



**HAL**  
open science

## Macroscopic quantities of collective brain activity during wakefulness and anesthesia

Adrián Ponce-Alvarez, Lynn Uhrig, Nikolas Deco, Camilo Miguel Signorelli,  
M. L. Kringelbach, Béchir Jarraya, Gustavo Deco

► **To cite this version:**

Adrián Ponce-Alvarez, Lynn Uhrig, Nikolas Deco, Camilo Miguel Signorelli, M. L. Kringelbach, et al.. Macroscopic quantities of collective brain activity during wakefulness and anesthesia. *Cerebral Cortex*, 2022, 32 (2), pp.298-311. 10.1093/cercor/bhab209 . hal-03579500

**HAL Id: hal-03579500**

**<https://hal.science/hal-03579500>**

Submitted on 2 Jun 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

# 1 **Macroscopic quantities of collective brain activity during wakefulness and** 2 **anesthesia**

## 3 **Authors**

4 Adrián Ponce-Alvarez<sup>1\*</sup>, Lynn Uhrig<sup>2</sup>, Nikolas Deco<sup>1</sup>, Camilo M. Signorelli<sup>2,3,1</sup>, Morten L.  
5 Kringelbach<sup>3,4,5</sup>, Béchir Jarraya<sup>2,6,7,8</sup>, Gustavo Deco<sup>1,9,10,11</sup>

## 7 **Affiliations**

8 1: Center for Brain and Cognition, Computational Neuroscience Group, Department of Information  
9 and Communication Technologies, Universitat Pompeu Fabra, Barcelona 08005, Spain.

10 2: NeuroSpin Center, Institute of BioImaging Commissariat à l’Energie Atomique, Gif-sur-Yvette  
11 91191, France.

12 3: Department of Psychiatry, University of Oxford, Oxford OX3 7JX, United Kingdom.

13 4: Center for Music in the Brain, Department of Clinical Medicine, Aarhus University, Aarhus  
14 8000, Denmark.

15 5: Life and Health Sciences Research Institute, School of Medicine, University of Minho, Braga  
16 4710-057, Portugal.

17 6: UniCog, INSERM, Gif-sur-Yvette 91191, France.

18 7: Université Paris-Saclay, UVSQ, Versailles 78000, France.

19 8: Neuromodulation Unit, Foch Hospital, 92150 Suresnes, France.

20 9: Institució Catalana de la Recerca i Estudis Avançats (ICREA), Barcelona 08010, Spain.

21 10: Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain  
22 Sciences, Leipzig 04103, Germany.

23 11: School of Psychological Sciences, Monash University, Melbourne, Clayton VIC 3800,  
24 Australia.

25 \* **Correspondence:** [adrian.ponce@upf.edu](mailto:adrian.ponce@upf.edu)

## 27 **Abstract**

28 The study of states of arousal is key to understand the principles of consciousness. Yet, how different  
29 brain states emerge from the collective activity of brain regions remains unknown. Here, we studied  
30 the fMRI brain activity of monkeys during wakefulness and anesthesia-induced loss of consciousness.  
31 Using maximum entropy models, we derived collective, macroscopic properties that quantify the  
32 system’s capabilities to produce work, to contain information and to transmit it, and that indicate a  
33 phase transition from critical awake dynamics to supercritical anesthetized states. Moreover,  
34 information-theoretic measures identified those parameters that impacted the most the network  
35 dynamics. We found that changes in brain state and in state of consciousness primarily depended on  
36 changes in network couplings of insular, cingulate, and parietal cortices. Our findings suggest that  
37 the brain state transition underlying the loss of consciousness is predominantly driven by the  
38 uncoupling of specific brain regions from the rest of the network.  
39

40 **This PDF includes:** Main Text with embodied Figures (5 figures), Appendix, and Supplementary  
41 Figures S1 to S8.

42

## 43 **Introduction**

44 Interesting phenomena in biological systems are usually collective behaviors emerging from the  
45 interactions among many constituents. Large-scale brain activity is not an exception: the brain's  
46 network continuously generates coordinated spontaneous patterns of activity among brain regions at  
47 multiple spatiotemporal scales (1-3). Changes in spontaneous brain activity are observed in different  
48 brain states, the study of which is essential to understand the organizing principles of brain activity.  
49 For instance, anesthesia has been used to transiently induce loss of consciousness and to investigate  
50 the neural correlates of the awake state. Previous studies showed that different anesthetics, acting on  
51 different molecular targets (4), similarly impact the strength and the structure of functional  
52 correlations (5-10), and their dependence on interareal anatomical connections (8, 11). However, how  
53 changes in local regions and subnetworks combine to affect the collective brain dynamics and to lose  
54 consciousness remains largely unknown. To answer this question, it is essential to precisely  
55 characterize the collective properties of different brain states and their dependence on parameters at  
56 the system's level. This dependence is likely not straightforward since, as for many complex systems,  
57 the system's behavior could be differently affected by changes in its parameters. In such a case, while  
58 some parameters can largely vary without affecting the system's behavior (so-called "sloppy"  
59 parameters), even small changes in some others can significantly modify it (12-14).

60 In recent years, statistical mechanics has proven to be more and more useful to describe collective  
61 neural activity. Statistical mechanics shows that the behaviors of complex systems can be captured  
62 by macroscopic properties, which emerge from the collective activity of the units, in a way largely  
63 independent of the microscopic details of the system. These emergent (macroscopic) behaviors can  
64 be classified into qualitatively different ordered or disordered phases. Of particular interest are  
65 dynamics poised close to phase transitions, or critical points, where order and disorder coexist.  
66 Theoretical reasoning shows that complex dynamics and optimal information processing are expected  
67 at critical points, making criticality a candidate unifying principle to account for the brain's inherent  
68 complexity necessary to process and represent its environment (15-18). Following this view, studies  
69 of whole-brain and local circuits dynamics have proposed that anesthesia shifts the dynamics from  
70 the critical point (19, 20). This is supported by the reduction of several measures of brain dynamics  
71 complexity under anesthesia (21-24). The global mechanisms underlying different conscious states  
72 have been recently investigated using an anatomically-constrained dynamical model with a global  
73 coupling parameter in combination with EEG recordings (25). However, it remains unknown which  
74 are the macroscopic properties and the relevant local/global parameters describing the transition of  
75 collective activity from the awake to anesthetized states. Indeed, different local/global network  
76 parameters are likely to jointly determine the different brain states and to differently contribute to the  
77 state transitions.

78 In this study, we addressed these questions by analyzing the brain's collective activity in different  
79 levels of arousal, i.e., during wakefulness and under anesthesia. Specifically, we analyzed resting-  
80 state fMRI dynamics of awake and anesthetized macaque monkeys (11). Five different anesthesia  
81 protocols, involving 3 different anesthetics (propofol, ketamine, and sevoflurane), were used to

82 induce moderate sedation or deep anesthesia. First, we derived efficient statistics that distinguished  
83 between awake and anesthetized brain states. Second, we used these statistics and the maximum  
84 entropy principle to model the brain's activity and to derive important emergent properties that  
85 described the different brain states. These emergent properties provided information about the  
86 system's physical state, and about its capability to produce work, to contain information and to  
87 transmit it. Finally, we investigated the dependence of collective activity on the different model  
88 parameters.

89

## 90 **Results**

91 We analyzed the resting-state fMRI dynamics of five rhesus macaques (*Macaca mulatta*) under  
92 different levels of arousal: wakefulness ( $n = 24$  scans), two levels of propofol sedation (light, LPP,  $n$   
93  $= 21$ , and deep, DPP,  $n = 23$ ), sedation through ketamine (KETA,  $n = 22$ ), and two types of  
94 sevoflurane anesthesia (SEV2,  $n = 18$ , and SEV4,  $n = 11$ ) (see Methods and [Appendix](#)). fMRI MION  
95 time-series were obtained for  $N = 82$  previously defined regions of interest (ROIs) (CoCoMac  
96 Regional Map parcellation). Each scan was 20 min long and was acquired in time frames of 2.4 s  
97 (i.e., 500 time frames).

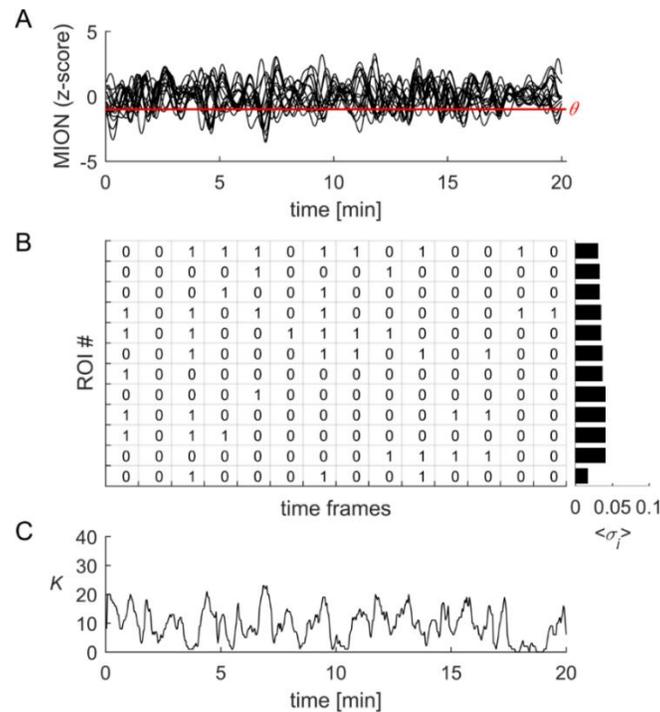
98

### 99 **Coupling to population reliably distinguished between awake and anesthetized brain states**

100 We were interested on collective patterns displayed among the  $N$  ROIs, for the six different  
101 experimental conditions. For this, we first binarized the z-scored time-series of each ROI,  $x_i(t)$ , by  
102 imposing a threshold  $\theta = -1$  (**Fig. 1A,B**, see Methods). Binarization of time-series has proven to  
103 effectively capture and compress fMRI large-scale dynamics ([26, 27](#)). We concentrated on different  
104 statistics that described the resulting binary data: the activation rate of each ROI, i.e.,  $\langle \sigma_i \rangle$ , the  
105 correlation between ROIs, i.e.,  $C_{ij} = \text{corr}(\sigma_i, \sigma_j)$ , and the population activity, i.e.,  $K(t) = \sum_{i=1}^N \sigma_i(t)$   
106 (**Fig. 1C**). We were particularly interested on the coupling of each ROI to the population activity,  
107 defined as:

$$108 \quad z_i = \text{corr}[\sigma_i(t), K_{\neq i}(t)], \quad (1)$$

109 where  $K_{\neq i}(t)$  is the sum activity of all but ROI  $i$ :  $K_{\neq i}(t) = \sum_{j \neq i} \sigma_j(t)$ . Recent findings showed that  
110 propofol anesthesia affects the coupling to global signal in human and rats ([28](#)). In the following we  
111 showed that the statistics  $\mathbf{z} = [z_1, \dots, z_N]$  provides, with only  $N$  parameters, a compact description of  
112 the binary collective activity and can be used to classify the brain states.

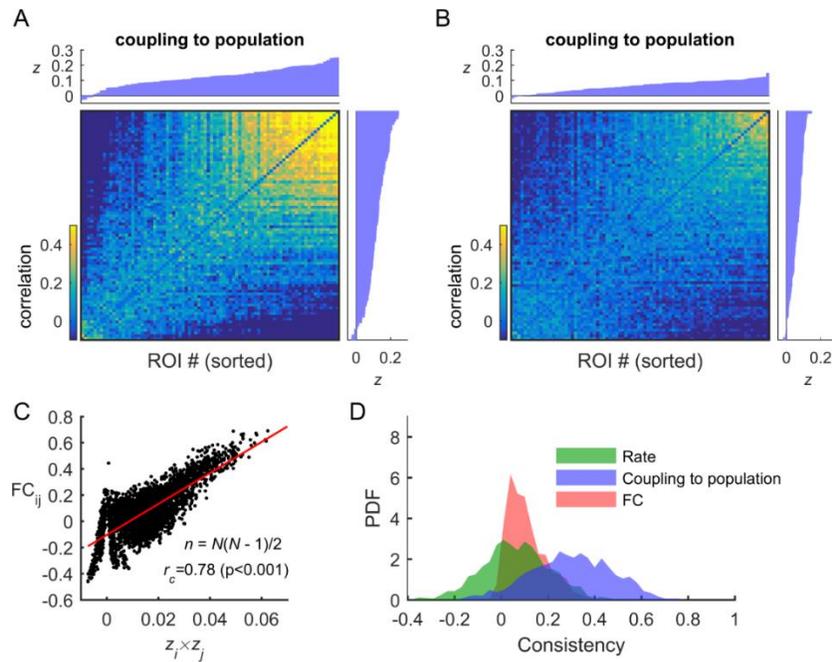


113

114 **Figure 1. Binarization and statistics.** (A) MION fMRI signals were z-scored and binarized by imposing a threshold  
 115 equal to the standard deviation, for each signal. (B) In each time bin of 2.4 s, the state of signal of ROI  $i$ , noted  $\sigma_i(t)$ ,  
 116 was equal to 1 if the MION signal for this ROI was lower than minus its standard deviation, or equal to 0 otherwise.  
 117 The average activity of ROI  $i$  was  $\langle \sigma_i \rangle = \sum_{t=1}^n \sigma_i(t)$ , where  $n$  is the number of time points. (C) The population  
 118 activity was defined as the sum of the binary activity of the  $N$  ROIs in each time bin  $t$ , i.e.,  $K(t) = \sum_{i=1}^N \sigma_i(t)$ .  
 119

