Cerebral Cortex, February 2022;32: 298–311

https://doi.org/10.1093/cercor/bhab209
Advance Access Publication Date: 7 July 2021
Original Article

ORIGINAL ARTICLE

Macroscopic Quantities of Collective Brain Activity during Wakefulness and Anesthesia

Adrián Ponce-Alvarez[1], Lynn Uhrig[2], Nikolas Deco[1], Camilo M. Signorelli[1,2,3], Morten L. Kringelbach[3,4,5], Béchir Jarraya[2,6,7,8] and Gustavo Deco[1,9,10,11]

[1]Computational Neuroscience Group, Department of Information and Communication Technologies, Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona 08005, Spain, [2]Life Science Division, NeuroSpin Center, Institute of BioImaging Commissariat à l'Energie Atomique, Gif-sur-Yvette 91191, France, [3]Department of Psychiatry, University of Oxford, Oxford OX3 7JX, UK, [4]Department of Clinical Medicine, Center for Music in the Brain, Aarhus University, Aarhus 8000, Denmark, [5]Department of Neurosciences, Life and Health Sciences Research Institute, School of Medicine, University of Minho, Braga 4710-057, Portugal, [6]UniCog, INSERM, Gif-sur-Yvette 91191, France, [7]Université Paris-Saclay, UVSQ, Versailles 78000, France, [8]Neuromodulation Unit, Foch Hospital, Suresnes 92150, France, [9]Institució Catalana de la Recerca i Estudis Avançats (ICREA), Barcelona 08010, Spain, [10]Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig 04103, Germany and [11]School of Psychological Sciences, Monash University, Melbourne, VIC 3800, Australia

Address correspondence to Adrián Ponce-Alvarez, Computational Neuroscience Group, Pompeu Fabra University, C/Ramon Trias Fargas, 25-27, Mercè Rodoreda 24, Office 24.334, Barcelona 08005, Spain. Email: adrian.ponce@upf.edu

Abstract

The study of states of arousal is key to understand the principles of consciousness. Yet, how different brain states emerge from the collective activity of brain regions remains unknown. Here, we studied the fMRI brain activity of monkeys during wakefulness and anesthesia-induced loss of consciousness. We showed that the coupling between each brain region and the rest of the cortex provides an efficient statistic to classify the two brain states. Based on this and other statistics, we estimated maximum entropy models to derive collective, macroscopic properties that quantify the system's capabilities to produce work, to contain information, and to transmit it, which were all maximized in the awake state. The differences in these properties were consistent with a phase transition from critical dynamics in the awake state to supercritical dynamics in the anesthetized state. Moreover, information-theoretic measures identified those parameters that impacted the most the network dynamics. We found that changes in the state of consciousness primarily depended on changes in network couplings of insular, cingulate, and parietal cortices. Our findings suggest that the brain state transition underlying the loss of consciousness is predominantly driven by the uncoupling of specific brain regions from the rest of the network.

Key words: brain states, Collective activity, consciousness, fMR, Imaximum entropy models, phase transitions

Introduction

Interesting phenomena in biological systems are usually collective behaviors emerging from the interactions among many constituents. Large-scale brain activity is not an exception: The brain's network continuously generates coordinated spontaneous patterns of activity among brain regions at multiple spatiotemporal scales (Biswal et al. 1995; Fox and Raichle 2007; Chang and Glover 2010). Changes in spontaneous brain activity are observed in different brain states; the study of which is essential to understand the organizing principles of brain activity. For instance, anesthesia has been used to transiently induce loss of consciousness and to investigate the neural correlates of the awake state. Previous studies showed that different anesthetics, acting on different molecular targets (Alkire et al. 2008), similarly impact the strength and the structure of functional correlations (Greicius et al. 2008; Ferrarelli et al. 2010; Boly et al. 2012; Guldenmund et al. 2013; Barttfeld et al. 2015; Uhrig et al. 2016), and their dependence on interareal anatomical connections (Barttfeld et al. 2015; Uhrig et al. 2018). However, how changes in local regions and subnetworks combine to affect the collective brain dynamics and to lose consciousness remains largely unknown. To answer this question, it is essential to precisely characterize the collective properties of different brain states and their dependence on parameters at the system's level. This dependence is likely not straightforward since, as for many complex systems, the system's behavior could be differently affected by changes in its parameters. In such a case, while some parameters can largely vary without affecting the system's behavior (the so-called sloppy parameters), even small changes in some others can significantly modify it (Machta et al. 2013; Panas et al. 2015; Ponce-Alvarez et al. 2020).

In recent years, statistical mechanics has proven to be more and more useful to describe collective neural activity. Statistical mechanics shows that the behaviors of complex systems can be captured by macroscopic properties, which emerge from the collective activity of the units, in a way largely independent of the microscopic details of the system. These emergent (macroscopic) behaviors can be classified into qualitatively different ordered or disordered phases. Of particular interest are dynamics poised close to phase transitions, or critical points, where order and disorder coexist. Theoretical reasoning shows that complex dynamics and optimal information processing are expected at critical points, making criticality a candidate-unifying principle to account for the brain's inherent complexity necessary to process and represent its environment (Chialvo 2010; Shew et al. 2011; Shew and Plenz 2013; Hidalgo et al. 2015). Following this view, studies of whole-brain and local circuit dynamics have proposed that anesthesia shifts the dynamics from the critical point (Tagliazucchi et al. 2016; Fekete et al. 2018). This is supported by the reduction of several measures of brain dynamics complexity under anesthesia (Casali et al. 2013; Hudetz et al. 2015; Schartner et al. 2015; Solovey et al. 2015). The global mechanisms underlying different conscious states have been recently investigated using an anatomically constrained dynamical model with a global coupling parameter in combination with EEG recordings (Lee et al. 2019). However, it remains unknown which are the macroscopic properties and the relevant local/global parameters describing the transition of collective activity from the awake to anesthetized states. Indeed, different local/global network parameters are likely to jointly determine the different brain states and to differently contribute to the state transitions.

In this study, we addressed these questions by analyzing the brain's collective activity in different levels of arousal, that is, during wakefulness and under anesthesia. Specifically, we analyzed resting-state fMRI dynamics of awake and anesthetized macaque monkeys (Uhrig et al. 2018). Five different anesthesia protocols, involving three different anesthetics (propofol, ketamine, and sevoflurane), were used to induce moderate sedation or deep anesthesia. First, we derived efficient statistics that distinguished between awake and anesthetized brain states. Second, we used these statistics and the maximum entropy principle to model the brain's activity and to derive important emergent properties that described the different brain states. These emergent properties provided information about the system's physical state and about its capability to produce work, to contain information, and to transmit it. Finally, we investigated the dependence of collective activity on the different model parameters.

Materials and Methods

Animals

This study included a total of five rhesus macaques (*Macaca mulatta*; four females, one male, 5–8 kg; 8–12 years of age). All procedures were conducted in accordance with the European convention for animal care (86–406) and the National Institutes of Health's Guide for the Care and Use of Laboratory Animals. Animal studies were approved by the institutional Ethical Committee (Comité d'Ethique en Expérimentation Animale, protocols #10–003 and #12–086).

Experimental Procedures

Monkeys received anesthesia either with propofol, ketamine, or sevoflurane (Uhrig et al. 2018). The details of the anesthesia protocols are described in the Supplementary Information. Monkeys were scanned on a 3-T horizontal scanner (Siemens Tim Trio; TR, 2400 ms; TE, 20 ms; and 1.5-mm^3 voxel size; 500 brain volumes per scan session). Before each scanning session, a contrast agent monocrystalline iron oxide nanoparticle (MION) was injected into the monkey's saphenous vein. MION isolates the regional cerebral blood volume (rCBV) component of the functional activity, with higher contrast-to-noise ratio as compared with blood-oxygen-level-dependent (BOLD) imaging (Vanduffel et al. 2001). Acquisition and preprocessing of functional images followed the standard steps described in (Barttfeld et al. 2015) and in the Supplementary Information. Time series were obtained for $N = 82$ previously defined cortical regions of interest (ROIs) (CoCoMac Regional Map parcellation).

Data Binarization

fMRI time series were binarized to study the data statistics and to learn two different families of maximum entropy models (MEMs). While binarization was required to construct the MEMs, transformation of continuous fMRI signals into discrete point processes has proven to effectively capture and compress fMRI large-scale dynamics (Tagliazucchi et al. 2012). Importantly, the fluctuations that cross the threshold do not merely represent noise. Indeed, it has been shown that the point process resulting from signal thresholding largely overlaps with deconvoluted fMRI BOLD signals using the hemodynamic response function and preserves the topology of the resting-state networks (RSNs) (Tagliazucchi et al. 2012). We here discretized the signals as

follows. For each scan, the z-scored time series of each ROI, $x_i(t)$ ($1 \leq i \leq N$), was binarized by imposing a threshold $\theta = -1$. The threshold has opposite sign compared with the one used by Tagliazucchi et al. (2012) because, in contrast to the BOLD signal, functional activations lead to increased rCBV that manifests a decrease in MION-based fMRI signal (Schölvinck et al. 2010). Two binarization procedures were used. The first method detects the threshold crossings: The binarized activity is $\sigma_i(t) = 1$ if $x_i(t) < \theta$ and $x_i(t-1) > \theta$, and $\sigma_i(t) = 0$ otherwise. The second method assigns the values 1 and -1 to all time points below or above the threshold, respectively: $\sigma_i(t) = 1$ if $x_i(t) < \theta$, and $\sigma_i(t) = -1$ otherwise. The first and the second procedure result in sparse and dense binary activity, respectively. We used the sparse and dense methods to construct coupling-MEMs and pairwise-MEM, respectively. This was to meet the assumptions of the model inference (see Supplementary Information).

k-Means Classification

We used k-means clustering to classify the scans based on different statistics. Let $\boldsymbol{v}^{(i)}$ be a vector calculated from scan i, for example, the vector containing all pairwise correlations among the ROIs. We used k-means to partition the collection of $\boldsymbol{v}^{(i)}$ into a prespecified number (k) of clusters. k-means minimizes the within-cluster variation, over all clusters. We used $k = 2$ to evaluate how well the scans corresponding to the awake state and those corresponding to the anesthetized states (independent of the anesthetic protocol) could be classified based on vectors $\boldsymbol{v}^{(i)}$. To classify the six different experimental conditions, we used $k = 6$. The classification performance was given by the proportion of correctly clustered scans. We used 100 random initial conditions of the k-means algorithm to obtain the average classification performance and its uncertainty.

