

# A mean field approach to model levels of consciousness from EEG recordings

Marco Alberto Javarone\*

*Department of Mathematics, University College London, London, UK*

Olivia Gosseries

*Coma Science Group, GIGA Consciousness University  
and University Hospital of Liege, Liege, Belgium*

Daniele Marinazzo

*University of Ghent, Ghent, Belgium*

Quentin Noirhomme

*Faculty of Psychology and Neuroscience Maastricht University, Maastricht, Netherlands*

Vincent Bonhomme

*GIGA - Consciousness, Anesthesia and Intensive Care Medicine Laboratory  
University and CHU University Hospital of Liege, Liege, Belgium  
Department of Anesthesia and Intensive Care Medicine CHU  
University Hospital of Liege and CHR Citadelle, Liege, Belgium*

Steven Laureys<sup>†</sup>

*Coma Science Group, GIGA Consciousness University  
and University Hospital of Liege, Liege, Belgium*

Srivas Chennu<sup>†</sup>

*University of Kent, Medway, UK  
University of Cambridge, Cambridge, UK*

(Dated: June 14, 2020)

## Abstract

We introduce a mean-field model for analysing the dynamics of human consciousness. In particular, inspired by the Giulio Tononi's Integrated Information Theory and by the Max Tegmark's representation of consciousness, we study order-disorder phase transitions on Curie-Weiss models generated by processing EEG signals. The latter have been recorded on healthy individuals undergoing deep sedation. Then, we implement a machine learning tool for classifying mental states using, as input, the critical temperatures computed in the Curie-Weiss models. Results show that, by the proposed method, it is possible to discriminate between states of awareness and states of deep sedation. Besides, we identify a state space for representing the path between mental states, whose dimensions correspond to critical temperatures computed over different frequency bands of the EEG signal. Beyond possible theoretical implications in the study of human consciousness, resulting from our model, we deem relevant to emphasise that the proposed method could be exploited for clinical applications.

---

\* marcojavarone@gmail.com

† these two authors contributed equally

## I. INTRODUCTION

Consciousness is one of the most complex and fascinating phenomena in the brain, attracting the interest of a variety of scholars, spanning from neuroscientists to mathematicians, and from physicists to philosophers [1–8]. In addition, consciousness, as well as other complex systems as those we find in biology, social science, finance and artificial intelligence [9–15], has strongly benefited from the introduction of cross-disciplinary approaches. Despite a huge number of investigations, a lot of its aspects and mechanisms still require to be clarified. Given these observations, we focus on the challenge of quantifying consciousness, putting attention on the transition between mental states. So, we introduce a method for generating Curie-Weiss models [16, 17] from EEG signals, and then we analyse its outcomes by a machine learning classifier. As discussed later, although the spirit of this work is mostly theoretical, we conceive a framework that could support clinicians in some relevant tasks, e.g. in calibrating the optimal amount of anaesthetic for patients, and in assessing the cognitive conditions of unresponsive individuals. Nowadays, a number of devices allow studying the brain structure and its dynamics. Usually, the choice of a specific tool reflects both clinical needs and patient conditions. Here, we use recordings obtained by EEG analyses for two main reasons, i.e. its cheaper cost compared to other technologies and its non-invasive nature. At the same time, it is worth to report that the EEG is less advanced than other diagnostic tools, as fMRI, that generate images of higher quality (e.g. higher spatial resolution). Notwithstanding, stimulated by the above reasons, we aim to improve as much as possible the value of the information content of EEG signals, concerning the dynamics of human consciousness. Before moving to the proposed method, we very briefly introduce two seminal works on this topic, the Integrated Information Theory (IIT hereinafter) developed by Giulio Tononi [18] and the Max Tegmark’s manuscript on the physical representation of consciousness as a state of matter [3]. Both works contain ideas and observations that inspired us during our investigation. The IIT is based on the core concept that human consciousness results from integrated information, generated by an ensemble of interacting elements. So, information emerges from collective action, and its content is extremely much richer than that one can obtain just by a simple summation of the individual contributions, like those provided by the elements belonging to the same ensemble. More in general, the concept of collective effect pervades the field of complex systems, as effectively explained by

Anderson in 'More is different' [19]. Under that light, Tegmark proposed a computational description of consciousness [3], trying to address the IIT by the language of Physics, and developing both a classical and a quantum representation of this phenomenon. From his work [3], we take into account the 'classical' description of IIT, achieved via the Ising model, where a 'conscious' regime emerges only within a very restricted range of values. Such restricted range refers to the collective phenomena occurring when a spin system gets close to its critical temperature, and it is in full agreement with the Damasio's observations [20] on the conditions required for the reaching of homeostasis (i.e. some physical parameters, in the brain, have to be kept within a narrow range of values). Therefore, following ideas and insights of the above-mentioned authors, we propose a method for building a mean-field model from EEG data. Then, we analyse its behaviour by implementing Monte Carlo simulations, whose outcomes are expected to provide the information we need to quantify the transitions between mental states of individuals undergoing deep sedation. In particular, we consider the critical temperature computed in the various realisations of the model, e.g. those achieved on varying the frequency of the EEG signal. Details about the proposed model and the method for classifying mental states are provided in the next section. Here we take the opportunity for highlighting that, despite the increasing interest for modelling brain dynamics by networked approaches (e.g. [21–23, 25–32]), the present investigation is based on the modelling and the analysis of the distribution of electrical activity recorded in the scalp. Moreover, to each frequency band of the signal, as  $\delta$  for 1 – 4 Hz and  $\theta$  for 4 – 7 Hz, corresponds a Curie-Weiss model whose interactions depend on the recorded phase differences across scalp locations. It is worth to note that previous studies showed that the  $\alpha$  (8 – 12 Hz) band can be useful for quantifying transitions between mental states (e.g. [31]), as well as other signal components. At the same time, most of these works consider networks generated by avoiding to include 'weak' interactions. Notably, to remove weak interactions a threshold needs to be defined, and such practice has received some fair criticisms [33]. Remarkably, to the best of our knowledge, the definition of a suitable threshold is currently based only on rules of thumb. So, it is important to remark that the proposed model does not require to filter out, or to cut off, weak interactions. Eventually, let us observe that while from a neuroscience perspective the human consciousness might be investigated considering the full set of frequency bands of an EEG signal, those of major interest seem to be the  $\delta$ , the  $\alpha$ , and the  $\beta$  band. However, due to the influence that has been reported

between the propofol, i.e. the drug administered in our individuals during the examination, and the behaviour of the  $\beta$  band [34, 35], we decided to take into account only the bands  $\delta$  and  $\alpha$ . A more detailed list of features, for analysing consciousness, can be found in [36]. Summarising, our goal is to quantify consciousness, looking also at potential clinical applications. Notably, the EEG signal, as currently processed, provides some information about the state of consciousness of a patient, but it has some limitations. For instance, assessing the level of unconsciousness, or understanding why some individuals report having been fully aware (even if, obviously, unable to communicate) during surgery, is currently difficult by inspecting only the EEG signal. Therefore, under the assumption that the latter might contain more information than those currently extracted, we propose a method to improve its content (see also [37]). Notably, in mathematical terms, the proposed model can be thought of as a more rich representation of the EEG signal, being mapped to a novel vector space that we define state space of mental states. In relation to that, the Curie-Weiss model represents the tool to generate that space of states, by identifying the critical temperatures of each individual during an examination. Here, while critical temperatures are computed to generate the state space of mental states, they have no meaning with what is occurring in the brain of individuals. It is also worth to remark that our choice of using a model (i.e. the Curie-Weiss) usually adopted to describe collective phenomena, as phase transitions, aims to build a direct link with the IIT framework, where the concept of collective behaviour is central. The remainder of the paper is organised as follows: Section II introduces the proposed model. Section III shows results of numerical simulations. Eventually, Section IV ends the paper providing an overall discussion on this investigation, from its goal to the main outcomes, and on some possible future developments.