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

129

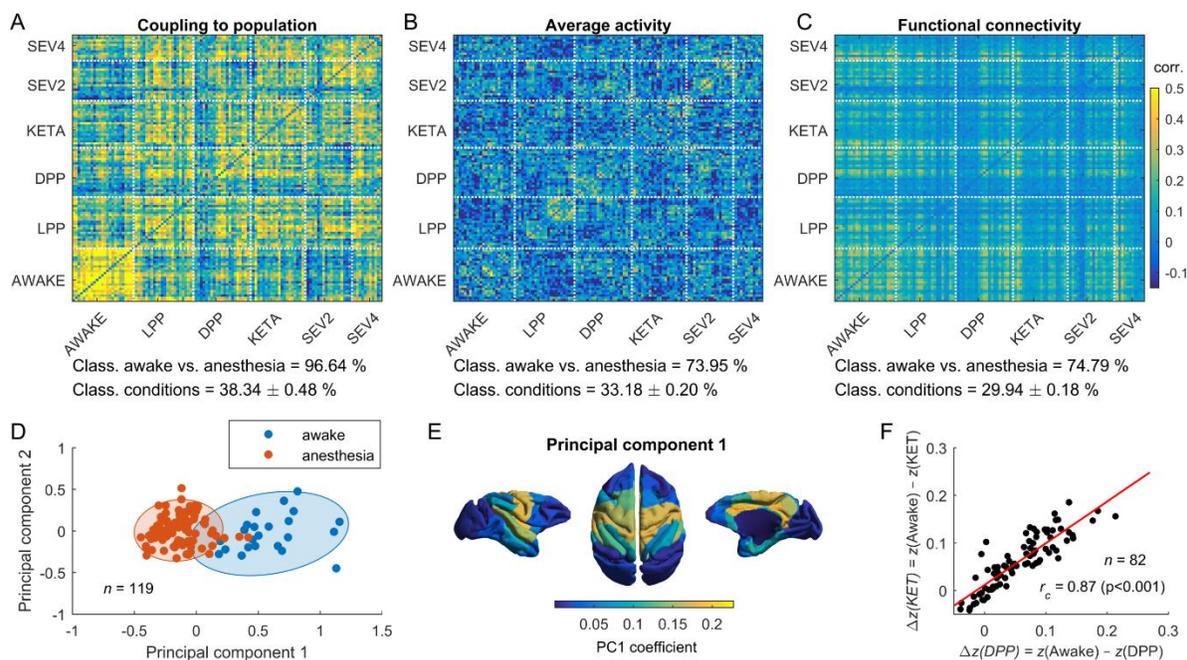


130

131 **Figure 2. Coupling to population predicts the functional connectivity and is consistent within experimental**  
 132 **conditions.** (A) *Top and left insets:* ROIs were ordered according to  $z_i$  averaged for each scan within a given  
 133 experimental condition (here for the awake condition). *Color map:* The average functional connectivity (FC) is shown  
 134 after ordering the ROIs according to  $z$ . (B) Same as (A) but for the deep propofol (DPP) anesthesia condition. (C)  
 135 The elements of the FC and the corresponding products of coupling to population ( $z_i z_j$ ) highly correlated. (D) We  
 136 tested whether  $\langle \sigma \rangle$ ,  $z$ , and FC were similar across scans within the same experimental condition. For example, for the  
 137 statistic  $z$ , we calculate the correlation of this  $N$ -dimensional variable for all pairs of scans belonging to the same  
 138 experimental condition and computed the distribution of correlation coefficients (*blue* distribution). High correlation  
 139 coefficients indicate that, within experimental conditions, scans yielded similar vector  $z$ . The same can be done for  
 140 the  $N$ -dimensional variable  $\langle \sigma \rangle$  (*green* distribution) and the vector of FC elements ( $N(N - 1)/2$  dimensions; *red*  
 141 distribution).

142

143 Furthermore, we found that the coupling to the population could be used to classify the awake and  
 144 anesthesia states with high accuracy (**Fig. 3A**). We tested this by using a classifier based on  $k$ -means  
 145 clustering (see Methods). Based on the statistic  $z$  we were able to classify the scans of two categories,  
 146 awake vs. anesthesia (independently of the anesthetic), with 96.6% of correct classifications (chance  
 147 level: 50%). This classification performance was higher than the one obtained using the statistics  $\mu$   
 148 and  $\rho$ , yielding 74.0% and 74.8% of correct classifications, respectively (**Fig. 3B,C**). Classification  
 149 among the six experimental conditions yielded lower performances but was higher for statistic  $z$  than  
 150 for  $\mu$  and  $\rho$ : 38.5%, 33.1%, and 29.9%, respectively (chance level: 16.7%). Similar differences in  
 151 classification performances for population couplings and functional correlations were obtained using  
 152 continuous (not thresholded) signals (**Fig. S1**). Altogether, these results show that the coupling to  
 153 population is a reliable marker to distinguish between awake and anesthetized brain states.



154

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

168

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

176

## 177 Modelling collective activity using maximum entropy models

178 Collective activity is ultimately described by the probability of each of the binary patterns  $\boldsymbol{\sigma} =$   
 179  $[\sigma_1, \dots, \sigma_N]$ . Estimating the distribution  $P(\boldsymbol{\sigma})$  over the  $2^N$  possible binary patterns from the data is  
 180 impractical with limited amount of observations, since for  $N = 82$  there are more than  $10^{24}$  possible

181 patterns. A useful technique to estimate  $P(\boldsymbol{\sigma})$  relies on the maximum entropy principle. Maximum  
182 entropy models (MEMs) find  $P(\boldsymbol{\sigma})$  by maximizing its entropy under the constraint that some  
183 empirical statistics are preserved (see Methods). As shown above, an interesting statistic for the  
184 present study is the coupling between the state of each binary signal,  $\sigma_i$ , and the population activity  
185  $K$ . The maximum entropy distribution that is consistent with the probability distribution  $P(K)$ , the  
186 average activations  $\langle \sigma_i \rangle$ , and the linear coupling between  $\sigma_i$  and  $K$ , i.e.,  $\langle \sigma_i K \rangle$  (which relates to  $z_i$ ),  
187 is given by the Boltzmann distribution  $P(\boldsymbol{\sigma}) = e^{-E(\boldsymbol{\sigma})}/Z$ , where  $E(\boldsymbol{\sigma})$  represents the energy of the  
188 pattern  $\boldsymbol{\sigma}$ , given as (29):

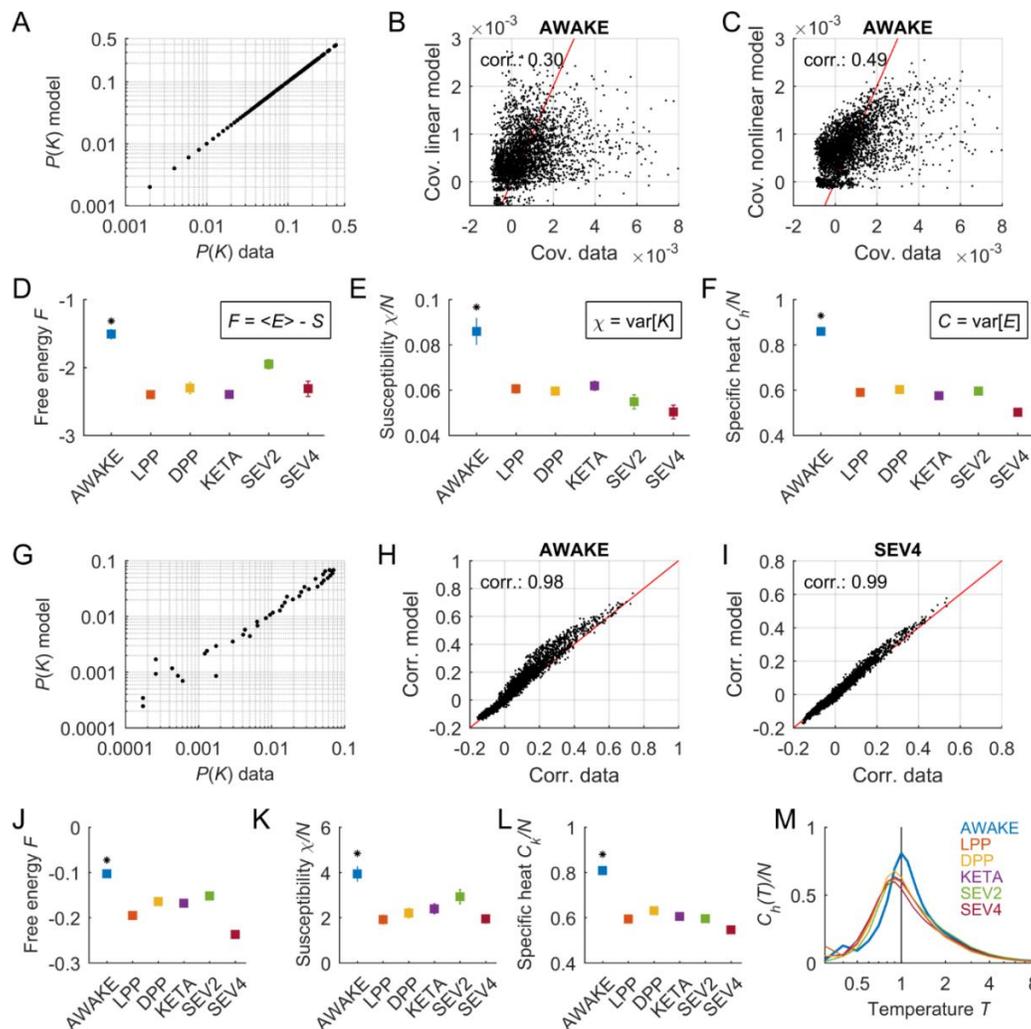
$$189 \quad E(\boldsymbol{\sigma}) = - \sum_{i=1}^N (\alpha_i + \beta_K + \gamma_i) \sigma_i. \quad (2)$$

190 The model parameters  $\alpha_i$ ,  $\beta_K$ , and  $\gamma_i$  are Lagrange multipliers associated to the constrained  
191 observables  $\langle \sigma_i \rangle$ ,  $P(K)$ , and  $\langle \sigma_i K \rangle$ , respectively. The normalizing constant  $Z$  is the partition function,  
192 given by  $Z = \sum_{\{\boldsymbol{\sigma}\}} e^{-E(\boldsymbol{\sigma})}$ , which contains information about useful statistics predicted by the model  
193 (see below). This model can be extended to include the non-linear coupling between  $\sigma_i$  and  $K$ . Indeed,  
194 the maximum entropy distribution that is consistent with the joint probability distributions of  $\sigma_i$  and  
195  $K$ , i.e.,  $P(\sigma_i, K)$ , yields the following energy function (29):

$$196 \quad E(\boldsymbol{\sigma}) = - \sum_{i=1}^N H_{i,K(\boldsymbol{\sigma})} \sigma_i, \quad (3)$$

197 where  $K(\boldsymbol{\sigma})$  is the number of active ROIs in pattern  $\boldsymbol{\sigma}$  and the parameters  $H_{i,K(\boldsymbol{\sigma})}$  represent the  
198 tendency of ROI  $i$  to activate when  $K(\boldsymbol{\sigma})$  ROIs are active. For both linear and non-linear coupling-  
199 MEMs the model parameters were inferred from the data using maximum likelihood (29). Notably,  
200 for the coupling-MEMs the partition function can be calculated directly — something that is generally  
201 not the case for most MEMs, since its calculation involves summing over all possible states.

202 We used these coupling-MEMs to fit the binary single-scan fMRI data for the different experimental  
203 conditions. The models accurately estimated the distribution of population activity  $P(K)$  (average  
204 Jensen-Shannon divergence  $D_{JS}$  between the model and data distributions:  $D_{JS} < 10^{-6}$  for both the  
205 non-linear and linear coupling-MEM; **Fig. 4A** and **Fig. S2**). Moreover, the models were able to  
206 moderately predict the covariances of the data (**Fig. 4B,C**), which were not used to constrain the  
207 models. Across the different datasets, the average correlation between the data and predicted  
208 covariances was  $r = 0.28 \pm 0.03$  for the linear coupling model and reached  $0.40 \pm 0.02$  for the non-  
209 linear coupling model (see also **Fig. S2**). Furthermore, scan-classification based on parameters  $\gamma_i$   
210 yielded 86% and 45% correct classifications between awake and anesthetized conditions and among  
211 the six experimental conditions, respectively (**Fig. S3A,B**). Using parameters  $\alpha_i$  the classifier  
212 performance decreased to 75% and 28%, respectively (**Fig. S3B**). Thus, the learned linear coupling-  
213 MEM showed consistent variations in parameters  $\gamma_i$  (associated to  $z_i$ ) across the different arousal  
214 states.



215

216 **Figure 4. Maximum entropy models indicate higher free energy and heat capacity during wakefulness than**  
 217 **during anesthesia. (A)** Fitting of  $P(K)$  using the non-linear coupling-MEM. Data and predictions from all scans  
 218 from the awake condition. **(B-C)** Fitting of covariances using the linear (B) and nonlinear (C) coupling-MEM for the  
 219 awake condition. **(D-F)** The free energy, the susceptibility, and the heat capacity were derived using non-linear  
 220 coupling-MEMs for the different conditions. Similar results were obtained using the linear model (see Fig. S4).  
 221 Squares and error bars indicate means and standard deviations across scans, respectively, and the asterisks indicate  
 222 significantly different values for the awake condition ( $p < 0.001$  one-way ANOVA followed by Tukey's post hoc  
 223 analysis). **(G-L)** same as (A-F) but using pairwise-MEMs. Error bars indicate standard errors across Monte Carlo  
 224 simulations of the models. Asterisks indicate significantly different values for the awake condition ( $p < 0.001$ , one-  
 225 way ANOVA followed by Tukey's post hoc analysis). **(M)** Heat capacity as a function of temperature. The peak of  
 226 heat capacity for  $T = 1$  indicates critical dynamics during wakefulness. The heat capacity peaked at  $T < 1$  for the  
 227 anesthetized conditions, indicating supercritical dynamics during anesthetized states.  
 228

229 **Collective activity indicated reduced free energy, susceptibility, and heat capacity under**  
 230 **anesthesia**

231 We can learn interesting features of collective activity using the estimated models. One important  
 232 quantity is the system's Helmholtz *free energy*, which is given by the difference between the average  
 233 energy ( $\langle E \rangle$ ) and the entropy ( $S$ ), i.e.,  $F = \langle E \rangle - S$ . The free energy quantifies the useful energy that

234 is obtainable from the system. Using the Boltzmann distribution, the free energy can be directly  
235 obtained from the partition function as  $F = -\ln(Z)$ . Thus, since  $Z$  is tractable for the coupling-MEMs,  
236 we can directly estimate  $F$ . We found that the free energy was significantly higher for the awake state  
237 compared to all anesthetized conditions for both the non-linear (**Fig. 4D**) and the linear (**Fig S4A**)  
238 coupling-MEM ( $p < 0.001$ , one-way ANOVA followed by Tukey's post hoc analysis). This result is  
239 both interesting and reasonable because it indicates that more useful energy can be extracted from the  
240 awake state than from the anesthetized state.

