Response function as a quantitative measure of consciousness in brain dynamics

Wenkang Du[1] and Haiping Huang[1,2,*]

[1]PMI Lab, School of Physics, Sun Yat-sen University, Guangzhou 510275, People's Republic of China

[2]Guangdong Provincial Key Laboratory of Magnetoelectric Physics and Devices, Sun Yat-sen University, Guangzhou 510275, People's Republic of China

(Received 9 September 2025; accepted 13 November 2025; published 3 December 2025)

Understanding the neural correlates of consciousness remains a central challenge in neuroscience. In this study, we investigate the relationship between consciousness and neural responsiveness by analyzing intracranial electrocorticography recordings from nonhuman primates across three distinct states: wakefulness, anesthesia, and recovery. Using a nonequilibrium recurrent neural network (RNN) model, we fit state-dependent cortical dynamics to extract the neural response function as a dynamics complexity indicator. Our findings demonstrate that the amplitude of the neural response function serves as a robust dynamical indicator of conscious state, consistent with the role of a linear response function in statistical physics. Notably, this aligns with our previous theoretical results showing that the response function in RNNs peaks near the transition between ordered and chaotic regimes—highlighting criticality as a potential principle for sustaining flexible and responsive cortical dynamics. Empirically, we find that during wakefulness, neural responsiveness is strong, widely distributed, and consistent with rich nonequilibrium fluctuations. Under anesthesia, response amplitudes are significantly suppressed, and the network dynamics become more chaotic, indicating a loss of dynamical sensitivity. During recovery, the neural response function is elevated, supporting the gradual re-establishment of flexible and responsive activity that parallels the restoration of conscious processing. Our work suggests that a robust, brain-state-dependent neural response function may be a necessary dynamical condition for consciousness, providing a principled framework for quantifying levels of consciousness in terms of nonequilibrium responsiveness in the brain.

DOI: 10.1103/6lx7-qsv7

I. INTRODUCTION

Understanding the neural basis of consciousness remains one of the most challenging questions in neuroscience [1–3]. Despite remarkable advances in neuroimaging and electrophysiology, reliably distinguishing conscious from unconscious brain states purely on the basis of measurable neural activity remains elusive [4–6]. Clinically, this gap limits our ability to monitor depth of anesthesia, diagnose disorders of consciousness, and understand the nature of brain-state transitions among wakefulness, sleep, anesthesia, and coma. Brain, as a hierarchical system of interacting neurons, lacks a clear theoretical foundation to determine what should be defined and measured in mathematical terms [7].

A growing number of works suggest that consciousness is not defined by static structural features alone, but by the brain dynamics far from equilibrium [8–12]. Empirical evidences now indicate that conscious states are underpinned by rich, flexible, and metastable patterns of neural activity, and thus, dynamics at the edge of instability are essential for maintaining consciousness [5]. For example, dynamical system approaches have revealed that cortical activity during wakefulness is poised near the edge of instability—a critical regime that maximizes responsiveness to external stimuli [13,14]. Complexity analysis showed that attractors in conscious and anesthesia-induced unconscious states exhibit significantly different shapes, which affects the information processing capability [5,15].

These empirical observations can be studied within the perspective of nonequilibrium statistical physics. Recent studies show that the human brain, like all living systems, fundamentally operates out of thermodynamic equilibrium [8,12]. Measures such as entropy production, broken detailed balance, and probability flux irreversibility have emerged as principled signatures of cognitive complexity and conscious awareness [9,11]. For instance, healthy conscious brain dynamics break time-reversal symmetry [16], generating a preferred direction in time (i.e., arrow of time), whereas proximity to equilibrium—as in anesthesia, nonrapid eye movement sleep, or neurodegenerative disorders like Alzheimer's disease—is accompanied by a loss of temporal irreversibility [17].

Our recent theoretical works further support this dynamic complexity by showing that in recurrent neural networks (despite random connections among neurons), the neural response function itself peaks exactly at the edge of chaos, where the system exhibits maximal sensitivity to perturbations [18,19]. Within an optimization-based framework

*Contact author: huanghp7@mail.sysu.edu.cn

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

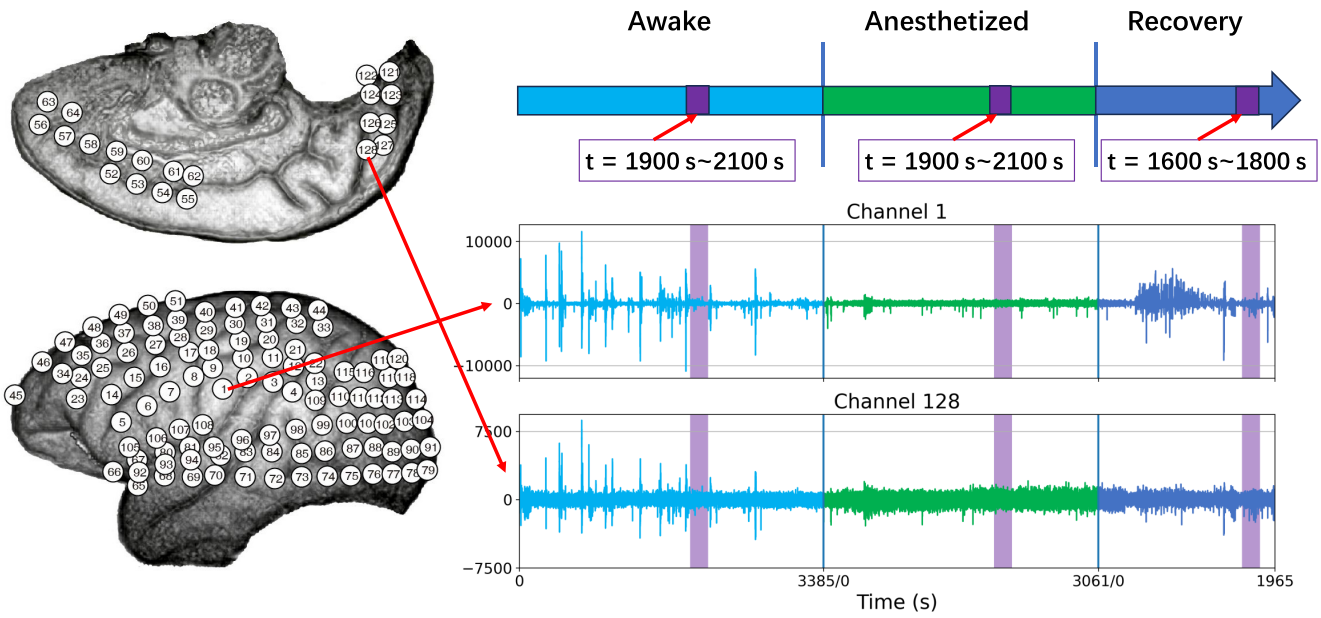


FIG. 1. Sketch of experimental data: two-dimensional ECoG electrode array and neural activity signals for monkey George [24]. Left: Spatial layout of the 128-channel ECoG grid covering frontal, parietal, and temporal cortices (image adapted from Ref. [24]). This map serves as the anatomical reference for interpreting the fitted RNN dynamics across cortical regions. Right: Representative neural activity traces from channels 1 and 128 during three behavioral states—wakefulness, anesthesia, and recovery. The purple shaded region marks the time window used for model fitting in this study (up to 200 s).