Maximum Entropy Models (MEMs)

MEMs estimate the probability distribution of all possible binary patterns, $P(\boldsymbol{\sigma})$, that matches the expectation of a set of data observables. Let $\mathcal{O}_1(\boldsymbol{\sigma}), \ldots, \mathcal{O}_L(\boldsymbol{\sigma})$ be the set of L data observables we seek to preserve. For example, if we were interested only on activation rates, $\langle \sigma_i \rangle$, we would need to consider N observables $\sigma_1, \ldots, \sigma_N$. Under the model distribution $P(\boldsymbol{\sigma})$, the observables' expectations are given as:

$$\langle \mathcal{O}_i \rangle_{\text{model}} = \sum_{\{\sigma\}} P(\boldsymbol{\sigma})\, \mathcal{O}_i(\boldsymbol{\sigma}), \tag{1}$$

and should fit those of the data, $\langle \mathcal{O}_i \rangle_{\text{data}} = \frac{1}{n}\sum_{t=1}^{n} \mathcal{O}_i(\boldsymbol{\sigma}^t)$, where $\boldsymbol{\sigma}^t$ is the observed pattern at time t ($1 \leq t \leq n$). We search for the model distribution $P(\boldsymbol{\sigma})$ that does fewer assumptions, that is, the one that has maximal entropy $S = \sum_{\{\sigma\}} P(\boldsymbol{\sigma}) \ln P(\boldsymbol{\sigma})$. Thus, the problem is equivalent to maximizing a function (the entropy) given some constraints on the expectation values of the observables, a problem that can be generally solved using Lagrange multipliers. The maximum entropy distribution has the general form:

$$P(\boldsymbol{\sigma}) = \frac{1}{Z} \exp\left(\sum_{i=1}^{L} \Omega_i \mathcal{O}_i(\boldsymbol{\sigma}) \right) = \frac{1}{Z} \exp\left(-E(\boldsymbol{\sigma}) \right), \tag{2}$$

Where $\boldsymbol{\Omega} = [\Omega_1, \ldots, \Omega_L]$ are the Lagrange multipliers enforcing the constraints, $E(\boldsymbol{\sigma}) = -\sum_{i=1}^{L} \Omega_i \mathcal{O}_i(\boldsymbol{\sigma})$ represents the energy of the pattern, and the normalizing factor $Z = \sum_{\{\sigma\}} \exp(-E(\boldsymbol{\sigma}))$ is the partition function (see Supplementary Information). We estimated different MEMs built on different constrained data observables.

Linear coupling-MEM

First, we considered the MEM that is consistent with the probability distribution $P(K)$, the average activations $\langle \sigma_i \rangle$, and the linear coupling between σ_i and K, that is, $\langle \sigma_i K \rangle$ (which relates to z_i). As shown in Gardella et al. (2016), the resulting energy function is given as: $E(\boldsymbol{\sigma}) = -\sum_{i=1}^{N} (\alpha_i + \beta_K + \gamma_i K)\sigma_i$. The model parameters $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_N]$, $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_N]$, and $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_N]$ are Lagrange multipliers associated to the constrained observables $\langle \sigma_i \rangle$, $P(K)$, and $\langle \sigma_i K \rangle$, respectively. The model has $3N + 1$ parameters.

Nonlinear coupling-MEM

The above model can be extended to include the nonlinear coupling between σ_i and K. The complete coupling between σ_i and K is provided by the joint probability distributions of σ_i and K, that is, $P(\sigma_i, K)$, which is the target observable of the nonlinear coupling-MEM. In this case, the energy is given as $E(\boldsymbol{\sigma}) = -\sum_{i=1}^{N} H_{i,K(\boldsymbol{\sigma})}\sigma_i$ (Gardella et al. 2016), where $K(\boldsymbol{\sigma})$ is the number of active ROIs in pattern $\boldsymbol{\sigma}$ and the parameters $H_{i,K(\boldsymbol{\sigma})}$ are associated to the constrained observables $\langle \sigma_i \delta_{K,k} \rangle$, where $P(K = k) = \langle \delta_{K,k} \rangle$ and $\delta_{K,k}$ is the Kronecker's delta. The parameters $H_{i,K(\boldsymbol{\sigma})}$ represent the tendency of ROI i to activate when $K(\boldsymbol{\sigma})$ ROIs are active. Relaxing the linearity assumption makes the model more complex, as it has $N(N+1)$ parameters. Moreover, in the nonlinear model, parameters jointly represent the interaction between K and σ_i, whereas the linear model allows to separate the terms related to global dynamics (K), local dynamics (σ_i), and their interaction. Indeed, the linear model is a special case of the nonlinear model with $H_{i,K(\boldsymbol{\sigma})} = \alpha_i + \beta_K + \gamma_i K$.

Pairwise-MEM

The third model we considered is the one that targets the activation rates ($< \sigma_i >$) and the pairwise correlations ($< \sigma_i \sigma_j >$) of the data. The resulting energy function of the maximum entropy distribution is given as $E(\boldsymbol{\sigma}) = -\sum_{i=1}^{N} h_i \sigma_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} J_{ij}\sigma_i\sigma_j$ (Schneidman et al. 2006; Tkačik et al. 2015). The model parameter h_i, called intrinsic bias, represents the intrinsic tendency of neuron i toward activation ($\sigma_i = +1$) or silence ($\sigma_i = -1$) and the parameter J_{ij} represents the effective interaction between neurons i and j.

The parameters of the coupling-MEMs and pairwise-MEMs were estimated from the data using likelihood (Gardella et al. 2016) and pseudolikelihood (Ezaki et al. 2017) maximization, respectively (see Supplementary Information).

Macroscopic Quantities

The analysis of the learned MEMs provides relevant properties of the collective activity. These quantities derive from the Boltzmann distribution and they are interpretable in the framework of statistical physics. The description and calculation of these quantities are presented in the Supplementary Information in detail. Briefly, we studied the system's Helmholtz *free energy*, *susceptibility*, and *heat capacity*. The free energy F is given by the difference between the average energy and the entropy, that is, $F = \langle E \rangle - S = -\ln(Z)$; it quantifies the useful energy that is obtainable from the system. The susceptibility χ relates to the diversity of population states, that is, $\chi = \text{var}(K)$, but, importantly, it also relates to the system's response to intrinsic

or external inputs (see Supplementary Information). The heat capacity C_h quantifies the diversity of accessible energy states, that is, $C_h = \mathrm{var}(E)$. The heat capacity measures the size of the dynamic repertoire of the system. Furthermore, a parameter T, which scales all model parameters ($\mathbf{\Omega} \rightarrow \mathbf{\Omega}/T$), can be introduced to study the effect of a change in the system's disorder ("temperature") on the repertoire of accessible energy states, that is, the function $C_h(T) = \mathrm{var}(E)/T^2$. This function is informative of the state of the system in terms of criticality: A maximum of the heat capacity close to $T_{\max} = 1$ suggests that the observed system is likely to be close to a critical state, whereas $T_{\max} < 1$ and $T_{\max} > 1$ indicate supercritical and subcritical dynamics, respectively (Tkačik et al. 2014, 2015; Hahn et al. 2017) (see Supplementary Information and Supplementary Fig. 7).

Fisher Information Matrix

We were interested in detecting which parameters have the strongest effect on the collective activity. For this, we studied the Fisher information matrix (FIM, noted $\mathbf{G}$) of the learned MEMs. The FIM represents the curvature of the log-likelihood of the model, $\log P(\boldsymbol{\sigma}|\mathbf{\Omega})$, with respect to the model parameters; that is, it quantifies the sensitivity of the model to changes in parameters. It is given as:

$$G_{kl} = -\sum_{\{\boldsymbol{\sigma}\}} P(\boldsymbol{\sigma}|\mathbf{\Omega}) \frac{\partial^2 \ln P(\boldsymbol{\sigma}|\mathbf{\Omega})}{\partial \Omega_k \partial \Omega_l}. \quad (3)$$

Where $1 \leq k, l \leq L$, where L is the number of parameters. As shown in Supplementary Information, the FIM is given by the second derivatives of the free energy:

$$G_{kl} = \frac{\partial^2 \ln Z}{\partial \Omega_k \partial \Omega_l} = -\frac{\partial^2 F}{\partial \Omega_k \partial \Omega_l}. \quad (4)$$

To quantify the sensitivity of the different parameters, we decomposed the FIM into eigenvectors, noted $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_L$, and measured the sensitivity of a given parameter i by its absolute contribution to the first eigenvector, that is, $|\nu_1(i)|$.

Statistical Analysis

We used one-way ANOVA followed by Tukey's post hoc analysis to compare the means of three or more distributions and Wilcoxon rank sum test to compare the medians of two distributions. We measured the dissimilarity between two distributions (i.e., data vs. model distribution) through the Jensen-Shannon divergence. Correlation matrices were analyzed using standard PCA. Statistical models (i.e., maximum entropy models) were estimated using likelihood and pseudolikelihood maximization.

Results

We analyzed the resting-state fMRI dynamics of five rhesus macaques (*Macaca mulatta*) under different levels of arousal: wakefulness ($n = 24$ scans), two levels of propofol sedation (moderate, MPP, $n = 21$, and deep, DPP, $n = 23$), sedation through ketamine (KETA, $n = 22$), and two types of sevoflurane anesthesia (SEV2, $n = 18$, and SEV4, $n = 11$) (see Materials and Methods and Supplementary Information). All the anesthesia protocols used here induced loss of consciousness (Uhrig et al. 2018; see also Supplementary Information). fMRI MION time series were obtained for $N = 82$ previously defined ROIs (CoCoMac Regional Map parcellation). Each scan was 20 min long and was acquired in time frames of 2.4 s (i.e., 500 time frames).

Coupling to Population Reliably Distinguished between Awake and Anesthetized Brain States

We were interested on collective patterns displayed among the N ROIs, for the six different experimental conditions. For this, we first binarized the z-scored time series of each ROI, $x_i(t)$, by imposing a threshold $\theta = -1$ (Fig. 1A,B, see Materials and Methods). Binarization of time series has proven to effectively capture and compress fMRI large-scale dynamics (Tagliazucchi et al. 2012; Watanabe et al. 2013). We concentrated on different statistics that described the resulting binary data: The activation rate of each ROI, that is, $\langle \sigma_i \rangle$, the correlation between ROIs, that is, $C_{ij} = \mathrm{corr}(\sigma_i, \sigma_j)$, and the population activity, that is, $K(t) = \sum_{i=1}^{N} \sigma_i(t)$ (Fig. 1C). We were particularly interested on the coupling of each ROI to the population activity, defined as:

$$z_i = \mathrm{corr}\left[\sigma_i(t), K_{\neq i}(t)\right], \quad (5)$$

Where $K_{\neq i}(t)$ is the sum activity of all ROIs but ROI i: $K_{\neq i}(t) = \sum_{j \neq i} \sigma_i(t)$. Recent findings showed that propofol anesthesia affects the coupling to global signal in humans and rats (Tanabe et al. 2020). In the following, we show that the statistics $\mathbf{z} = [z_1, \ldots, z_N]$ provides, with only N parameters, a compact description of the binary collective activity and can be used to classify the brain states.

