BCCCD, BDDDA, BDDDB, BDDDC, CAAAB, CAAAC, CAAAD, CBBBA, CBBBC, CBBBD, CDDDA, CDDDB, CDDDC, DAAAB, DAAAC, DAAAD, DBBBA, DBBBC, DBBBD, DCCCA, DCCCB, DCCCD

n=6: ABBBBA, ABBBBC, ABBBBD, ACCCCA, ACCCCB, ACCCCD, ADDDDA, ADDDDB, ADDDDC, BAAAAB, BAAAAC, BAAAAD, BCCCCA, BCCCCB, BCCCCD, BDDDDA, BDDDDB, BDDDDC, CAAAAB, CAAAAC, CAAAAD, CBBBBC, CBBBBD, CDDDDA, CDDDDB, CDDDDC, DAAAAB, DAAAAC, DAAAAD, DBBBBA, DBBBBC, DBBBBD, DCC-CCA, DCCCCB, DCCCCD

n=7: ABBBBBC, ACCCCCA, ACCCCCD, ADDDDDC, BAAAAAB, BAAAAAC, BAAAAAD, BCCCCCA, BCCCCCD, BDDDDDA, BDDDDDB, BDDDDDC, CAAAAAB, CAAAAAC, CDDDDDA, CDDDDDB, DAAAAAB, DAAAAAC, DBBBBBA, DBBBBBC, DCCCCCA, DCCCCCD

n=8, n=9: (none)

SZOs List 3: Global Verbalization SZOs in clock mode (canonical EEG microstates)

n=1, n=2: (none)

n=3: ABA, ABC, ABD, ACA, ACB, ACD, ADA, ADB, ADC, BAB, BAC, BAD, BCA, BCB, BCD, BDA, BDB, BDC, CAB, CAC, CAD, CBA, CBC, CBD, CDA, CDB, CDC, DAB, DAC, DAD, DBA, DBC, DBD, DCA, DCB, DCD

n=4: ABBA, ABBC, ABBD, ACCA, ACCB, ACCD, ADDA, ADDB, ADDC, BAAB, BAAC, BAAD, BCCA, BCCB, BCCD, BDDA, BDDB, BDDC, CAAB, CAAC, CAAD, CBBA, CBBC, CBBD, CDDA, CDDB, CDDC, DAAB, DAAC, DAAD, DBBA, DBBC, DBBD, DCCA, DCCB, DCCD

n=5: ABBBA, ABBBC, ABBBD, ACCCA, ACCCB, ACCCD, ADDDA, ADDDB, ADDDC, BAAAB, BAAAC, BAAAD, BCCCA, BCCCB, BCCCD, BDDDA, BDDDB, BDDDC, CAAAB, CAAAC, CAAAD, CBBBA, CBBBC, CBBBD, CDDDA, CDDDB, CDDDC, DAAAB, DAAAC, DAAAD, DBBBA, DBBBC, DBBBD, DCCCA, DCCCB, DCCCD

n=6: ABBBBA, ABBBBC, ABBBBD, ACCCCA, ACCCCB, ACCCCD, ADDDDA, ADDDDB, ADDDDC, BAAAAB, BAAAAC, BAAAAD, BCCCCA, BCCCCB, BCCCCD, BDDDDA, BDDDDB, BDDDDC, CAAAAB, CAAAAC, CAAAAD, CBBBBA, CBBBBC, CBBBBD, CDDDDA, CDDDDB, CDDDDC, DAAAAB, DAAAAC, DAAAAD, DBBBBA, DBBBBC, DBBBBD, DCCCCB, DCCCCD

n=7: ABBBBBD, ACCCCCA, ACCCCCB, ACCCCCD, ADDDDDA, ADDDDDC, BAAAAAD, BCCCCCA, BCCCCCD, BDDDDDA, BDDDDDB, BDDDDDC, CAAAAAD, CBBBBBC, CDDDDDA, CDDDDDB, CDDDDDC, DBBBBBC, DBBBBBD, DCCCCCD

n=8, n=9: (none)

SZOs List 4: Global Visualization SZOs in clock mode (canonical EEG microstates)