for nonequilibrium dynamics, this critical regime naturally emerges as the point where the responsive property of the network is maximized, indicating a deep link between dynamical instability and the network's capacity for flexible information processing. This is also consistently supported by an analysis of cortical electrodynamics [14] showing that the information richness occurs at the edge of chaos, and the conscious brain states are located near the edge of chaos. Another study focusing on a clinical index of conscious level (but on recurrent neural networks) also identifies this picture [20].

In this work, we verify this theoretical concept in real brain dynamics by building a data-driven recurrent neural network, where the couplings between neurons (or brain regions) are learned from a predictive processing of the time series collected from experiments [21]. The recurrent neural network (RNN) model fits high-density electrocorticography (ECoG) recordings from the cortex of nonhuman primates during reversible loss and recovery of consciousness. The main methodology is based on a predictive learning principle [22]. After the network is reconstructed, we obtain the functional connectivity among all involved brain regions. As derived by our previous theory [18], the response function is measured with a tiny current perturbation to the neural dynamics equation. Within the statistical physics framework, we provide a principled and mechanistic approach to quantify how cortical dynamics sustain responsiveness far from equilibrium, and how this capacity deteriorates under anesthesia and re-emerges during recovery. While our modeling currently focuses on primate data, the methodology and analysis may lay the foundation for future extensions to human datasets and for understanding how the responsiveness of nonequilibrium brain dynamics is correlated with conscious awareness (such as cortical activity during wakefulness, nonrapid eye movement sleep, and rapid eye movement sleep [23]).

II. MODEL AND METHODOLOGY

In this section, we describe the network model used to fit the observed time series data. We first give an overall introduction to the training dataset, and then to the RNN model and training method. Finally, we introduce the response function as a brain complexity index and show how to measure this function from the reconstructed networks.

A. Data description

In this study, we analyzed ECoG recordings from a publicly available dataset of nonhuman primates provided by the Neurotycho project [21,24]. The data were collected at the Laboratory for Adaptive Intelligence, Brain Science Institute, RIKEN, from four male monkeys under different states of consciousness, including wakefulness, deep sleep, and anesthesia induced by anesthetic drugs (propofol, ketamine, and ketamine combined with medetomidine). The ECoG array consists of 128 electrodes covering the prefrontal cortice, primary motor and somatosensory cortices, parietal and temporal cortices, and visual cortices, providing high-density cortical coverage. ECoG signals were recorded at a sampling rate of 1 kHz. A sketch of these experimental data is given in Fig. 1.

In the current work, we carried out the analysis of experimental data from a single representative monkey (George), focusing on three distinct recording sessions that capture the transition across conscious wakefulness, induction of anesthesia (loss of consciousness), and recovery (responsive consciousness). While we present detailed results for this ex-

ample, the same analytical framework and conclusions apply consistently to other settings in the same experiment. A key feature of this dataset is the reversible nature of the anesthetic procedure, allowing us to study transitions across distinct levels of consciousness on the same animal.

The time series data are preprocessed in the following way. We applied a simple standardization procedure by subtracting the mean and dividing by the standard deviation for each channel within the time window under consideration.

B. Dynamics model and training

To capture the dynamical properties of cortical activity in different brain states, we fit the ECoG signal using an RNN model [22,25]. The continuous-time dynamics for each unit x_i are given by

$$\tau \frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} J_{ij}\,\phi(x_j) + h_i, \tag{1}$$

where τ is the neuronal time constant, J_{ij} denotes the recurrent connectivity weight from neuron j to i, $\phi(\cdot)$ is a nonlinear activation function (tanh in this work), and h_i represents a weak external sensory perturbation if necessary. In our following analyses, we fix $\tau = 1.0$ (but it can be learned as well); in the case of optimizing τ, an additional parameter $\alpha \equiv \Delta t/\tau$, where Δt is a small time increment used to discretize Eq. (1), can be updated to minimize the training cost specified later. We confirm that allowing τ to be optimized besides $\mathbf{J}$ does not qualitatively change the fitted dynamics or response-function estimates.

To fit ECoG time series, Eq. (1) is discretized with a small time increment Δt as follows:

$$x_i(t+\Delta t) = (1-\Delta t)\,x_i(t) + \Delta t \sum_{j=1}^{N} J_{ij}\,\phi(x_j(t)), \tag{2}$$

where $N = 128$ corresponds to the number of ECoG channels and $\Delta t = 0.001$ corresponds to the sampling interval (1 kHz in raw data). Note that during fitting, the external perturbation is absent.

To capture the dynamic property of interactions among units, the model is trained on short segments: Each training batch contains $T = 2000$ time points (equivalent to 2 s of ECoG data). For each experimental condition (awake, anesthesia, recovery), we repeatedly sample nonoverlapping segments from the empirically identified steady-state portions of the recording. These segments are evenly distributed to span over time windows of 100 and 200 s for a robust estimate of the population response.

The network parameters $\mathbf{J}$ are trained to minimize the discrepancy between the observed trajectory $\mathbf{X}$ and the model prediction $\hat{\mathbf{X}}$ using a mean squared error loss with an ℓ_2-norm regularization:

$$\mathcal{L} = \frac{1}{T}\sum_{t} \|\hat{\mathbf{X}}(t) - \mathbf{X}(t)\|^2 + \lambda_{\text{reg}}\|\mathbf{J}\|_F^2, \tag{3}$$

where $\|\cdot\|_F$ denotes the Frobenius norm and λ_{reg} sets the regularization strength (for the current analysis, we set $\lambda_{\text{reg}} = 10^{-10}$ making the two terms in $\mathcal{L}$ balanced during training). The optimization is performed using the Adam algorithm with a learning rate 0.1. For each minibatch, the RNN recursively predicts the neural state at the next time point, and the neural couplings are updated to minimize the reconstruction error [22].