The couplings to the population were highly predictive of the functional correlations (Fig. 2A–C). Indeed, the product $\eta_{ij} = z_i \times z_j$ highly correlated with the functional correlation (FC) between the fMRI time series of ROIs (i, j) (corr.: 0.65–0.78, $P < 0.001$). Moreover, we found that the vector $\mathbf{z}$ correlated across scans within the same experimental conditions, with the average correlation coefficient being equal to 0.3 ± 0.01 (Fig. 2D, *blue* distribution). This correlation was significantly higher ($P < 0.001$, $F_{(2,3533)} = 976.5$, one-way ANOVA followed by Tukey's post hoc analysis) than correlations obtained using the vectors representing the average activities and the correlations, that is, vectors $\boldsymbol{\mu} = [\langle \sigma_1 \rangle, \ldots, \langle \sigma_N \rangle]$ and $\boldsymbol{\rho} = [C_{1,2}, C_{1,3}, \ldots, C_{N-1,N}]$, respectively (Fig. 2D, *green* and *red* distributions, corr.: 0.06 ± 0.01 and 0.11 ± 0.01).

Furthermore, we found that the coupling to the population could be used to classify the awake and anesthesia states with high accuracy (Fig. 3A). We tested this by using a classifier based on k-means clustering (see Materials and Methods). Based on the statistic $\mathbf{z}$, we were able to classify the scans of two categories, awake versus anesthesia (independently of the anesthetic), with 96.6% of correct classifications (chance level: 50%). This classification performance was higher than the one obtained using the statistics $\boldsymbol{\mu}$ and $\boldsymbol{\rho}$, yielding 74.0% and 74.8% of correct classifications, respectively (Fig. 3B,C). Classification among the six experimental conditions yielded lower performances but was higher for statistic $\mathbf{z}$ than for $\boldsymbol{\mu}$ and $\boldsymbol{\rho}$: 38.3%, 33.2%, and 29.9%, respectively (chance level: 16.7%). Similar differences in classification performances for population couplings and functional correlations were obtained using continuous (not thresholded) signals (Supplementary Fig. 1). Altogether, these results show that the coupling to population is a reliable marker to distinguish between awake and anesthetized brain states.

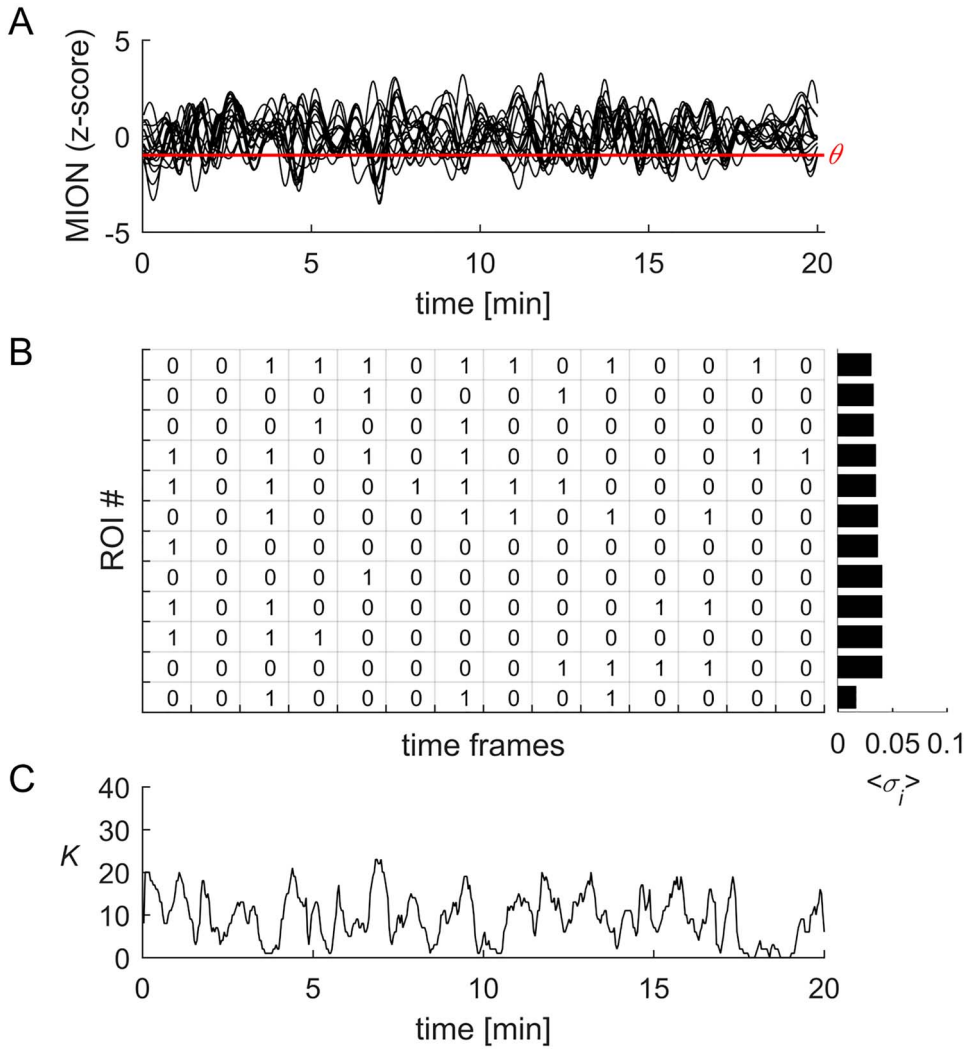

Figure 1. Binarization and statistics. (*A*) MION fMRI signals were z-scored and binarized by imposing a threshold equal to minus the standard deviation, for each signal (see Materials and Methods). (*B*) In each time bin of 2.4 s, the state of signal of ROI i, noted $\sigma_i(t)$, was equal to 1 if the MION signal for this ROI was lower than minus its standard deviation, or equal to 0 otherwise. The average activity of ROI i was $\langle\sigma_i\rangle = \frac{1}{n}\sum_{t=1}^{n}\sigma_i(t)$, where n is the number of time points. (*C*) The population activity was defined as the sum of the binary activity of the N ROIs in each time bin t, that is, $K(t) = \sum_{i=1}^{N}\sigma_i(t)$.

To examine which ROIs contributed the most to distinguish between the awake state and anesthesia based on $\mathbf{z}$, we performed PCA on the collection of z-scored vectors $\mathbf{z}$. The first principal component was sufficient to separate the awake and anesthesia conditions (Fig. 3*D*). This component had strong coefficients for brain regions located in the cingulate, parietal, intraparietal, insular cortices, and the hippocampus (Fig. 3*E*). Overall, changes in average couplings to the population with respect to awake values were similar for all anesthetics (Fig. 3*F*). We next asked how these changes affect the collective properties of brain dynamics.

Modeling Collective Activity Using Maximum Entropy Models

Collective activity is ultimately described by the probability of each of the binary patterns $\boldsymbol{\sigma} = [\sigma_1, \ldots, \sigma_N]$. Estimating the distribution $P(\boldsymbol{\sigma})$ over the 2^N possible binary patterns from the data is impractical with limited amount of observations, since for $N = 82$ there are more than 10^{24} possible patterns. A useful technique to estimate $P(\boldsymbol{\sigma})$ relies on the maximum entropy principle. Maximum entropy models (MEMs) find $P(\boldsymbol{\sigma})$ by maximizing its entropy under the constraint that some empirical statistics are preserved (see Materials and Methods). As shown above, an interesting statistic for the present study is the coupling between the state of each binary signal, σ_i, and the population activity K. The maximum entropy distribution that is consistent with the probability distribution $P(K)$, the average activations $\langle\sigma_i\rangle$, and the linear coupling between σ_i and K, that is, $\langle\sigma_i K\rangle$ (which relates to z_i), is given by the Boltzmann distribution $P(\boldsymbol{\sigma}) = e^{-E(\boldsymbol{\sigma})}/Z$, where $E(\boldsymbol{\sigma})$ represents the energy of the pattern $\boldsymbol{\sigma}$, given as (Gardella et al. 2016):

$$E(\boldsymbol{\sigma}) = -\sum_{i=1}^{N}(\alpha_i + \beta_K + \gamma_i K)\,\sigma_i. \tag{6}$$

The model parameters α_i, β_K, and γ_i are Lagrange multipliers associated to the constrained observables $\langle\sigma_i\rangle$, $P(K)$,