241 Two other important statistical quantities can be derived from the model, namely the *susceptibility*  
242 and the *heat capacity*. The susceptibility  $\chi$  relates to the diversity of population states, while the heat  
243 capacity  $C_h$  quantifies the diversity of accessible energy states. Specifically, the susceptibility and the  
244 heat capacity of the model are given by the variances of the population activity and the energy,  
245 respectively, i.e.,  $\chi = \text{var}(K)$  and  $C_h = \text{var}(E)$ . We found that  $\chi$  and  $C_h$  were significantly higher  
246 for the awake state compared to all anesthetized conditions for both linear and non-linear coupling-  
247 MEM (non-linear model: **Fig. 4E,F**, linear model: **Fig S4B,C**;  $p < 0.001$ , one-way ANOVA followed  
248 by Tukey's post hoc analysis). This indicates that the awake system had larger population fluctuations  
249 and a larger repertoire of energy states than the system under anesthesia.

250 We next tested whether the same differences in these statistical quantities were found using MEM  
251 constraint by other statistics. To build the models we estimated the maximum entropy distribution  
252  $P(\boldsymbol{\sigma})$  under the constraint that the activation rates ( $\langle \sigma_i \rangle$ ) and the pairwise correlations ( $\langle \sigma_i \sigma_j \rangle$ )  
253 are preserved. The energy of the Boltzmann distribution that is consistent with these expectation  
254 values is given by (30, 31):

$$255 \quad E(\boldsymbol{\sigma}) = - \sum_{i=1}^N h_i \sigma_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N J_{ij} \sigma_i \sigma_j . \quad (4)$$

256 In this pairwise-MEM, the parameter  $h_i$ , called intrinsic bias, represents the intrinsic tendency of ROI  
257  $i$  towards activation or silence and the parameter  $J_{ij}$  represents the effective interaction between ROIs  
258  $i$  and  $j$ . The estimation of the model parameters  $\boldsymbol{\Omega} = \{\mathbf{h}, \mathbf{J}\}$  was achieved through a pseudo-likelihood  
259 maximization (32) (see Methods). Since this model requires the precise estimation of  $\langle \sigma_i \sigma_j \rangle$ , it  
260 cannot be fitted to single-scan data and, for this reason, we used concatenated data from each  
261 experimental condition. The pairwise-MEM accurately predicted the observed correlations and, to a  
262 lower extent, it predicted the distribution of population activity  $P(K)$  (average correlation fit:  $r =$   
263  $0.985 \pm 0.002$ ; average  $D_{JS} = 0.006 \pm 0.002$ ; **Fig. 4G-I** and **Fig. S2**) — this is expected, since  $P(K)$   
264 was not used to constrain the model. We found that biases and couplings parameters were changed  
265 for different states, with some parameters increasing or decreasing, and with a reduction of the  
266 variance of couplings in the anesthetized states (**Fig. S5A-D**). Moreover, coupling parameters showed  
267 a higher correlation with the anatomical connectivity (or brain connectome) in the anesthetized states  
268 than in the awake state (**Fig. S5E**).

269 Using this model, we calculated the collective statistical quantities for the different experimental  
270 conditions. Since in the pairwise-MEM the partition function is not tractable, we calculated  $F$ ,  $\chi$  and  
271  $C_h$  using Monte Carlo simulations (see Methods). Consistent with the above results, we found that  
272 the awake system had larger available energy (free energy, **Fig. 4J**, see also **Fig. S6**), larger  
273 population fluctuations (susceptibility, **Fig. 4K**) and larger repertoire of states (heat capacity, **Fig.**  
274 **4L**) than the system under anesthesia. Thus, the different versions of the MEM used here indicate the  
275 same results concerning the statistical properties of awake and anesthetized states. Furthermore, as  
276 shown in the [Appendix](#), the susceptibility can be viewed as a measure of the network response to a  
277 small stimulus. Consequently, we found that application of an external stimulus elicited larger and  
278 more diverse responses for the pairwise-MEM corresponding to the awake state than for the models  
279 corresponding to the anesthetized states (**Fig. S7**).

280

### 281 **Awake collective activity displayed critical dynamics that were shifted to a super-critical regime** 282 **under anesthesia**

283 The pairwise-MEM can be used to assess the physical state of the system. Indeed, by introducing a  
284 scaling parameter  $T$ , analogous to the temperature in statistical physics, one can obtain relevant  
285 features of the collective dynamics. For this, we scaled all model parameters as  $\Omega \rightarrow \Omega/T$  and  
286 calculated the heat capacity as a function of  $T$ , given by  $C_h(T) = \text{var}[E]/T^2$ . The “temperature”  $T$   
287 controls the level of disorder and its effects can be understood by examining the system’s energy  
288 levels (**Fig. S8**). Briefly, at low temperatures, interactions dominate over fluctuations making the  
289 system predominantly silent and ordered. In contrast, at high temperatures, the system is disordered  
290 and relatively uncoupled because fluctuations dominate over interactions. Both low and high  
291 temperatures lead to a low  $C_h$ . However, for a specific temperature  $T_{\max}$ , order and disorder coexist  
292 in the system and  $C_h$  is maximal as expected for critical dynamics (33, 34). Thus, a maximal heat  
293 capacity at  $T_{\max} = 1$  (corresponding to the model learned from the data) suggests that the system  
294 operated close to a critical state (whereas  $T_{\max} < 1$  and  $T_{\max} > 1$  indicates super-critical and sub-  
295 critical dynamics, respectively).

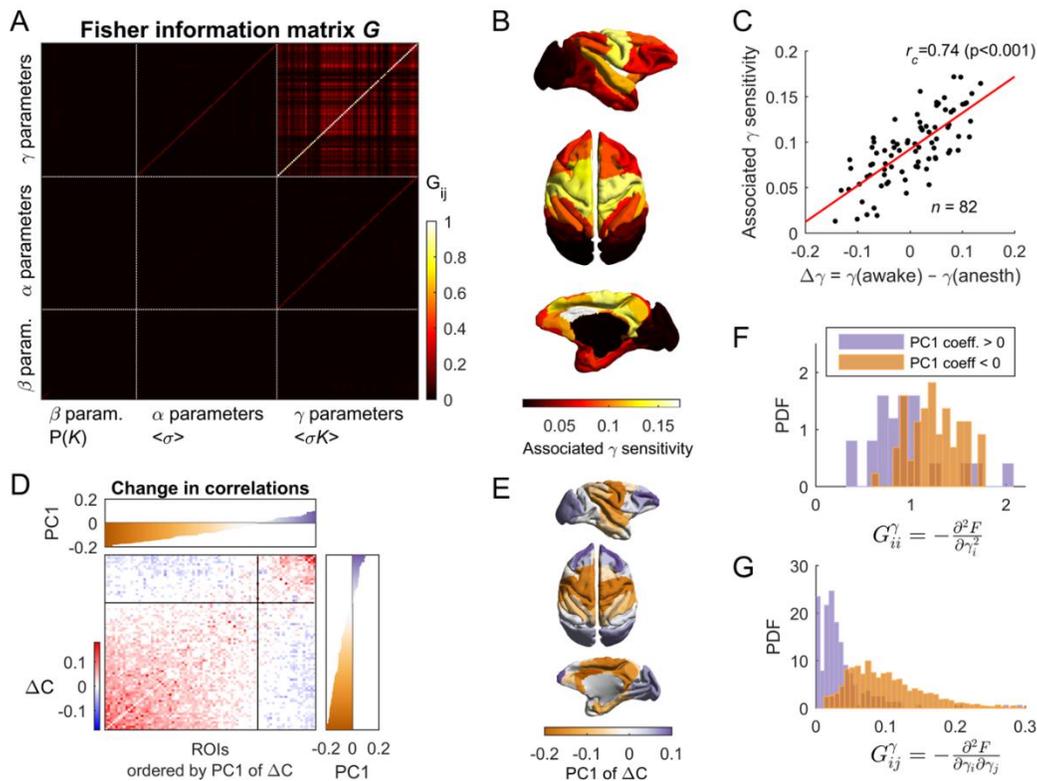
296 We found that the heat capacity curve was maximal for a temperature equal to 1 for the awake state,  
297 while it peaked at  $T_{\max} < 1$  for the anesthetized conditions (**Fig. 4M**). These results suggest that the  
298 awake state displayed critical dynamics, while dynamics under anesthesia were super-critical, which  
299 indicates that the anesthetics had a disconnection effect.

300

### 301 **Couplings to population relate to the sensitive parameters of the system**

302 We next evaluated how the different parameters affected the model’s collective behavior. In general,  
303 changes in parameters can differently affect the system’s behavior, with some parameters (called  
304 “stiff” parameters) effectively modifying it, while others have little effect on it (“sloppy” parameters)  
305 (12). We used an information-theoretical approach based on the Fisher Information Matrix (FIM,  
306 noted  $G$ ) to detect the parameters that have a strong effect on the collective activity (see Methods).

307 The FIM measures the change in the model log-likelihood  $P(\sigma|\Omega)$  with respect to changes in the  
 308 model parameters  $\Omega$ . As demonstrated in the [Appendix](#), the FIM relates to the second derivatives of  
 309 the free energy with respect to the model parameters, i.e.,  $G_{ij} = -\frac{\partial^2 F}{\partial \Omega_i \partial \Omega_j}$ . This relation provides a  
 310 direct link between a macroscopic quantity (the free energy) describing the collective dynamics of  
 311 the different brain states and the underlying model parameters. For the linear model, the parameters  
 312 that contributed the most to the FIM were the parameters  $\gamma_i$  (**Fig 5A**). This explains how changes in  
 313 couplings to population, as observed between awake and anesthetized states, effectively change the  
 314 collective state of the system, leading to the observed shift from critical to supercritical dynamics.  
 315



316

317 **Figure 5. Model sensitivity to the different parameters.** (A) Fisher information matrix (FIM) calculated using the  
 318 linear coupling-MEM built using the data concatenated across scans from the awake condition. The FIM measures  
 319 how much the model log-likelihood changes with respect to changes in the model parameters, i.e.,  $\Omega = \{\alpha, \beta, \gamma\}$   
 320 for the linear coupling-MEM. (B) For each scan, we decomposed the FIM in eigenvectors and the mean contribution to  
 321 first eigenvector of parameters  $\gamma$  was represented on the brain's image. This represents the sensitivity of collective  
 322 activity on the parameters  $\gamma$  associated to the different ROIs. (C) The ROIs with larger reduction of parameter  $\gamma$  in  
 323 anesthetized states with respect to the awake state were those with strongest sensitivity. (D) Change of pairwise  
 324 correlations between awake and anesthetized states:  $\Delta C = C_{\text{awake}} - \langle C_{\text{anesth}} \rangle$ . ROIs were ordered according to their  
 325 contribution to the first eigenvector of  $\Delta C$  (top and right insets). Two groups of ROIs were detected according to their  
 326 positive or negative contribution to this eigenvector, respectively, with both groups reducing the correlation between  
 327 awake and anesthesia. (E) First eigenvector of  $\Delta C$  represented in the brain. (F-G) The two groups of ROIs had  
 328 significantly different associated sensitivity ( $p < 0.001$ , Wilcoxon rank sum test), as measured by the FIM values  
 329 associated to parameters  $\gamma$ .

330

331 To evaluate the importance of each of the parameters, we defined the parameter's sensitivity as its  
332 absolute contribution to the first eigenvector of the FIM (see Methods). The regions with the largest  
333 associated sensitivity for parameter  $\gamma$  were located in the cingulate, parietal, and insular cortices (**Fig.**  
334 **5B**). Those that contributed the least were visual and prefrontal cortices. Interestingly, the regions  
335 presenting larger reductions of  $\gamma$  between awake and anesthesia tended to be those with higher  
336 associated sensitivity (corr: 0.74,  $p < 0.001$ ; **Fig. 5C**).

337 Finally, we further examined how changes in pairwise correlations between awake and anesthesia  
338 related to changes in parameters of different sensitivity. We analyzed the average difference of  
339 correlation ( $\Delta C$ ) between awake and anesthesia. Two groups of ROIs were clearly separated  
340 according to their positive or negative contribution to the first eigenvector of the matrix  $\Delta C$ ,  
341 respectively (**Fig. 5D**). Those that contributed positively were prefrontal and visual cortices, and those  
342 that contributed negatively were the cingulate, parietal, and insular cortices (**Fig. 5E**). Both groups  
343 presented a reduction of correlations under anesthesia, but prefrontal and visual cortices were related  
344 to parameters of low sensitivity (**Fig. 5F-G**). Hence, although prefrontal and visual areas changed  
345 their correlations, these changes were related to parameters that had a low impact on collective  
346 dynamics.

347

## 348 **Discussion**

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

359

### 360 **Population couplings and network sensitivity**

361 Previous research at the microcircuit level showed that neurons differ in their coupling to the  
362 population activity, with neurons that activate most often when many others are active and neurons  
363 that tend to activate more frequently when others are silent (35). Using the FIM analysis to detect  
364 sloppy and stiff parameters, it has been shown that these different types of neurons have a different  
365 impact on the network activity, different stimulus response properties, and different involvement in  
366 cortical state transitions (14). We showed that brain regions coupled differently to the rest of the

367 whole-brain network, consistent with previous findings (28), and, furthermore, that these couplings  
368 primarily determined the collective activity (i.e., they were associated to the stiff parameters of the  
369 model) and varied across arousal brain states. Overall, we proposed that the distribution of sensitivity  
370 of brain regions and their functional role could be a general principle of neural networks at different  
371 scales.