C. Response function measurement

To quantify the network's dynamical responsiveness, we estimate the linear response function using a standard perturbative approach, by borrowing concepts from physics [18,26]. For the neural dynamics, the response function is formally defined as

$$\chi(t) = \frac{\partial}{\partial h}\langle\phi(\mathbf{x}(t))\rangle, \tag{4}$$

where h is applied at the zero time point, or the moment when the perturbation is applied is set to the starting time point. In practice, $h_i = h$ is applied to all units (a homogeneous perturbation). $\langle\cdot\rangle$ means an ensemble average. Then, the dynamics run after the perturbation according to the following ordinary differential equation:

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{J}\,\phi(\mathbf{x}). \tag{5}$$

The mean population output $\langle\phi(\mathbf{x}(t))\rangle$ can then be computed by considering many simulated trajectories. The empirical response function is then estimated by the slope of the output-perturbation relation:

$$\chi(t) \approx \frac{\Delta\langle\phi(\mathbf{x}(t))\rangle}{\Delta h}. \tag{6}$$

Once the perturbation is sufficiently weak, as assumed in physics, there exists a linear relationship between evoked response and the perturbation [27]. In this sense, the intrinsic difference between spontaneous and evoked population activity leads to a quantitative measure for the detection of awareness.

For each reconstructed functionally connected network, numerical integration is performed using a Runge-Kutta solver with initial conditions set by the recorded ECoG state. To ensure robustness, the final response values are computed as averages across multiple steady-state windows (100–200 s) and multiple perturbation amplitudes (h from 0.0001 to 0.0005). Together, these procedures yield an interpretable estimate of how neural responsiveness varies across brain states, which will be verified in the following, providing a mechanistic link between nonequilibrium cortical dynamics and signatures of conscious state, as implied by our previous theoretical work [18].

III. RESULTS AND DISCUSSION

In this section, we show the results of our methodology applied to the monkey's experimental data, focusing on the transition between conscious and unconscious brain states.

A. RNN model reveals state-dependent network structures

We first assessed how well the RNN model captures the local dynamics of cortical activity in the three consecutive brain

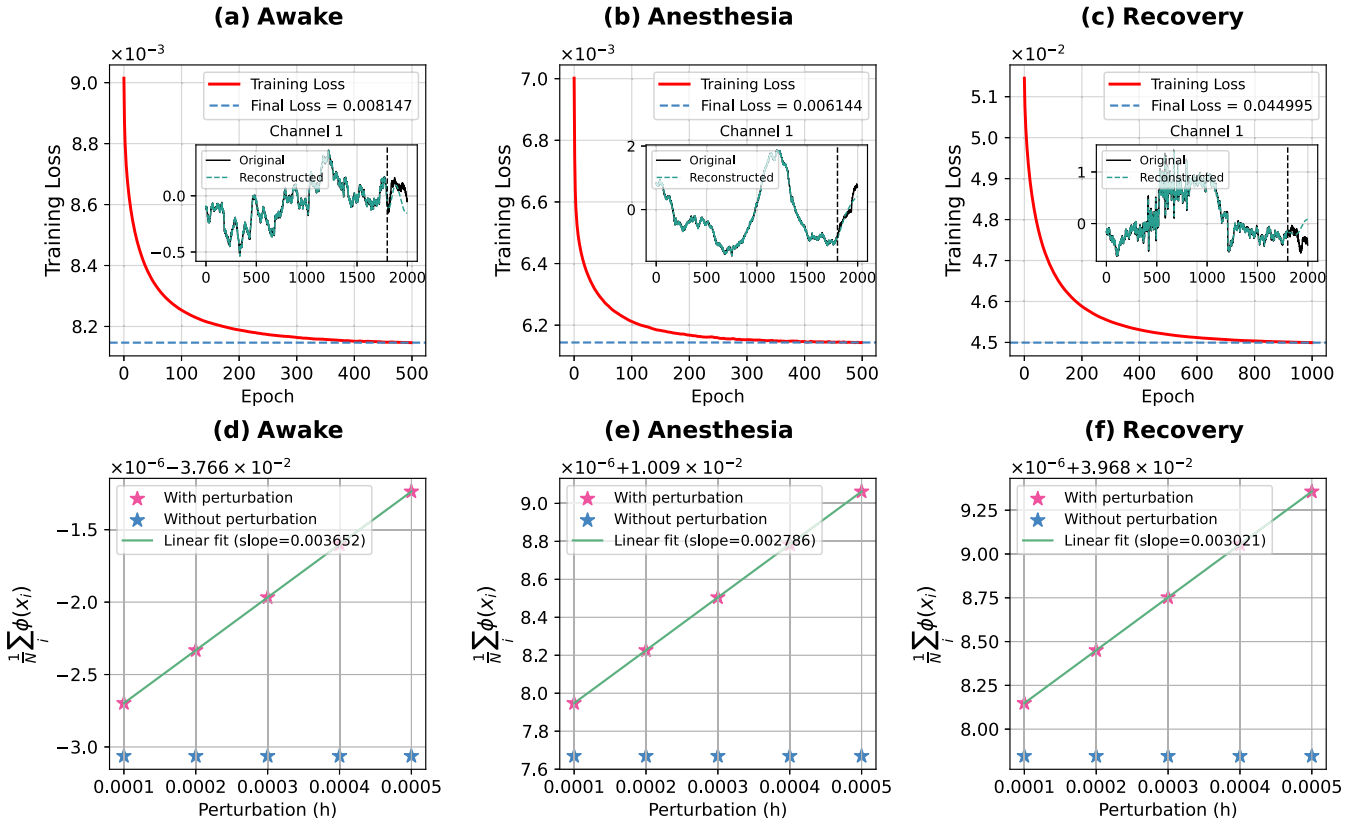


FIG. 2. Fitted RNN results across three brain states: awake, anesthesia, and recovery. (a)–(c) Each panel shows the average training loss within 100-s fitting windows (models were fitted every 2 s). Insets compare the reconstructed and target trajectories. At the 1800th time step (indicated by the vertical dashed line), the reconstructed RNN initialized from the target trajectory runs for the next 200 time steps, in order to evaluate the prediction accuracy. (d)–(f) The measured mean population activity $\langle \phi(\mathbf{x}(t)) \rangle$ as a function of perturbation amplitude h (applied at $t = 0.005$ s, i.e., five time steps away from the moment when the perturbation is applied) for the three different brain states. The results are average over all inferred recurrent neural networks [see panels (a)–(c)]. The corresponding slope reflects the strength of the neural response function defined in the main text.

states with different levels of consciousness: awake, anesthesia, and recovery. The mean reconstruction loss for each 2-s ECoG segment was computed and then averaged over longer time windows (100 s each). As shown in Figs. 2(a)–2(c), the RNN learns very well the target time series in the brain. However, if the learned RNN is initialized from the target [e.g., at the 1800th time step; see the insets of Figs. 2(a)–2(c)], the later dynamics deviates from the target, which can be explained by the chaotic nature of the brain dynamics, as also evidenced in a recent theoretical study [22]. How the fitted RNNs respond to small external perturbations is also shown in Fig. 2, and we will come to discuss these observed differences later.