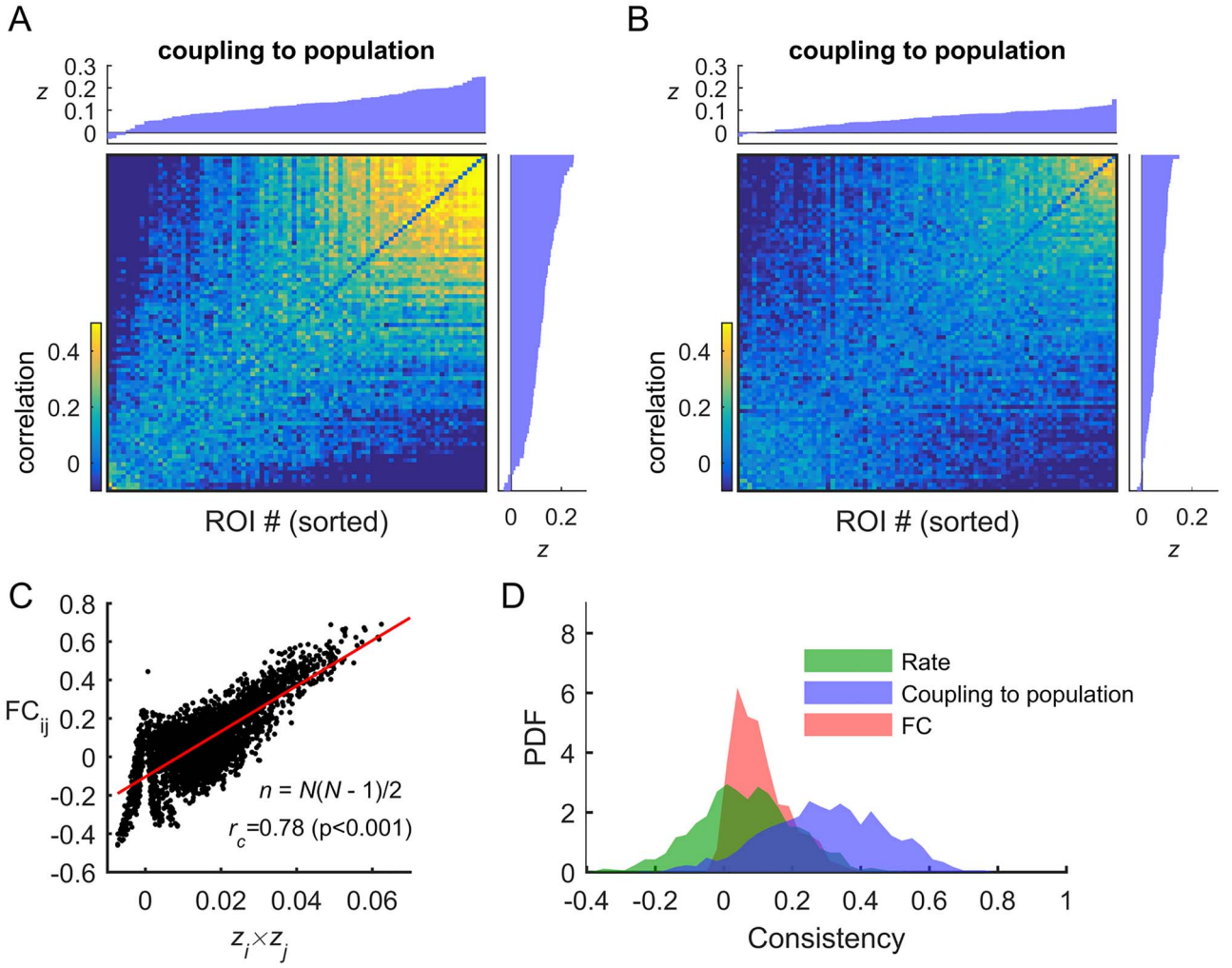

Figure 2. Coupling to population predicts the functional connectivity and is consistent within experimental conditions. (*A*) Relation between population couplings and pairwise correlations. *Color map*: The average functional connectivity (FC) is shown after ordering the ROIs according to z_i averaged for each scan within a given experimental condition (here for the awake condition). *Top and right insets*: ordered values of **z**. (*B*) Same as (*A*) but for the deep propofol (DPP) anesthesia condition. (*C*) Relation between the elements of the FC and the corresponding products of coupling to population ($z_i z_j$) (correlation: 0.78, $P < 0.001$). (*D*) Consistency of the different statistics. We tested whether $\langle\sigma\rangle$, z, and FC were similar across scans within the same experimental condition. For example, for the statistic z, we calculate the correlation of this *N*-dimensional variable for all pairs of scans belonging to the same experimental condition and computed the distribution of correlation coefficients (*blue* distribution). High correlation coefficients indicate that, within experimental conditions, scans yielded similar vector **z**. The same can be done for the *N*-dimensional variable $\langle\sigma\rangle$ (*green* distribution) and the vector of FC elements ($N(N-1)/2$ dimensions; *red* distribution).

and $\langle\sigma_i K\rangle$, respectively. The normalizing constant Z is the partition function, given by $Z = \sum_{\{\sigma\}} e^{-E(\sigma)}$, which contains information about useful statistics predicted by the model (see below). This model can be extended to include the nonlinear coupling between σ_i and *K*, that is, by targeting the joint probability distributions of σ_i and *K* (see Materials and Methods).

For both linear and nonlinear coupling-MEMs, the model parameters were inferred from the single-scan data using maximum likelihood (Gardella et al. 2016) (see Materials and Methods). Notably, for the coupling-MEMs, the partition function can be calculated directly—something that is generally not the case for most MEMs, since its calculation involves summing over all possible states.

We used these coupling-MEMs to fit the binary single-scan fMRI data for the different experimental conditions. The models accurately estimated the distribution of population activity *P*(*K*) (average Jensen-Shannon divergence D_{JS} between the model and data distributions: $D_{JS} < 10^{-6}$ for both the linear and nonlinear coupling-MEM; Fig. 4A and Supplementary Fig. 2). Moreover, the models were able to moderately predict the covariances of the data (Fig. 4B,C), which were not used to constrain the models. Across the different datasets, the average correlation between the data and predicted covariances was $r = 0.31 \pm 0.03$ for the linear coupling model and reached 0.44 ± 0.02 for the nonlinear coupling model (see also Supplementary Fig. 2). Furthermore, scan classification based on parameters γ_i yielded 86% and 45% correct classifications between awake and anesthetized conditions and among the six experimental conditions, respectively (Supplementary Fig. 3A,B). Using parameters α_i, the classifier performance decreased to 75% and 28%, respectively (Supplementary Fig. 3B). Thus, the learned linear coupling-MEM showed

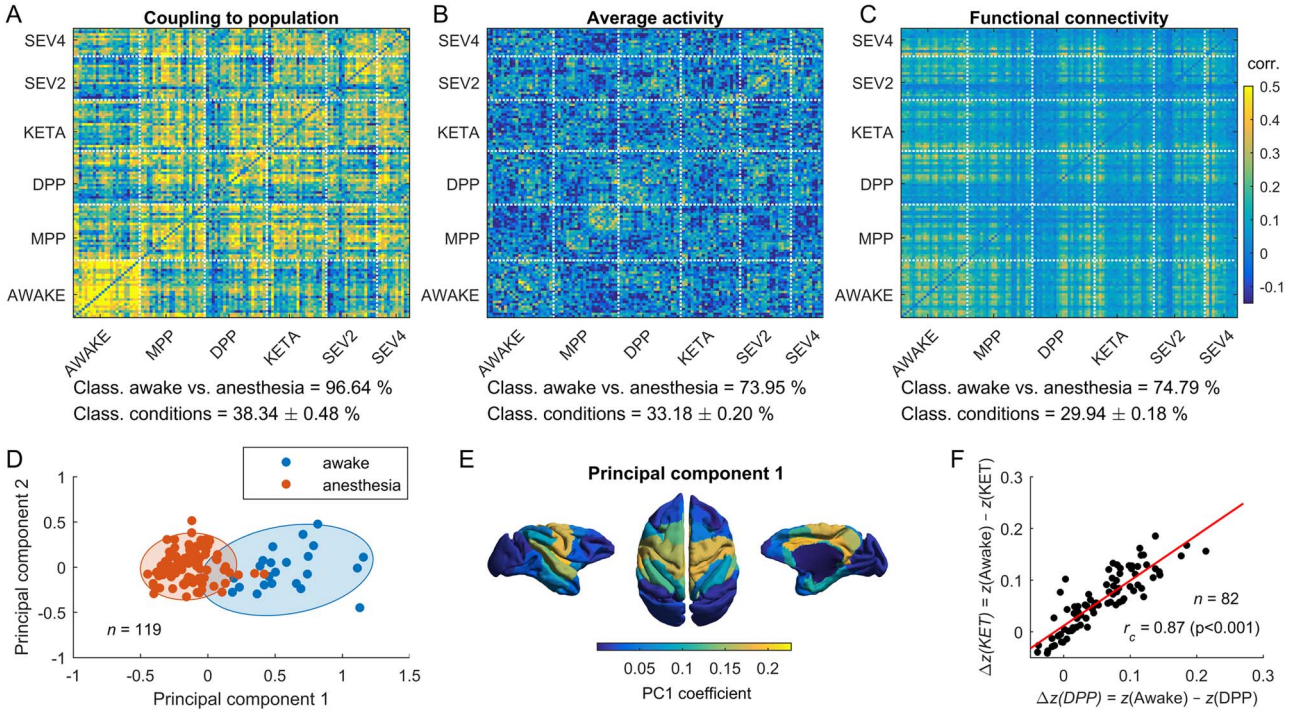

Figure 3. Coupling to population predicts the state of the brain. (*A*–*C*) Correlation matrix comparing the statistics z, $\langle\sigma\rangle$, and FC among all scans. For example, in panel (*A*), the element (k, l) of the matrix represents the correlation between the coupling to population vector $\mathbf{z}$ of scans k and l. Coupling to population clearly separated awake and anesthesia data. Using k-means, we evaluated how well the different statistics could be used to classify the awake and anesthetized conditions (chance level: 50%). The classification performance using the coupling to population statistic was 96.64%, which was significantly higher than using the mean activity (73.95%) or the functional connectivity (74.79%). Classification of the six experimental conditions was generally lower, but higher for z than for $\langle\sigma\rangle$ and FC (38% vs. 33% and 29%, chance level: 16.67%). (*D*) PCA analysis showed that z vectors separated the awake and anesthetized conditions along the first principal component (PC1). Each dot represents a scan. (*E*) The absolute coefficient of PC1 associated to each ROI. (*F*) During anesthesia, z was reduced compared with wakefulness for most of the ROIs. Changes from awake baseline, $\Delta z(i) = z(\text{Awake}) - z(i)$, where z was averaged over scans, were highly correlated for the different anesthetics (with correlation coefficients ranging from 0.85 to 0.93). The panel shows the comparison between Δz(DPP) and Δz(KET).

consistent variations in parameters γ_i (associated to z_i) across the different arousal states.

Collective Activity Indicated Reduced Free Energy, Susceptibility, and Heat Capacity under Anesthesia

We can learn interesting features of collective activity using the estimated models. One important quantity is the system's Helmholtz *free energy*, which is given by the difference between the average energy ($\langle E\rangle$) and the entropy (S), that is, $F = \langle E\rangle - S$. The free energy quantifies the useful energy that is obtainable from the system. Using the Boltzmann distribution, the free energy can be directly obtained from the partition function as $F = -\ln(Z)$. Thus, since Z is tractable for the coupling-MEMs, we can directly estimate F. We found that the free energy was significantly higher for the awake state compared with all anesthetized conditions for both the linear (Fig. 4*D*) and the nonlinear (Supplementary Fig. 4*A*) coupling-MEMs ($P < 0.001$, one-way ANOVA followed by Tukey's post hoc analysis). This result is both interesting and reasonable because it indicates that more useful energy can be extracted from the awake state than from the anesthetized state.

Two other important statistical quantities can be derived from the model, namely the "susceptibility" and the "heat capacity". The susceptibility χ relates to the diversity of population states, while the heat capacity C_h quantifies the diversity of accessible energy states. Specifically, the susceptibility and the heat capacity of the model are given by the variances of the population activity and the energy, respectively, that is, $\chi = \text{var}(K)$ and $C_h = \text{var}(E)$. We found that χ and C_h were significantly higher for the awake state compared with all anesthetized conditions for both linear and nonlinear coupling-MEMs (linear model: Fig. 4*E*,*F*, nonlinear model: Supplementary Fig. 4*B*,*C*; $P < 0.001$, one-way ANOVA followed by Tukey's post hoc analysis). This indicates that the awake system had larger population fluctuations and a larger repertoire of energy states than the system under anesthesia.