## II. MODEL

In this section, we describe a framework for classifying mental states, whose variation is represented as a phase transition. In statistical mechanics, the most simple representation of phase transitions is achieved by the Ising model, and the latter has been used in [3] for showing how, in terms of information content, the set of states reachable by that model, at the critical temperature, contains suitable candidates for representing states of consciousness. Therefore, we aim to evaluate with data (i.e. EEG recordings) whether that theoretical

insight can be exploited for quantifying consciousness and performing classification tasks. It is worth to add that the Ising model has been used also by other authors for investigating different dynamics of the brain (see for instance [21, 23, 24, 38–40]). As above mentioned, the EEG signal can be decomposed into frequency bands and different measures can be adopted for its analysis, usually selected according to specific needs. For our purposes, a particularly useful parameter is the weighted Phase Lag Index [32] (wPLI hereinafter) that quantifies the correlation between pairs of sensors in the scalp. In general, considering two time series  $a(t)$  and  $b(t)$ , the wPLI is defined as

$$wPLI = \left| \frac{\sum_{t=i}^n |im(P_{ab,t})| \operatorname{sgn} im(P_{ab,t})}{\sum_{t=i}^n |im(P_{ab,t})|} \right| \quad (1)$$

where  $\operatorname{sgn}$  indicates the signum function,  $im$  indicates the imaginary part, and  $P_{ab,t}$  the complex cross-spectral density, of  $a(t)$  and  $b(t)$ , at time  $t$ . Notably, given the power spectral densities of the two signals  $P_{aa}$  and  $P_{bb}$ , i.e. the distribution of the power across the frequency components of  $a(t)$  and  $b(t)$ , respectively, the cross-spectral density quantifies the correlation between them. Then, for each frequency band, the interactions of the resulting Curie-Weiss are computed by scalar multiplication of the wPLI index with the relative power of the signal. A quick inspection of 1 shows that the wPLI and the power of the signal are strongly correlated. However, we found beneficial to combine them for realising the mean-field model. In doing so, inspired by the Tegmark’s approach for studying the IIT by a simple physical system, and remaining in the land of the classical physics, we build a Curie-Weiss from EEG recordings assigning a spin to each sensor and computing interactions by the combination of the indexes above mentioned (i.e. wPLI and Power). Thus, starting with randomly assigned values of spins ( $\sigma \pm 1$ ), and quenching interactions [17, 41], we study the dynamics of the system towards equilibrium. Following this method, it is possible to compute a critical temperature for each frequency band of the EEG signal, as  $T_c^\alpha$  for the  $\alpha$  band. However, since this signal varies over time, node interactions can vary as well. Here, actually, the variation of interactions is expected to be useful for detecting variations of mental states. For instance, as reported in [31], networks built using the wPLI index show variations as individuals undergo sedation and then recover to their original conscious state. To tackle this aspect, the EEG signal is sampled into four different points labelled as  $C$ ,  $S$ ,  $DS$ ,  $R$ , representing consciousness, sedation, deep sedation, and recovery, respectively. Mental states  $C$  and  $DS$  are both classified as states of consciousness, in agreement with

[42], while  $S$  and  $R$  are labelled as transition states. Also,  $C$  and  $DS$  can be viewed as two equilibrium states (although the deep sedation, i.e.  $DS$ , in our individuals has been induced by a drug). Let us now proceed to the formal definition of spin interactions. Recalling that the EEG signal is decomposed into 5 main frequency bands and that we extract 4 samples per recording, for each individual we can generate up to 20 mean-field models. Since the wPLI quantifies the correlation between pairs of nodes, we indicate with  $wPLI_{i,j}^x$  the correlation between sensors  $i$  and  $j$  in the  $x$ th frequency band. Accordingly, the interaction term  $J$  reads

$$J_{i,j}^x(s) = P^x(s) \cdot wPLI_{i,j}^x(s) \quad (2)$$

with  $s$  sample (or mental state) and  $P^x(s)$  power of the  $x$ th band for that specific sample. Thus, the Hamiltonian of the system is

$$H = -\frac{1}{N} \sum_{i,j=1, j \neq i}^N J_{i,j} \sigma_i \sigma_j \quad (3)$$

with  $N$  number of sensors and  $\sigma$  spin assigned to them. While the spin is a quantum property of particles, it finds large utilisation in non-physical models, typically for representing binary features. For instance, in social dynamics the spin can represent a binary opinion [43], in evolutionary game theory a strategy [44, 45], and in neural models it can indicate firing (+1) and resting (−1) states [46]. The proposed model does not define explicitly a connection between the value of the spin and the underlying neural activity since the spin dynamics is implemented only to obtain theoretical insights about the activity distribution recorded in the scalp. In particular, the studying of the order-disorder phase transitions, occurring in each model realisation, allows computing the set of critical temperatures  $T_c$  associated to every mental state. To this end, for each configuration (i.e. an individual in a given state), spin interactions  $J$  are considered as *quenched*, so one can analyse the dynamics of spins starting from a random distribution. Here, since our efforts are directed towards quantifying the human consciousness, we put the attention on the  $\alpha$  and  $\delta$  bands that, according to previous clinical studies (e.g. [47, 48]), seem to be quite relevant for investigating this complex phenomenon, as well as others as psychotic disorders [49]. The described method allows observing the motion that individuals take in the space of mental states. Such motion is defined along a path on a bidimensional plane, whose axes are  $(T_c^\alpha, T_c^\delta)$ . Moreover, since recordings terminate once individuals recover their original cognitive state, the resulting

path forms a closed cycle. Then, we implement a Machine Learning tool to exploit the resulting paths (or cycles) for classification tasks, as for identifying the correct label of a point (e.g.  $DS$ ) in the mental state space. In particular, a classifier able to assign a label to each point takes as input vectors with components  $(T_c^\alpha, T_c^\delta)$ . As a huge literature suggests, classifiers can be realised by many algorithms, e.g. neural networks [50]. However, given the small size of our dataset, we implemented a Support Vector Machine (SVM hereinafter) [50] —see also A. Summarising, starting from EEG recordings, the proposed model generates bidimensional vectors, whose entries correspond to the critical temperatures computed in the  $\alpha$  and  $\delta$  frequency bands, state by state. These vectors constitute the input of an SVM designed to discriminate between the two states of consciousness  $C$  and  $DS$ .