We next discuss the functional connection patterns learned from the brain dynamics across the three different brain states. As shown in Figs. 3–5, the three stages exhibit clear differences in statistics of functional connectivity. Compared to the awake condition, the anesthesia condition yields sparser connections with weaker strengths, reducing feedback and thus blocking global communications among brain regions [23]. Nevertheless, the recovery condition yields a broader distribution of coupling, being of salient strength and heterogeneity; these well-tuned feedbacks are able to support

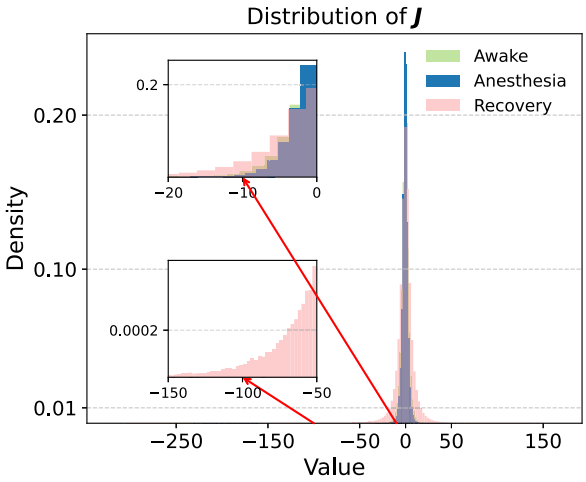


FIG. 3. Distribution of the learned recurrent weight matrix $\mathbf{J}$ across the three brain states: awake, anesthesia, and recovery. All learned coupling weights obtained from 100-s fitting windows (as in Fig. 2) are combined and shown in a single plot, with different colors representing different brain states. The inset provides an enlarged view that highlights subtle differences in the weight distributions among the three states.

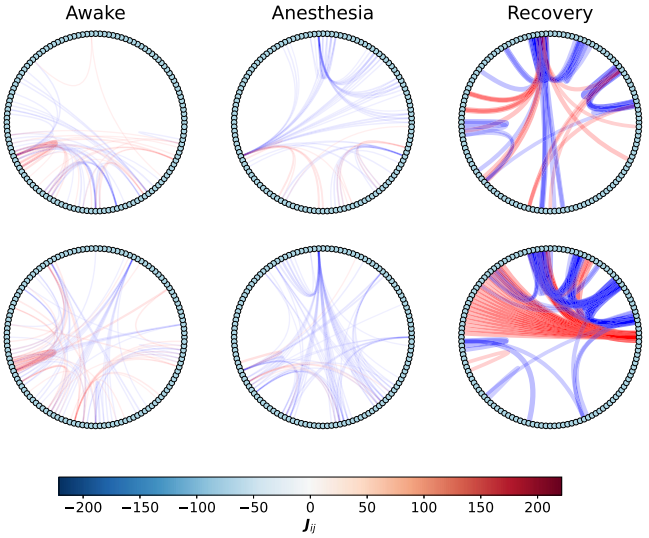


FIG. 4. Connection pattern of **J** for the three brain states: awake, anesthesia, and recovery. Results are shown for the first 2-s segment from the 100-s fitting window. Shown are the strongest 1% of connections in absolute value among the 128 channels. Positive weights are shown in red and negative weights in blue. The first row displays the upper-triangular part of **J**, and the second row shows the lower-triangular part, highlighting how coupling patterns evolve across different cortical conditions.

consciousness or awareness, showing how consciousness is re-established, which will be further confirmed by the precise connection pattern and the dynamics response property in the next section.

The learned recurrent weight matrices **J** are apparently asymmetric, showing consistent differences across three sessions (see Figs. 4 and 5). Visualizations of the upper and lower triangles of the 128 × 128 coupling matrix reveal that the awake and recovery conditions bear relatively strong and dense but highly heterogeneous couplings, consistent with a rich, flexible, and responsive dynamics (discussed below). During anesthesia, the connection weights become weaker and sparser, indicating a globally less fine-tuned feedback, while the recovery shows an emergence of stronger weights and a structurally organized connection pattern. For the awake, especially recovered brain state, a structurally well-organized pattern of strong positive and negative couplings supports the high-dimensional dynamics and robust responsiveness. Anesthesia flattens this structure, making connections sparse and suppressing global communication [23]. Recovery restores this structural heterogeneity, signaling the gradual return of responsiveness as quantified by $\chi(t)$. Results shown in Figs. 4 and 5 also suggest that in terms of the connection pattern, the recovery state after the loss of consciousness is distinct from the awake state just before the anesthesia, which deserves future systematic analysis.

We also verify how the connection pattern varies across different segments from which the functional connectivity is learned. As shown in Fig. 6, the qualitative property of each connection pattern holds across different segments within each individual brain state. To see how these connection patterns impact the dynamics, we also do a control experiment where the original learned couplings for a given inferred network are pruned once their absolute values are above a threshold (set to 1 here, yet a slight increase of the threshold does not affect the qualitative conclusion). We then compare the controlled dynamics with the original one with an intact topology. Results are shown in Fig. 7, which demonstrates that the pruned strong feedbacks through the coupling weights play a vital role in maintaining the high-dimensional chaotic attractors. In the absence of these strong feedbacks, the dynamical regime shifts to a low-dimensional fixed point, and thus the flexible and collective computational function offered by the high-dimensional brain activity disappears. This observation highlights the important role of criticality even in real brain data analysis, as the concept was previously explored in theoretical and computational frameworks [18,28–30].