We next extend these results by estimating MEMs constrained by pairwise correlations (Schneidman et al. 2006; Tkačik et al. 2015). Pairwise-MEMs can be mapped to Ising models, allowing to assess the state of the observed system in terms of criticality and to study its response to external stimuli. To build the models, we estimated the maximum entropy distribution $P(\boldsymbol{\sigma})$ under the constraint that the activation rates ($< \sigma_i >$) and the pairwise correlations ($< \sigma_i\sigma_j >$) are preserved. The energy of the Boltzmann distribution that is consistent with these expectation values is given by (Schneidman et al. 2006; Tkačik et al. 2015):

$$E(\boldsymbol{\sigma}) = -\sum_{i=1}^{N} h_i\sigma_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} J_{ij}\sigma_i\sigma_j. \quad (7)$$

In this pairwise-MEM, the parameter h_i, called intrinsic bias, represents the intrinsic tendency of ROI i toward activation

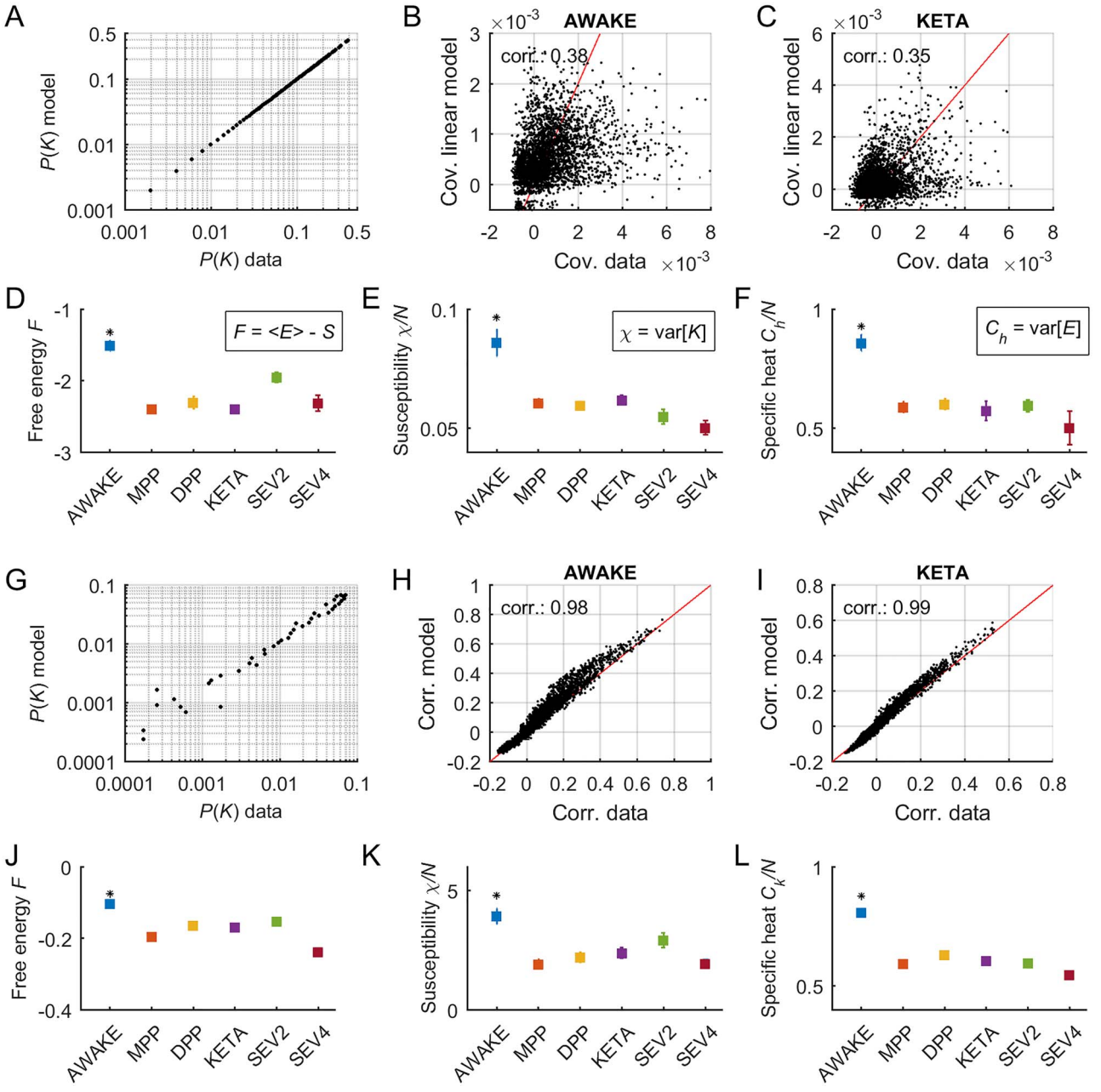

Figure 4. Maximum entropy models indicate higher free energy and heat capacity during wakefulness than during anesthesia. (*A*) Fitting of *P*(*K*) using the linear coupling-MEM. Data and predictions from all scans from the awake condition. (*B*, *C*) Fitting of covariances using the linear coupling-MEM for the awake (*B*) and ketamine (*C*) conditions. (*D*–*F*) The free energy, the susceptibility, and the heat capacity were derived using linear coupling-MEMs for the different conditions. Similar results were obtained using the nonlinear model (see Supplementary Fig. 4). Squares and error bars indicate means and standard deviations across scans, respectively, and the asterisks indicate significantly different values for the awake condition ($P < 0.001$ one-way ANOVA followed by Tukey's post hoc analysis). (*G*–*L*) same as (*A*–*F*) but using pairwise-MEMs. Error bars indicate standard errors across Monte Carlo simulations of the models. Asterisks indicate significantly different values for the awake condition ($P < 0.001$, one-way ANOVA followed by Tukey's post hoc analysis).

or silence and the parameter J_{ij} represents the effective interaction between ROIs i and j. The estimation of the model parameters $\mathbf{\Omega} = \{\boldsymbol{h}, \boldsymbol{J}\}$ was achieved through a pseudolikelihood maximization (Ezaki et al. 2017) (see Materials and Methods). Since this model requires the precise estimation of $< \sigma_i \sigma_j >$, it cannot be fitted to single-scan data and, for this reason, we used concatenated data from each experimental condition. The pairwise-MEM accurately predicted the observed correlations and, to a lower extend, it predicted the distribution of population activity *P*(*K*) (average correlation fit: $r = 0.985 \pm 0.002$; average $D_{JS} = 0.006 \pm 0.002$; Fig. 4*G*–*I* and Supplementary Fig. 2)—this is expected since *P*(*K*) was not used to constrain the model. We found that biases and coupling parameters were changed for different states, with some parameters increasing or decreasing, and with a significant reduction of the variance of couplings in the anesthetized states (Supplementary Fig. 5*A*–*D*).

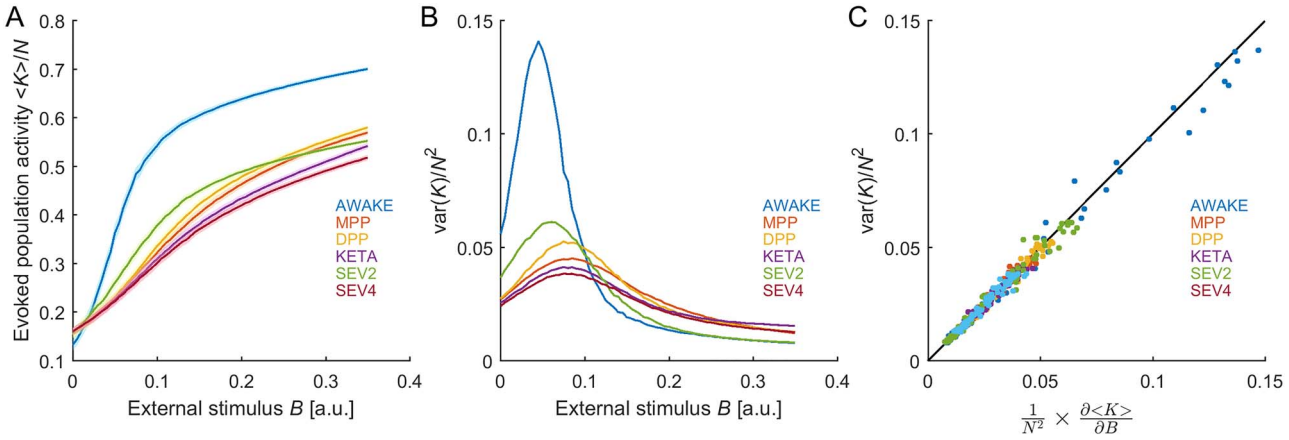

Figure 5. Population response to an external stimulus predicted by the pairwise-MEMs. (*A*) A global external stimulus was applied to the learned pairwise-MEMs. The stimulus *B* added a term to the energy as: $E(\sigma) + \sum_{i=1}^{N} B\sigma_i$. We performed Monte Carlo simulations (100 trials of 5.10^4 steps) for different values of *B* to examine the mean population activity $\langle K \rangle$ as a function of the external stimulus. In this case, the work *W* produced by changing the external stimulus from B_1 to B_2, that is, $W = -\int_{B_1}^{B_2} \langle K \rangle dB$, relates to the variation of the free energy ΔF (see Supplementary Information). (*B*) Variance of the population activity. In these simulations, inactive ROIs ($\sigma_i = -1$) were set to 0, so that *K* represents the number of active ROIs. (*C*) As shown in equation (S4) in the Supplementary Information, the derivative of the population response is equal to the variance of population activity, that is, $\chi = \partial \langle K \rangle / \partial B = \mathrm{var}(K)$. Thus, the higher the population response to a stimulus, the higher the variability of the population activity.

Moreover, coupling parameters (*J*) showed a higher correlation with the anatomical connectivity (or brain connectome) in the anesthetized states than in the awake state (Supplementary Fig. 5E).

Using this model, we calculated the collective statistical quantities for the different experimental conditions. Since in the pairwise-MEM the partition function is not tractable, we calculated *F*, χ, and C_h using Monte Carlo simulations (see Materials and Methods). Consistent with the above results obtained for coupling-MEMs, we found that the awake system had larger available energy (free energy, Fig. 4J, see also Supplementary Fig. 6), larger population fluctuations (susceptibility, Fig. 4K), and larger repertoire of states (heat capacity, Fig. 4L) than the system under anesthesia. Thus, the different versions of the MEM used here indicate the same results concerning the statistical properties of awake and anesthetized states.

As shown in detail in the Supplementary Information, the susceptibility can be viewed as a measure of the network response to a stimulus. Indeed, if an external global perturbation *B* adds a term to the energy, that is, $E(\boldsymbol{\sigma}) + \sum_{i=1}^{N} B\sigma_i$, the susceptibility is given by the derivative of $\langle K \rangle$ with respect to *B*, that is, $\chi = \partial \langle K \rangle / \partial B$ (Fig. 5A–C). It can be shown that $\chi = \partial \langle K \rangle / \partial B = \mathrm{var}(K)$ (see Supplementary Information); thus, the higher the network response to a variation of the external stimulus, the higher its variability (Fig. 5C). We found that application of an external stimulus elicited larger and more diverse responses for the pairwise-MEM corresponding to the awake state than for the models corresponding to the anesthetized states.

Awake Collective Activity Displayed Critical Dynamics that Were Shifted to a Supercritical Regime under Anesthesia

The pairwise-MEM can be used to assess the physical state of the system. Indeed, by introducing a scaling parameter *T*, analogous to the temperature in statistical physics, one can obtain relevant features of the collective dynamics. For this, we scaled all model parameters as $\boldsymbol{\Omega} \rightarrow \boldsymbol{\Omega}/T$ and calculated the heat capacity as a function of *T*, given by $C_h(T) = \mathrm{var}[E]/T^2$. The "temperature" *T* controls the level of disorder and its effects can be understood by examining the system's energy levels (Supplementary Fig. 7). Briefly, at low temperatures, interactions dominate over fluctuations making the system predominantly silent and ordered. In contrast, at high temperatures, the system is disordered and relatively uncoupled because fluctuations dominate over interactions. Both low and high temperatures lead to a low C_h. However, for a specific temperature T_{max}, order and disorder coexist in the system and C_h is maximal as expected for critical dynamics (Tkačik et al. 2014; Hahn et al. 2017). Thus, a maximal heat capacity at $T_{max} = 1$ (corresponding to the model learned from the data) suggests that the system operated close to a critical state (whereas $T_{max} < 1$ and $T_{max} > 1$ indicates supercritical and subcritical dynamics, respectively).