### III. RESULTS

The proposed model has been tested with a dataset of EEG signals obtained by recording 8 healthy volunteers [51] undergoing sedation induced by propofol —see Appendix B for details. So, each recording started with individuals in the conscious state and terminated after their complete recovery. Four main mental states can be identified: awareness, sedation, deep sedation, and recovery, and for each of them, one sample is extracted from the recording. The resulting amount of samples (based on 173 sensors), available for all 5 frequency bands, allows generating 20 Curie-Weiss configurations per individual. However, for the reasons reported above, only the  $\alpha$  and  $\delta$  bands are considered, therefore the number of configurations in this investigation is limited to 8. A model configuration has specific values of spin interactions. Let us remind that spins are the mathematical representation of the sensors in the scalp, and their value (i.e.  $\pm 1$ ) is randomly assigned as below described. Instead, the interactions are computed by Eq. 2. Before proceeding showing results of the simulation, we compare the average value of interactions obtained by using the above equation (i.e. Eq. 2) with those (average) values that can be achieved by using the spectral power ( $P$ ) and by the  $wPLI$ , individually. This comparison is performed over the 4 mental states and the outcomes have been averaged over all individuals. Results are shown in Figure 1, which reports also the related standard error of the mean, i.e.  $SEM = \frac{\sigma}{\sqrt{N}}$  with  $\sigma$  standard deviation and  $N$  number of individuals. A quick inspection of Figure 1 suggests that our method enhances in good extent the difference between the states  $C$  and  $DS$ , embedding



the contribution of the spectral power and that of the wPLI index. Let us highlight that our

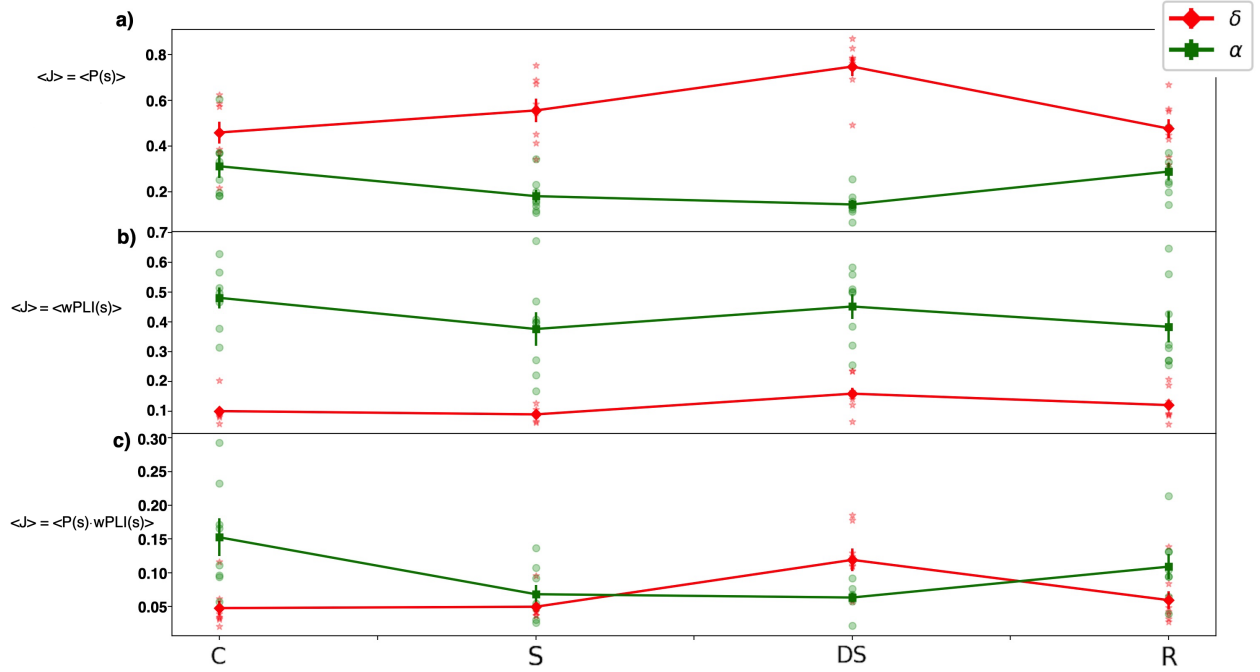


FIG. 1. Average value of the interaction terms, over 4 different mental states, in Curie-Weiss built using three different approaches. In particular, the interactions have been defined by using **a)** the power spectrum  $\langle J \rangle = \langle P(s) \rangle$ , **b)** the wPLI index, i.e. the correlation  $\langle J \rangle = \langle wPLI(s) \rangle$  between pairs of sensors, and **c)** the scalar product of the power with wPLI index (i.e.  $\langle J \rangle = \langle P(s) \cdot wPLI(s) \rangle$ ). The legend indicates the considered frequency bands: red line for the  $\delta$  and the green line for the  $\alpha$ , and the related diamonds and squares represent the average value for each label. The error bars have been calculated using the standard error of the mean (SEM). Then, semi-transparent points (i.e. red stars for  $\delta$  and green circles for  $\alpha$ ) represent the single individuals.

goal is now to extract information related to the mental states of individuals by processing the resulting Curie-Weiss models. So, the next step is focused on the identification of the critical temperature for the different configurations of the Curie-Weiss model. This process allows us to build a 'state space' of mental states, where we can quantify, and also visualise, the path followed by individuals across the clinical examination. Then, the paths of mental states are used to train a simple machine learning tool whose goal is to classify conscious states of our individuals.

### A. Mental States I: Building the State Space

All numerical simulations have been performed on Curie-Weiss configurations related to every single individual, i.e. spin interactions have not been averaged as (instead) it has been done for the analysis shown in Figure 1. While spin interactions  $J$  are *quenched*, we study the dynamics of spins that, at the beginning of each simulation, are randomly set to  $\pm 1$ . Notably, simulations are implemented for studying order-disorder phase transitions, and more specifically for identifying the critical temperature of each configuration. For that purpose, a useful parameter is the absolute value of the average magnetization of the system. The latter strongly depends on the system temperature  $T$ , despite its definition does not include the temperature explicitly, and it reads

$$\langle M \rangle = \sum_i^N \frac{|\sigma_i|}{N} \quad (4)$$

It is worth to observe that beyond identifying the critical temperature  $T_c$  for each case, it is possible to assess if  $T_c \approx \langle J \rangle$ . The latter, as further explained later, can be a relevant feature for designing clinical applications based on the proposed model. Figure 2 reports the results obtained on one individual, randomly chosen among those available. To assess whether a temperature is 'critical', the variance of  $M$  (indicated as  $\sigma^2$ ) is analysed in function of the inverse temperature  $\beta$  since the highest value of  $\sigma^2$  is reached at  $T = T_c$  —see inset of Figure 2. Then, once computed the critical temperature for all considered cases, we focus on the numerical difference between these values and the average interaction term (i.e.  $\langle J \rangle$ ) of each Curie-Weiss realisation. In doing so, we found that approximating the  $T_c$  with the  $\langle J \rangle$  gives small errors, limited to the 12.5% of  $T_c$  (see for instance the lines red and green, indicating the  $\langle J \rangle$  and the  $T_c$ , respectively, in figure 2). Finally, results of the mean-field model are shown in Figure 3. Notably, plotting the values of the critical temperatures computed in the band  $\delta$  ( $T_c^\delta$ ) versus those computed in the band  $\alpha$  ( $T_c^\alpha$ ) allows us to visualise the path from the initial point  $C$  to the point  $DS$ , and that of return. So, we have now a state space that contains 'mental paths', whose evolution (or motion) can be further analysed. Before proceeding to the classification of mental states, we highlight that the order-disorder phase transition studied in the resulting Curie-Weiss model has not a biological meaning in this context. Notably, it is a process simulated to extract information from the resulting model, that we assume can be useful for understanding the dynamics of