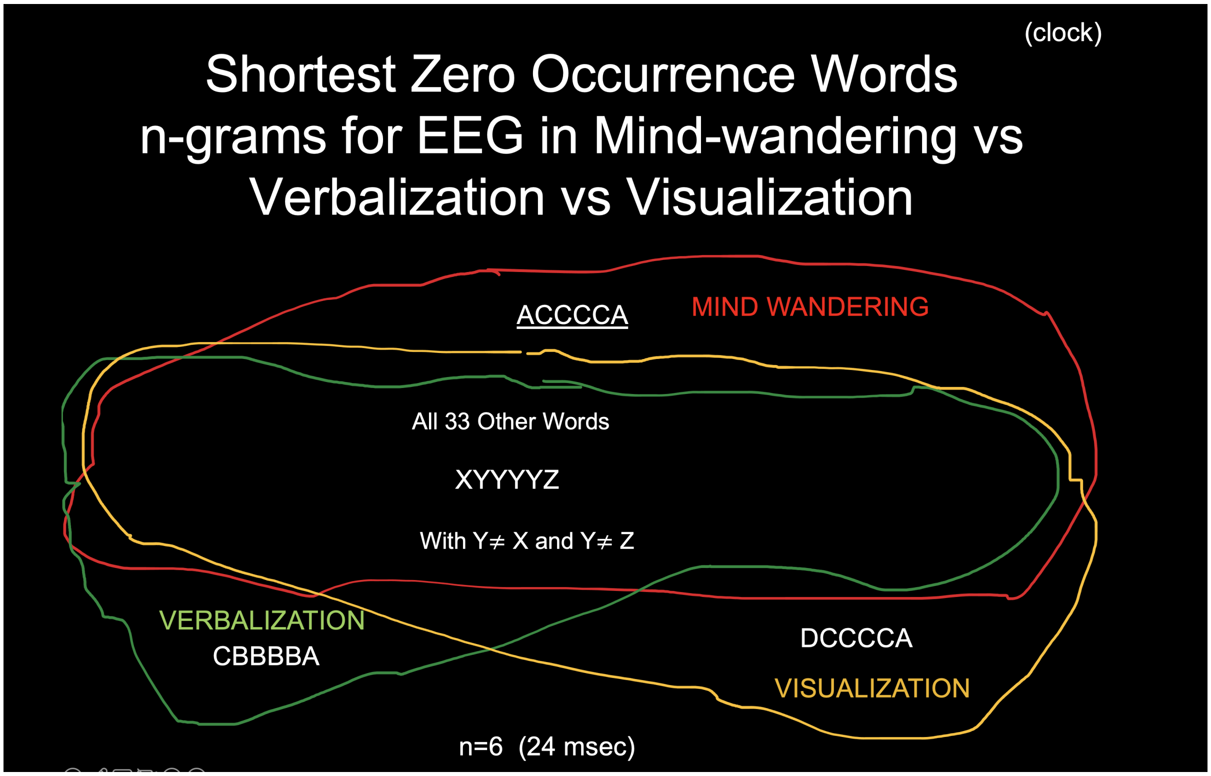

Figure 56: For length $n = 6$ n-grams in clock mode EEG microstate sequences of duration 24 ms, all there are 33 shortest zero occurrence words of length six that ooccur in none of the cognition process conditions MW, Ver and Vis in the eyes-closed EEG dataset with the four canonical microstates. The words AC^4A does not occur in MW, CB^4A does not occur in Ver and DC^4A does not occur in Vis, but each of these three occurs in the other two modes.

No 1- and 2-grams were shortest zero occurrence words in clock mode. This just means that all microstates occur in each cognitive condition and all possible transitions occur and all possible microstates last at least 8 ms in each condition (of course we know they last longer). For n-grams of lengths 3, 4 and 5 in clock mode (12-20 ms), no words of the form $XY^{n-2}Z$ occur for X, Y, Z EEG microstates from the canonical set with $X \neq Y \neq Z$ as these would corresponding to microstate Y occurring with duration of 16 ms or less. For $n = 6$, only 3 n-grams occur in two cognitive processing conditions among mind-wandering, verbalization and visualization, but not the other mode as shown Venn diagram in Figure 56, These correspond to 16 ms microstates C and A flanked by certain non-identical letters. In principle occurrence of two of these words would could only occur in the reminaing cognitive mode. However, these are group level results and the microstates are so short that it is possible occurrences of such words are due to noise and chance rather than characterizing certain cognitive modes.

With 7-grams, all SZOs are microstates of this type corresponding to a microstate of duration 20 ms. Two such words occurred for all three cognitive modes among the participants, while 9 occurred for no condition at all. None of these 9 are palindromes, i.e., short microstates Y embedding betwenn two microstates of the same type X. The remaining 25 such words appear in the either one or two SZO set for the three modes. Therefor no

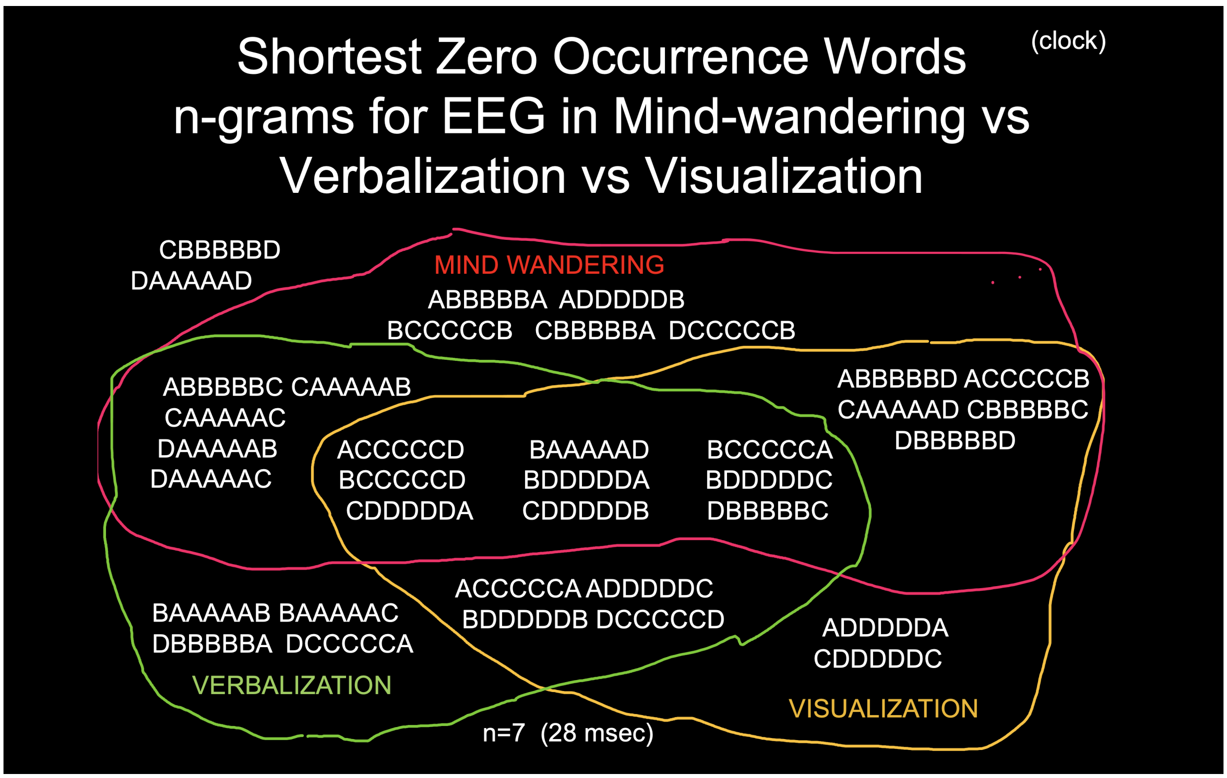

Figure 57: For length $n = 7$, n-grams in clock mode EEG microstate sequences, where duration is 28 ms, there are 9 shortest zero occurrence (SZOs) words that occur in none of the cognition process conditions (MW, Ver and Vis) in the eyes-closed EEG dataset with four canonical microstates. The Venn diagram shows for which cognitive modes words of the form XY^5Z are SZOs. All 34 SZOs are of this form, but two such words CD^5D and DA^5D do occur for all conditions.