We found that the heat capacity curve was maximal for a temperature equal to 1 for the awake state, while it peaked at $T_{max} < 1$ for the anesthetized conditions (Fig. 6). These results suggest that the awake state displayed critical dynamics, while dynamics under anesthesia were supercritical, which indicates that the anesthetics had a disconnection effect.

Couplings to Population Relate to the Sensitive Parameters of the System

We next evaluated how the different parameters affected the model's collective behavior. In general, changes in parameters can differently affect the system's behavior, with some parameters (called "stiff" parameters) effectively modifying it, while others have little effect on it ("sloppy" parameters) (Machta et al. 2013). We used an information-theoretical approach based on the Fisher Information Matrix (FIM, noted **G**) to detect the parameters that have a strong effect on the collective activity (see Materials and Methods). The FIM measures the change in the model log-likelihood $P(\boldsymbol{\sigma}|\boldsymbol{\Omega})$ with respect to changes in the model parameters $\boldsymbol{\Omega}$. As demonstrated in the Supplementary Information, the FIM relates to the second derivatives of the free energy with respect to the model parameters, that is, $G_{ij} = -\frac{\partial^2 F}{\partial \Omega_i \partial \Omega_i}$.

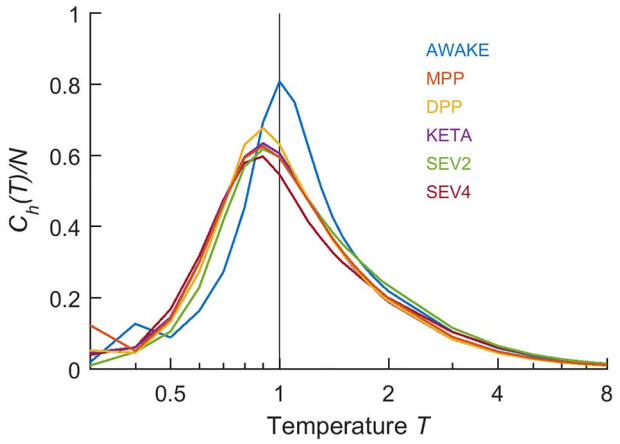

Figure 6. Heat capacity as a function of temperature. Heat capacity curves $C_h(T)$ were calculated for varying temperature parameter T for each condition. The peak of heat capacity for $T = 1$ indicates critical dynamics during wakefulness. The heat capacity peaked at $T < 1$ for the anesthetized conditions, indicating supercritical dynamics during anesthetized states.

This relation provides a direct link between a macroscopic quantity (the free energy) describing the collective dynamics of the different brain states and the underlying model parameters. For the linear model, the parameters that contributed the most to the FIM were the parameters γ_i (Fig. 7A). This explains how changes in couplings to population, as observed between awake and anesthetized states, effectively change the collective state of the system, leading to the observed shift from critical to supercritical dynamics.

To evaluate the importance of each of the parameters, we defined the parameter's sensitivity as its absolute contribution to the first eigenvector of the FIM (see Materials and Methods). The regions with the largest associated sensitivity for parameter γ were located in the cingulate, parietal, and insular cortices (Fig. 7B). Those that contributed the least were visual and prefrontal cortices. Interestingly, the regions presenting larger reductions of γ between awake and anesthesia tended to be those with higher associated sensitivity (corr: 0.74, $P < 0.001$; Fig. 7C).

We further examined how changes in pairwise correlations between awake and anesthesia related to changes in parameters of different sensitivity. We analyzed the average difference of correlation (ΔC) between awake and anesthesia. The matrix ΔC was decomposed into eigenvectors. Two groups of ROIs were clearly separated according to their positive or negative contribution to the first eigenvector of the matrix ΔC, respectively (Fig. 7D). Those that contributed positively were prefrontal and visual cortices, and those that contributed negatively were the cingulate, parietal, and insular cortices (Fig. 7E). Both groups presented a reduction of correlations under anesthesia; however, we found that the ROIs belonging to different groups were related to parameters that differently impacted the network dynamics. Indeed, prefrontal and visual cortices were related to parameters of significantly lower sensitivity than cingulate, parietal, and insular cortices (Fig. 7F,G, $P < 0.001$, Wilcoxon rank sum test). Hence, although prefrontal and visual areas changed their correlations, these changes were related to parameters that had a low impact on collective dynamics.

Finally, we found that the sensitivity of the ROIs was predictive of the response of the entire network to a local external stimulation (see Supplementary Fig. 8). Indeed, when externally stimulating single ROIs, we found the network response increases with the associated sensitivity of the locally stimulated ROI.

Discussion

In this study, we analyzed the fMRI binary collective activity of monkeys during wakefulness and under anesthesia. We showed that the coupling between each brain region and the rest of the population provides an efficient statistic that discriminates between awake and anesthetized states. We built MEMs based on this and other statistics to derive macroscopic properties that described the different brain states, such as the free energy F, the susceptibility χ, and the heat capacity C_h. All these quantities were maximized in the awake state. By studying the heat capacity curve $C_h(T)$ as a function of a scaling parameter controlling the disorder of the system, we showed that awake critical dynamics were shifted to supercritical ones under anesthesia. Finally, using the FIM, we showed that changes in brain state were primarily dependent on changes in the couplings to population that were associated to sensitive model parameters and to specific brain regions.

Population Couplings and Network Sensitivity

Previous research at the microcircuit level showed that neurons differ in their coupling to the population activity, with neurons that activate most often when many others are active and neurons that tend to activate more frequently when others are silent (Okun et al. 2015). Using the FIM analysis to detect sloppy and stiff parameters, it has been shown that these different types of neurons have a different impact on the network activity, different stimulus response properties, and different involvement in cortical state transitions (Ponce-Alvarez et al. 2020). Similar to these previous observations at the microcircuit level, we here showed that brain regions coupled differently to the rest of the whole-brain network, that these couplings primarily determined the collective activity (i.e., they were associated to the stiff parameters of the model), and that they were consistently different during wakefulness and anesthesia. Notably, the couplings with largest associated sensitivity, or "stiffness," were those that changed the most between brain states (Fig. 7C). Thus, we proposed that population couplings and their associated sensitivity are key to understand state transitions of neural activity at different scales.

Using principal components, we detected the combination of ROIs that contributed the most to distinguish between the awake and the anesthetized states based on their population couplings. The brain regions that changed their population coupling from the awake state to the anesthetized state were the cingulate, parietal, and insular cortices (Fig. 3E). Notably, the model parameters associated to the couplings of these regions were among those impacting the most the collective dynamics (Fig. 7B,C). Our results suggest that anesthesia modified some important local/global parameters that effectively induced a change of brain state. Our results highlight the key role of the parietocingulate cortex in the mechanism of anesthesia-induced loss of consciousness. Previous studies have shown that the parietal cortex (Kaisti et al. 2002; Uhrig et al. 2016) and the cingulate cortex (Luppi et al. 2019) are most strongly affected by anesthetics. These cortices also present alterations in brain injury–induced unconsciousness in humans (Juengling et al.

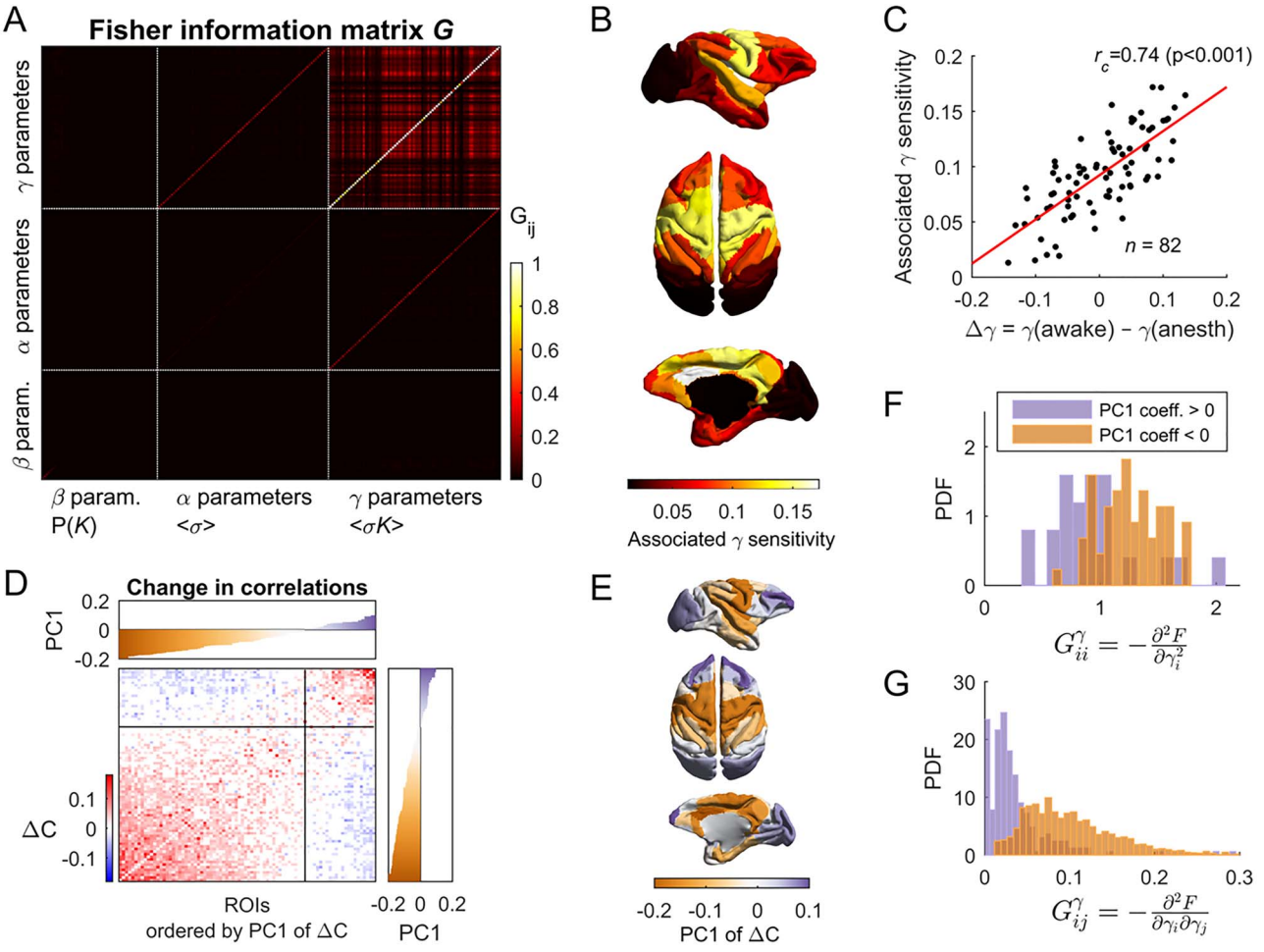

Figure 7. Model sensitivity to the different parameters. (A) Fisher information matrix (FIM) calculated using the linear coupling-MEM built using the data concatenated across scans from the awake condition. The FIM measures how much the model log-likelihood changes with respect to changes in the model parameters, that is, $\mathbf{\Omega} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}\}$ for the linear coupling-MEM. (*B*) Parameter sensitivity associated to population couplings. For each scan, we decomposed the FIM in eigenvectors and the mean contribution to first eigenvector of parameters γ was represented on the brain's image. This represents the sensitivity of collective activity on the parameters γ associated to the different ROIs. (*C*) The ROIs with larger reduction of parameter γ in anesthetized states with respect to the awake state were those with strongest sensitivity. (*D*) Change of pairwise correlations between awake and anesthetized states: $\Delta C = C_{\text{awake}} - \langle C_{\text{anesth}} \rangle$. ROIs were ordered according to their contribution to the first eigenvector of ΔC (top and right insets). Two groups of ROIs were detected according to their positive (labeled in *blue* in the insets) or negative (labeled in *brown* in the insets) contribution to this eigenvector, respectively, with both groups reducing the correlation between awake and anesthesia (*red* colors indicate pairs of nodes for which $\Delta C_{ij} > 0$; *blue* colors indicate pairs of nodes for which $\Delta C_{ij} < 0$). (*E*) First eigenvector of ΔC represented in the brain; ROIs with positive contribution to the first eigenvector are labeled in *blue*, those with negative contribution to it are labeled in *brown*. (*F*, *G*) The two groups of ROIs had significantly different associated sensitivity ($P < 0.001$, Wilcoxon rank sum test), as measured by the FIM values associated to parameters γ.

