An AI-Driven Model of Consciousness, Its Disorders, and Their Treatment

Daniel Toker[1,2*], Jasmine A. Thum[3], Jing Guang[4], Hiroyuki Miyamoto[5,6,7], Kazuhiro Yamakawa[5,8], Paul M. Vespa[3], Caroline Schnakers[9], Ausaf A. Bari[3], Andrew Hudson[10,11], Nader Pouratian[12], Martin M. Monti[2,3]

[1*]Neurology Department, University of California, Los Angeles, Los Angeles, CA, USA.
[2]Psychology Department, University of California, Los Angeles, Los Angeles, CA, USA.
[3]Department of Neurosurgery, University of California, Los Angeles, Los Angeles, CA, USA.
[4]Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel.
[5]Laboratory for Neurogenetics, RIKEN Center for Brain Science, Saitama, Japan.
[6]PRESTO, Japan Science and Technology Agency, Saitama, Japan.
[7]International Research Center for Neurointelligence, University of Tokyo, Tokyo, Japan.
[8]Department of Neurodevelopmental Disorder Genetics, Institute of Brain Science, Nagoya City University Graduate School of Medical Science, Nagoya, Japan.
[9]Research Institute, Casa Colina Hospital and Centers for Healthcare, Pomona, CA, USA.
[10]Department of Anesthesiology, Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, CA, USA.
[11]Department of Anesthesiology and Perioperative Medicine, University of California, Los Angeles, Los Angeles, CA, USA.
[12]Department of Neurological Surgery, UT Southwestern Medical Center, Dallas, TX, USA.

*Corresponding author(s). E-mail(s): danieltoker@g.ucla.edu;

Abstract

Understanding the neural signatures of consciousness and the mechanisms underlying its disorders, such as coma and unresponsive wakefulness syndrome, remains a critical challenge in neuroscience. In this study, we present a novel computational approach for the in silico discovery of neural correlates of consciousness, the mechanisms driving its disorders, and potential treatment strategies. Inspired by generative adversarial networks, which have driven recent advancements in generative artificial intelligence (AI), we trained deep neural networks to detect consciousness across multiple brain areas and species, including humans. These networks were then integrated with a genetic algorithm to optimize a brain-wide mean-field model of neural electrodynamics. The result is a realistic simulation of conscious brain states and disorders of consciousness (DOC), which not only recapitulates known mechanisms of unconsciousness but also predicts novel causes expected to lead to these conditions. Beyond simulating DOC, our model provides a platform for exploring therapeutic interventions, specifically deep brain stimulation (DBS), which has shown promise in improving levels of awareness in DOC in over five decades of study. We systematically applied simulated DBS to various brain regions at a wide range of frequencies to identify an optimal paradigm for reigniting consciousness in this cohort. Our findings suggest that in addition to previously studied thalamic and pallidal stimulation, high-frequency stimulation of the subthalamic nucleus, a relatively underexplored target in DOC, may hold significant promise for restoring consciousness in this set of disorders.

1 Introduction

The mechanisms that underpin consciousness and the transitions between conscious and comatose states remain some of the most fundamental and unresolved questions in neuroscience. Despite significant advances in neuroimaging and electrophysiology, the field still lacks a comprehensive understanding of how brain activity supports conscious awareness or collapses into unconscious states, such as coma. This gap in knowledge is particularly striking given the clinical need to improve outcomes for patients with disorders of consciousness (DOC), including coma and the vegetative states (or unresponsive wakefulness syndrome) [1, 2]. Some therapeutic options, such as deep brain stimulation (DBS), show promise in treating these conditions [3–8]; however, a mechanistic understanding of optimal targets, patterns of stimulation, and how these interventions impact brain dynamics is still limited.