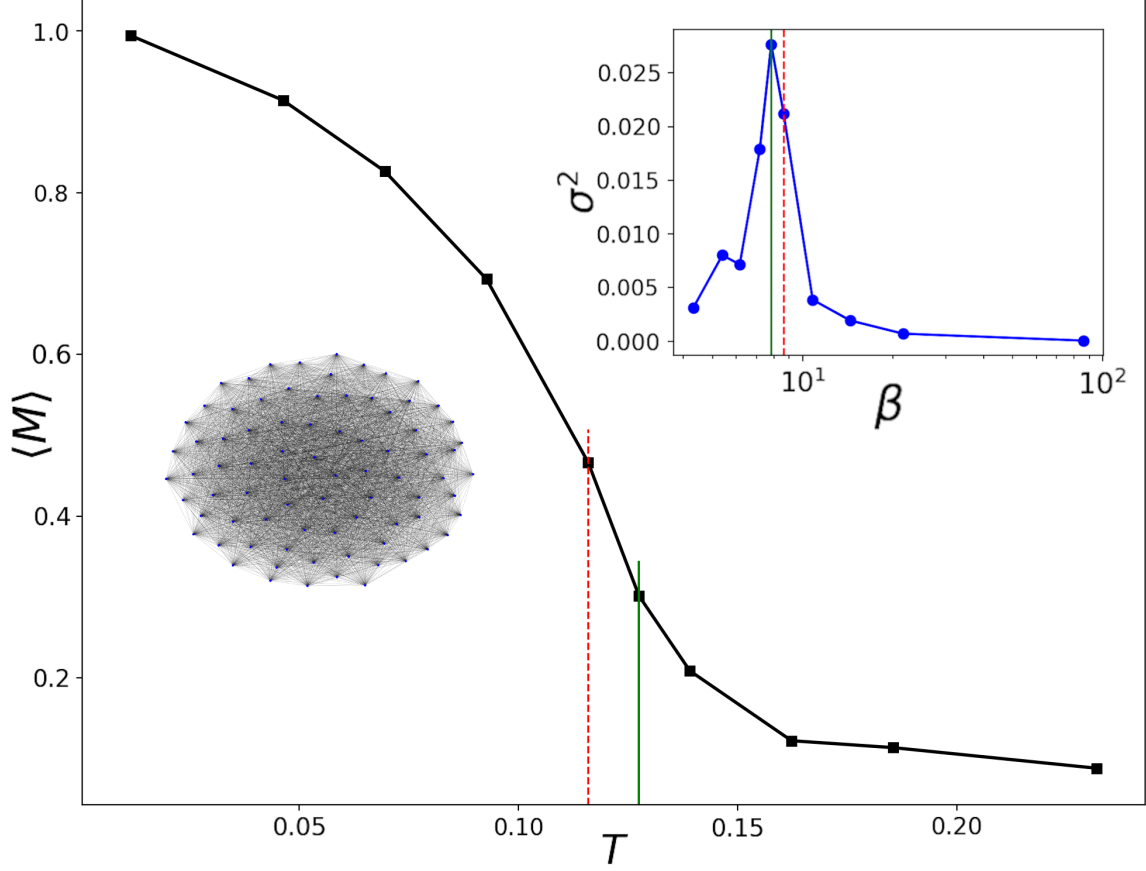


FIG. 2. Average magnetization in function of the system temperature. A pictorial of a Curie-Weiss model is shown close to the main line, while the inset shows the variance of  $M$  in function of the inverse temperature  $\beta$  (on a semi-logarithmic scale). The critical temperature is indicated by the green line (both in the main picture and in the inset), while the average interaction term for the system ( $\langle J \rangle$ ) is indicated by a red dotted line (both in the main picture and in the inset).

mental states. In addition, the value of critical temperatures has been computed on a finite size system, while phase transitions occur in the thermodynamic limit. Therefore, further investigations with bigger systems can be useful also to evaluate if the critical temperatures we obtained in our analyses are much different from those one would find scaling the size of the model.

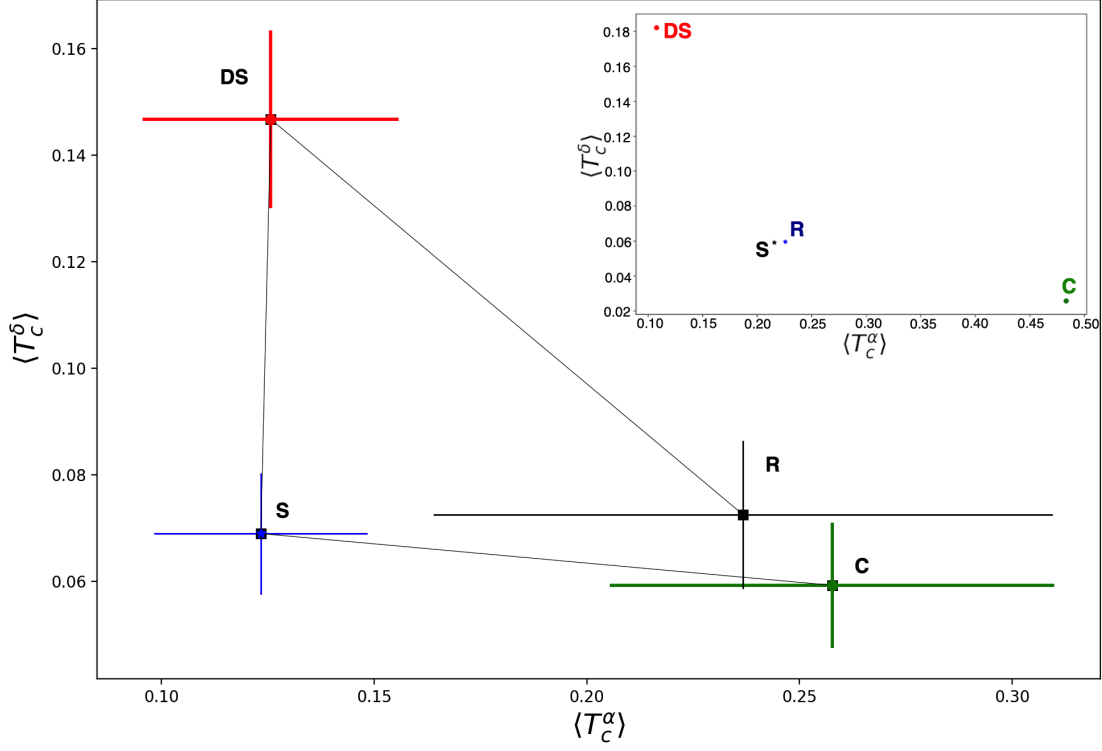


FIG. 3. Diagram  $\langle T_c^\alpha \rangle, \langle T_c^\delta \rangle$ , with an inset showing the results obtained on a single individual (randomly chosen). Notably, the points in the diagram (and in the inset) represent the 4 different mental states during the clinical experiment:  $C$ ,  $S$ ,  $DS$ , and  $R$ , i.e. consciousness, sedation, deep sedation, and recovery, respectively. The black line indicates the path followed by each individual, undergoing sedation and then recovering to the initial state.