B. Neural response function quantifies the level of consciousness

To probe how the fitted RNNs respond to small external perturbations within each conscious state, we computed the mean population response function $\chi(t)$ by applying constant weak input perturbations and measuring the resulting relationship between mean activity and the perturbation. For each experimental session, Fig. 2 illustrates this linear regime where the neural response function can be read off. The approximately linear trends in these plots demonstrate that the fitted dynamics yield well-defined response functions, with the slope of each line quantifying the network's global sensitivity to external input, which we identify as a metaphor of responsiveness in brain dynamics. In particular, the anesthesia

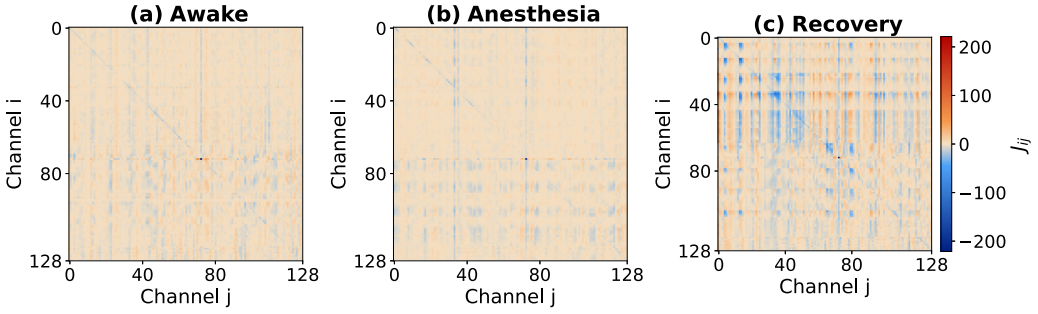


FIG. 5. Full recurrent weight matrices **J** for the three brain states: awake, anesthesia, and recovery. Each panel shows the entire structured **J** matrix (including all elements) corresponding to the same fitted segment as in Fig. 4, visualized as a heat map. Positive weights are shown in red and negative weights in blue.

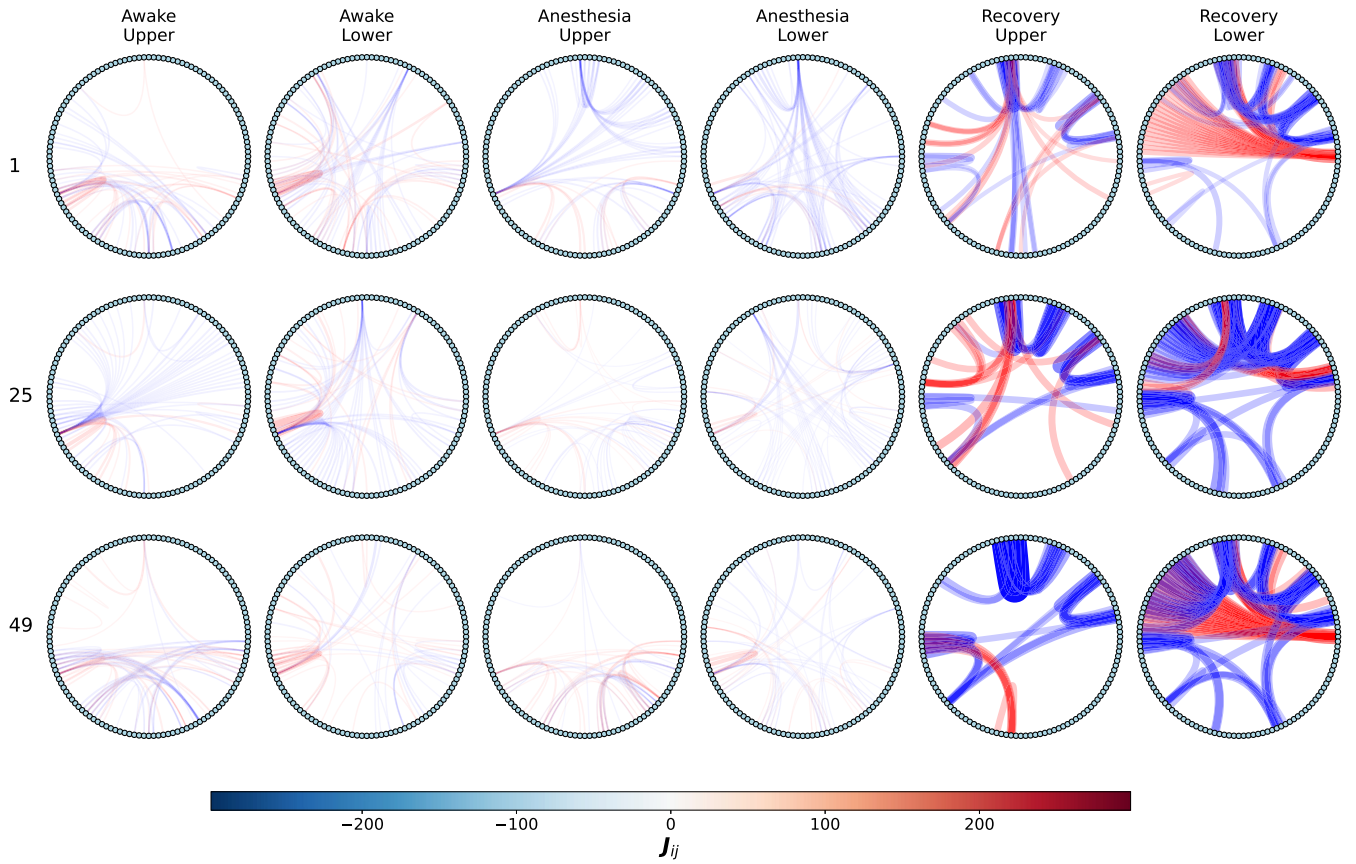


FIG. 6. Comparison of recurrent connection patterns across the three brain states—awake, anesthesia, and recovery. For each brain state, the plots display the recurrent connections **J** obtained from the first, middle, and last 2-ssegments (indicated by the numbers on the left-hand side) within the 100-s fitting window (as in Fig. 4), allowing direct comparison of inferred functional connectivity structures across time. For each condition, the upper- and lower-triangular parts of the weight matrix are separately displayed.

condition shows slightly smaller slopes, indicating weaker responsiveness, whereas the awake and recovery conditions produce larger linear coefficients, consistent with more flexible nonequilibrium dynamics. These within-state analyses establish that the RNN models capture a well-defined linear relationship between network response and input perturbations, providing a principled estimate of global dynamical sensitivity as also observed in a toy random model [18].