The "mesocircuit hypothesis" provides a valuable framework for understanding DOC and recovery of consciousness [9]. Drawing on a range of neuroimaging and interventional studies, the mesocircuit hypothesis proposes that reactivation of an anterior forebrain mesocircuit is essential for regaining consciousness following severe brain injury. The model focuses on the central thalamus, which serves as a key hub with broad projections to the cortex and striatum, making it particularly vulnerable to the effects of widespread deafferentation often seen in severe brain injury. Deafferentation of the central thalamus is posited to reduce excitatory input to both the cortex and the striatum, resulting in disfacilitation—or a passive hyperpolarization—across the

anterior forebrain, which leads to diminished firing rates and a reduction in overall cerebral metabolic activity.

Traditional accounts of the mesocircuit hypothesis emphasize that the resulting decline in striatal activity disinhibits the globus pallidus interna (GPi), which then actively inhibits the central thalamus, further driving the thalamocortical system into a state of global metabolic depression [9]. However, recent updates to the hypothesis highlight a potentially key role of the globus pallidus externa (GPe), a region traditionally seen as part of the motor-suppressing indirect pathway of the basal ganglia, as a key regulator of arousal [10]. Emerging evidence indicates that the GPe has direct projections to both the cortex and the thalamic reticular nucleus (TRN), positioning it to exert more nuanced control over thalamocortical and corticocortical rhythms [10, 11]. The GPe may play a direct role in modulating cortical arousal by regulating inhibitory output to the cortex and the TRN, which is essential for maintaining the thalamic sensory gating that underpins conscious awareness [12]. In addition, neuroanatomical work shows that the complexity of neural response to transcranial magnetic stimulation (i.e., Perturbational Complexity Index) in DOC patients is inversely proportional to the amount of atrophy in the GPe [13]. This updated framework underscores the GPe's involvement in sleep-wake regulation and electrocortical arousal regulation [14] and suggests it could be an important target for interventions aimed at restoring consciousness in DOC patients. Direct testing of these ideas, however, has been challenging due to the limited availability of animal models that reliably replicate DOC. Recent advancements, however, including including a porcine model [15], offer vital opportunities for future experimental validation.

A promising alternative avenue for directly testing and expanding upon the mesocircuit hypothesis lies in neural field theory. Neural field theory offers a powerful framework for modeling large-scale brain activity by averaging neural properties such as firing rates and membrane potentials across spatially extended neural populations. Neural field models capture spatially extended neural electrodynamics across continuous sheets of neural tissue, enabling the modeling of phenomena like electric wave propagation and synchrony across cortical and subcortical regions. By retaining both spatial and temporal information, neural field theory bridges the gap between detailed neural microstructure and the global patterns observable in neural electrophysiology recording techniques like electroencephalography (EEG), magnetoencephalography (MEG), electrocorticography (ECoG), and local field potentials (LFPs) [16]. This approach can therefore enable in silico "dissection" of the relationships between brain anatomy and function on the one hand, and the large-scale electrodynamics associated with consciousness and coma on the other.

Inspired by generative adversarial networks [17], we here took this understanding further by training deep neural networks to recognize signatures of consciousness in multiple species, including humans, in both cortical and subcortical structures. These networks were then applied in tandem with a genetic algorithm to optimize the parameters of a mean-field model of electrodynamics in the thalamo-cortical-basal ganglia system. This cross-talk between deep learning and genetic optimization allowed us to simulate realistic conscious brain states. Crucially, our approach incorporated empirical constraints, ensuring that the model displayed firing rates, intra- and inter-regional

phase-amplitude coupling, power spectra, proximity to criticality [18–21], and thalamocortical information transfer patterns [19] closely matching those seen in actual brain recordings during conscious states across multiple mammalian species.

Beyond simulating conscious states, we also used this framework to simulate disorders of consciousness. We trained a deep neural network to accurately predict varying levels of consciousness in both acute coma and chronic disorders of consciousness, and validated its ability to detect induced pharmacological coma in unseen, independent data from humans, rats, and monkeys. We then used this trained network to tune our brain-wide mean-field model to generate LFPs consistent with a comatose state. This allowed us to explore, in silico, the causes of unconsciousness following severe brain injury. Our analysis not only successfully retrodicts several well-characterized drivers of DOC, but also points to novel potential causes of pathological unconsciousness.