## B. Mental States II: Classification

Each path obtained by the previous method characterises an individual, thus we use this representation for implementing a Machine Learning tool for classifying mental states. In our investigation, paths are composed of 4 points in bi-dimensional state space. Therefore, we can identify a vector, whose entries are the critical temperatures computed for the bands  $\alpha$  and  $\delta$ , for each mental state. The goal is to assess whether these vectors are useful for generating a confident boundary able to separate different mental states. So, for the sake of simplicity, we build a classifier for discriminating between the state  $C$  and the state  $DS$ . Due to the small size of the dataset, we use an SVM implemented by a kernel based radial basis function (as commonly adopted in classification tasks). The outcomes are shown in

Figure 4, where the axes refer to the two critical temperatures of each individual (i.e. one per frequency band), not to their average values (as in the main plot of Figure 3). Let us

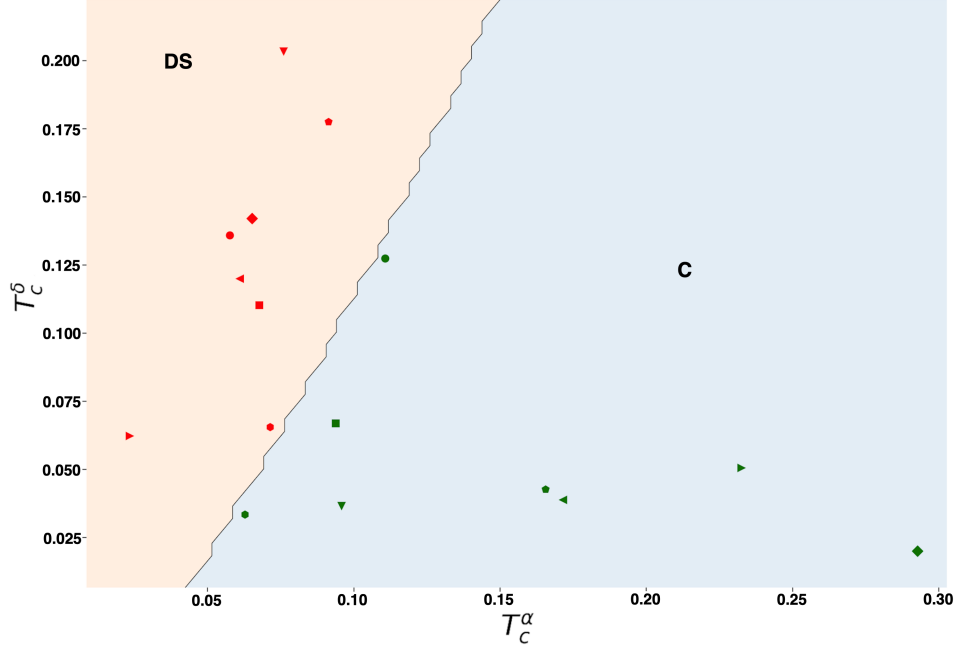


FIG. 4. Results of SVM applied to the EEG dataset on the plane  $T_c^\alpha$  vs  $T_c^\delta$ . Different symbols refer to different individuals. The green colour indicates the conscious state ( $C$ ), while the red colour indicates deep sedation ( $DS$ ). The black line identifies the edge between the two classes, i.e.  $C$  and  $DS$ .

briefly describe the procedure for training and testing the SVM. Notably, given a dataset of 8 individuals, the training has been performed on 7 out of them, and the testing on the excluded one. This process was then repeated excluding each time one different individual. In doing so, we can evaluate 8 different testing phase results. Following the above procedure, now a further analysis compares the results that an SVM achieves when fed with vectors obtained by three different approaches. Notably, in all cases, vectors resulted from a mean-field model (with the method above described), but its interactions can be defined by Eq. 2, or by the power spectrum and the wPLI, individually. This comparison, shown in Figure 5, is quite relevant because we want to evaluate in which extent the utilisation of Eq. 2 provides an actual benefit for our purposes. Remarkably, we found that our method produced the smallest error rate in the task of classifying mental states —see Figure 6. In particular, the SVM trained and tested with our method did only one error, over 8 tests (i.e. 12%), while

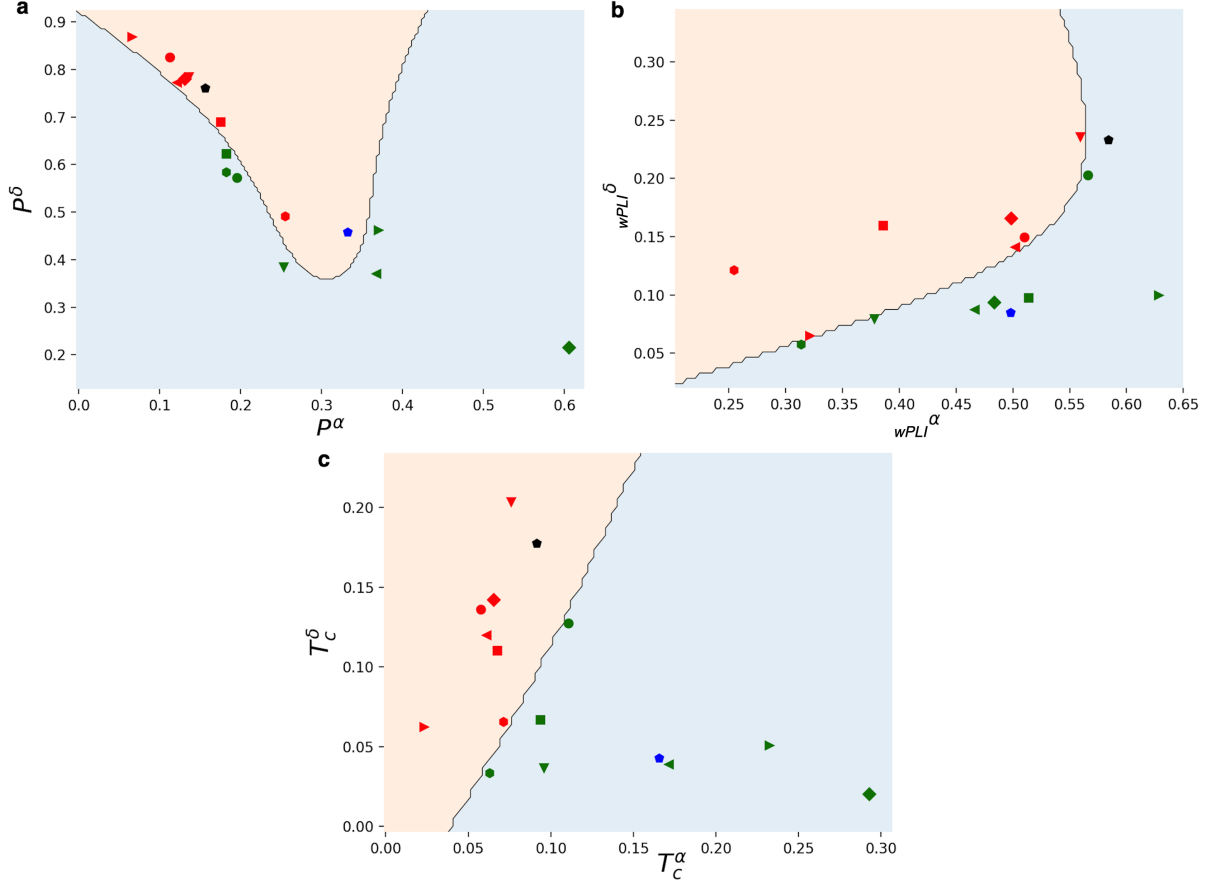


FIG. 5. Comparison between the results of SVM model built with data coming from three different methods: **a)** Using spectral power; **b)** Using wPLI; **c)** Critical temperatures obtained by using Eq. 2. Different symbols refer to different individuals. The green colour indicates the conscious state (*C*) and red one that of deep sedation (*DS*). The black line identifies the edge between the two classes, i.e. *C* and *DS* in the mental state space. Then, points with the same symbol, but a different colour (e.g. blue and black), indicate an error in the classification process.

those trained with data coming from the two other methods did more errors (i.e. 33% and 25% by using power spectrum and wPLI, respectively). Before to conclude this section, we deem important to mention that actual measures might be used to compare outcomes of different models, as the AUC. At the same time in this investigation, considering the number of samples, we found beneficial to evaluate the error rate.