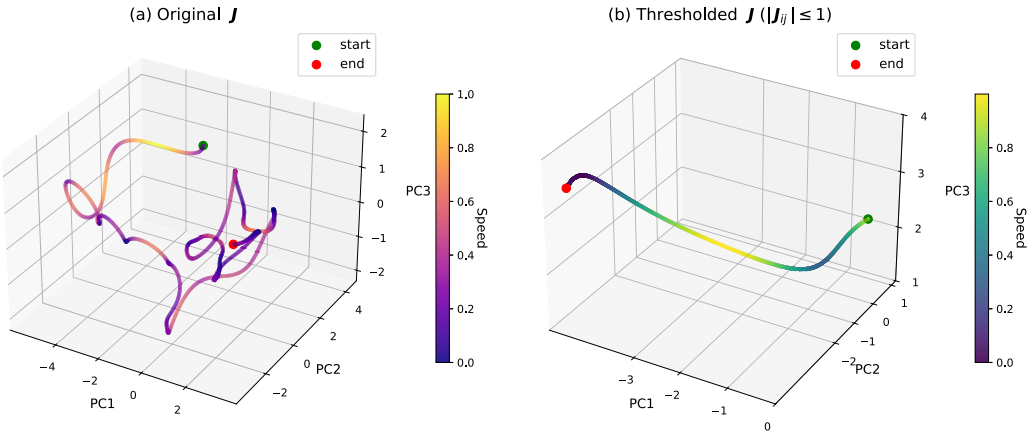


FIG. 7. Neural trajectories around the edge of instability in recurrent neural networks learned from a 2-s segment. The trajectories are colored by their speeds (the magnitude of the velocity $\dot{\mathbf{x}}$). (a) Trajectories in three-dimensional PCA (principal component analysis) space generated from the original fitted recurrent weights (awake state), showing irregular and high-dimensional patterns consistent with chaotic dynamics. (b) Trajectories after setting all recurrent weights greater than 1 in absolute value to zero, showing convergence toward a low-dimensional fixed point. This dynamics visualization in both unperturbed and perturbed settings manifests the edge of instability, further supporting our response function angle and the Lyapunov exponent analysis.

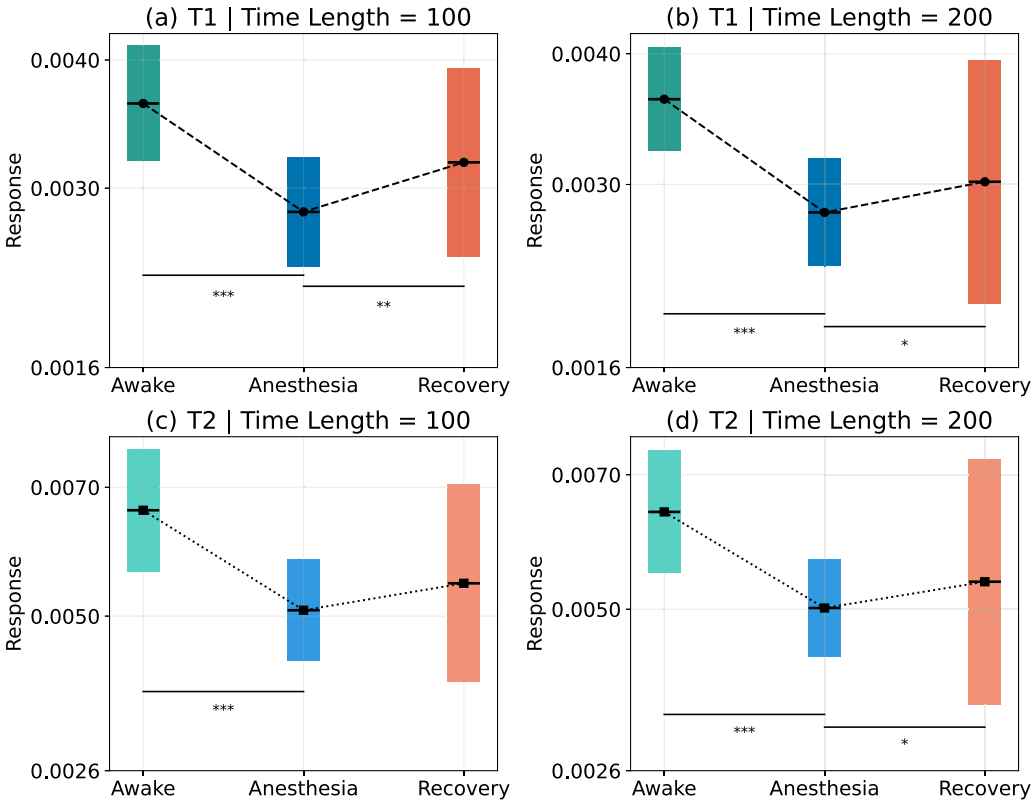


FIG. 8. Comparison of neural response functions across brain states: awake, anesthesia, and recovery. Average response slopes $\chi(t)$ at $t = 0.005$ s (five time steps, T1 in the plot) are shown for models fitted on all nonoverlapping segments of 100-s [panel (a)] and 200-s [panel (b)] windows. Average response slopes $\chi(t)$ at $t = 0.01$ s (ten time steps, T2 in the plot) are shown for models fitted on all nonoverlapping segments of 100-s [panel (c)] and 200-s [panel (d)] windows. Bars represent mean slopes across segments, and error bars indicate standard deviations. Statistical significance between states was assessed with two-sample t-tests (the symbols "$***$", "$**$", and "$*$" mean $p < 0.001$, 0.01, and 0.05, respectively). Note that in the panel (c), the p value for the case of anesthesia vs recovery is 0.0938, and thus there is no symbol shown in the plot. The anesthesia state shows consistently weaker and more concentrated responses, while awake and recovery states show stronger and more variable responsiveness.

To directly compare how this responsiveness varies across conscious states, we next collect these slopes over multiple steady-state time windows and visualize their distributions to verify an indicator of state transition in terms of our theoretically grounded response function. As shown in Fig. 8, the anesthesia state consistently shows lower mean responsiveness, highlighting the reduced capacity of the cortex to propagate perturbations under general anesthesia [23]. In

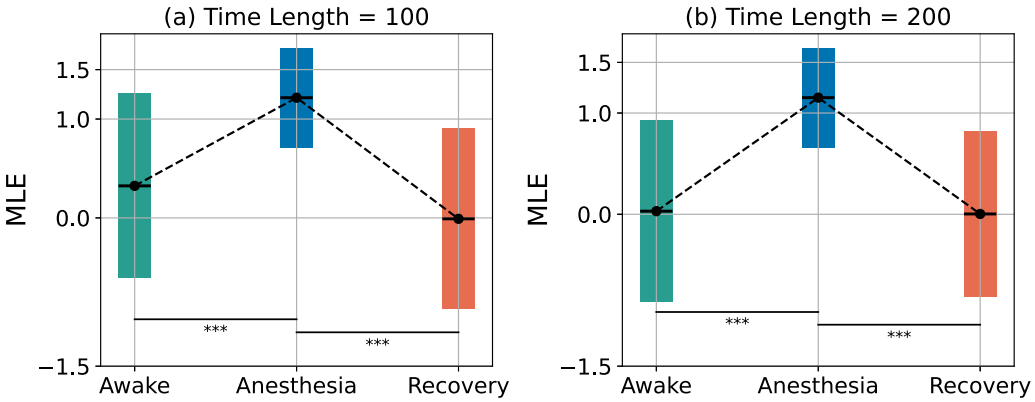


FIG. 9. Comparison of dynamic complexity across brain states: awake, anesthesia, and recovery, quantified by the maximal Lyapunov exponent (MLE) using the orbit separation method. (a) MLEs for models fitted on all nonoverlapping segments of 100-s windows. (b) MLEs for models fitted on all nonoverlapping segments of 200-s windows. The symbols "$***$" means $p < 0.001$ for two-sample t-tests. The error bars reflect the variability across many fitted segments. The anesthesia state consistently exhibits the highest MLE, indicating greater dynamic instability compared with awake and recovery states. Other settings are the same as in Fig. 8.