Our simulations of DOC also enabled us to explore, in silico, how brain stimulation might modulate levels of consciousness in these conditions. Given the track record of DBS specifically in treating neurological conditions [22], its ability to directly modulate deep brain regions essential for consciousness, and its consistent preliminary success in DOC [3–8], we focused on testing potential DBS interventions (rather than other stimulation modalities) for DOC. Our results not only replicate prior empirical work on thalamic and pallidal DBS for DOC, but also point to the subthalamic nucleus (STN) as a promising novel target for improving awareness in this cohort.

In sum, we present a new computational approach for simulating both conscious and comatose brain states, describe the neural mechanisms revealed by our model, and leverage the model to discover novel potential interventions to restore consciousness in DOC.

2 Results

We present our results in three main sections. First, we show that it is possible to use an adversarial network-like approach, leveraging electrophysiology data from multiple species and conditions, to develop a realistic in silico model of the conscious brain. Second, we use a similar approach to simulate brain dynamics more typical of pathological unconsciousness, which recapitulates known DOC electrophysiology phenotypes (e.g., large-amplitude delta waves and alpha coma [23]). We show that our AI-driven modeling of DOC correctly retrodicts well-known network disturbances associated with unconsciousness following severe brain injury (e.g., corticothalamic disconnection [9], cortical hypoactivity [24–27], and pallidal hyperactivity [27]), but also predicts new potential drivers of DOC (e.g., faster conduction velocities between surviving interneurons). Finally, we leverage the developed DOC model to explore different potential neurorestorative interventions and show that the model correctly retrodicts the efficacy of DBS to thalamus and globus pallidus [3–8], but also predicts a novel target, the subthalamic nucleus, as likely the most effective site of intervention for restoring conscious brain dynamics.

2.1 Simulating the Conscious Brain With an Adversarial-Like Approach

To simulate the electrodynamics of the conscious brain using a genetic adversarial network-like approach, we first trained deep convolutional neural networks (DCNNs) on electrophysiology recordings from multiple species and brain regions (Fig. 1A). These included recordings from three brain regions of interest: the cortex, thalamic projection nuclei, and GPe (see Introduction). Given that cortical electrophysiology is the most accessible modality in DOC patients via scalp EEG, the cortex was a primary focus of our deep learning approach. Our first DCNN was therefore trained on cortical data from acute coma patients, chronic DOC patients, awake African green monkeys (*Chlorocebus aethiops*), awake human Parkinson's disease (PD) and essential tremor (ET) patients, and awake healthy human controls (see Methods).

Using a linear mixed-effects model, we assessed the performance of the network in predicting the level of consciousness in held-out validation EEG data from 22 acute coma patients, showing a significant association between predicted and actual Glasgow Coma Scale (GCS; [28]) scores ($p<0.001$) (Fig. 1B). In held-out validation data from chronic DOC patients, a similar analysis revealed a significant relationship between actual Coma Recovery Scale Revised (CRS-R; [29]) scores and the AI-predicted level of consciousness ($p<0.001$) (Fig. 1C).

The model also successfully generalized to detect loss of consciousness in anesthetic coma in independent test data. A mixed-effects model comparing AI-predicted consciousness during wake and anesthesia across *Chlorocebus aethiops*, human ET patients, and human PD patients revealed a significant effect of state, with wakefulness showing higher predicted consciousness levels ($p<0.001$) (Fig. 1D). Notably, all wake data were held-out validation data, while anesthesia data were independent test data. Pairwise comparisons further demonstrated significant differences in AI-predicted levels of consciousness between wake and anesthesia states in both *Chlorocebus aethiops* ($p<0.001$) and human ET patients ($p<0.001$), though not in human PD patients ($p = 0.36$), likely due to limited data (Fig. 1D). Analysis of validation EEG data from healthy, awake human controls led to AI predictions of near-maximal GCS scores (>14.7) (Fig. 1E). Furthermore, the trained deep neural network effectively identified reduced consciousness in independent test data from Long-Evans rats ($p< 0.001$) and from *Macaca fuscata* under propofol anesthesia. In *Macaca fuscata*, pairwise comparisons showed significantly lower predicted levels of consciousness during propofol anesthesia compared to the awake eyes-open state ($p = 0.019$) and the awake eyes-closed state ($p = 0.016$). Importantly, no significant difference in predicted consciousness was observed between the awake eyes-open and eyes-closed states ($p = 0.78$) (Fig. 1F). Collectively, these results demonstrate the accuracy of our cortical consciousness detector in differentiating waking from comatose states.