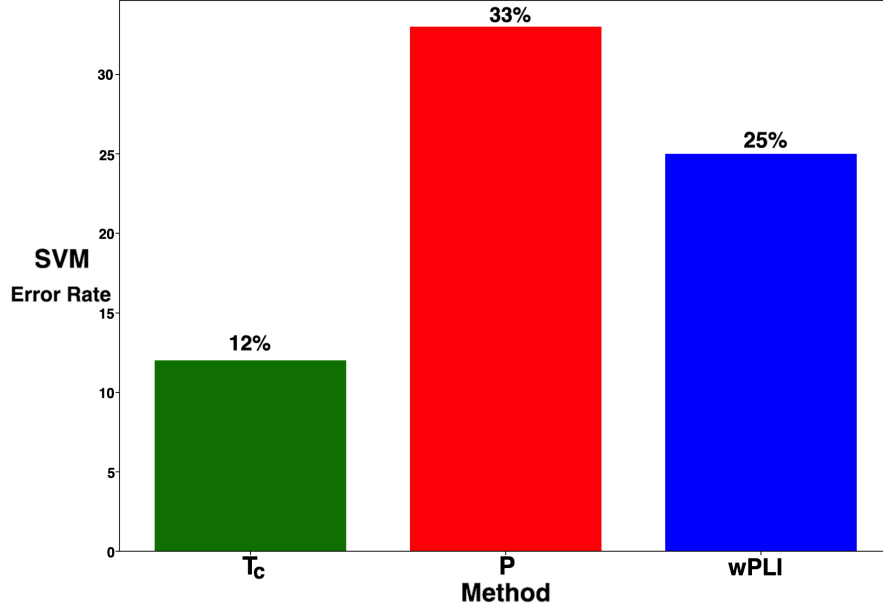


FIG. 6. SVM error rate computed by data coming from three different methods: Critical temperatures ( $T_c$ ) obtained by Eq. 2, Spectral power, and  $wPLI$ . Note: the lesser the better.

#### IV. DISCUSSION AND CONCLUSION

In this manuscript, we propose a method for quantifying human consciousness and classifying mental states using EEG signals (see also [36, 52]). Notably, we introduce a mean-field model of the distribution of electrical activity in the brain, whose outcomes are used for training a Machine Learning classifier. Inspired by the Tegmark's work [3] about the IIT, we study order-disorder phase transitions occurring in Curie-Weiss models whose interactions depend on the phase differences across scalp locations. This analysis allows computing the critical temperatures achieved for different configurations of the model, i.e. on varying the mental state and the considered frequency band of the signal. Here, the interactions between spins are computed by the scalar product of the power spectrum with the  $wPLI$  index, for each specific band. The benefits coming from this choice are reported in Figure 5 and Figure 6. Remarkably, the critical temperatures, computed by means of numerical simulations, allow defining a state space where we can observe the path of mental states. It is worth to mention that, according to Giulio Tononi [42], consciousness emerges also during dreamlike phases of the sleep. Therefore, considering previous investigations stating that some individuals reported dreamlike activity, during an induced deep sedation [53],

both awareness and deep sedation in principle should be considered as conscious states (see also [54]). The path between awareness and deep sedation shows, in the middle, the transition states ( $S$  and  $R$ ). Thus, summarising, we have two conscious states and two transition states. It is relevant to clarify that we are not using the formal meaning of 'transition state', as usually adopted in stochastic models (e.g. the voter model [55]), otherwise also the 'deep sedation' state would be defined a transition state since it lasts only for the duration of the drug effect. At the same time, these considerations could be extended further since, for instance, states of coma could be properly classified as absorbing states, and so on. Hence, our definitions have not that level of formality, despite we find interesting to investigate more on this. So, once defined a mental state space, we use an SVM for tracing the boundary between states  $C$  and  $DS$ . We remind that the dataset for performing the investigation has been obtained with 8 individuals wearing EEG sensors. Then, the EEG recordings have been used for building the mean-field model (i.e. Curie-Weiss) and generating a training dataset. The actual choice of the frequency bands, i.e.  $\alpha$  and  $\delta$ , actually depends also on the clinical settings of examinations (e.g. the utilisation of propofol for inducing the sedation). Results suggest that the critical temperatures are useful to perform classification tasks, discriminating between the two conscious states. Therefore, thinking about the potential use of the proposed model, at clinical level, we conceive a framework composed of two elements: one devised for computing the average interaction term in different bands, that, as we proved, approximates the critical temperature in the related Curie-Weiss configuration, and the other based on an SVM (or on another Machine Learning algorithm). The fact that the average interaction can be approximated by the critical temperature, in principle, is not too surprising. Notably, the critical temperature of the 2D Ising model is  $T_c = 1$  for  $J = 1$ . However, it has been worth, for the reasons above described, to confirm that hypothesis. Beyond the theoretical analysis, which requires further investigations to confirm our achievements, the framework that we conceive could support, after appropriate testing, clinicians in different scenarios. For instance, it could be useful for defining the optimal amount of drug for sedating a patient, or for classifying the level of unconsciousness of unresponsive patients. Also, it is interesting to observe that the mean-field model realised by EEG recordings might be conceptually related to the 'classical' Tegmark's description of consciousness (i.e. that based on the Ising model). Notably, it would be useful to evaluate how to obtain only one model, across the different bands, and to study its behaviour at dif-



ferent temperatures. Furthermore, we highlight the possibility to extend further this work trying to improve the connection with the IIT framework, in order to develop mathematical tools able to analyse the human consciousness from a perspective supported by the Tononi's insights. Finally, we deem interesting to provide a comment on the method implemented to represent the path of mental states. Notably, as above reported, each state is described by a vector of critical temperatures (one per frequency band). So, like for other kinds of models, as those based on network theory, the proposed approach can find application in various contexts, in particular when the strength of interaction among the elements of a system is relevant. In our case, the interaction strength clearly shows a time dependency, however that is not a mandatory requirement to implement and analyse a system by means of a Curie-Weiss model. Moreover, the method might provide interesting insights also when built considering interactions defined by meaningful semantic relations among elements of a system, i.e. abstracting from a physical system and therefore increasing its potential applicability to a much wider set of cases. We conclude emphasising that our results indicate that the EEG signal might be further exploited both for obtaining a deeper understanding of human consciousness, and for implementing novel tools to support clinicians in many complex and critical activities. To this aim, there are some important aspects to consider in future investigations. Firstly, here we focused on two frequency bands ( $\alpha$  and  $\delta$ ), however it is important to assess if our choice is the optimal one. Notably, although previous literature suggests that these two bands are particularly relevant for the human consciousness, finding the way to exploit also other frequency bands might be useful (as before mentioned, the choice can depend also on the anesthetic drug). Then, interactions among sensors have been identified by means of the wPLI index. However, also other correlation measures could be taken into account. Eventually, we identified transient states, i.e. those states between  $C$  and  $DS$ . We deem that their role, dynamics and properties need to be further analysed, as well as a classification algorithm for their detection could be useful.