contrast, both the awake and recovery conditions display higher mean response functions, reflecting richer metastable nonequilibrium dynamics that can sustain information processing. Interestingly, the recovery condition just after the loss of consciousness displays a significant variability in neural response function, and moreover, the responsiveness is still weaker than that of the awake condition. This asymmetric property is interesting and deserves further theoretical studies of underlying mechanisms.

Finally, we explore the dynamic complexity of fitted RNNs for three separate sessions in Fig. 9. The results demonstrate that the RNNs' dynamics reflect the significant difference across three sessions. More precisely, the unconscious brain state induced by anesthesia bears a higher dynamic instability characterized by maximal Lyapunov exponent (obtained by the orbit separation method [22,31]), consistent with an empirical study on the relationship between the brain complexity and the dynamic complexity [14]. Figure 9 also shows the asymmetric role of awake and recovery conditions, consistent with the above connection pattern and response analyses.

Taken together, these results demonstrate that the fitted RNN framework captures both local dynamical properties (via loss and weight distribution) and global responsiveness (via the response function). The intrinsic differences across states reinforce the view that neural responsiveness is a necessary dynamical signature of consciousness [32,33]. By putting these analyses within a nonequilibrium dynamical framework, our work provides a quantitative and mechanistic link between the empirical ECoG signals and theoretical models of cortical criticality and perturbation propagation. To conclude, these analyses illustrate how local model fitting, recurrent coupling, and perturbation sensitivity collectively reveal state-dependent signatures of consciousness in large-scale cortical dynamics.

IV. CONCLUDING REMARKS

How to provide a behavior evaluation of consciousness level is an important frontier, as the quantitative measure helps to characterize brain complexity of consciousness, e.g., clinical treatment of prolonged disorders of consciousness, unresponsive wakefulness syndrome, and even detection of awareness in understanding the nature of consciousness. Inspired by our previous theoretical works, we argued in this work that the response function of nonequilibrium dynamics in the neural circuits (here modeled by a recurrent neural network) serves as a natural and simple measurable signature of consciousness, as we observe a salient change of this quantity when analyzing the different stages of ECoG dynamics (e.g., waking-anesthesia-recovery transition). Another dominant measure in clinical applications is the perturbational complexity index, which provides a direct measure of the spatiotemporal complexity of the evoked responses of the brain to a perturbation such as transcranial magnetic stimulation [6,34]. Computing this index requires multiple steps of delicate analysis (see details in previous works [4,6]). It is thus interesting to compare our method with this complexity index in the human dataset.

Responsiveness studied in this paper is merely a necessary signature of consciousness, while this signature can be mathematically formulated using functionally connected neurons or brain regions. In the future, one has to introduce more biological constraints into the hierarchical brain system, study the multiscale brain dynamics, and finally demonstrate the sufficient condition to support the maximal information processing ability around a self-organized criticality. The scientific endeavor of putting a data-driven model of brain dynamics within a mathematical framework would be valuable to refine existing theories of consciousness [35,36], or to provide criteria for justifying or refuting the existence of consciousness in artificial intelligence (AI) systems [37].

ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China (Grant No. 12475045), Guangdong Provincial Key Laboratory of Magnetoelectric Physics and Devices (Grant No. 2022B1212010008), and Guangdong Basic and Applied Basic Research Foundation (Grant No. 2023B1515040023).

DATA AVAILABILITY

The data that support the findings of this article are openly available [38]; embargo periods may apply.

[1] S. Dehaene and J.-P. Changeux, Experimental and theoretical approaches to conscious processing, Neuron **70**, 200 (2011).

[2] T. Bayne, A. K. Seth, and M. Massimini, Are there islands of awareness? Trends Neurosci. **43**, 6 (2020).

[3] M. Lenharo, The quest to detect consciousness—in all its possible forms, Nature (London) **643**, 1172 (2025).

[4] A. G. Casali, O. Gosseries, M. Rosanova, M. Boly, S. Sarasso, K. R. Casali, S. Casarotto, M.-A. Bruno, S. Laureys, G. Tononi, and M. Massimini, A theoretically based index of consciousness independent of sensory processing and behavior, Sci. Transl. Med. **5**, 198ra105 (2013).

[5] G. Solovey, L. M Alonso, T. Yanagawa, N. Fujii, M. O Magnasco, G. A Cecchi, and A. Proekt, Loss of consciousness is associated with stabilization of cortical activity, J. Neurosci. **35**, 10866 (2015).

[6] R. Comolatti, A. Pigorini, S. Casarotto, M. Fecchio, G. Faria, S. Sarasso, M. Rosanova, O. Gosseries, M. Boly, O. Bodart, D. Ledoux, J.-F. Brichant, L. Nobili, S. Laureys, G. Tononi, M. Massimini, and A. G. Casali, A fast and general method to empirically estimate the complexity of brain responses to transcranial and intracranial stimulations, Brain Stimul. **12**, 1280 (2019).

[7] H. Huang, Eight challenges in developing theory of intelligence, Front. Comput. Neurosci **18**, 1388166 (2024).

[8] C. W. Lynn, E. J. Cornblath, L. Papadopoulos, M. A. Bertolero, and D. S. Bassett, Broken detailed balance and entropy pro-

duction in the human brain, Proc. Natl. Acad. Sci. USA **118**, e2109889118 (2021).

[9] Y. Sanz Perl, H. Bocaccio, C. Pallavicini, I. Pérez-Ipiña, S. Laureys, H. Laufs, M. Kringelbach, G. Deco, and E. Tagliazucchi, Nonequilibrium brain dynamics as a signature of consciousness, Phys. Rev. E **104**, 014411 (2021).

[10] J. A. Galadí, S. S. Pereira, Y. S. Perl, M. L. Kringelbach, I. Gayte, H. Laufs, E. Tagliazucchi, J. A. Langa, and G. Deco, Capturing the non-stationarity of whole-brain dynamics underlying human brain states, NeuroImage **244**, 118551 (2021).