372 Using principal components, we detected the combination of ROIs that contributed the most to  
373 distinguish between the awake and the anesthetized states based on their population couplings. The  
374 brain regions that changed their population coupling from the awake state to the anesthetized state  
375 were the cingulate, parietal, and insular cortices (**Fig. 3E**). Notably, the model parameters associated  
376 to the couplings of these regions were among those impacting the most the collective dynamics (**Fig.**  
377 **5B-C**). Our results suggest that anesthesia modified some important local/global parameters that  
378 effectively induced a change of brain state. Our results highlight the key role of the parieto-cingulate  
379 cortex in the mechanism of anesthesia-induced loss of consciousness. Previous studies have shown  
380 that the parietal cortex (4, 9, 36) and the cingulate cortex (37) are most strongly affected by  
381 anesthetics. These cortices also present alterations in brain injury-induced unconsciousness in  
382 humans (37, 38). Moreover, consistent with our results, it has been shown that the insula plays an  
383 important role in awareness and is a potential neural correlate of consciousness (39).

384 Interestingly, although some brain regions, such as visual and prefrontal cortices, had different  
385 correlations between awake and anesthesia, they were associated to parameters with low impact on  
386 collective activity. This highlights the importance of studying not only the change in statistics  
387 between brain states but also their sensitivity on network dynamics. Consistent with our findings, a  
388 recent study of neuronal activity from several brain regions and in different arousal states (40) shows  
389 that parieto-basal ganglia circuits predicted the state of consciousness, while prefrontal activity failed.  
390 In addition, it has been proposed that the prefrontal cortex is mostly involved in the report of  
391 consciousness, rather than in the conscious experiences *per se* (41).

392

### 393 **Macroscopic thermodynamic quantities**

394 Using the MEMs, we learned interesting collective properties describing the different brain states.  
395 We measured the susceptibility that quantifies the diversity of spontaneous population fluctuations.  
396 The susceptibility can be viewed as a measure of the network response to a vanishing stimulus (**Fig**  
397 **S7**, see also the [Appendix](#)). Thus, the higher susceptibility observed in the awake state, compared to  
398 the anesthetized states, is consistent with Transcranial Magnetic Stimulation (TMS) studies showing  
399 that stimulation elicits a more diverse and complex response in the awake state than in low-level  
400 states of consciousness, such as sleep, anesthesia, and coma (42-45). Our study predicts that the  
401 network response to a localized stimulation would covary with the population couplings and the  
402 associated parameter sensitivities of brain region.

403 Moreover, the models also allowed the estimation of the system's heat capacity, a measure that  
404 quantifies the extent of the accessible dynamical repertoire. Indeed, a maximal heat capacity not only

405 indicates that the system can display a large number of energy states, but also that these states are  
406 distinguishable (**Fig S8**). Thus, a large heat capacity indicates a large capacity to represent  
407 information in numerous separable states. The observed reduction of heat capacity in the anesthetized  
408 states is consistent with previous studies showing that the repertoire of correlation states is limited  
409 during anesthesia (11). Furthermore, by varying a scaling parameter analogous to temperature, the  
410 resulting heat capacity curves suggest that awake dynamics were critical, while anesthetized  
411 dynamics were supercritical, consistent with previous predictions (19, 20). The model used here gives  
412 an intuitive interpretation of the transition between critical to supercritical dynamics. Indeed, in the  
413 pairwise-MEM supercritical dynamics are associated with a regime in which random fluctuations  
414 dominate over interactions, which is consistent with a disconnection of *effective* couplings. It is  
415 important to note that the scale parameter  $T$  is only introduced to assess the state, i.e., subcritical,  
416 critical, or supercritical, of the observed system (the one given for  $T = 1$ , for which the pairwise-  
417 MEM fits the data). This does not mean that differences between awake and anesthetized states are  
418 due to a global reduction of interactions and biases, instead different arousal states yielded different  
419 biases and couplings (**Fig S5**) which, in combination, resulted in a change of the system's state. This  
420 means that the anesthesia reconfigured the system and not only scaled its parameters. Consistently,  
421 we found that effective couplings correlated more with the anatomical connections for the  
422 anesthetized states than for the awake state, an effect that has been observed in empirical data (8, 11)  
423 and cannot be explained by changes in global connectivity alone (25).

424 Lastly, we measured the Helmholtz free energy of the estimated models. The free energy measures  
425 the useful energy that can be extracted from the system to the environment, i.e., its ability to produce  
426 work. Reasonably, the awake state led to higher free energy than the anesthetized states. Another  
427 important property of free energy is that its change with respect to the model parameters is equal to  
428 the Fisher information and, thus, it relates to the sensitivity of collective dynamics on these  
429 parameters. This result provides a direct link between the sensitivity of parameters and the change of  
430 a macroscopic quantity, the free energy, the behavior of which is known to characterize the phase  
431 transition (46). For the linear coupling-MEM, we showed that the couplings to population ( $z$ ) were  
432 associated to the parameters that have the strongest impact on collective activity. Consistently, we  
433 found that that  $z$  was an efficient observable to classify the arousal states which collective dynamics  
434 were qualitatively different (in terms of criticality and supercriticality). Thus, these results give a  
435 coherent theoretical justification of the relevance of the statistic  $z$  to characterize the brain states and  
436 to estimate their free energy. Altogether, our findings represent a significant step in the understanding  
437 of brain states, resulting in a coherent explanation of the transition from awake to anesthesia: the  
438 phase transition between brain states is driven by those parameters that change the free energy, which  
439 are the “stiff” parameters of the systems and which relate to population couplings.

440

#### 441 **Implications for studies on pathological low-level states of consciousness**

442 An interesting extension of this work could be to study brain dynamics in coma using the present  
443 statistical mechanics framework. Loss of consciousness due to anesthesia or coma share common

444 features: complexity of dynamics and neural communication are generally reduced in low-level states  
445 of consciousness (23, 47). Consequently, estimates of complexity of human brain activity have been  
446 used to assess the depth of anesthesia (48, 49) and to predict the recovery of consciousness in  
447 vegetative patients (50). Reduction of complexity is consistent with a deviation from critical  
448 dynamics when consciousness is lost. Since the coupling-MEMs can be fitted to data from single  
449 scans and, as shown here, their parameters change in different brain states, future investigation could  
450 use these models and the statistic  $z$  in the case of disorders of consciousness.

451

## 452 **Methods**

### 453 **Animals**

454 This study included a total of five rhesus macaques (*Macaca mulatta*; 4 females, 1 male, 5–8 kg; 8–  
455 12 years of age). All procedures were conducted in accordance with the European convention for  
456 animal care (86-406) and the National Institutes of Health’s Guide for the Care and Use of Laboratory  
457 Animals. Animal studies were approved by the institutional Ethical Committee (Comité d’Ethique en  
458 Expérimentation Animale, protocols #10-003 and #12-086).

### 459 **Experimental procedures**

460 Monkeys received anesthesia either with propofol, ketamine, or sevoflurane (11). The details of the  
461 anesthesia protocols are described in the Appendix. Monkeys were scanned on a 3-T horizontal  
462 scanner (Siemens Tim Trio; TR, 2,400 ms; TE, 20 ms; and 1.5-mm<sup>3</sup> voxel size; 500 brain volumes  
463 per scan session). Before each scanning session, a contrast agent monocrystalline iron oxide  
464 nanoparticle (MION) was injected into the monkey’s saphenous vein. Acquisition and preprocessing  
465 of functional images followed the standard steps described in (8) and in the Appendix. Time-series  
466 were obtained for  $N = 82$  previously defined cortical regions of interest (ROIs) (CoCoMac Regional  
467 Map parcellation).

### 468 **Data binarization**

469 fMRI time series were binarized to study the data statistics and to learn two different families of  
470 maximum entropy models (MEMs). While binarization was required to construct the MEMs,  
471 transformation of continuous fMRI signals into discrete point processes has proven to effectively  
472 capture and compress fMRI large-scale dynamics (26). Indeed, it has been shown that point process  
473 resulting from signal thresholding largely overlaps with deconvoluted fMRI signals using the  
474 hemodynamic response function and preserve the topology of the resting state networks (RSNs) [24].  
475 We here discretized the signals as follows. For each scan, the  $z$ -scored time-series of each ROI,  $x_i(t)$   
476 ( $1 \leq i \leq N$ ), was binarized by imposing a threshold  $\theta = -1$ . Two binarization procedures were used.  
477 The first method detects the threshold crossings: the binarized activity is  $\sigma_i(t) = 1$  if  $x_i(t) < \theta$  and  
478  $x_i(t-1) > \theta$ , and  $\sigma_i(t) = 0$  otherwise. The second method assigns the values 1 and  $-1$  to all time  
479 points below or above the threshold, respectively:  $\sigma_i(t) = 1$  if  $x_i(t) < \theta$ , and  $\sigma_i(t) = -1$  otherwise.

480 The first and the second procedure result in sparse and dense binary activity, respectively. We used  
481 the sparse and dense methods to construct coupling-MEMs and pairwise-MEM, respectively. This  
482 was to meet the assumptions of the model inference (see [Appendix](#)).

### 483 ***k*-means classification**

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

### 493 **Maximum entropy models (MEMs)**

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

$$499 \quad \langle \mathcal{O}_i \rangle_{\text{model}} = \sum_{\{\boldsymbol{\sigma}\}} P(\boldsymbol{\sigma}) \mathcal{O}_i(\boldsymbol{\sigma}), \quad (5)$$

500 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*  
501 ( $1 \leq t \leq n$ ). We search for the model distribution  $P(\boldsymbol{\sigma})$  that does less assumptions, i.e., the one that  
502 has maximal entropy  $S = -\sum_{\{\boldsymbol{\sigma}\}} P(\boldsymbol{\sigma}) \ln P(\boldsymbol{\sigma})$ . Thus, the problem is equivalent to maximizing a function  
503 (the entropy) given some constraints on the expectation values of the observables, a problem that can  
504 be generally solved using Lagrange multipliers. The maximum entropy distribution has the general  
505 form:

$$506 \quad P(\boldsymbol{\sigma}) = \frac{1}{Z} \exp \left( \sum_{i=1}^L \Omega_i \mathcal{O}_i(\boldsymbol{\sigma}) \right) = \frac{1}{Z} \exp(-E(\boldsymbol{\sigma})), \quad (6)$$

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

511 *Linear coupling-MEM.* First, we considered the MEM that is consistent with the probability  
512 distribution  $P(K)$ , the average activations  $\langle \sigma_i \rangle$ , and the linear coupling between  $\sigma_i$  and  $K$ , i.e.,  $\langle \sigma_i K \rangle$   
513 (which relates to  $z_i$ ). As shown in [Gardella et al. \(29\)](#) the resulting energy function is given as:  $E(\boldsymbol{\sigma}) =$   
514  $-\sum_{i=1}^N (\alpha_i + \beta_K + \gamma_i) \sigma_i$ . The model parameters  $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_N]$ ,  $\boldsymbol{\beta} = [\beta_0, \dots, \beta_N]$ , and  $\boldsymbol{\gamma} =$   
515  $[\gamma_1, \dots, \gamma_N]$  are Lagrange multipliers associated to the constrained observables  $\langle \sigma_i \rangle$ ,  $P(K)$ , and  $\langle \sigma_i K \rangle$ ,  
516 respectively.

517 *Non-linear coupling-MEM.* The above model can be extended to include the non-linear coupling  
518 between  $\sigma_i$  and  $K$ . The complete coupling between  $\sigma_i$  and  $K$  is provided by the joint probability  
519 distributions of  $\sigma_i$  and  $K$ , i.e.,  $P(\sigma_i, K)$ , which is the target observable of the non-linear coupling-  
520 MEM. In this case, the energy is given as  $E(\boldsymbol{\sigma}) = -\sum_{i=1}^N H_{i,K(\boldsymbol{\sigma})} \sigma_i$  (29), where  $K(\boldsymbol{\sigma})$  is the number  
521 of active ROIs in pattern  $\boldsymbol{\sigma}$  and the parameters  $H_{i,K(\boldsymbol{\sigma})}$  are associated to the constrained observables  
522  $\langle \sigma_i \delta_{K,k} \rangle$ , where  $P(K = k) = \langle \delta_{K,k} \rangle$  and  $\delta_{K,k}$  is the Kronecker's delta. The linear model is a special  
523 case of this model with  $H_{i,K(\boldsymbol{\sigma})} = \alpha_i + \beta_K + \gamma_i$ .

524 *Pairwise-MEM.* The third model we considered is the one that targets the activation rates ( $\langle \sigma_i \rangle$ )  
525 and the pairwise correlations ( $\langle \sigma_i \sigma_j \rangle$ ) of the data. The resulting energy function of the maximum  
526 entropy distribution is given as  $E(\boldsymbol{\sigma}) = -\sum_{i=1}^N h_i \sigma_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N J_{ij} \sigma_i \sigma_j$  (30, 31). The model  
527 parameter  $h_i$ , called intrinsic bias, represents the intrinsic tendency of neuron  $i$  towards activation  
528 ( $\sigma_i = +1$ ) or silence ( $\sigma_i = -1$ ) and the parameter  $J_{ij}$  represents the effective interaction between  
529 neurons  $i$  and  $j$ .

530 The parameters of the coupling-MEMs and pairwise-MEMs were estimated from the data using  
531 likelihood (29) and pseudo-likelihood (32) maximization, respectively (see [Appendix](#)).