## ACKNOWLEDGMENTS

This work was supported by the Belgian National Funds for Scientific Research (FRS-FNRS), the University and University Hospital of Liege, the fund Generet, the King Baudouin Foundation, the BIAL Foundation, the AstraZeneca foundation, the Belgian Federal

Science Policy Office (BELSPO) in the framework of the PRODEX Programme, the Center-TBI project (FP7-HEALTH- 602150), the Public Utility Foundation ‘Université Européenne du Travail’, “Fondazione Europea di Ricerca Biomedica”, the Mind Science Foundation, and the European Commission. SL is research director at F.R.S-FNRS.

---

- [1] Tononi, G., Edelman, G.M.: Consciousness and Complexity. *Science* **282-5395** 1846–1851 (1998)
- [2] Blackmore, S., Troscianko, E.T.: Consciousness. An Introduction. *Routledge* (2018)
- [3] Tegmark, M.: Consciousness as a State of Matter. *Chaos, Solitons & Fractals* **76** 238–270 (2015)
- [4] Zhou, D.W., Mowrey, D.D., Tang, P., Xu, Y.: Percolation Model of Sensory Transmission and Loss of Consciousness Under General Anesthesia. *PRL* **115** 108103 (2015)
- [5] Fingelkurts, Andrew A., Fingelkurts, Alexander A., Neves, Carlos F.H.: Consciousness as a phenomenon in the operational architectonics of brain organization: Criticality and self-organization considerations. *Chaos, Solitons & Fractals* **55** 13–31 (2013)
- [6] Fingelkurts, A. A., Fingelkurts, A. A., Bagnato, S., Boccagni, C., Galardi, G.: Do we need a theory-based assessment of consciousness in the field of disorders of consciousness? *Frontiers in Human Neuroscience* **8** 402 (2014)
- [7] Hameroff, S.R., Penrose, R.: Conscious events as orchestrated space-time selections. *Journal of Consciousness Studies* **3-1** 36–53 (1996)
- [8] Solms, M., Friston, K.: How and Why Consciousness Arises: Some Considerations From Physics and Physiology. *Journal of Consciousness Studies* **25(5-6)** 202–238 (2018)
- [9] Mitchell, M.: Complexity: a guided tour. *Oxford University Press* (2009)
- [10] Kaneko, K.: Life: An Introduction to Complex Systems Biology. *Springer* (2006)
- [11] Flack, J.C.: Life’s information hierarchy. In From Matter to Life: Information and Causality (eds: SI Walker, PCW Davies, GFR Ellis). *Cambridge University Press* 283—302 (2017)
- [12] Vespignani, A.: Modelling dynamical processes in complex socio-technical systems. *Nature Physics* **8** 32–39 (2012)
- [13] Rickles, D.: Econophysics and the Complexity of Financial Markets. *Philosophy of Complex Systems* **10** 531–565 (2011)

- [14] Solé, R., Moses, M., Forrest, S.: Liquid brains, solid brains. *Trans. R. Soc. B* **374** 20190040 (2019)
- [15] Friston, K.: The free-energy principle: a unified brain theory?. *Nature reviews neuroscience* **11(2)** 127 (2010)
- [16] Kochmanski, M., Paszkiewicz, M., Wolski, S.: Curie-Weiss magnet: a simple model of phase transition. *European Journal of Physics* **34** 1555 (2013)
- [17] Genovese, G., Barra, A.: A mechanical approach to mean field spin models. *Journal of Mathematical Physics* **50** 053303 (2009)
- [18] Tononi, G.: An information integration theory of consciousness. *BMC Neuroscience* **5** 4 (2004)
- [19] Anderson, P.W.: More Is Different *Science* **177** 4047 (1972)
- [20] Damasio, A.: Self Comes to Mind: Constructing the Conscious Brain. *New York, Vintage* (2010)
- [21] Fraiman, D., et al.: Ising-like dynamics in large-scale functional brain networks. *Physical Review E* **79** 061922 (2009)
- [22] Betzel, R.F., Bassett, D.S.: Multi-scale brain networks. *NeuroImage* **160** 73–83 (2017)
- [23] Gu, S., Cieslak, M., Baird, B. et al.: The Energy Landscape of Neurophysiological Activity Implicit in Brain Network Structure. *Scientific Reports* **8** 2507 (2018)
- [24] Stramaglia, S., Pellicoro, M., Angelini, L., Amico, E., Aerts, H., Cortés, J.M., Laureys, S., Marinazzo, D.: Ising model with conserved magnetization on the human connectome: Implications on the relation structure-function in wakefulness and anesthesia. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **27** 108103 (2017)
- [25] Bassett, D.S., Bullmore, E.T.: Small-world brain networks revisited. *The Neuroscientist* **23-5** 499–516 (2017)
- [26] Papo, D., Buldú, J.M., Boccaletti, S., Bullmore, E.T.: Complex network theory and the brain. *Phil. Trans. R. Soc. B* **369-1653** 20130520 (2014)
- [27] De Domenico, M., Sasai, S., Arenas, A.: Mapping Multiplex Hubs in Human Functional Brain Networks. *Frontiers in Neuroscience* **10** 326 (2016)
- [28] Mastrandrea, R., Gabrielli, A., Piras, F. et al.: Organization and hierarchy of the human functional brain network lead to a chain-like core. *Scientific Reports* **7** 4888 (2017)
- [29] van Diessen, E., et al.: Opportunities and methodological challenges in EEG and MEG resting state functional brain network research. *Clinical Neurophysiology* **126(8)** 1468–1481 (2015)