[11] M. Gilson, E. Tagliazucchi, and R. Cofré, Entropy production of multivariate Ornstein-Uhlenbeck processes correlates with consciousness levels in the human brain, Phys. Rev. E **107**, 024121 (2023).

[12] R. Nartallo-Kaluarachchi, M. L. Kringelbach, G. Deco, R. Lambiotte, and A. Goriely, Nonequilibrium physics of brain dynamics, Physics Reports **1152**, 1 (2026).

[13] L. M. Alonso, A. Proekt, T. H. Schwartz, K. O. Pryor, G. A. Cecchi, and M. O. Magnasco, Dynamical criticality during induction of anesthesia in human ECoG recordings, Front. Neural Circuits **8**, (2014).

[14] D. Toker, I. Pappas, J. D. Lendner, J. Frohlich, D. M. Mateos, S. Muthukumaraswamy, R. Carhart-Harris, M. Paff, P. M. Vespa, M. M. Monti *et al.*, Consciousness is supported by near-critical slow cortical electrodynamics, Proc. Natl. Acad. Sci. USA **119**, e2024455119 (2022).

[15] S. L. Eagleman, D. Chander, C. Reynolds, N. T. Ouellette, and M. B. MacIver, Nonlinear dynamics captures brain states at different levels of consciousness in patients anesthetized with propofol, PLoS One **14**, e0223921 (2019).

[16] L. A. de la Fuente, F. Zamberlan, H. Bocaccio, M. Kringelbach, G. Deco, Y. S. Perl, C. Pallavicini, and E. Tagliazucchi, Temporal irreversibility of neural dynamics as a signature of consciousness, Cereb. Cortex **33**, 1856 (2022).

[17] J. Cruzat, R. Herzog, P. Prado, Y. Sanz-Perl, R. Gonzalez-Gomez, S. Moguilner, M. L. Kringelbach, G. Deco, E. Tagliazucchi, and A. Ibañez, Temporal irreversibility of large-scale brain dynamics in Alzheimer's disease, J. Neurosci. **43**, 1643 (2023).

[18] J. Qiu and H. Huang, An optimization-based equilibrium measure describing fixed points of non-equilibrium dynamics: Application to the edge of chaos, Commun. Theor. Phys. **77**, 035601 (2025).

[19] W. Du and H. Huang, Synaptic plasticity alters the nature of the chaos transition in neural networks, Phys. Rev. E **112**, 054208 (2025).

[20] D. Mastrovito, Y. H. Liu, L. Kusmierz, E. Shea-Brown, C. Koch, and S. Mihalas, Transition to chaos separates learning regimes and relates to measure of consciousness in recurrent neural networks, bioRxiv (2024).

[21] T. Yanagawa, Z. C. Chao, N. Hasegawa, and N. Fujii, Large-scale information flow in conscious and unconscious states: An ECoG study in monkeys, PLoS One **8**, e80845 (2013).

[22] Z. Yu and H. Huang, Network reconstruction may not mean dynamics prediction, Phys. Rev. E **111**, 034308 (2025).

[23] M. Massimini, F. Ferrarelli, R. Huber, S. K. Esser, H. Singh, and G. Tononi, Breakdown of cortical effective connectivity during sleep, Science **309**, 2228 (2005).

[24] https://neurotycho.org/.

[25] H. Sompolinsky, A. Crisanti, and H. J. Sommers, Chaos in random neural networks, Phys. Rev. Lett. **61**, 259 (1988).

[26] U. Marini, B. Marconi, A. Puglisi, L. Rondoni, and A. Vulpiani, Fluctuation–dissipation: Response theory in statistical physics, Phys. Rep. **461**, 111 (2008).

[27] W. Zou and H. Huang, Introduction to dynamical mean-field theory of randomly connected neural networks with bidirectionally correlated couplings, SciPost Phys. Lect. Notes 79 (2024).

[28] N. Bertschinger and T. Natschläger, Real-time computation at the edge of chaos in recurrent neural networks, Neural Comput. **16**, 1413 (2004).

[29] K. Krishnamurthy, T. Can, and D. J. Schwab, Theory of gating in recurrent neural networks, Phys. Rev. X **12**, 011011 (2022).

[30] T. Kanamaru, T. K. Hensch, and K. Aihara, Maximal memory capacity near the edge of chaos in balanced cortical E-I networks, Neural Comput. **35**, 1430 (2023).

[31] J. C. Sprott, *Chaos and Time-Series Analysis* (Oxford University Press, Oxford, 2003).

[32] A. Sarracino, O. Arviv, O. Shriki, and L. de Arcangelis, Predicting brain evoked response to external stimuli from temporal correlations of spontaneous activity, Phys. Rev. Res. **2**, 033355 (2020).

[33] G. Deco, J. Cruzat, J. Cabral, E. Tagliazucchi, H. Laufs, N. K. Logothetis, and M. L. Kringelbach, Awakening: Predicting external stimulation to force transitions between different brain states, Proc. Natl. Acad. Sci. USA **116**, 18088 (2019).

[34] D. O. Sinitsyn, A. G. Poydasheva, I. S. Bakulin, L. A. Legostaeva, E. G. Iazeva, D. V. Sergeev, A. N. Sergeeva, E. I. Kremneva, S. N. Morozova, D. Y. Lagoda, S. Casarotto, A. Comanducci, Y. V. Ryabinkina, N. A. Suponeva, and M. A. Piradov, Detecting the potential for consciousness in unresponsive patients using the perturbational complexity index, Brain Sci. **10**, 917, (2020).

[35] J. F. Storm, P. C. Klink, J. Aru, W. Senn, R. Goebel, A. Pigorini, P. Avanzini, W. Vanduffel, P. R. Roelfsema, M. Massimini, M. E. Larkum, and C. M. A. Pennartz, An integrative, multiscale view on neural theories of consciousness, Neuron **112**, 1531 (2024).

[36] A. K. Seth and T. Bayne, Theories of consciousness, Nat. Rev. Neurosci. **23**, 439 (2022).

[37] P. Butlin, R. Long, E. Elmoznino, Y. Bengio, J. Birch, A. Constant, G. Deane, S. M. Fleming, C. Frith, X. Ji, R. Kanai, C. Klein, G. Lindsay, M. Michel, L. Mudrik, M. A. K. Peters, E. Schwitzgebel, J. Simon, and R. VanRullen, Consciousness in artificial intelligence: Insights from the science of consciousness, arXiv:2308.08708.

[38] W. Du, Code for "Response function as a qualitative measure of consciousness in brain dynamics", 2025, https://github.com/Wenkang-Du/Response-function-of-consciousness.