## 532 **Macroscopic quantities**

533 The analysis of the learned MEMs provides relevant properties of the collective activity. These  
534 quantities derive from the Boltzmann distribution and they are interpretable in the framework of  
535 statistical physics. The description and calculation of these quantities are presented in the [Appendix](#)  
536 in detail. Briefly, we studied the system's Helmholtz *free energy*, *susceptibility*, and *heat capacity*.  
537 The free energy  $F$  is given by the difference between the average energy and the entropy, i.e.  $F =$   
538  $\langle E \rangle - S = -\ln(Z)$ ; it quantifies the useful energy that is obtainable from the system. The  
539 susceptibility  $\chi$  relates to the diversity of population states, i.e.  $\chi = \text{var}(K)$ , but, importantly, it also  
540 relates to the system's response to intrinsic or external inputs (see [Appendix](#) and [Fig. S7](#)). The heat  
541 capacity  $C_h$  quantifies the diversity of accessible energy states, i.e.  $C_h = \text{var}(E)$ . The heat capacity  
542 measures the size of the dynamic repertoire of the system. Furthermore, a parameter  $T$ , that scales all  
543 model parameters ( $\boldsymbol{\Omega} \rightarrow \boldsymbol{\Omega}/T$ ), can be introduced to study the effect of a change in the system's  
544 disorder ("temperature") on the repertoire of accessible energy states, i.e., the function  $C_h(T) =$   
545  $\text{var}(E)/T^2$ . This function is informative of the state of the system in terms of criticality: a maximum  
546 of the heat capacity close to  $T_{\text{max}} = 1$  suggests that the observed system is likely to be close to a

547 critical state, whereas  $T_{\max} < 1$  and  $T_{\max} > 1$  indicate super-critical and sub-critical dynamics,  
548 respectively (31, 33, 34) (see Appendix and Fig. S8).

### 549 Fisher information matrix

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

$$554 \quad G_{kl} = - \sum_{\{\boldsymbol{\sigma}\}} P(\boldsymbol{\sigma}|\boldsymbol{\Omega}) \frac{\partial^2 \ln P(\boldsymbol{\sigma}|\boldsymbol{\Omega})}{\partial \Omega_k \partial \Omega_l}. \quad (7)$$

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

$$557 \quad G_{kl} = \frac{\partial^2 \ln Z}{\partial \Omega_k \partial \Omega_l} = - \frac{\partial^2 F}{\partial \Omega_k \partial \Omega_l}. \quad (8)$$

558 To quantify the sensitivity of the different parameters we decomposed the FIM into eigenvectors,  
559 noted  $\mathbf{v}_1, \dots, \mathbf{v}_L$ , and measured the sensitivity of a given parameter  $i$  by its absolute contribution to  
560 the first eigenvector, i.e.  $|\mathbf{v}_1(i)|$ .

### 561 Statistical analysis

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

568

569 **Acknowledgements:** APA, BJ, and GD received funding from the FLAG-ERA JTC (PCI2018-  
570 092891). GD acknowledges funding from the European Union's Horizon 2020 FET Flagship Human  
571 Brain Project under Grant Agreement 785907 HBP SGA1, SGA2 and SGA3, the Spanish Ministry  
572 Research Project PSI2016-75688-P (AEI/FEDER), the Catalan Research Group Support 2017 SGR  
573 1545, and AWAKENING (PID2019-105772GB-I00, AEI FEDER EU) funded by the Spanish  
574 Ministry of Science, Innovation and Universities (MCIU), State Research Agency (AEI) and  
575 European Regional Development Funds (FEDER). MLK is supported by the ERC Consolidator  
576 Grant: CAREGIVING (n. 615539), Center for Music in the Brain, funded by the Danish National  
577 Research Foundation (DNRF117), and Centre for Eudaimonia and Human Flourishing funded by the  
578 Pettit and Carlsberg Foundations. BJ received funding from Fondation Bettencourt-Schueller,

579 Université Paris-Saclay (UVSQ), Fondation de France, Collège de France and from INSERM. CMS  
580 was supported by Comisión Nacional de Investigación Ciencia y Tecnología (CONICYT, currently  
581 ANID) through Programa Formacion de Capital Avanzado (PFCHA), Doctoral scholarship Becas  
582 Chile: CONICYT PFCHA/DOCTORADO BECAS CHILE/2016 - 72170507.

583

584 **Competing interests:** No competing interests declared.

585

586 **Author Contributions:** APA designed research, analyzed the data, studied the models, and wrote  
587 the manuscript; LU performed the experiments and curated the data; ND studied the implementation  
588 of the models; CMS curated data; MK analyzed the data and provided data visualization codes; BJ  
589 designed and supervised the experiments; GD designed and supervised research. All authors  
590 discussed the results and contributed to the editing of the manuscript.

591

592

## 593 **References**

594 [1] B. Biswal, F.Z. Yetkin, V.M. Haughton, J.S. Hyde, Functional connectivity in the motor cortex  
595 of resting human brain using echo-planar MRI. *Magn. Reason. Med.* **34**(4), 537–541 (1995).

596 [2] M.D. Fox, M.E. Raichle, Spontaneous fluctuations in brain activity observed with functional  
597 magnetic resonance imaging. *Nat. Rev. Neurosci.* **8**, 700–711 (2007).

598 [3] C. Chang, G.H. Glover, Time-frequency dynamics of resting-state brain connectivity measured  
599 with fMRI. *Neuroimage* **50**, 81–98 (2010).

600 [4] M.T. Alkire, A.G. Hudetz, G. Tononi, Consciousness and anesthesia. *Science* **322**, 876–80 (2008).

601 [5] F. Ferrarelli et al., Breakdown in cortical effective connectivity during midazolam-induced loss  
602 of consciousness. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2681–2686 (2010).

603 [6] M. Boly et al., Connectivity changes underlying spectral EEG changes during propofol-induced  
604 loss of consciousness. *J. Neurosci.* **32**, 7082–90 (2012).

605 [7] P. Guldenmund et al., Thalamus, brainstem and salience network connectivity changes during  
606 propofol-induced sedation and unconsciousness. *Brain Connect.* **3**, 273–85 (2013).

607 [8] P. Barttfeld et al., Signature of consciousness in the dynamics of resting-state brain activity. *Proc.*  
608 *Natl. Acad. Sci. U.S.A.* **112**, 887–92 (2015).

609 [9] L. Uhrig, D. Janssen, S. Dehaene, B. Jarraya, Cerebral responses to local and global auditory  
610 novelty under general anesthesia. *Neuroimage* **141**, 326–40 (2016).

- 611 [10] M.D. Greicius et al., Persistent default-mode network connectivity during light sedation. *Hum.*  
612 *Brain. Mapp.* **29**, 839–47 (2008).
- 613 [11] L. Uhrig et al. Resting-state Dynamics as a Cortical Signature of Anesthesia in Monkeys.  
614 *Anesthesiology* **129**(5), 942–958 (2018).
- 615 [12] B.B. Machta, R. Chachra, M.K. Transtrum, J.P. Sethna, Parameter space compression underlies  
616 emergent theories and predictive models. *Science* **342**, 604–607 (2013).
- 617 [13] D. Panas et al., Sloppiness in spontaneously active neuronal networks. *J. Neurosci.* **35**, 8480–  
618 8492 (2015).
- 619 [14] A. Ponce-Alvarez, G. Mochol, A. Hermoso-Mendizabal, J. de la Rocha, G. Deco, Cortical state  
620 transitions and stimulus response evolve along stiff and sloppy parameter dimensions, respectively.  
621 *eLife* **9**, e53268 (2020).
- 622 [15] D. Chialvo, Emergent complex neural dynamics: the brain at the edge. *Nature Physics* **6**, 744–  
623 750 (2010).
- 624 [16] W.L. Shew, H. Yang, S. Yu, R. Roy, D. Plenz, Information capacity and transmission are  
625 maximized in balanced cortical networks with neuronal avalanches. *J. neurosci.* **31**, 55–63 (2011).
- 626 [17] W.L. Shew, D. Plenz, The functional benefits of criticality in the cortex. *Neuroscientist* **19**, 88–  
627 100 (2013).
- 628 [18] J. Hidalgo et al., Information-based fitness and the emergence of criticality in living systems.  
629 *Proc. Natl Acad. Sci. USA* **111**, 10095–10100 (2015).
- 630 [19] E. Tagliazucchi et al., Large-scale signatures of unconsciousness are consistent with a departure  
631 from critical dynamics. *J. R. Soc. Interface* **13**(114), 20151027 (2016).
- 632 [20] T. Fekete, D.B. Omer, K. O'Hashi, A. Grinvald, C. van Leeuwen, O. Shriki, Critical dynamics,  
633 anesthesia and information integration: Lessons from multi-scale criticality analysis of voltage  
634 imaging data. *NeuroImage* **183**, 919–933 (2018).
- 635 [21] A.G. Casali et al., A theoretically based index of consciousness independent of sensory  
636 processing and behavior. *Sci. Transl. Med.* **5**, 198ra105 (2013).
- 637 [22] A.G. Hudetz, X. Liu, S. Pillay S, Dynamic repertoire of intrinsic brain states is reduced in  
638 propofol-induced unconsciousness. *Brain Connect.* **5**, 10–22 (2015).
- 639 [23] M. Schartner, A. Seth, Q. Noirhomme, M. Boly, M.A. Bruno, S. Laureys, A. Barrett, Complexity  
640 of Multi-Dimensional Spontaneous EEG Decreases during Propofol Induced General Anaesthesia.  
641 *PLoS ONE* **10**(8), e0133532 (2015).

- 642 [24] G. Solovey et al., Loss of consciousness is associated with stabilization of cortical activity. *J.*  
643 *Neurosci.* **35**, 10866–10877 (2015).
- 644 [25] H. Lee et al., Relationship of critical dynamics, functional connectivity, and states of  
645 consciousness in large-scale human brain networks. *NeuroImage* **188**, 228-238 (2019).
- 646 [26] E. Tagliazucchi, P. Balenzuela, D. Fraiman, D.R. Chialvo, Criticality in large-scale brain fMRI  
647 dynamics unveiled by a novel point process analysis. *Front. Physiol.* **3**, 15 (2012).
- 648 [27] T. Watanabe et al., A pairwise maximum entropy model accurately describes resting-state human  
649 brain networks. *Nat. Commun.* **4**, 1370 (2013).
- 650 [28] S. Tanabe et al., Altered global brain signal during physiologic, pharmacologic, and pathologic  
651 states of unconsciousness in humans and rats. *Anesthesiology* **132**, 1392–1406 (2020).
- 652 [29] C. Gardella, O. Marre, T. Mora, A Tractable Method for Describing Complex Couplings between  
653 Neurons and Population Rate. *eNeuro* **3**(4), 0160–15 (2016).
- 654 [30] E. Schneidman, M.J. Berry, R. Segev, W. Bialek, Weak pairwise correlations imply strongly  
655 correlated network states in a neural population. *Nature* **440**, 1007–1012 (2006).
- 656 [31] G. Tkačik et al., Thermodynamics and signatures of criticality in a network of neurons. *Proc.*  
657 *Natl. Acad. Sci.* **112**, 11508–11513 (2015).
- 658 [32] T. Ezaki, T. Watanabe, M. Ohzeki, N. Masuda (2017) Energy landscape analysis of  
659 neuroimaging data. *Philos. Trans. R. Soc. A* **375**, 20160287 (2017).
- 660 [33] G. Tkačik, O. Marre, D. Amodei, E. Schneidman, W. Bialek, M.J. Berry II, Searching for  
661 collective behavior in a large network of sensory neurons. *PLoS Comput. Biol.* **10**, e1003408  
662 (2014).
- 663 [34] G. Hahn et al., Spontaneous cortical activity is transiently poised close to criticality. *PLoS*  
664 *Comput. Biol.* **13**, e1005543 (2017).
- 665 [35] M. Okun et al., Diverse coupling of neurons to populations in sensory cortex. *Nature* **521**, 511–  
666 515. (2015).
- 667 [36] K.K. Kaisti et al., Effects of surgical levels of propofol and sevoflurane anesthesia on cerebral  
668 blood flow in healthy subjects studied with positron emission tomography. *Anesthesiology* **96**, 1358–  
669 1370 (2002).
- 670 [37] A. Luppi et al., E.A. Stamatakis, Consciousness-specific dynamic interactions of brain  
671 integration and functional diversity. *Nature communications* **10**(1), 4616 (2019).

- 672 [38] F.D. Juengling, J. Kassubek, H.J. Huppertz, T. Krause, T. Els, Separating functional and  
673 structural damage in persistent vegetative state using combined voxel-based analysis of 3-D MRI and  
674 FDG-PET. *J. Neurol. Sci.* **228**, 179–184 (2005).
- 675 [39] A.D. Craig, How do you feel — now? The anterior insula and human awareness, *Nature Reviews*  
676 *Neuroscience* **10**, 59–70 (2009).
- 677 [40] M. Afrasiabi et al., More than just front or back: Parietal-striatal-thalamic circuits predict  
678 consciousness level. Biorxiv: 10.1101/2020.04.07.030429 (9 April 2020).
- 679 [41] J.F. Storm et al., Consciousness regained: disentangling mechanisms, brain systems, and  
680 behavioral responses. *J. Neurosci.* **37**, 10882–10893 (2017).
- 681 [42] M. Massimini, F. Ferrarelli, S. Sarasso, G. Tononi, Cortical mechanisms of loss of  
682 consciousness: insight from TMS/EEG studies. *Arch. Ital. Biol.* **150**, 44–55 (2012).
- 683 [43] S. Sarasso et al., Quantifying cortical EEG responses to TMS in (un)consciousness. *Clin. EEG*  
684 *Neurosci.* **45**, 40–49 (2014).
- 685 [44] S. Sarasso et al., Consciousness and complexity during unresponsiveness induced by propofol,  
686 xenon, and ketamine. *Curr. Biol.* **25**, 3099–3105 (2015).
- 687 [45] M. Napolitani et al., Transcranial magnetic stimulation combined with high-density EEG in  
688 altered states of consciousness. *Brain Inj.* **28**, 1180–1189 (2014).
- 689 [46] V.L. Ginzburg, L.D. Landau, On the theory of superconductors. *Zh Eksp Teor Fiz* **20**, 1064–  
690 1082. English translation in: Landau LD (1965) Collected papers, Oxford: Pergamon Press.
- 691 [47] J. Sitt et al., Large scale screening of neural signatures of consciousness in patients in a vegetative  
692 or minimally conscious state. *Brain* **137**(8), 2258–2270 (2014).
- 693 [48] X.S. Zhang, R.J. Roy, E.W. Jensen, EEG complexity as a measure of depth of anesthesia for  
694 patients. *IEEE Trans. Biomed. Eng.* **48**, 1424–1433 (2001).
- 695 [49] S. Singh, S. Bansal, G.Kumar, I. Gupta, J.R. Thakur, Entropy as an indicator to measure depth  
696 of anaesthesia for laryngeal mask airway (LMA) insertion during sevoflurane and propofol  
697 anaesthesia. *J. Clin. Diagn. Res.* **11**, UC01–UC03 (2017).
- 698 [50] M. Sarà, F. Pistoia, P. Pasqualetti, F. Sebastiano, P. Onorati, P.M. Rossini, Functional isolation  
699 within the cerebral cortex in the vegetative state: a nonlinear method to predict clinical outcomes.  
700 *Neurorehabil. Neural Repair* **25**, 35–42 (2011).
- 701 [51] S. Frey, D.N. Pandya, M.M. Chakravarty, L. Bailey, M. Petrides, D.L. Collins, An MRI based  
702 average macaque monkey stereotaxic atlas and space (MNI monkey space). *Neuroimage* **55**(4), 1435–  
703 1442 (2011).