2005; Luppi et al. 2019). Moreover, consistent with our results, it has been shown that the insula plays an important role in awareness and is a potential neural correlate of consciousness (Craig 2009).

Interestingly, although some brain regions, such as visual and prefrontal cortices, had different correlations between awake and anesthesia, they were associated to parameters with low impact on collective activity. This highlights the importance of studying not only the change in statistics between brain states but also their sensitivity on network dynamics. Consistent with our findings, a recent study of neuronal activity from several brain regions and in different arousal states (Afrasiabi et al. 2021) shows that parietobasal ganglia circuits predicted the state of consciousness, while prefrontal activity failed. In addition, it has been proposed that the prefrontal cortex is mostly involved in the report of consciousness, rather than in the conscious experiences *per se* (Storm et al. 2017).

Macroscopic "Thermodynamic" Quantities

Using the MEMs, we learned interesting collective properties describing the different brain states. We measured the susceptibility that quantifies the diversity of spontaneous population fluctuations. The susceptibility can be also viewed as a measure of the network response to a vanishing global stimulus (Fig. 5, see also the Supplementary Information). Thus, the higher susceptibility observed in the awake state, compared with the anesthetized states, is consistent with transcranial magnetic stimulation (TMS) studies showing that stimulation elicits a more diverse and complex response in the awake state than in

low-level states of consciousness, such as sleep, anesthesia, and coma (Massimini et al. 2012; Napolitani et al. 2014; Sarasso et al. 2014, 2015). Moreover, our study predicts that localized stimulation of ROIs with strong associated parameter sensitivity would elicit a large network response (Supplementary Fig. 8). This model prediction could be experimentally tested by combining intracranial stimulation and the present FIM analysis.

The models also allowed the estimation of the system's heat capacity, a measure that quantifies the extent of the accessible dynamical repertoire. Indeed, a maximal heat capacity not only indicates that the system can display a large number of energy states but also that these energy states are distinguishable (Supplementary Fig. 7). Thus, a large heat capacity indicates a large capacity to represent information in numerous separable states. The observed reduction of heat capacity in the anesthetized states is consistent with previous studies showing that the repertoire of correlation states is limited during anesthesia (Uhrig et al. 2018) and maximized in conscious states (Guevara Erra et al. 2016; Perez Velazquez et al. 2019). Furthermore, by varying a scaling parameter analogous to temperature, the resulting heat capacity curves suggest that awake dynamics were critical, while anesthetized dynamics were supercritical, consistent with previous predictions (Tagliazucchi et al. 2016; Fekete et al. 2018). The model used here gives an intuitive interpretation of the transition between critical to supercritical dynamics. Indeed, in the pairwise-MEM, supercritical dynamics are associated with a regime in which random fluctuations dominate over interactions, which is consistent with a disconnection of *effective* couplings. It is important to note that the scale parameter T is only introduced to assess the state, that is, subcritical, critical, or supercritical, of the observed system (the one given for $T = 1$, for which the pairwise-MEM fits the data). This does not mean that differences between awake and anesthetized states are due to a global reduction of interactions and biases; instead, different arousal states yielded different biases and couplings (Supplementary Fig. 5), which, in combination, resulted in a change of the system's state. This means that the anesthesia reconfigured the system and not only scaled its parameters. Interestingly, we observed a significant reduction of the variance of pairwise couplings in the anesthetized states (Supplementary Fig. 5*A–D*), which is consistent with previous work showing that the number configurations of functional connections are maximized in conscious states (Guevara Erra et al. 2016; Perez Velazquez et al. 2019). Moreover, we found that effective couplings correlated more with the anatomical connections for the anesthetized states than for the awake state (Supplementary Fig. 5*E*), an effect that has been observed in empirical data (Barttfeld et al. 2015; Uhrig et al. 2018) and cannot be explained by changes in global connectivity alone (Lee et al. 2019).

Lastly, we measured the Helmholtz free energy of the estimated models. The free energy measures the useful energy that can be extracted from the system to the environment, that is, its ability to produce work. Reasonably, the awake state led to higher free energy than the anesthetized states. Another important property of free energy is that its change with respect to the model parameters is equal to the Fisher information and, thus, it relates to the sensitivity of collective dynamics on these parameters. This result provides a direct link between the sensitivity of parameters and the change of a macroscopic quantity, the free energy, the behavior of which is known to characterize the phase transition (Ginzburg and Landau 1965). For the linear coupling-MEM, we showed that the couplings to population (z) were associated to the parameters that have the strongest impact on collective activity. Consistently, we found that z was an efficient observable to classify the arousal states that collective dynamics were qualitatively different (in terms of criticality and supercriticality). Thus, these results give a coherent theoretical justification of the relevance of the statistic z to characterize the brain states and to estimate their free energy. Altogether, our findings represent a significant step in the understanding of brain states, resulting in a coherent explanation of the transition from awake to anesthesia: The phase transition between brain states is driven by those parameters that change the free energy, which are the "stiff" parameters of the systems and which relate to population couplings.

Implications for Studies on Pathological Low-Level States of Consciousness

The present study reports different fMRI statistics and derived model properties for awake and anesthesia states. Notably, we obtained similar results for the different anesthetic agents used here. Specifically, we found that population couplings in the different anesthesia conditions were very homogeneous (Fig. 3 and Supplementary Fig. 1), that is, fairly independent of the anesthesia protocol. This homogeneity was also observed in the derived macroscopic quantities (Fig. 4*D–F* and *J–M*). Since all the anesthesia protocols used here induced loss of consciousness (Uhrig et al. 2018; see also Supplementary Information), this suggests that our findings relate to the state of (un)consciousness rather than to the specific anesthetic agents.

An interesting extension of this work could be to study brain dynamics in coma using the present statistical mechanics framework. Loss of consciousness due to anesthesia or coma shares common features: Complexity of dynamics and neural communication are generally reduced in low-level states of consciousness (Sitt et al. 2014; Schartner et al. 2015). Consequently, estimates of complexity of human brain activity have been used to assess the depth of anesthesia (Zhang et al. 2001; Singh et al. 2017) and to predict the recovery of consciousness in vegetative patients (Sarà et al. 2011). Reduction of complexity is consistent with a deviation from critical dynamics when consciousness is lost. Deviations from criticality can be quantified using pairwise MEMs, although these models can only be estimated using concatenated data from multiple scans. On the other hand, an important advantage of coupling-MEMs is that they can be fitted to single scans and, thus, they can evaluate interindividual differences by analyzing the resulting parameters. Moreover, the tractability of the partition function of these models allows to directly calculate the free energy. We believe that combination of the models presented here, together with the measure z, represents promising tools to study different disorders of consciousness.

Supplementary Material

Supplementary material can be found at *Cerebral Cortex* online.

Author Contributions

A.P.A. designed research, analyzed the data, studied the models, and wrote the manuscript; L.U. performed the experiments and curated the data; N.D. studied the implementation of the

models; C.M.S. curated data; M.K. analyzed the data and provided data visualization codes; B.J. designed and supervised the experiments; G.D. designed and supervised research. All authors discussed the results and contributed to the editing of the manuscript.

Notes

A.P.A., B.J., and G.D. received funding from the FLAG-ERA JTC (PCI2018-092891). G.D. acknowledges funding from the European Union's Horizon 2020 FET Flagship Human Brain Project under Grant Agreement 785907 HBP SGA1, SGA2, and SGA3, the Spanish Ministry Research Project PSI2016-75688-P (AEI/FEDER), the Catalan Research Group Support 2017 SGR 1545, and AWAKENING (PID2019-105772GB-I00, AEI FEDER EU) funded by the Spanish Ministry of Science, Innovation and Universities (MCIU), State Research Agency (AEI) and European Regional Development Funds (FEDER). M.L.K. is supported by the ERC Consolidator Grant: CAREGIVING (n. 615539), Center for Music in the Brain, funded by the Danish National Research Foundation (DNRF117), and Centre for Eudaimonia and Human Flourishing funded by the Pettit and Carlsberg Foundations. BJ received funding from Fondation Bettencourt-Schueller, Université Paris-Saclay (UVSQ), Fondation de France, Collège de France and from INSERM. CMS was supported by Comisión Nacional de Investigación Ciencia y Tecnología (CONICYT, currently ANID) through Programa Formacion de Capital Avanzado (PFCHA), Doctoral scholarship Becas Chile: CONICYT PFCHA/DOCTORADO BECAS CHILE/2016—72170507. *Conflict of interests*: No competing interests declared.

References

Afrasiabi M, Redinbaugh MJ, Phillips JM, Kambi NA, Mohanta S, Raz A, Haun AM, Saalmann YB. 2021. Consciousness depends on integration between parietal cortex, striatum, and thalamus. *Cell Syst*. 12(4):363–373.

Alkire MT, Hudetz AG, Tononi G. 2008. Consciousness and anesthesia. *Science*. 322:876–880.

Barttfeld P, Uhrig L, Sitt JD, Sigman M, Jarraya B, Dehaene S. 2015. Signature of consciousness in the dynamics of resting-state brain activity. *Proc Natl Acad Sci USA*. 112:887–892.

Biswal B, Yetkin FZ, Haughton VM, Hyde JS. 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn Reason Med*. 34(4): 537–541.

Boly M, Moran R, Murphy M, Boveroux P, Bruno M-A, Noirhomme Q, Ledoux D, Bonhomme V, Brichant J-F, Tononi G, et al. 2012. Connectivity changes underlying spectral EEG changes during propofol-induced loss of consciousness. *J Neurosci*. 32:7082–7090.