SZOs of length 8 to 10 at the global level for canonical microstates. This indicates the occurrence of all four microstates with durations 24-32 ms flanked by all possible other microstates in all combinations in all cognitive conditions. That is, at least with EEG processing done using KeyPy global field power (GFP peaks) can be so close together that the occurrence of such short microstates in all cognitive modes in the dataset is observed.

Results for data-driven n-grams shortewt zero occurrence words were similar although the words occurring were not exactly the same (Section C.6).

Distances (SZOs sets, Clock, Global - Canoncial EEG Microstates). Jaccard distances between sets of shortest zero occurrence (SZOs) words for each cognitive mode are zero for words of each length n from 1 to 9, except for $n = 6$ and $n = 7$. Tables of these distances are presented here together with the area of spanned by the SZO sets using the distances as sides of a triangle.

Area = 0.001	**MW**	**Ver**	**Vis**
MW	0	0.056	0.056
Ver	0.056	0	0.056
Vis	0.056	0.056	0

Table 62: SZO set distances in clock mode for length 6 canonical EEG microstate sequences

Area = 0.130	**MW**	**Ver**	**Vis**
MW	0	0.562	0.534
Ver	0.563	0	0.552
Vis	0.533	0.551	0

Table 63: SZO set distances in clock mode for length 7 canonical EEG microstate sequences

Area in these tables is the area of the triangle computed using the Jaccard distances between the three SZO sets and is non-zero exactly when the SZO sets discriminate between the three modes. The maximum discrimination would be achieved when all distances are 1 between different modes (no overlap between non-empty SZO sets) and the area is $\frac{\sqrt{3}}{4} = \sim 0.433$ [7]

Observations. All possible words occur in peak mode, when looked at globally and except for AA, BB, CC, and DD in length 2 all other words occur in event mode. It is only in length 3 to 7 in clock mode where we can find other words that never globally occur.

All of them are of the structure $XY^{L-2}(X|Z)$, where X, Y and Z can be any microstates such that $X \neq Y \neq Z$ and L is the word length. That is, these correspond to very short duration single EEG microstates in clock mode that are not possible. This does not however apply to all short words of this type, as there is $36 = 3 \times 4 \times 3$ possible words of this type, where the in-between microstate is

[7] Area is computed according the Heron's formula from the lengths a, b, c of the three sides of the triangle in this space as $\sqrt{s(s-a)(s-b)(s-c)}$, where $s = \frac{a+b+c}{2}$. Although this assumes the Jaccard distances embed in Euclidean space, one can use the computed area as an index of discrimination in any case.

different from the first and from the last, while the first and last can be the same. But not all of the SZO sets contain all 36 words, more specifically the sets for lengths 6 and 7 contain only 35 words as we will see below.

In lengths 3 to 5, the sets of words that never occur are the same, containing impossibly short microstates for 36 possible words.

In length 6 the distances between the sets are only approximately 0.05, due to only threee words not shared by the SZO sets for MW, Ver and Vis, as each set contains all 33 possible words of the form XY^4Z (with Y different from X and Z) plus one more word not contained by the other two SZO sets:

- the SZO set for MW contains *ACCCCA*,
- the SZO set for Ver contains *CBBBBA*,
- the SZO set for Vis contains *DCCCCA*.

In other words, using canonical EEG microstates, during MW, *ACCCCA* does not occur while it can in the other modes. During Ver, *CBBBBA* does not occur but can in the other two modes; while in Vis, *DCCCCA* does not occur but can in the other modes. This could also be interpreted as follows: short duration of EEG microstate C (24 msec) is not possible in MW if A occurs both pre and post C, and similarly for other cognitive processing conditions, short duration B is not possible in Ver if preceded by C and followed by A, while in Vis short duration of C cannot occur if both preceded by D and followed by A. Note these conclusions are limited to the present data-set and may not hold for other EEG microstate clustering methods.

In length 7, corresponding to 28 msec, the sets of short zero-occurrence words contain only 24, 22 and 20 elements respectively of the 36 possible words XY^5Z with Y different from X and Z for MW, Ver and Vis, accounting for 34 distinct works, and naturally the distances are much greater: about 0.5. The remaining two words of this length, CB^5D and DA^5D occur in all cognitive modes. The difference (what is in the first set, but not the second one, or what occurs in second mind-mode, but does not in first one) between MW and Ver is

$ABBBBBD, DCCCCCB, ABBBBBA, BCCCCCB, CBBBBBA, ADDDDDB,$
$DBBBBBD, CAAAAAD, ACCCCCB, CBBBBBC$

while between Ver and MW

$DCCCCCD, BDDDDDB, DCCCCCA, BAAAAAB, BAAAAAC, ADDDDDC,$
$DBBBBBA, ACCCCCA.$

Between MW and Vis

$ADDDDDB, DCCCCCB, ABBBBBA, BCCCCCB, ABBBBBC, CBBBBBA,$
$CAAAAAC, CAAAAAB, DAAAAAB, DAAAAAC$

and Vis and MW:

$DCCCCCD, BDDDDDB, CDDDDDC, ADDDDDC, ADDDDDA, ACCCCCA.$

Finally Ver and Vis

$ABBBBBC, DCCCCCA, BAAAAAB, BAAAAAC, CAAAAAC, CAAAAAB,$
$DBBBBBA, DAAAAAB, DAAAAAC$

and Vis and Ver

ABBBBBD, *CDDDDDC*, *CBBBBBC*, *DBBBBBD*, *ADDDDDA*, *CAAAAAD*, *ACCCCCB*.

Intersection of SZO sets: Nine 7-grams never occur in any mode:
ACCCCCD BAAAAAD BCCCCCA BCCCCCD BDDDDDA BDDDDDC CDDDDDA CDDDDDB DBBBBBC

The following 7-grams never occurring in one of three modes only:
SZOs for Vis only : ADDDDDA CDDDDDC
SZOs for Ver only: BAAAAAB BAAAAAC DBBBBBA DCCCCCA
SZOs for MW only: ABBBBBA ADDDDDB BCCCCCB CBBBBBA DCCCCCB

C.3 Canonical Individual Clock Data

In this section we look at individual canonical data in clock mode. For individual participants there are many more entries, therefore we will illustrate the results for participant 1 and summarize the results for others.