- 704 [52] K. Shen et al., A macaque connectome for large-scale network simulations in TheVirtualBrain.  
705 *Scientific Data* **6**, 123 (2019).
- 706 [53] N.T : Markov et al., A weighted and directed interareal connectivity matrix for macaque  
707 cerebral cortex. *Cereb. Cortex* **24**, 17–36 (2014).
- 708 [54] C. Chatelain, D. Karevski, Probability distributions of the work in the 2D-Ising model, *J. Stat.*  
709 *Mech.: Theory Exp.* **18**, 06005 (2006).
- 710 [55] G. Tkacik, E. Schneidman, M.J. Berry II, W. Bialek (2006) Ising models for networks of real  
711 neurons. arXiv:q-bio/0611072
- 712 [56] J. Besag, Statistical analysis of non-lattice data. *J. R. Stat. Soc. D* **24**, 179–195 (1975).
- 713

## 714 **Supplementary Materials**

715

716 **Includes:**

717 Appendix

718 Supplementary Figures S1 to S8

719

## 720 **Appendix**

721

### 722 **Anesthesia Protocols**

723 Monkeys received anesthesia either with propofol, ketamine, or sevoflurane (11). Levels of  
724 anesthesia were defined using a clinical score and continuous electroencephalography monitoring.  
725 Under ketamine, deep propofol anesthesia and deep sevoflurane anesthesia, monkeys stopped  
726 responding to all stimuli, reaching a state of general anesthesia (11). The 3 animals scanned at a deep  
727 level of ketamine anesthesia received an intramuscular (i.m.) injection of ketamine (20 mg/kg i.m.,  
728 Virbac, France), followed by a continuous intravenous infusion of ketamine (15-16 mg/kg/h i.v.) to  
729 maintain anesthesia. For propofol anesthesia, monkeys were scanned under moderate propofol  
730 sedation and deep propofol anesthesia. The awake monkeys were injected with an intravenous (i.v.)  
731 propofol bolus (5-7.5 mg/kg i.v.; Fresenius Kabi, France) to induce anesthesia, followed by target-  
732 controlled infusion (Alaris PK Syringe pump, CareFusion, CA, USA) of propofol (moderate propofol  
733 sedation: 3.7-4.0 microg/ml; deep propofol anesthesia 5.6-7.2 microg/ml). During the ketamine and  
734 moderate propofol sedation, a muscle-blocking agent was administered (cisatracrium, 0.15 mg/kg  
735 bolus i.v. followed by continuous i.v. infusion at a rate of 0.18 mg/kg/h, GlaxoSmithKline, France)  
736 to avoid artifacts related to potential movements during magnetic resonance imaging (MRI)  
737 acquisition. For sevoflurane anesthesia, monkeys were scanned under moderate and deep sevoflurane  
738 anesthesia. Monkeys received an intramuscular (i.m.) injection of ketamine (20 mg/kg i.m., Virbac,  
739 France) for induction of anesthesia, followed by sevoflurane anesthesia (moderate sevoflurane  
740 anesthesia: I/E: 2,2/2,1 vol% or deep sevoflurane anesthesia: I/E: 4,4/4,0 vol%) (Abbott, France). For  
741 all the anesthesia experiments, monkeys were intubated and ventilated. Heart rate, non-invasive blood  
742 pressure (systolic/diastolic/mean), oxygen saturation (SpO<sub>2</sub>), respiratory rate, end-tidal CO<sub>2</sub>  
743 (EtCO<sub>2</sub>), and cutaneous temperature were monitored (Maglife, Schiller, France) and recorded online  
744 (Schiller, France).

745

### 746 **fMRI Data Acquisition**

747 Monkeys were scanned on a 3-T horizontal scanner (Siemens Tim Trio; TR, 2,400 ms; TE, 20 ms;  
748 and 1.5-mm<sup>3</sup> voxel size; 500 brain volumes per scan session). Before each scanning session, a contrast  
749 agent monocrySTALLINE iron oxide nanoparticle (MION, Feraheme; AMAG Pharmaceuticals; 10  
750 mg/kg, i.v.) was injected into the monkey's saphenous vein (8). Preprocessing of functional images  
751 followed the standard steps described in (8), normalized to the anatomical template of the monkey

752 MNI space (51), and band-pass filtered in the frequency range of interest (0.0025–0.05 Hz). Time-  
753 series were obtained for  $N = 82$  previously defined cortical regions of interest (ROIs) (CoCoMac  
754 Regional Map parcellation). Scans that presented signs of artifacts in time-series or power spectral  
755 density were discarded. The procedure was based on the visual inspection of the time series for all  
756 the nodes, the Fourier transform of each signal. A total of 119 scans were kept for subsequent  
757 analyses, corresponding to different levels of arousal: wakefulness ( $n = 24$  scans), two levels of  
758 propofol sedation (light, LPP,  $n = 21$ , and deep, DPP,  $n = 23$ ), ketamine anesthesia (KETA,  $n = 22$ ),  
759 and two types of sevoflurane anesthesia (SEV2,  $n = 18$ , and SEV4,  $n = 11$ ).

760

## 761 **Anatomical connectivity**

762 We used a fully weighted whole-cortex macaque structural connectivity matrix (connectome) derived  
763 by combining the information from fiber-tracing and tractography (52). The connectome is publicly  
764 available here: <https://zenodo.org/record/1471588#.X44C6dAzY2x>. Briefly, the tractography  
765 algorithm was optimized to best reproduce the weighted but partial-cortex tracer connectome from  
766 Markov et al. (53), before estimating whole-cortex connectome weights. The directed connectome  
767 weights between ROIs of the CoCoMac parcellation were given as the number of streamlines detected  
768 between them, divided by the total number of streamlines that were sent from the seed. Tractography-  
769 derived structural connectivity matrices were averaged across nine macaque monkeys. For details see  
770 Shen et al. (52).

771

## 772 **Macroscopic quantities**

773 The analysis of the learned MEMs provides relevant properties of the collective activity. These  
774 quantities derive from the Boltzmann distribution and they are interpretable in the framework of  
775 statistical physics. Using the distribution  $P(\sigma)$  the mean energy and the entropy are given as:

$$776 \quad \langle E \rangle = \sum_{\{\sigma\}} P(\sigma) E(\sigma), \quad (\text{S1})$$

$$777 \quad S = - \sum_{\{\sigma\}} P(\sigma) \ln P(\sigma). \quad (\text{S2})$$

778 An important quantity describing the system is the Helmholtz free energy defined as:  $F = \langle E \rangle - S$ .  
779 Using the above expressions for the mean energy and the entropy we get:

$$780 \quad F = \sum_{\{\sigma\}} P(\sigma) [E(\sigma) + \ln P(\sigma)] = -\ln(Z) \sum_{\{\sigma\}} P(\sigma) = -\ln(Z). \quad (\text{S3})$$

781 Thus, the free energy can be obtained either by calculating the mean energy and the entropy or directly  
782 by means of the partition function. These two strategies can be more or less difficult depending on

783 the model. For the coupling-MEMs, we measured  $F$  for each scan through  $Z$ , since  $Z$  is tractable and  
 784 the models can be fitted to single-scan data. For the pairwise-MEM, for which  $Z$  is not tractable, we  
 785 performed Metropolis Monte Carlo simulations of the model ( $10^6$  steps) to estimate  $\langle E \rangle$  and  $S$ . The  
 786 Monte Carlo simulations were repeated ten times to estimate uncertainties on these quantities. The  
 787 free energy quantifies the useful energy that can be extracted from the system, called ‘work’ in  
 788 physics (54).

789 Another important quantity is the susceptibility  $\chi$  that measures the fluctuations of the population  
 790 activity, i.e.  $\chi = \text{var}(K)$ . The susceptibility can be also viewed as a measure of the population  
 791 response to a perturbation applied to all units. Assuming that the perturbation  $B$  adds a term to the  
 792 energy, i.e.,  $E(\boldsymbol{\sigma}) + \sum_{i=1}^N B\sigma_i$ , the susceptibility is given by the derivative of  $\langle K \rangle$  with respect to  $B$ ,  
 793 i.e.,  $\chi = \partial\langle K \rangle / \partial B$  (see **Fig S7**). Here, we were interested in the so-called zero-field susceptibility  
 794 obtained for  $B = 0$ . Using the Boltzmann distribution and noting that  $\langle K \rangle = \frac{1}{Z} \frac{\partial Z}{\partial B}$ , it can be shown  
 795 that  $\partial\langle K \rangle / \partial B$  is equal to the variance of  $K$ , and thus:

$$796 \quad \chi = \left. \frac{\partial\langle K \rangle}{\partial B} \right|_{B=0} = \langle K^2 \rangle - \langle K \rangle^2 = \text{var}(K). \quad (\text{S4})$$

797 The zero-field susceptibility measures the spontaneous fluctuations of the population activity. This  
 798 quantity can be measured directly in the coupling-MEMs through the estimated distribution  $P(K)$ . In  
 799 the pairwise-MEM, the variance of  $K$  can be estimated using Metropolis Monte Carlo simulations  
 800 (ten simulations of  $10^6$  steps). In addition, note that  $\langle K \rangle = \frac{1}{Z} \frac{\partial Z}{\partial B} = \frac{\partial \ln Z}{\partial B} = -\frac{\partial F}{\partial B}$ , which implies the work  
 801  $W$  produced by changing the external stimulus from  $B_1$  to  $B_2$ , i.e.,  $W = -\int_{B_1}^{B_2} \langle K \rangle dB$ , relates to the  
 802 variation of the free energy  $\Delta F$ .

803 A last important quantity is the heat capacity. The heat capacity  $C_h$  quantifies the diversity of  
 804 accessible energy states of the system, i.e.,  $C_h = \text{var}(E)$ . A useful related measure is the variation of  
 805 the heat capacity as a function of a scale parameter  $T$ , analogous to temperature in statistical physics.  
 806 This parameter acts as a scaling factor for all model parameters as  $\boldsymbol{\Omega} \rightarrow \boldsymbol{\Omega}/T$ . The ‘‘temperature’’  $T$   
 807 controls the level of disorder and its effects can be understood by examining the system’s energy  
 808 levels (**Fig. S8**). This creates a family of scaled models in which  $T = 1$  corresponds to the MEM  
 809 that was fitted to the data. The heat capacity as a function of  $T$  is given by  $C_h(T) = \text{var}(E)/T^2$  and  
 810 provides useful features of the system. Indeed, it is known that a maximum of the heat capacity close  
 811 to  $T = 1$  suggests that the observed system is likely to be close to a critical state, whereas  $T_{\text{max}} < 1$   
 812 and  $T_{\text{max}} > 1$  indicate super-critical and sub-critical dynamics, respectively (31, 33, 34). Hence, the  
 813 curve  $C_h(T)$  can be used as a tool to assess criticality. The heat capacity measures the size of the  
 814 dynamic repertoire of the system. It not only provides a measure of the system’s complexity, but also  
 815 assess whether the complexity is maximized and whether any reduction of complexity is due to a  
 816 transition to subcritical or supercritical regimes (both regimes result in a decrease of complexity with  
 817 respect to criticality, but through different mechanisms, see **Fig. S8**). For instance, if  $T_{\text{max}} \neq 1$ , this  
 818 means that the system can be re-scaled to increase the complexity of the model dynamics (e.g., if

819  $T_{\max} < 1$ ,  $T$  needs to decrease to reach the maximum heat capacity, indicating supercritical  
820 dynamics).

821 Furthermore it can be shown that the entropy  $S$  of the system is equal the integral of the function  
822  $C_h(T)/T$  from  $T = 0$  to  $T = 1$  (31, 33). This is a useful strategy to calculate the entropy that we  
823 used to compute the free energy when direct access to the partition function was not feasible (i.e., for  
824 the pairwise-MEM). In our study, we calculated  $C_h(T)$  by estimating the variance of the energies  
825 through Monte Carlo simulations of the pairwise-MEM for different  $T$  (five simulations of  $5 \cdot 10^6$  steps  
826 for each  $T$ ).

827 Finally, note that the energies are equal to the patterns' minus log probabilities, or “surprise”, minus  
828 the free energy (a constant), i.e.,  $E(\boldsymbol{\sigma}) = -\ln P(\boldsymbol{\sigma}) + \ln Z$ . Thus, the variance of the energy (heat  
829 capacity) measures the range of surprises of the different collective states. A large heat capacity  
830 allows the system to represent sensory events that occur with a wide range of likelihoods (energy  
831 states that are distributed, numerous, and separable) (55).

832

### 833 Fisher information matrix and free energy

834 We were interested in detecting which parameters have the strongest effect on the collective activity.  
835 To measure how distinguishable two models, with parameters  $\boldsymbol{\Omega}$  and  $\boldsymbol{\Omega} + \delta\boldsymbol{\Omega}$ , are based on their  
836 predictions, we used the Fisher information matrix (FIM). Indeed, the Kullback-Leibler divergence  
837 between the two models can be written as:

$$838 \quad D_{KL}(\boldsymbol{\Omega}; \boldsymbol{\Omega} + \delta\boldsymbol{\Omega}) = G_{kl} \delta\Omega_k \delta\Omega_l + \mathcal{O}(\delta\Omega^3), \quad (\text{S5})$$

839 where  $1 \leq k, l \leq L$ , where  $L$  is the number of parameters, and the matrix  $\mathbf{G}$  is the FIM matrix given  
840 by:

$$841 \quad G_{kl} = - \sum_{\{\boldsymbol{\sigma}\}} P(\boldsymbol{\sigma}|\boldsymbol{\Omega}) \frac{\partial^2 \ln P(\boldsymbol{\sigma}|\boldsymbol{\Omega})}{\partial \Omega_k \partial \Omega_l}. \quad (\text{S6})$$

842 The FIM represents the curvature of the log-likelihood of the model,  $\log P(\boldsymbol{\sigma}|\boldsymbol{\Omega})$ , with respect to the  
843 model parameters, i.e., it quantifies the sensitivity of the model to changes in parameters.