Casali AG, Gosseries O, Rosanova M, Boly M, Sarasso S, Casali KR, Casarotto S, Bruno M-A, Laureys S, Tononi G, et al. 2013. A theoretically based index of consciousness independent of sensory processing and behavior. *Sci Transl Med*. 5: 198ra105.

Chang C, Glover GH. 2010. Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage*. 50:81–98.

Chialvo D. 2010. Emergent complex neural dynamics: the brain at the edge. *Nat Phys*. 6:744–750.

Craig AD. 2009. How do you feel — now? The anterior insula and human awareness. *Nature Reviews Neuroscience*. 10:59–70.

Ezaki T, Watanabe T, Ohzeki M, Masuda N. 2017. Energy landscape analysis of neuroimaging data. *Philos Trans R Soc A*. 375:20160287.

Fekete T, Omer DB, O'Hashi K, Grinvald A, van Leeuwen C, Shriki O. 2018. Critical dynamics, anesthesia and information integration: lessons from multi-scale criticality analysis of voltage imaging data. *Neuro Image*. 183:919–933.

Ferrarelli F, Massimini M, Sarasso S, Casali A, Riedner BA, Angelini G, Tononi G, Pearce RA. 2010. Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness. *Proc Natl Acad Sci USA*. 107:2681–2686.

Fox MD, Raichle ME. 2007. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat Rev Neurosci*. 8:700–711.

Gardella C, Marre O, Mora T. 2016. A tractable method for describing complex couplings between neurons and population rate. *eNeuro*. 3(4):0160–0115.

Ginzburg VL, Landau LD. 1965. On the theory of superconductors. *Zh Eksp Teor Fiz*. 20:1064–1082.

Greicius MD, Kiviniemi V, Tervonen O, Vainionpää V, Alahuhta S, Reiss AL, Menon V. 2008. Persistent default-mode network connectivity during light sedation. *Hum Brain Mapp*. 29(7):839–847.

Guldenmund P, Demertzi A, Boveroux P, Boly M, Vanhaudenhuyse A, Bruno M-A, Gosseries O, Noirhomme Q, Brichant J-F, Bonhomme V, et al. 2013. Thalamus, brainstem and salience network connectivity changes during propofol-induced sedation and unconsciousness. *Brain Connect*. 3(3):273–285.

Guevara Erra R, Mateos DM, Wennberg R, Velazquez JP. 2016. Statistical mechanics of consciousness: maximization of information content of network is associated with conscious awareness. *Physical Review E*. 94(5):052402.

Hahn G, Ponce-Alvarez A, Monier C, Benvenuti G, Kumar A, Chavane F, Deco G, Frégnac Y. 2017. Spontaneous cortical activity is transiently poised close to criticality. *PLoS Comput Biol*. 13(5):e1005543.

Hidalgo J, Grilli J, Suweis S, Muñoz MA, Banavar JR, Maritan A. 2015. Information-based fitness and the emergence of criticality in living systems. *Proc Natl Acad Sci USA*. 111:10095–10100.

Hudetz AG, Liu X, Pillay S. 2015. Dynamic repertoire of intrinsic brain states is reduced in propofol-induced unconsciousness. *Brain Connect*. 5(1):10–22.

Juengling FD, Kassubek J, Huppertz HJ, Krause T, Els T. 2005. Separating functional and structural damage in persistent vegetative state using combined voxel-based analysis of 3-D MRI and FDG-PET. *J Neurol Sci*. 228(2):179–184.

Kaisti KK, Metsähonkala L, Teräs M, Oikonen V, Aalto S, Jääskeläinen S, Hinkka S, Scheinin H. 2002. Effects of surgical levels of propofol and sevoflurane anesthesia on cerebral blood flow in healthy subjects studied with positron emission tomography. *Anesthesiology*. 96(6): 1358–1370.

Lee H, Golkowski D, Jordan D, Berger S, Ilg R, Lee J, Mashour GA, Lee U, ReCCognition Study Group. 2019. Relationship of critical dynamics, functional connectivity, and states of consciousness in large-scale human brain networks. *Neuro Image*. 188:228–238.

Luppi AI, Craig MM, Finoia P, Williams GB, Naci L, Menon DK, Emmanuel A. 2019. Consciousness-specific dynamic interactions of brain integration and functional diversity. *Nat Commun*. 10(1):4616.

Machta BB, Chachra R, Transtrum MK, Sethna JP. 2013. Parameter space compression underlies emergent theories and predictive models. *Science*. 342(6158):604–607.

Massimini M, Ferrarelli F, Sarasso S, Tononi G. 2012. Cortical mechanisms of loss of consciousness: insight from TMS/EEG studies. *Arch Ital Biol*. 150(2–3):44–55.

Napolitani M, Bodart O, Canali P, Seregni F, Casali AG, Laureys S, Rosanova M, Massimini M, Gosseries O. 2014. Transcranial magnetic stimulation combined with high-density EEG in altered states of consciousness. *Brain Inj*. 28(9):1180–1189.

Okun M, Steinmetz NA, Cossell L, Iacaruso MF, Ko H, Barthó P, Moore T, Hofer SB, Mrsic-Flogel TD, Carandini M, et al. 2015. Diverse coupling of neurons to populations in sensory cortex. *Nature*. 521(7553):511–515.

Panas D, Amin H, Maccione A, Muthmann O, van Rossum M, Berdondini L, Hennig MH. 2015. Sloppiness in spontaneously active neuronal networks. *J Neurosci*. 35(22):8480–8492.

Perez Velazquez JL, Mateos DM, Guevara Erra R. 2019. On a simple general principle of brain organization. *Front Neurosci*. 13:1106.

Ponce-Alvarez A, Mochol G, Hermoso-Mendizabal A, de la Rocha J, Deco G. 2020. Cortical state transitions and stimulus response evolve along stiff and sloppy parameter dimensions, respectively. *Elife*. 9:e53268.

Sarà M, Pistoia F, Pasqualetti P, Sebastiano F, Onorati P, Rossini PM. 2011. Functional isolation within the cerebral cortex in the vegetative state: a nonlinear method to predict clinical outcomes. *Neurorehabil. Neural Repair*. 25(1):35–42.

Sarasso S, Rosanova M, Casali AG, Casarotto S, Fecchio M, Boly M, Gosseries O, Tononi G, Laureys S, Massimini M. 2014. Quantifying cortical EEG responses to TMS in (un)consciousness. *Clin EEG Neurosci*. 45:40–49.

Sarasso S, Boly M, Napolitani M, Gosseries O, Charland-Verville V, Casarotto S, Rosanova M, Casali AG, Brichant J-F, Boveroux P, et al. 2015. Consciousness and complexity during unresponsiveness induced by propofol, xenon, and ketamine. *Curr Biol*. 25:3099–3105.

Schartner M, Seth A, Noirhomme Q, Boly M, Bruno M-A, Laureys S, Barrett A. 2015. Complexity of multi-dimensional spontaneous EEG decreases during Propofol induced general anaesthesia. *PLoS One*. 10(8):e0133532.

Schneidman E, Berry MJ, Segev R, Bialek W. 2006. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*. 440:1007–1012.

Schölvinck ML, Maier A, Ye FQ, Duyn JH, Leopold DA, Scholvinck ML, Maier A, Ye FQ, Duyn JH, Leopold DA, et al. 2010. Neural basis of global resting-state fMRI activity. *Proc Natl Acad Sci USA*. 107:10238–10243.

Shew WL, Yang H, Yu S, Roy R, Plenz D. 2011. Information capacity and transmission are maximized in balanced cortical networks with neuronal avalanches. *J Neurosci*. 31:55–63.

Shew WL, Plenz D. 2013. The functional benefits of criticality in the cortex. *Neuroscientist*. 19:88–100.

Singh S, Bansal S, Kumar G, Gupta I, Thakur JR. 2017. Entropy as an indicator to measure depth of anaesthesia for laryngeal mask airway (LMA) insertion during sevoflurane and propofol anaesthesia. *J Clin Diagn Res*. 11:UC01–UC03.

Sitt JD, King J-R, El Karoui I, Rohaut B, Faugeras F, Gramfort A, Cohen L, Sigman M, Dehaene S, Naccache L. 2014. Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain*. 137(8):2258–2270.

Solovey G, Alonso LM, Yanagawa T, Fujii N, Magnasco MO, Cecchi GA, Proekt A. 2015. Loss of consciousness is associated with stabilization of cortical activity. *J Neurosci*. 35:10866–10877.

Storm JF, Boly M, Casali AG, Massimini M, Olcese U, Pennartz CMA, Wilke. 2017. Consciousness regained: disentangling mechanisms, brain systems, and behavioral responses. *J Neurosci*. 37:10882–10893.

Tagliazucchi E, Balenzuela P, Fraiman D, Chialvo DR. 2012. Criticality in large-scale brain fMRI dynamics unveiled by a novel point process analysis. *Front Physiol*. 3:15.

Tagliazucchi E, Chialvo DR, Siniatchkin M, Amico E, Brichant JF, Bonhomme V, Noirhomme Q, Laufs H, Laureys S. 2016. Large-scale signatures of unconsciousness are consistent with a departure from critical dynamics. *J R Soc Interface*. 13(114):20151027.

Tanabe S, Huang Z, Zhang J, Chen Y, Fogel S, Doyon J, Wu J, Xu J, Zhang J, Qin P, et al. 2020. Altered global brain signal during physiologic, pharmacologic, and pathologic states of unconsciousness in humans and rats. *Anesthesiology*. 132:1392–1406.

Tkačik G, Marre O, Amodei D, Schneidman E, Bialek W, Berry MJ II. 2014. Searching for collective behavior in a large network of sensory neurons. *PLoS Comput Biol*. 10:e1003408.

Tkačik G, Mora T, Marre O, Amodei D, Palmer SE, Berry MJ, Bialek W. 2015. Thermodynamics and signatures of criticality in a network of neurons. *Proc Natl Acad Sci USA*. 112:11508–11513.

Uhrig L, Janssen D, Dehaene S, Jarraya B. 2016. Cerebral responses to local and global auditory novelty under general anesthesia. *Neuroimage*. 141:326–340.

Uhrig L, Sitt JD, Jacob A, Tasserie J, Barttfeld P, Dupont M, Dehaene S, Jarraya B. 2018. Resting-state dynamics as a cortical signature of anesthesia in monkeys. *Anesthesiology*. 129(5):942–958.

Vanduffel W, Fize D, Mandeville JB, Nelissen K, Van Hecke P, Rosen BR, Tootell RB, Orban GA. 2001. Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys. *Neuron*. 32:565–577.

Watanabe T, Hirose S, Wada H, Imai Y, Machida T, Shirouzu I, Konishi S, Miyashita Y, Masuda N. 2013. A pairwise maximum entropy model accurately describes resting-state human brain networks. *Nat Commun*. 4:1370.

Zhang XS, Roy RJ, Jensen EW. 2001. EEG complexity as a measure of depth of anesthesia for patients. *IEEE Trans Biomed Eng*. 48:1424–1433.