SZOs for Individual Clock Data (canonical).

n=2 : (none)

n=3 : ABA, ABC, ABD, ACA, ACB, ACD, ADA, ADB, ADC, BAB, BAC, BAD, BCA, BCB, BCD, BDA, BDB, BDC, CAB, CAC, CAD, CBA, CBC, CBD, CDA, CDB, CDC, DAB, DAC, DAD, DBA, DBC, DBD, DCA, DCB, DCD

n=4 : ABBA, ABBC, ABBD, ACCA, ACCB, ACCD, ADDA, ADDB, ADDC, BAAB, BAAC, BAAD, BCCA, BCCB, BCCD, BDDA, BDDB, BDDC, CAAB, CAAC, CAAD, CBBA, CBBC, CBBD, CDDA, CDDB, CDDC, DAAB, DAAC, DAAD, DBBA, DBBC, DBBD, DCCA, DCCB, DCCD

n=5 : ABBBA, ABBBC, ABBBD, ACCCA, ACCCB, ACCCD, ADDDA, ADDDB, ADDDC, BAAAB, BAAAC, BAAAD, BCCCA, BCCCB, BCCCD, BDDDA, BDDDB, BDDDC, CAAAB, CAAAC, CAAAD, CBBBA, CBBBC, CBBBD, CDDDA, CDDDB, CDDDC, DAAAB, DAAAC, DAAAD, DBBBA, DBBBC, DBBBD, DCCCA, DCCCB, DCCCD

n=6 : ABBBBA, ABBBBC, ABBBBD, ACCCCA, ACCCCB, ACCCCD, ADDDDA, ADDDDB, ADDDDC, BAAAAB, BAAAAC, BAAAAD, BCCCCA, BCCCCB, BCCCCD, BDDDDA, BDDDDB, BDDDDC, CAAAAB, CAAAAC, CAAAAD, CBBBBA, CBBBBC, CBBBBD, CDDDDA, CDDDDB, CDDDDC, DAAAAB, DAAAAC, DAAAAD, DBBBBA, DBBBBC, DBBBBD, DCCCCA, DCCCCB, DCCCCD

n=7 :ABBBBBA, ABBBBBC, ABBBBBD, ACCCCCA, ACCCCCB, ACCCCCD, ADDDDDA, ADDDDDB, ADDDDDC, BAAAAAB, BAAAAAC, BAAAAAD, BCCCCCA, BCCCCCB, BCCCCCD, BDDDDDA, BDDDDDB, BDDDDDC, CAAAAAB, CAAAAAC, CAAAAAD, CBBBBBA, CBBBBBC, CBBBBBD, CDDDDDA, CDDDDDB, CDDDDDC, DAAAAAB, DAAAAAC, DAAAAAD, DBBBBBA, DBBBBBC, DBBBBBD, DCCCCCA, DCCCCCB, DCCCCCD

n=8 : ABBBBBBA, ABBBBBBC, ABBBBBBD, ACCCCCCA, ACCCCCCB, ADDDDDDA, ADDDDDDB, ADDDDDDC, BAAAAAAB, BAAAAAAC, BCCCCCCA, BCCCCCCD, BDDDDDDA, BDDDDDDB, CAAAAAAB, CAAAAAAC, CBBBBBBA, CBBBBBBC, CBBBBBBD, CDDDDDDC, DAAAAAAC, DBBBBBBC, DBBBBBBD, DCCCCCCA, DCCCCCCB

n=9 : ABBBBBBBA, ACCCCCCCA, BCCCCCCCB, CAAAAAAAB, CBBBBBBBA, CDDDDDDDA, DAAAAAAAB

n=10 : BAAAAAAAAB, BDDDDDDDDB

SZOs List 5: Individual MW in clock mode for participant 1

n=2 : (none)

n=3 : ABA, ABC, ABD, ACA, ACB, ACD, ADA, ADB, ADC, BAB, BAC, BAD, BCA, BCB, BCD, BDA, BDB, BDC, CAB, CAC, CAD, CBA, CBC, CBD, CDA, CDB, CDC, DAB, DAC, DAD, DBA, DBC, DBD, DCA, DCB, DCD

n=4 : ABBA, ABBC, ABBD, ACCA, ACCB, ACCD, ADDA, ADDB, ADDC, BAAB, BAAC, BAAD, BCCA, BCCB, BCCD, BDDA, BDDB, BDDC, CAAB, CAAC, CAAD, CBBA, CBBC, CBBD, CDDA, CDDB, CDDC, DAAB, DAAC, DAAD, DBBA, DBBC, DBBD, DCCA, DCCB, DCCD

n=5 : ABBBA, ABBBC, ABBBD, ACCCA, ACCCB, ACCCD, ADDDA, ADDDB, ADDDC, BAAAB, BAAAC, BAAAD, BCCCA, BCCCB, BCCCD, BDDDA, BDDDB, BDDDC, CAAAB, CAAAC, CAAAD, CBBBA, CBBBC, CBBBD, CDDDA, CDDDB, CDDDC, DAAAB, DAAAC, DAAAD, DBBBA, DBBBC, DBBBD, DCCCA, DCCCB, DCCCD

n=6 : ABBBBA, ABBBBC, ABBBBD, ACCCCA, ACCCCB, ACCCCD, ADDDDA, ADDDDB, ADDDDC, BAAAAB, BAAAAC, BAAAAD, BCCCCA, BCCCCB, BCCCCD, BDDDDA, BDDDDB, BDDDDC, CAAAAB, CAAAAC, CAAAAD, CBBBBA, CBBBBC, CBBBBD, CDDDDA, CDDDDB, CDDDDC, DAAAAB, DAAAAC, DAAAAD, DBBBBA, DBBBBC, DBBBBD, DCCCCA, DCCCCB, DCCCCD

n=7 :ABBBBBA, ABBBBBC, ABBBBBD, ACCCCCA, ACCCCCB, ACCCCCD, ADDDDDA, ADDDDDB, ADDDDDC, BAAAAAB, BAAAAAC, BAAAAAD, BCCCCCA, BCCCCCB, BCCCCCD, BDDDDDA, BDDDDDB, BDDDDDC, CAAAAAB, CAAAAAC, CAAAAAD, CBBBBBA, CBBBBBC, CBBBBBD, CDDDDDA, CDDDDDB, CDDDDDC, DAAAAAB, DAAAAAC, DAAAAAD, DBBBBBA, DBBBBBC, DBBBBBD, DCCCCCA, DCCCCCB, DCCCCCD