844 Note that the FIM relates to the free energy  $F = -\ln(Z)$ . Indeed, using the Boltzmann distribution,  
845 we have:

$$846 \quad \frac{\partial^2 \ln P}{\partial \Omega_k \partial \Omega_l} = \frac{\partial}{\partial \Omega_k \partial \Omega_l} [-E(\boldsymbol{\sigma}) - \ln(Z)]. \quad (\text{S7})$$

847 In MEMs the energy is given by the parameters (i.e., Lagrange multipliers) and the target observables  
 848 as:  $E(\boldsymbol{\sigma}) = \sum_i \Omega_i \mathcal{O}_i(\boldsymbol{\sigma})$ . Thus, all second derivatives of the energy with respect to the parameters are  
 849 zero,  $\frac{\partial^2 E}{\partial \Omega_k \partial \Omega_l} = 0$ , and we have:

$$850 \quad \frac{\partial^2 \ln P}{\partial \Omega_k \partial \Omega_l} = - \frac{\partial^2 \ln Z}{\partial \Omega_k \partial \Omega_l}. \quad (\text{S8})$$

851 Since the right-hand term of equation 17 does not depend on  $\boldsymbol{\sigma}$ , equation 15 gives:

$$852 \quad G_{kl} = \frac{\partial^2 \ln Z}{\partial \Omega_k \partial \Omega_l} = - \frac{\partial^2 F}{\partial \Omega_k \partial \Omega_l}. \quad (\text{S9})$$

853

854

### 855 **Estimation of maximum entropy models.**

856 Linear and non-linear coupling-MEMs were estimated using the method described in (29). Briefly,  
 857 Newton's method was used to maximize the log-likelihood  $LL$ , by iteratively updating the parameters  
 858 as:

$$859 \quad \boldsymbol{\Omega}^{(k+1)} = \boldsymbol{\Omega}^{(k)} - a \mathbf{G}^{-1} \nabla LL, \quad (\text{S10})$$

860 Where  $\boldsymbol{\Omega}^{(k)}$  is the vector of parameters at iteration  $k$ ,  $a$  is a learning rate,  $\mathbf{G}$  is the FIM with parameter  
 861  $\boldsymbol{\Omega}^{(k)}$ , and  $\nabla LL$  is the gradient of  $LL$  with respect to the model parameters  $\boldsymbol{\Omega}^{(k)}$ , i.e.,  $\nabla LL_i = \frac{\partial LL}{\partial \Omega_i}$ . It can  
 862 be shown that  $\nabla LL_i = \langle \mathcal{O}_i \rangle_{\text{data}} - \frac{\partial \ln Z}{\partial \Omega_i}$ . This method is feasible because, for coupling-MEM, the  
 863 partition function can be analytically obtained from the model parameters and thus providing  $\mathbf{G}$  and  
 864  $\nabla LL$ ; see details in (29). The algorithm was stopped when the estimation errors for all constraint  
 865 observables became lower than  $10^{-6}$ . We found that this algorithm correctly converged for the sparse  
 866 binarization. The code for learning the coupling-MEMs is available at:  
 867 [https://github.com/ChrisGill/MaxEnt\\_Model\\_Population\\_Coupling](https://github.com/ChrisGill/MaxEnt_Model_Population_Coupling).

868 The pairwise-MEM was learned using pseudo-likelihood maximization (32). This method  
 869 approximates the likelihood function:

$$870 \quad L(\boldsymbol{\Omega}) \approx \prod_{t=1}^n \prod_{i=1}^N \tilde{P}(\sigma_i | \boldsymbol{\Omega}, \sigma_{\neq i}), \quad (\text{S11})$$

871 where  $n$  is the number of time points of the data and  $\tilde{P}(\sigma_i | \boldsymbol{\Omega}, \sigma_{\neq i})$  is the conditional Boltzmann  
 872 distribution for a unit  $\sigma_i$ .

873 This was done by updating the biases and couplings as:

874 
$$h_i^{(k+1)} = h_i^{(k)} + a[\langle \sigma_i \rangle_{\text{data}} - \langle \sigma_i \rangle_{\text{mf}}], \quad (\text{S12})$$

875 
$$J_{ij}^{(k+1)} = J_{ij}^{(k)} + a[\langle \sigma_i \sigma_j \rangle_{\text{data}} - \langle \sigma_i \sigma_j \rangle_{\text{mf}}], \quad (\text{S13})$$

876 Where  $k$  denotes the updating iteration (up to  $5 \cdot 10^3$ ) and  $\langle \cdot \rangle_{\text{mf}}$  are the expected values using the  
877 mean-field approximation:

878 
$$\langle \sigma_i \rangle_{\text{mf}} = \frac{1}{n} \sum_{t=1}^n \tanh \left( h_i + \sum_{j \neq i} J_{ij} \sigma_j(t) \right), \quad (\text{S14})$$

879 
$$\langle \sigma_i \sigma_j \rangle_{\text{mf}} = \frac{1}{n} \sum_{t=1}^n \sigma_j(t) \tanh \left( h_i + \sum_{k \neq i} J_{ik} \sigma_k(t) \right), \quad (\text{S15})$$

880 where  $\sigma_i(t)$  is the activity of ROI  $i$  (taking values 1 or  $-1$ ). The estimator of the maximum pseudo-  
881 likelihood approaches the maximum-likelihood estimator for  $n \rightarrow \infty$  (56). For this reason, pairwise-  
882 MEMs were fitted to concatenated data for each of the six experimental conditions. Also, since this  
883 method uses a mean-field approximation, it is accurate when the nodes of the network receive many  
884 inputs, thus we used the dense binarization scheme (see Methods). The code for learning the pairwise-  
885 MEMs is available at:

886 <https://royalsocietypublishing.org/doi/suppl/10.1098/rsta.2016.0287>.

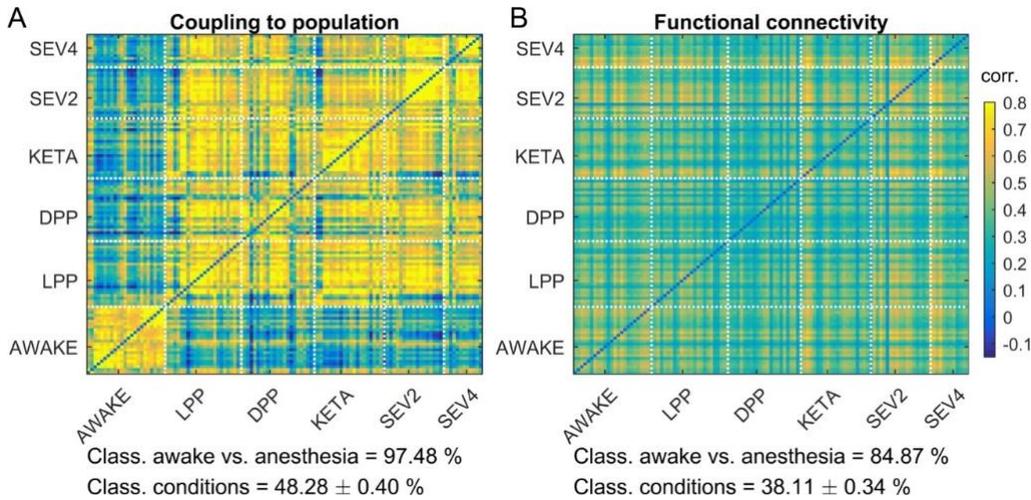
887

888

## 889 Supplementary Figures

890

891



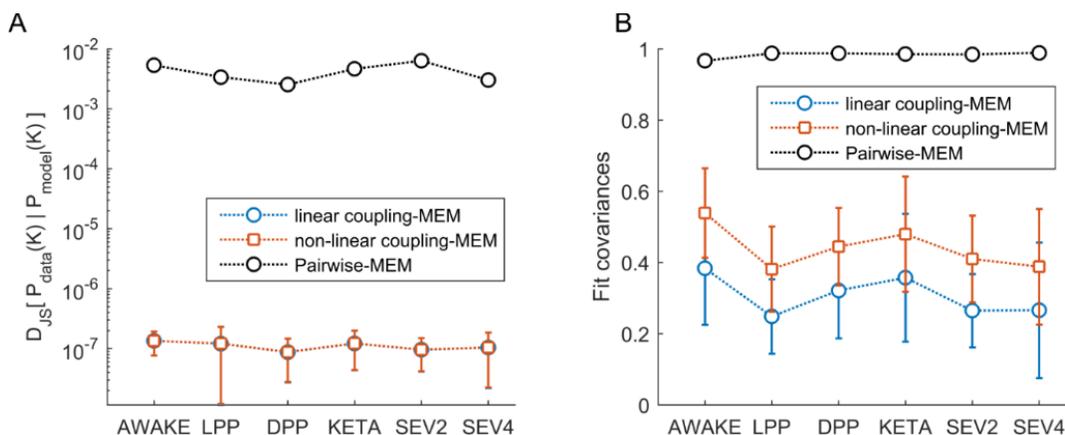
892

893 **Figure S1. Scan-classification using continuous fMRI signals. A-B)** Correlation matrix comparing  
894 the statistics  $z$  and functional correlations, calculated using continuous fMRI signals, among all scans.  
895 For example, in panel (A), the element  $(k, l)$  of the matrix represents the correlation between the  
896 coupling to population vector  $\mathbf{z}$  of scans  $k$  and  $l$ . Coupling to population clearly separated awake and  
897 anesthesia data. Using  $k$ -means, we evaluated how well the different statistics could be used to  
898 classify the awake and anesthetized conditions (chance level: 50%). The classification performance  
899 using the coupling to population statistic was 97.48%, that was significantly higher than using the  
900 functional connectivity (84.87%). Classification of the six experimental conditions was generally  
901 lower, but higher for  $z$  than for functional connectivity (48% vs. 38.11%, chance level: 16.67%).

902

903

904



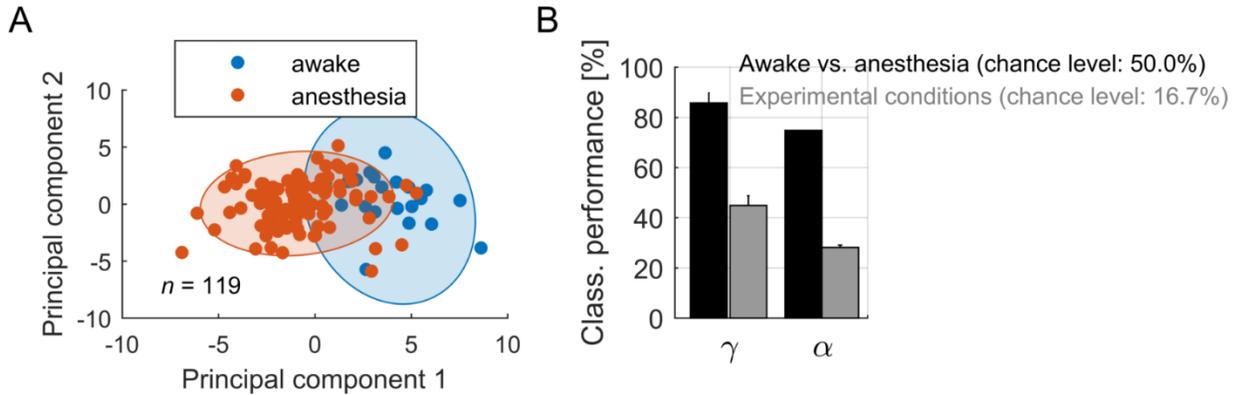
905

906 **Figure S2. Goodness-of-fit and prediction accuracy of MEMs.** We calculated the goodness-of-fit  
 907 and prediction accuracy of the three types MEMs. For the linear and non-linear coupling-MEMs, the  
 908 goodness-of-fit was given by the fit of the  $P(K)$  distribution, which was the target observable of the  
 909 MEMs. For these models, the fit of covariances represent predictions of the models, because  
 910 covariances were not used to construct them. In contrast, for the pairwise-MEM, the goodness-of-fit  
 911 was given by the covariance fit (since covariances were the target of the model's learning step) and  
 912 the fit of  $P(K)$  was a model prediction. **A)** Jensen-Shannon divergence ( $D_{JS}$ ) between the model and  
 913 data  $P(K)$  distributions, for the three types of MEMs (lower values of  $D_{JS}$  indicate better  
 914 approximation of  $P(K)$ ). ANOVA: linear coupling-MEM:  $F_{(5,118)} = 1.3$ ,  $p = 0.289$ ; non-linear  
 915 coupling-MEM:  $F_{(5,118)} = 1.2$ ,  $p = 0.298$ . For the pairwise-MEM, ANOVA was not applicable since  
 916 models were learned using concatenated data. **B)** Covariance fit for the three types of MEMs.  
 917 ANOVA: linear coupling-MEM:  $F_{(5,118)} = 11.4$ ,  $p < 0.001$ ; non-linear coupling-MEM:  $F_{(5,118)} = 16.6$ ,  
 918  $p < 0.001$ . Error bars indicate SEM.