To expand beyond the cortex, we also trained additional networks to detect consciousness from thalamic and GPe LFP recordings. The thalamic model, trained on human ET patient and Long-Evans rat data, achieved high accuracy in detecting consciousness in human validation data (receiver operating characteristic AUC = 0.98), with slightly lower performance in rats (AUC = 0.69) (Fig. 1G). Similarly, the GPe network, trained on pallidal LFPs from human PD patients and *Chlorocebus aethiops*,

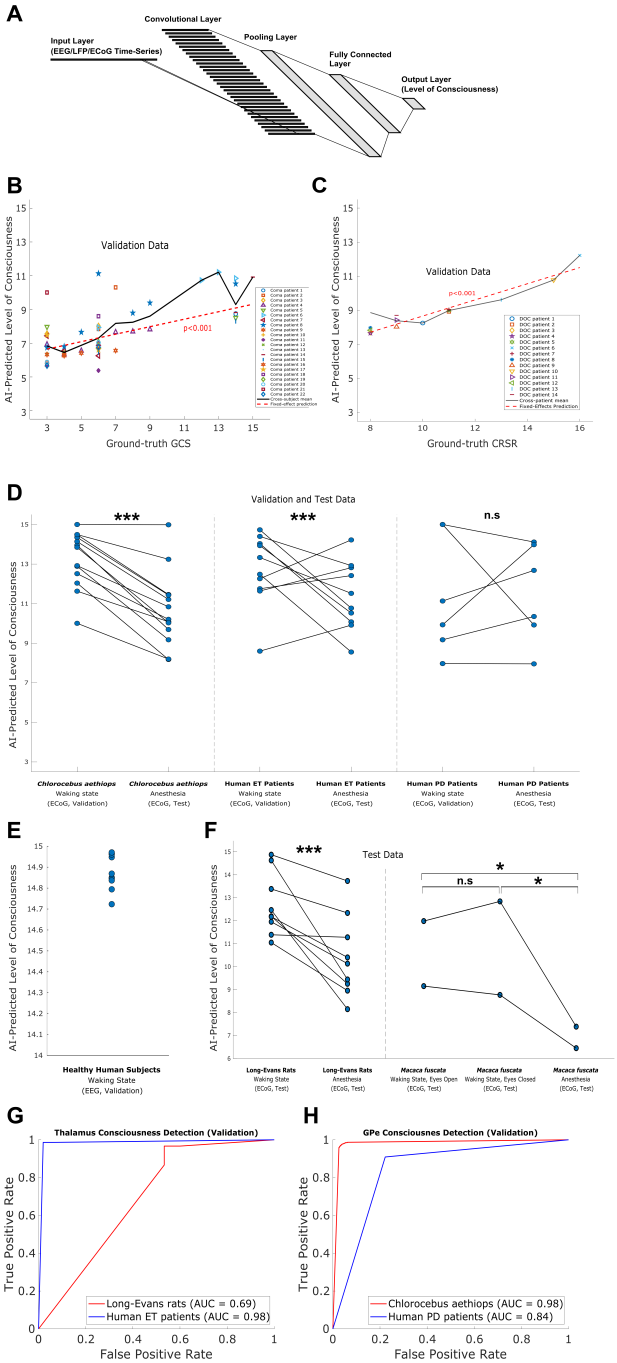

Fig. 1 Performance of neural networks trained to detect consciousness across multiple species and brain regions. In all subplots, a single marker corresponds to the average predicted level of consciousness across all available trials for a single patient, subject, or animal. (A) We trained deep neural networks on electrophysiology recordings from the cortex, thalamus, and external globus pallidus (GPe) to predict levels of consciousness across species and electrophysiology recording modalities. (B) A linear mixed-effects model showed a significant association between empirical GCS and AI-predicted levels of consciousness in held-out validation EEG data from acute coma patients ($p<0.001$). (C) In chronic DOC patients, a mixed-effects model likewise revealed a significant association between ground-truth CRSR and AI-predicted GCS in held-out validation data ($p<0.001$). (D) The model successfully generalized to detect loss of consciousness under anesthesia in independent test data. AI-predicted levels of consciousness were significantly higher during wakefulness compared to anesthesia across *Chlorocebus aethiops*, human ET, and PD patients ($p<0.001$), with pairwise differences observed in both *Chlorocebus aethiops* ($p<0.001$) and ET patients ($p<0.001$), but not in PD patients ($p = 0.36$). (E) Healthy, conscious human controls had AI-predicted GCS scores greater than 14.7. (F) In independent test data from Long-Evans rats, the DCNN also detected significantly reduced consciousness under propofol anesthesia ($p<0.001$ for Long-Evans rats). Similarly, in independent test data in *Macaca fuscata*, predicted levels of consciousness were significantly lower during propofol anesthesia compared to the awake eyes-open state ($p = 0.019$) and the awake eyes-closed state ($p = 0.016$). No significant difference was observed between the awake eyes-open and eyes-closed states ($p = 0.78$). (G) Receiver operating characteristic (ROC) analysis of the thalamic consciousness detector yielded an AUC of 0.98 for human ET patients and 0.69 for Long-Evans rats. (H) ROC analysis of the GPe consciousness detector showed an AUC of 0.98 for *Chlorocebus aethiops* and 0.84 for human PD patients.