n=8 : ABBBBBBA, ABBBBBBC, ABBBBBBD, ACCCCCCA, ACCCCCCD, ADDDDDDB, ADDDDDDC, BAAAAAAB, BAAAAAAC, BAAAAAAD, BCCCCCCD, BDDDDDDA, BDDDDDDB, BDDDDDDC, CAAAAAAB, CAAAAAAC, CAAAAAAD, CBBBBBBA, CBBBBBBC, CBBBBBBD, CDDDDDDA, DAAAAAAB, DAAAAAAC, DAAAAAAD, DBBBBBBA, DBBBBBBC, DCCCCCCA, DCCCCCCB

n=9 : ACCCCCCCB, BAAAAAAAC, BCCCCCCCB, BDDDDDDDB, CAAAAAAAB, CBBBBBBBD, DAAAAAAAB

n=10 : BAAAAAAAAB, BDDDDDDDDB

SZOs List 6: Individual Ver in clock mode for participant 1

n=2 : (none)

n=3 : ABA, ABC, ABD, ACA, ACB, ACD, ADA, ADB, ADC, BAB, BAC, BAD, BCA, BCB, BCD, BDA, BDB, BDC, CAB, CAC, CAD, CBA, CBC, CBD, CDA, CDB, CDC, DAB, DAC, DAD, DBA, DBC, DBD, DCA, DCB, DCD

n=4 : ABBA, ABBC, ABBD, ACCA, ACCB, ACCD, ADDA, ADDB, ADDC, BAAB, BAAC, BAAD, BCCA, BCCB, BCCD, BDDA, BDDB, BDDC, CAAB, CAAC, CAAD, CBBA, CBBC, CBBD, CDDA, CDDB, CDDC, DAAB, DAAC, DAAD, DBBA, DBBC, DBBD, DCCA, DCCB, DCCD

n=5 : ABBBA, ABBBC, ABBBD, ACCCA, ACCCB, ACCCD, ADDDA, ADDDB, ADDDC, BAAAB, BAAAC, BAAAD, BCCCA, BCCCB, BCCCD, BDDDA, BDDDB, BDDDC, CAAAB, CAAAC, CAAAD, CBBBA, CBBBC, CBBBD, CDDDA, CDDDB, CDDDC, DAAAB, DAAAC, DAAAD, DBBBA, DBBBC, DBBBD, DCCCA, DCCCB, DCCCD

n=6 : ABBBBA, ABBBBC, ABBBBD, ACCCCA, ACCCCB, ACCCCD, ADDDDA, ADDDDB, ADDDDC, BAAAAB, BAAAAC, BAAAAD, BCCCCA, BCCCCB, BCCCCD, BDDDDA, BDDDDB, BDDDDC, CAAAAB, CAAAAC, CAAAAD, CBBBBA, CBBBBC, CBBBBD, CDDDDA, CDDDDB, CDDDDC, DAAAAB, DAAAAC, DAAAAD, DBBBBA, DBBBBC, DBBBBD, DCCCCA, DCCCCB, DCCCCD

n=7 :ABBBBBA, ABBBBBC, ABBBBBD, ACCCCCA, ACCCCCB, ACCCCCD, ADDDDDA, ADDDDDB, ADDDDDC, BAAAAAB, BAAAAAC, BAAAAAD, BCCCCCA, BCCCCCB, BCCCCCD, BDDDDDA, BDDDDDB, BDDDDDC, CAAAAAB, CAAAAAC, CAAAAAD, CBBBBBA, CBBBBBC, CBBBBBD, CDDDDDA, CDDDDDB, CDDDDDC, DAAAAAC, DAAAAAD, DBBBBBA, DBBBBBC, DBBBBBD, DCCCCCA, DCCCCCB, DCCCCCD

n=8 : ABBBBBBA, ABBBBBBD, ACCCCCCA, ACCCCCCB, ACCCCCCD, ADDDDDDA, ADDDDDDB, ADDDDDDC, BAAAAAAB, BAAAAAAC, BAAAAAAD, BCCCCCCB, BCCCCCCD, BDDDDDDB, BDDDDDDC, CAAAAAAC, CAAAAAAD, CBBBBBBA, CBBBBBBC, CBBBBBBD, CDDDDDDA, CDDDDDDB, CDDDDDDC, DAAAAAAB, DAAAAAAC, DAAAAAAD, DBBBBBBC, DCCCCCCB, DCCCCCCD

n=9 : ABBBBBBBA, ABBBBBBBC, ADDDDDDDB, BAAAAAAAB, BAAAAAAAC, BDDDDDDDC, DAAAAAAAC, DBBBBBBBC, DCCCCCCCA, DCCCCCCCB n=10 : (none)

SZOs List 7: Individual Vis in clock mode for participant 1

Distances (Canoncial EEG Microstates). Jaccard distances between sets of shortest zero occurrence (SZOs) words for each cognitive mode are zero for words of each length n from 1 to 9, except for $n = 6$ and $n = 7$. Tables of these distances are presented here together with the area of spanned by the SZO sets using the distances as sides of a triangle.

Area = 0	**MW**	**Ver**	**Vis**
MW	0	0	0.028
Ver	0	0	0.028
Vis	0.028	0.028	0

Table 64: SZO set distances for participant 1 in clock mode for L7

Area = 0.062	**MW**	**Ver**	**Vis**
MW	0	0.394	0.457
Ver	0.394	0	0.324
Vis	0.457	0.324	0

Table 65: SZO set distances for participant 1 in clock mode for L8

Area = 0.314	**MW**	**Ver**	**Vis**
MW	0	0.728	0.938
Ver	0.728	0	0.938
Vis	0.938	0.938	0

Table 66: SZO set distances for participant 1 in clock mode for L9

Area = 0	**MW**	**Ver**	**Vis**
MW	0	0	1.000
Ver	0	0	1.000
Vis	1.000	1.000	0

Table 67: SZO set distances for Participant 1 in clock mode for L10

Observations. The non-occurring words have the same structure as when looked at clock mode globally. Compared to global data, the distances can be greater and are also measured for lengths 8, 9 and 10, where the global intersections are empty. Especially at lengths 9 and 10, the distinctions between mind modes are very strong with very different sets for each mind-mode.