919

920

921

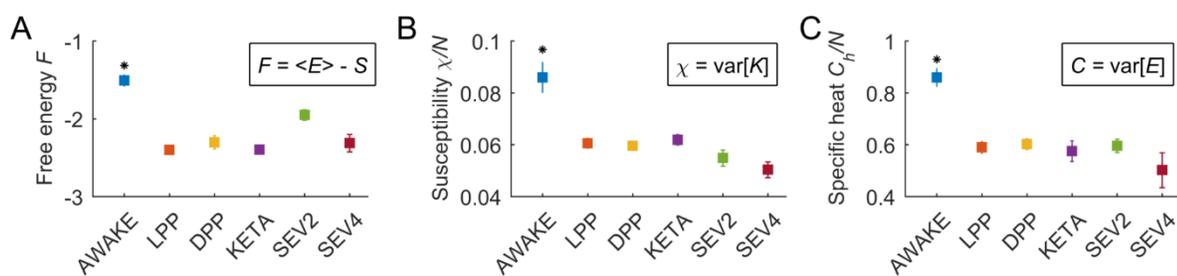


922

923 **Figure S3. Parameters  $\gamma$  predict the state of the brain.** A) PCA analysis showed that the  
924 parameters  $\gamma = [\gamma_1, \dots, \gamma_N]$  from the linear coupling-MEM separated the awake and anesthetized  
925 conditions along the first principal component. Each point represents a scan. B) Using  $k$ -means, we  
926 evaluated how well the different parameters  $\gamma$  and  $\alpha$  could be used to classify the awake and  
927 anesthetized conditions (chance level: 50%). The classification performance based on  $\gamma$  and  $\alpha$   
928 parameters was  $85.6 \pm 4.4\%$ , and  $75.8 \pm 0.0\%$ , respectively. Classification of the six experimental  
929 conditions was  $45.0 \pm 4.0\%$ , and  $27.9 \pm 0.9\%$  based on  $\gamma$  and  $\alpha$  parameters, respectively.

930

931

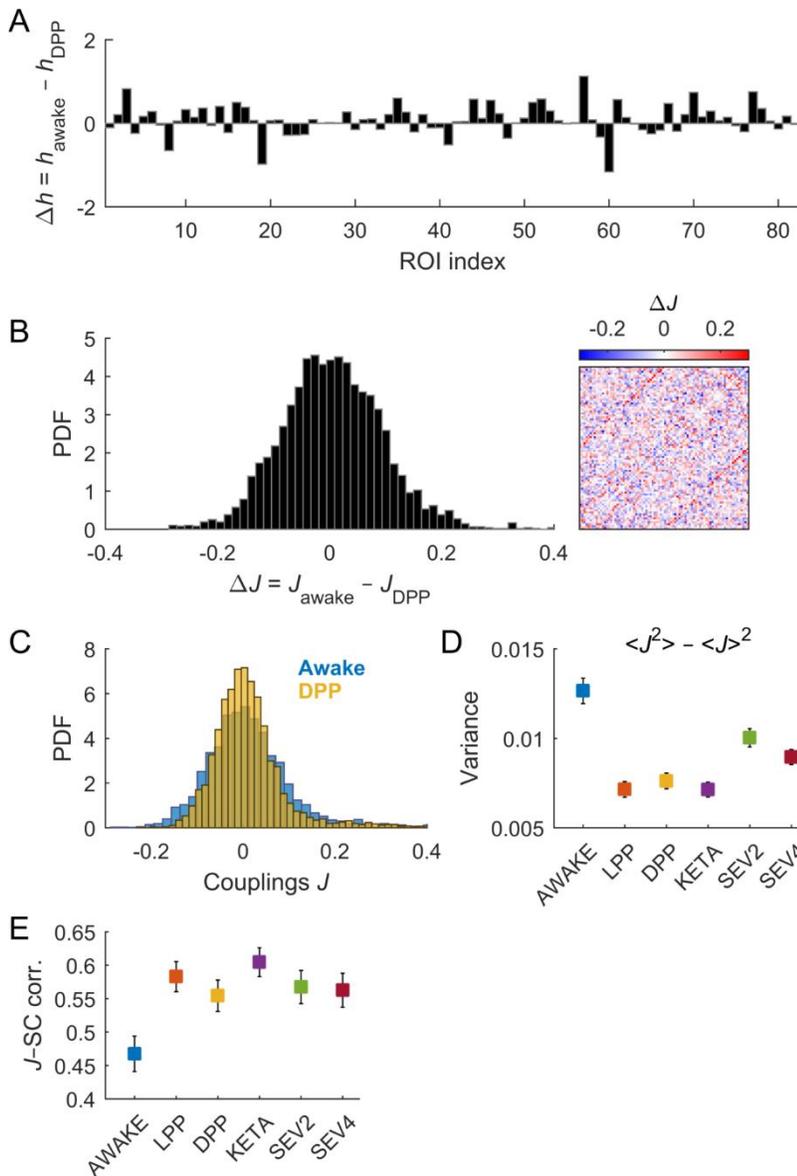


932

933 **Figure S4. Macroscopic quantities using the linear coupling-MEM.** **A)** Free energy  $F$  for the  
934 different conditions,  $F = \langle E \rangle - S = -\ln Z$ . **B)** Susceptibility  $\chi$  for the different conditions,  $\chi =$   
935  $\text{var}(K)$ . **C)** Heat capacity  $C$  for the different conditions,  $C_h = \text{var}(E)$ . In panels (A) to (B), we the  
936 linear model was used. Squares and error bars indicate means and standard deviations across scans,  
937 respectively, and the asterisks indicate significantly different values for the awake condition ( $p <$   
938  $0.001$  one-way ANOVA followed by Tukey's post hoc analysis).

939

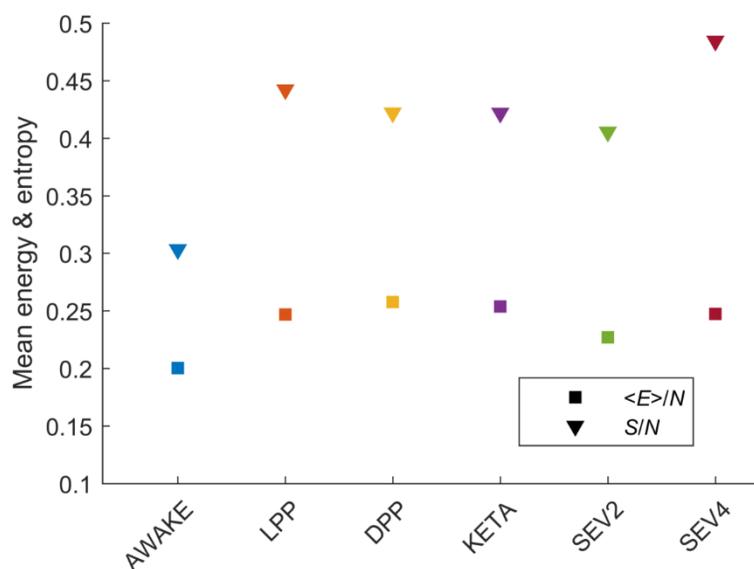
940



941

942 **Figure S5. Change of the pairwise-MEM parameters  $h$  and  $J$  across brain states.** **A)** Change in  
 943 bias model parameters ( $h$ ) between the awake and deep propofol (DPP) conditions for each ROI. **B)**  
 944 Distribution of the change in couplings ( $J$ ) between the awake and deep propofol (DPP) conditions  
 945 for all pairs of ROIs. Inset: the matrix represents the change of the couplings between awake and  
 946 DPP, i.e.,  $\Delta J_{ij} = J_{ij}^{\text{awake}} - J_{ij}^{\text{DPP}}$ . **C)** Distribution of couplings  $J$  in the awake and DPP conditions. **D)**  
 947 Variance of couplings  $J$  for each experimental condition. Error bars indicate bootstrap uncertainties  
 948 (500 repetitions). **E)** Pearson correlation between the coupling matrix  $J$  and the structural  
 949 connectivity, for each experimental condition. The error bars indicate the correlation coefficient's  
 950 95% confidence interval.

951



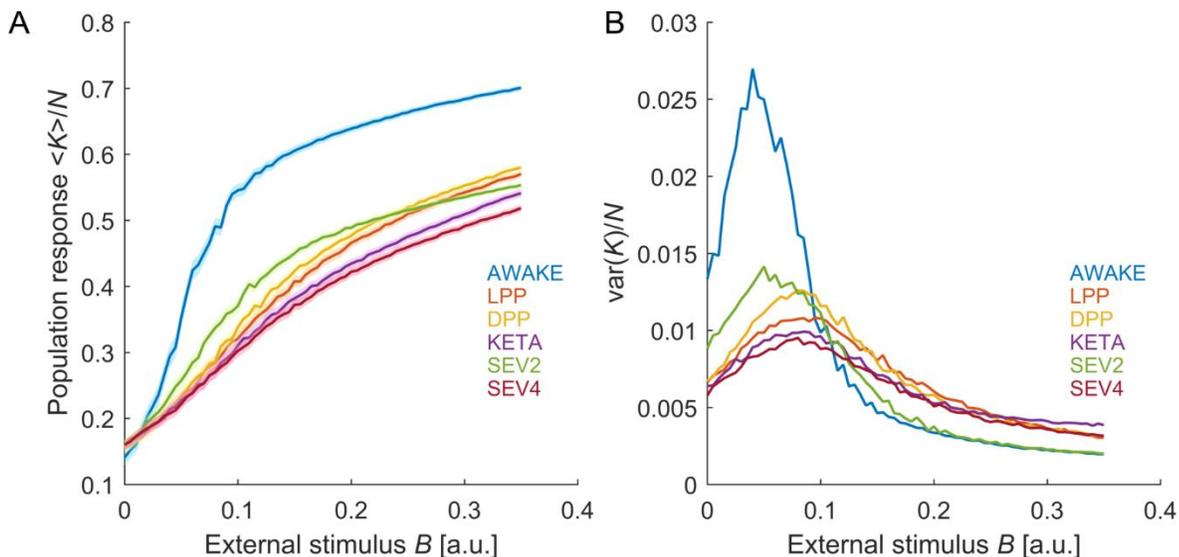
952

953 **Figure S6. Mean energy and entropy for the pairwise-MEM.** The mean energy  $\langle E \rangle$  and the entropy  
954  $S$  of the learned pairwise-MEMs were calculated using Monte Carlo simulations ( $10^6$  steps). To  
955 calculate the entropy, we used the heat capacity as a function of the scaled parameter  $T$ , analogous to  
956 temperature (see Methods).

957

958

959



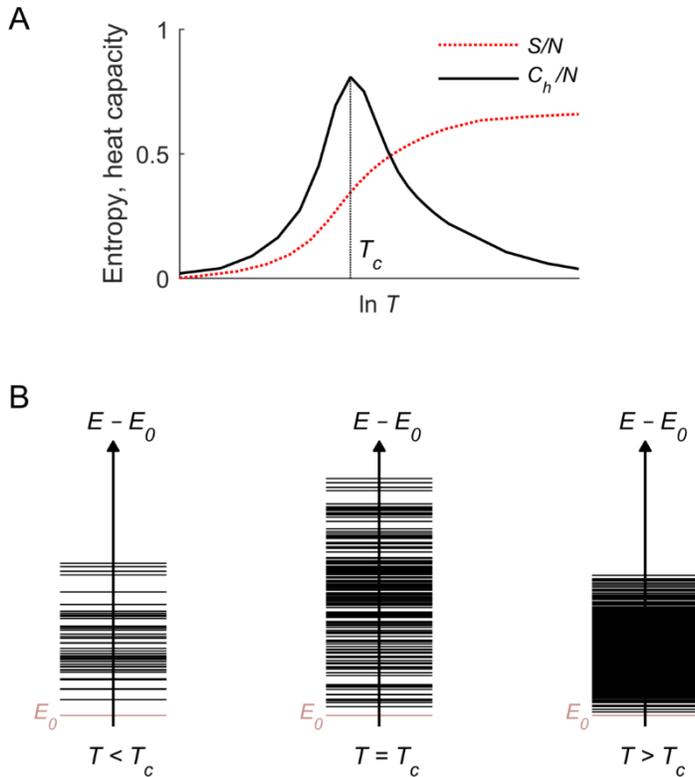
960

961 **Figure S7. Population response to an external stimulus predicted by the pairwise-MEMs.** **A)** An  
962 external stimulus was applied to the learned pairwise-MEMs. The stimulus  $B$  added a term to the  
963 energy as:  $E(\sigma) + \sum_{i=1}^N B\sigma_i$ . We performed Monte Carlo simulations (100 trials of  $5 \cdot 10^4$  steps) for  
964 different values of  $B$  to examine the mean population response  $\langle K \rangle$  as a function of the external  
965 stimulus. In this case, the work  $W$  produced by changing the external stimulus from  $B_1$  to  $B_2$ , i.e.,  
966  $W = - \int_{B_1}^{B_2} \langle K \rangle dB$ , relates to the variation of the free energy  $\Delta F$ . **B)** Variance of the population  
967 activity. In these simulations, inactive ROIs ( $\sigma_i = -1$ ) were set to 0, so that  $K$  represents the number  
968 of active ROIs.

969

970

971



972

973 **Figure S8. Heat capacity and energy levels.** **A)** Heat capacity  $C_h$  and entropy  $S$  of the pairwise-  
 974 MEM, as a function of the scaling parameter  $T$ , analogous to temperature in statistical physics. The  
 975 temperature  $T$  controls the disorder of the system (i.e., the entropy  $S(T)$  increases as a function of  $T$ ).  
 976 If the model learned from the data (i.e.,  $T = 1$ ) maximized the heat capacity, then its dynamics are  
 977 critical. If the temperature that maximizes the heat capacity is larger than 1, the dynamics of the learn  
 978 model are subcritical (the temperature must increase to reach the maximum). Finally, if the  
 979 temperature that maximizes the heat capacity is lower than 1, the dynamics of the learn model are  
 980 supercritical (the temperature must decrease to reach the maximum). **B)** Observed energies  $E$  for  
 981 subcritical, critical, and supercritical dynamics.  $E_0$  correspond to the energy of the state for which all  
 982 units are silent. For subcritical dynamics, the few visited energy levels are sparsely distributed, and  
 983 the variance of energies (i.e., heat capacity) is relatively low. For supercritical dynamics, many energy  
 984 levels are densely distributed, and their variance of energy is also relatively low. For critical  
 985 dynamics, the energy levels are numerous and separable, leading to a maximal variance of energies.

986

987