- [30] De Vico Fallani, F., Richiardi, J., Chavez, M., Achard, S.: Graph analysis of functional brain networks: practical issues in translational neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369(1653)** 20130521 (2014)
- [31] Chennu, S., O'Connor, S., Adapa, R., Menon, D.K., Bekinschtein, T.A.: Brain Connectivity Dissociates Responsiveness from Drug Exposure during Propofol-Induced Transitions of Consciousness. *PLoS Comput Biol* **12(1)** e1004669 (2016)
- [32] Imperatori, L.S., et al.: EEG functional connectivity metrics wPLI and wSMI account for distinct types of brain functional interactions. *bioRxiv* 450270 (2018)
- [33] Papo, D., Zanin, M., Buldú, J.M.: Reconstructing functional brain networks: have we got the basics right? *Frontiers in Human Neuroscience* **8** 107 (2014)
- [34] Boussen, S., Spiegler, A., Benar, C., Carrère, M., Bartolomei, F., Metellus, P., Voituriez, R., Velly, L., Bruder, N., Trébuchon, A.: Time rescaling reproduces EEG behavior during transition from propofol anesthesia-induced unconsciousness to consciousness. *Scientific Reports* **8** 6015 (2018)
- [35] McCarthy, M.M., Brown, E.N., Kopell, N.: Potential Network Mechanisms Mediating Electroencephalographic Beta Rhythm Changes during Propofol-Induced Paradoxical Excitation. *Journal of Neuroscience* **28(50)** 13488–13504 (2008)
- [36] Engemann, D.A., et al.: Robust EEG-based cross-site and cross-protocol classification of states of consciousness. *Brain* **141-11** 3179–3192 (2018)
- [37] Juel, B.E., et al.: Distinguishing Anesthetized from Awake State in Patients: A New Approach Using One Second Segments of Raw EEG. *Frontiers in Human Neuroscience* **12** 40 (2018)
- [38] Lynn, C.W., Papadopoulos, L., Lee, D.D., Bassett, D.S.: Surges of Collective Human Activity Emerge from Simple Pairwise Correlations. *Physical Review X* **9** 011022 (2019)
- [39] Das, T.K., et. al: Highlighting the Structure-Function Relationship of the Brain with the Ising Model and Graph Theory. *BioMed research international* **2014** 237898 (2014)
- [40] Abeyasinghe, P.M., et. al.: Role of Dimensionality in Predicting the Spontaneous Behavior of the Brain Using the Classical Ising Model and the Ising Model Implemented on a Structural Connectome. *Brain Connectivity* **8-7** 444–455 (2018)
- [41] Javarone, M.A., Marinazzo, D.: Dilution of Ferromagnets via a Random Graph-Based Strategy. *Complexity* **2845031** 11 (2018)

- [42] Alkire, M.T., Hudetz, A.G., Tononi, G.: Consciousness and anesthesia. *Science* **322**(5903) 876—880 (2008)
- [43] Martins, André, C. R., Galam, S.: Building up of individual inflexibility in opinion dynamics. *Physical Review E* **87-4** 042807 (2013)
- [44] Perc, M., et al.: Statistical physics of human cooperation. *Physics Reports* **687** 1–51 (2017)
- [45] Javarone, M.A.: Statistical Physics and Computational Methods for Evolutionary Game Theory. *Springer* (2018)
- [46] Amit, D.J.: Modeling Brain Function: The World of Attractor Neural Networks. *Cambridge University Press* (1992)
- [47] Lee, M., Baird, B., Gosseries, O. et al.: Connectivity differences between consciousness and unconsciousness in non-rapid eye movement sleep: a TMS–EEG study. *Scientific Reports* **9** 5175 (2019)
- [48] Golkowski, D., et al.: Simultaneous EEG–PET–fMRI measurements in disorders of consciousness: an exploratory study on diagnosis and prognosis. *Journal of Neurology* **264-9** 1986–1995 (2017)
- [49] Howells, F.M., Temmingh, H.S., Hsieh, J.H., van Dijen, A.V., Baldwin, D.S., Stein, D.J.: Electroencephalographic delta/alpha frequency activity differentiates psychotic disorders: a study of schizophrenia, bipolar disorder and methamphetamine-induced psychotic disorder. *Transl Psychiatry* **8:1** 75 (2018)
- [50] Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. *MIT Press* **5** 4 (2016)
- [51] Murphy, M., et al.: Propofol Anesthesia and Sleep: A High-Density EEG Study *Sleep* **34**(3) 283–291, (2011)
- [52] Patlatzoglou, K. et al.: Deep Neural Networks for Automatic Classification of Anesthetic-Induced Unconsciousness. *Brain Informatics* 216–225 (2018)
- [53] Gyuláházi, J., Redl, P., Karányi, Z., Varga, K., Fülesdi, B.: Dreaming under anesthesia: is it a real possibility? Investigation of the effect of preoperative imagination on the quality of postoperative dream recalls. *BMC anesthesiology* **16:1** 53 (2014)
- [54] Yoo, J., Kwon, J., Choe, Y.: Predictable internal brain dynamics in EEG and its relation to conscious states. *Front Neurobot* **8** 18 (2014)
- [55] Liggett, T.M.: Stochastic Models of Interacting Systems. *The Annals of Probability* **25-1** 1–29 (1997)

## Appendix A: Data Classification

Here we briefly present the algorithm used for classifying mental states, i.e. the Support Vector Machine (SVM). The latter receives as input vectors generated by using the output of the Curie-Weiss model. In doing so, the input vectors contain the critical temperature computed for each frequency band, across the different mental states. For instance, considering the bands  $\delta$  and  $\alpha$ , a vector representing one individual in the conscious state contains the two related critical temperatures, i.e.  $V_i(c) = [Tc^\delta, Tc^\alpha]$ . The collection of these vectors is used for training and testing the SVM. Notably, since an SVM is a supervised learning model, it requires a training set for the learning process and, usually, such training set is based on a fraction (randomly selected) of the dataset under investigation. For instance, a generic dataset  $D$  contains  $n$  elements, i.e.  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ , so that for each input vector  $x_i$  one has the associated output  $y_i$ , i.e. a label or class of belonging (in our case  $x_i$  is a temperature vector, while  $y_i$  is the corresponding mental state). The dimension of the input vector depends on the number of features which describe the system, so in our case corresponds to the number of considered frequency bands. Accordingly, a binary classification task can be implemented considering as outputs two possible values, e.g.  $+1$  and  $-1$ . Similarly, for a multi-label case, one has to identify proper values for the output. Then, an SVM aims to identify the 'maximum-margin hyperplane' that separates input vectors according to their corresponding output. In the bidimensional case, the hyperplane is a simple line. The SVM can be both linear and non-linear, depending on the function (defined as Kernel) used to separates vectors into the related classes. Also, the closest vectors to the hyperplane are called 'support vectors'. For instance, in the linear case, a bidimensional hyperplane can be identified by the equation  $\hat{w}\hat{x} + b = 0$ , with  $\hat{w}$  vector of weights,  $b$  (small constant) bias, and  $\hat{x}$  feature space of the input vectors. Therefore, the SVM tries to identify the most suitable  $\hat{w}$  and  $b$  to maximise the margin, i.e. the distance, between vectors belonging to the two (or more) different classes. Following the above example, in a binary classification, input vectors located above the computed hyperplane belong to one class (e.g.  $+1$ ), while those located below the hyperplane belong to the other class (e.g.  $-1$ ).

## Appendix B: Clinical Information

The full description of the experimental setup implemented to acquire the dataset, used in this investigation, is reported in [51]. However, for the sake of completeness, here we provide some relevant details related to the participants and to the experimental protocol. First of all, we emphasise that the investigation was approved by the Ethics Committee of the Faculty of Medicine of the University of Liege. Also, the group of participants, with mean age  $22 \pm 2$  y, includes 4 males. All participants gave written informed consent, so an appropriate examination, including anamnesis, has been performed to assess potential issues related to the anaesthesia, as pregnancy, mental illness, drug addiction, asthma, and so on. Experiments are based on fifteen-minute spontaneous 256-electrode hd-EEG recordings, that can be divided into 4 different states: normal wakefulness, sedation (slower response to command), loss of consciousness with clinical unconsciousness (i.e. no response to command), and recovery of consciousness. The total experimental procedure lasted about 5 hours, including both the positioning of the 256-electrode hd-EEG cap and the recovery time of participants.