C.4 Canonical Individual Peak Data

SZOs for Indvidual Peak Data for a Single Participant (canonical).

n=2, n=3, n=4 : (none)

n=5 : AABAC, ADABB, ADBBA, BAADB, BACDB, BADAB, BDADB, BDCAB, BDCDA, CBDCB, CDBAC

n=6 : AADADB, AADCDA, AADDAB, ABDDAB, ACADDB, ACCADA, ADAADB, ADBCCA, ADDACB, ADDCBB, ADDCCB, BADDAC, BADDDA, BCDDBB, BCDDBC, BCDDCA, BDACDC, BDDABB, BDDBDB, CBADDC, CCACDA, CCBDDA, CDCDCB, CDDDBB

n=7 : ADDBDDA, ADDCCDA, ADDDBDC, ADDDDBB, BAADDDA, BADADDD, BCADDDC, BCDDBDB, BCDDCCA, BDDADDB, BDDCCDA, BDDDDAB, CAADDDC, CCDCCDB, CDDDACD, CDDDCCB, CDDDDBB, CDDDDCB

n=8 : ADDADDDB, ADDDBDDB, ADDDDADB, ADDDDDCB, BDDDDDAC, BDDDDDCA, BDDDDDDB, CDCDDDDA, CDDADDDA, CDDADDDB, CDDDCDCC, CDDDDCDA

n=9 : ACDDDDDDC, ADDDDDDAB, BDDDDDCDA, BDDDDDCDC, CCDDDDDDB, CDADDDDDB, CDDDDADDB, CDDDDDCDC, CDDDDDDAB, CDDDDDDBB

n=10 : ADDDDDDDAC, BDDDDDDDDA, CCDDDDDDDB, CDDDDADDDA, CDDDDADDDC, CDDDDDDDBA, CDDDDDDDBB

SZOs List 8: Individual MW in peak mode for participant 1

n=2, n=3 : (none)

n=4 : BAAB

n=5 : AACAA, AADBC, ADBDB, ADCBA, ADDCB, BBCCA, BCBDA, BDCDB

n=6 : AADDDB, ABDADC, ABDDDB, ACCDCA, BBDDDB, BBDDDC, BCBDDB, BDCCDA, BDCDCA, BDDACA, BDDACB, BDDADB, BDDCCB, BDDDBA, CABDDC, CADCDA, DDCCBA

n=7 : ABDDDDC, ACCCCDA, ACCCDCD, ADACCCC, ADDCDDB, ADDDDCB, BCDDDDB, BDCCDDA, BDCCDDB, BDDACDC, BDDADDB, BDDCDDB, BDDDDCB, CBCDDDC, CCADDDC, CDDADDB, CDDBDDA, CDDDCCB, DDDBDAC

n=8 : ACDDCDDC, ADADDDDA, ADDDCDDB, ADDDDCDA, BDADDDDA, BDADDDDC, BDDDDCDC, BDDDDDDA, CCCCCCDC, CDDCDCCC, CDDDCCDA, CDDDDBDA, CDDDDCCA

n=9 : AADDDDDDA, ABDDDDDDC, ADCDDDDDC, ADDDDCDDB, BDDDDDDCA, CDCCDDDDA, CDDCDDDDA

n=10 : ADDCCDDDDA, BCDDDDDDDC, CDDDCDDDDA, CDDDDCDDCC, CDDDDDDDCA

SZOs List 9: Individual Ver in peak mode for participant 1

n=2, n=3, n=4 : (none)

n=5 : ABCDB, ABDAA, ADBAB, ADBCA, BACBC, BBAAB, BCBBA, CADBB, CBCCA, CDBBA

n=6 : ABCCDC, ACCCAB, ACDCCA, ACDDDB, ADCCDB, ADCDCB, ADDBDA, BADBAC, BCDCDA, BCDDAB, BCDDAC, BCDDDA, BDBCCA, BDBDDA, BDCDDB, BDDCCC, BDDCDA, BDDDBC, CADCDC, CBDADD, CBDDCC, CBDDDA, CDADAA, CDDBBA, DBBDBB, DCACDA, DCBADA, DDCACD

n=7 : ACCDDDB, ADCDDDA, ADDADDC, BAADDDC, BCCCCCB, BCDDDCB, BCDDDDA, BDAADDC, BDDCDDA, BDDDDAB, BDDDDCA, CCDDDCB

n=8 : ACDDCDDA, ACDDDDDA, ACDDDDDC, ADDDADDC, BDDDDCCA, CDBDDDDC, CDDDCDDA, CDDDDDAB

n=9 : BCDDDDDDC, CADDDDDDB

n=10 : BDDDDDDDDA, CDADDDDDDB, CDDDDDDCCA, CDDDDDDCDC

SZOs List 10: Individual Vis in peak mode for participant 1

Distances (Individual SZO sets, Peak Mode - canonical).

Area = 0	**MW**	**Ver**	**Vis**
MW	0	1.0	0
Ver	1.0	0	1.0
Vis	0	1.0	0

Table 68: SZO set distances for Participant 1 in peak mode for L4

Area = 0.433	**MW**	**Ver**	**Vis**
MW	0	1.0	1.0
Ver	1.0	0	1.0
Vis	1.0	1.0	0

Table 69: SZO set distances for Participant 1 in peak mode for L5

Area = 0.433	**MW**	**Ver**	**Vis**
MW	0	1.0	1.0
Ver	1.0	0	1.0
Vis	1.0	1.0	0

Table 70: SZO set distances for Participant 1 in peak mode for L6

Area = 0.406	**MW**	**Ver**	**Vis**
MW	0	0.943	0.966
Ver	0.943	0	1.0
Vis	0.966	1.0	0

Table 71: SZO set distances for participant 1 in peak mode for L7

Area = 0.433	**MW**	**Ver**	**Vis**
MW	0	1.0	1.0
Ver	1.0	0	1.0
Vis	1.0	1.0	0

Table 72: SZO set distances for participant 1 in peak mode for L8

Area = 0.433	MW	Ver	Vis
MW	0	1.0	1.0
Ver	1.0	0	1.0
Vis	1.0	1.0	0

Table 73: SZO set distances for Participant 1 in peak mode for L9

Area = 0.401	MW	Ver	Vis
MW	0	1.0	0.9
Ver	1.0	0	1.0
Vis	0.9	1.0	0

Table 74: SZO set distances for participant 1 in peak mode for L10

Observations. The structure of the words is much more varied and the distances between the sets much greater compared to clock mode. This explains why the intersections for all individual participants to create global data are empty.

C.5 Canonical Individual Event Data

SZOs in Individual Event Data (Examples from a Single Participant).

n=2 : AA, BB, CC, DD
n=3, n=4, n=5 : (none)
n=6 : ABCDCB, CBDBAB
n=7 : ABADADC, ABCDCDA, BADCADB, BADCDAB, BCDBDAC, BCDCBDA, BDCDACB, CADCBCB
n=8 : ACDBDCDB, ADBDCDCB, ADCDCADC, BADADCDB, BADCDBDC, BCADADAB, CDACDCDB
n=9 : ACDCDCDCB, DACDCDADA
n=10 : (none)

SZOs List 11: Individual MW in event mode for participant 1

n=2 : AA, BB, CC, DD
n=3, n=4, n=5 : (none)
n=5 : BDCBA, CBADB
n=6 : BADBDB, BCDACA, BCDBDB, BCDCBA, BDBCDB
n=7 : ACACDAB, BACDADB, BCDCDAB, BCDCDCB, BDADACA, CADADAB
n=8 : ACDCDADB, ADACDCDB, ADCDCDAB, BCACDCDB, BDACDCDA, BDADADAB, BDCDACDB, DACDCDCB
n=9 : ACDCDADCA, BDACDCDCA
n=10 : (none)

SZOs List 12: Individual Ver in event mode for participant 1

n=2 : AA, BB, CC, DD
n=3, n=4, n=5 : (none)

n=6 : ABDBAC, ADCABA, BACBDC, BCDACB, BCDBAB, CADCBA, DBCBDA

n=7 : ADBCDCB, BCDADBA, BDCADCB, BDCDCBA, CACDCAB, CDACBDA

n=8 : BCDCDADA

n=9, n=10 : (none)

SZOs List 13: Individual Vis in event mode for participant 1

Distances (Individual SZO sets, Event Mode - canonical.

Area = 0	**MW**	**Ver**	**Vis**
MW	0	1.0	0
Ver	1.0	0	1.0
Vis	0	1.0	0

Table 75: SZO set distances for participant 1 in event mode for L5

Area = 0.433	**MW**	**Ver**	**Vis**
MW	0	1.0	1.0
Ver	1.0	0	1.0
Vis	1.0	1.0	0

Table 76: SZO set distances for Participant 1 in event mode for L6, L7, L8, L9

C.6 Data-driven Global Data and Shortest Zero Occurrence Words

For the data-driven EEG microstate sequences, a largely similar situation holds for n-grams that do not occur. **SZOs for data-driven global data.** For peak and event mode at the global level, the situation with data-driven EEG microstates is exactly the same as for the canonical microstates. There are non-occurring words in peak mode at all and for event mode the only shortest zero-occurrence words are *AA*, *BB*, *CC*, *DD* which cannot occur by definition of event mode. That is, each possible sequence of length up to 10 occurs for at least some participant.

For clock mode, there are no SZOs of length 1 or 2, but for $3 \leq n \leq 5$ the words of the form $XY^{n-2}Z$ with $X \neq Y$ and $Y \neq Z$ are all SZOs. Since the microstate Y is flanked in this words by a microstate distinct from Y on each side, this can be interpreted as there is no EEG microstate of duration than $4(n-2)$ milliseconds, i.e., no microstate lasting than 20 ms or less. This is exactly the same as we saw for the canonical EEG microstate sequences.

For $n = 8$ and $n = 9$ there are no SZOs in any condition in clock mode. This implies $XY^{n-2}Z$ do occur for each of mindwandering, visualization and verbalization, i.e., microstate Y occurs with durations 24 ms and 28 ms preceded and followed by any possible microstates X and Z, respectively,

For the lengths $n = 6$ and $n = 7$, however, just as for the canonical microstates, the data-driven microstates include SZOs, words of the form $XY^{n-2}Z$ that do not occur and these are different for different cognitive processing conditions:

n=6: ABBBBA, ABBBBC, ABBBBD, ACCCCB, ACCCCD, ADDDDA, ADDDDB, BAAAAB, BAAAAC, BAAAAD, BCC-CCA, BCCCCB, BDDDDA, BDDDDB, BDDDDC,
CAAAAB, CAAAAC, CAAAAD, CBBBBA, CBBBBC, CBBBBD, CDDDDA, CDDDDB, CDDDDC, DAAAAB, DAAAAC, DAAAAD, DBBBBA, DBBBBC, DBBBBD, DCCCCA, DCCCCB, DCCCCD

n=7: ABBBBBA, ABBBBBC, ABBBBBD, ACCCCCB, ACCCCCD, ADDDDDB, BAAAAAB, BAAAAAD, BCCCCCA, BCCCCCD, BDDDDDA, BDDDDDC, CAAAAAB, CAAAAAC, CAAAAAD, CBBBBBA, CBBBBBD, CDDDDDA, CDDDDDC, DAAAAAB, DAAAAAC, DBBBBBA, DCCCCCA, DCCCCCB

SZOs List 14: Global Shortest Zero-Occurrence words in MW in clock mode (data-driven)

SZOs List (Data-driven) : MW in clock mode for all participants.

n=6: ABBBBA, ABBBBC, ABBBBD, ACCCCA, ACCCCB, ACCCCD, ADDDDA, ADDDDB, ADDDDC, BAAAAB, BAAAAC, BAAAAD, BCCCCA, BCCCCB, BCCCCD, BDDDDA, BDDDDB, BDDDDC, CAAAAB, CAAAAC, CAAAAD, CBBBBA, CBBBBC, CBBBBD, CDDDDA, CDDDDB, CDDDDC, DAAAAB, DAAAAC, DAAAAD, DBBBBA, DBBBBC, DBBBBD, DCCCCA, DCCCCB, DCCCCD

n=7: ABBBBBA, ABBBBBC, ABBBBBD, ACCCCCA, ACCCCCB, ADDDDDA, ADDDDDB, BAAAAAC, BCCCCCD, BDDDDDC, CAAAAAC, CAAAAAD, CDDDDDB, DAAAAAB, DAAAAAC, DAAAAAD, DBBBBBA, DCCCCCA, DCCCCCD

SZOs List 15: Global Shortest Zero-Occurrence words in Ver in clock mode (data-driven)

SZOs List (Data-driven) : Verbalization in clock mode for all participants.

n=6: ABBBBA, ABBBBC, ABBBBD, ACCCCB, ACCCCD, ADDDDA, ADDDDB, BAAAAB, BAAAAC, BAAAAD, BCCCCA, BCCCCB, BDDDDA, BDDDDB, BDDDDC, CAAAAB, CAAAAC, CAAAAD, CBBBBA, CBBBBC, CBBBBD, CD-DDDA, CDDDDB, CDDDDC, DAAAAB, DAAAAC, DAAAAD, DBBBBA, DBBBBC, DBBBBD, DCCCCA, DCCCCB, DCCCCD

n=7: ABBBBBA, ABBBBBC, ABBBBBD, ACCCCCB, ACCCCCD, ADDDDDB, BAAAAAB, BAAAAAD, BCCCCCA, BCCCCCD, BDDDDDA, BDDDDDC, CAAAAAB, CAAAAAC, CAAAAAD, CBBBBBA, CBBBBBD, CDDDDDA, CDDDDDC, DAAAAAB, DAAAAAC, DBBBBBA, DCCCCCA, DCCCCCB

SZOs List 16: Global Shortest Zero-Occurrence words in Vis in clock mode (data-driven)

SZOs List (Data-driven) : Visualization in clock mode for all participants.

Note the for $n = 6$ all 36 words of the form $XY^{n-2}Z$ do not occur in the verbalization condition. The SZO sets for MW and Vis share 32 such words, but the SZO set for Vis has also DAAAAD, while the SZO set for MW has also ACCCCA, ADDDDC, and BCCCCD.

For $n = 7$, the SZO sets for MW, Vis and Ver have 24, 13, and 19 words, respectively, for a total of 33 distinct words. The remaining three words of this length, BC^5B, DB^5C, and DB^5D occur in all cognitive processing conditions.

Comparison of Canonical and Data-Driven SZO sets. Note the words occurring in cognitive modes in data-driven sequences are different from the ones canonical microstates.

In the data-driven case, EEG microstate topographies bearing the labels A, B, C, and D were generated independently for different subjects and condition, so are do not represent fixed topolographies.[8]

The 9 words never occurring in any cognitive mode for canonical sequences and the two words never occurring for data-driven sequences (AB^5D and BD^5C) have only BD^5C in common. We remark that none

[8]It wold be possible to use populatioin-level clustering to obtain fixed EEG microstate topographies, but that method was not employed here.

these words are palindromes. One might speculate that palindromes are more likely to occur than non-palindromes, e.g., if noise creates appears as a GFP peak during the course of a microstate X and it is classified as another microstate Y this would result in a n-gram of the form XY^kX, a palindrome. Thus one might expect it to be easier for a non-palindrome XY^Z with $X \neq Z$ to be a shortest zero occurrence word than a palindrome.

Of the 36 words of length 7, if we consider the Venn diagram with eight 8 regions for the three SZO sets, 25% are in the same region for both canonical and data-driven sequences: BD^5C in the intersection of all 3 SZO sets, AC^5A and DC^5D in Ver and Vis SZO sets only; CB^4 in the MW SZO set only; BA^5C in the Ver SZO set only; and AB^5C, CA^5C, DA^5B and DC^5C in the MW and Ver SZO sets only. Perhaps this might indicate stability of their non-occurrence under change of EEG microstate clustering method.

Distances (Data-Driven EEG Microstates).

For data-driven EEG microstates, as before, at group level, Jaccard distances between sets of shortest zero occurrence (SZOs) words for each cognitive mode are zero for words of each length n from 1 to 9, except for $n = 6$ and $n = 7$ in clock mode. The Jaccard distances between SZO sets of the three cognitive and triangle area are recorded here for these two word lengths. If the area is non-zero, then potentially the occurrence of SZOs could distinguish among the cognitive modes.

Area < .0001	**MW**	**Ver**	**Vis**
MW	0	0.083	0.111
Ver	0.083	0	0.028
Vis	0.111	0.028	0

Table 77: SZO set distances in clock mode for length 6 canonical EEG microstate sequences

Area = 0.208	**MW**	**Ver**	**Vis**
MW	0	0.567	0.767
Ver	0.567	0	0.814
Vis	0.767	0.814	0

Table 78: SZO set distances in clock mode for length 7 canonical EEG microstate sequences

C.7 Discussion

Differences in non-occurring words in canonical and data-driven EEG microstates were broadly similar at group-level. The only non-trivial words that were SZOs in clock-, peak- or event- mode in the various cognitive mode conditions were very short EEG microstates in clock mode flanked by letters from other microstates. These had durations much shorter than the average distance between gloabl field power (GFP) peaks which occur on average every 50ms. The KeyPy software used interpolates EEG microstates between these peaks, and other EEG software toolkits may suppress microstates of very short duration as noise. While we saw differences in the SZO sets of different cognitive modes, it is unclear at present whether not these SZO words are due to noise. At subject-level more shortest-zero occurrence words are found, but this time not only in clock mode but also for peak and event mode too. These appear to differ between individuals and within individuals between cognitive modes. However, their significance remains unclear and it also not clear whether most of these words would no longer arise as shortest zero occurrence words given more data.

It remains to test whether SZO can help classify cognitive states by eliminating states in which they can occur leaving only those states where they may occur. Even if this is not the case for a population level, it might still be the case for individuals reflecting inter-individual variability that we generally observe in EEG. Also, this may be possible for certain individuals but not others (as when the areas of the SZO distance triangles are zero as in some cases shown above).