showed strong performance in detecting consciousness in held-out validation data from both monkeys (AUC = 0.98) and humans (AUC = 0.84) (Fig. 1H). These results confirm the ability of our DCNNs to detect consciousness based on subcortical LFPs.

After training, our three consciousness detectors (cortical, thalamic, and pallidal) were used in an adversarial-like approach to generate a realistic simulation of the thalamo-cortical-basal ganglia system in a conscious state. Specifically, these networks contributed to the loss function of a genetic optimization algorithm, which tuned the parameters of a mean-field model of large-scale electrodynamics in the cortex, thalamic nuclei, striatum, globus pallidus, and subthalamic nucleus. Note that the anatomical connectivity of the model is based on the known structural connections between cortical and subcortical regions (see [19, 30] for more details), but that the strength of any particular structural connection could be modified by the genetic optimization algorithm. The loss function of the genetic optimization also included outputs from additional deep neural networks trained to differentiate real cortical electrophysiology from synthetic signals and detect seizures (see Methods). Empirical constraints, such as region-specific firing rates, phase-amplitude coupling, cross-frequency information transfer, and weakly chaotic dynamics, were also added to the loss function. Once model parameters were identified which produced simulations classified as real, conscious, and non-seizing, additional discriminator DCNNs were trained to differentiate between simulated and real electrophysiology from the cortex, thalamus, and GPe during conscious states. The outputs of these discriminator networks were then added to the loss function, and genetic optimization was repeated until these networks failed to discriminate between real and simulated LFPs from the conscious brain (Fig. 2A)

This iterative approach led to a realistic simulation of the conscious brain: simulated LFPs from the cortex, thalamic projection nuclei, and GPe were all classified as real and conscious (predicted GCS=15) by the trained networks, with mean firing rates in each region falling within empirically tested bounds (Supplementary Table 1). Fig. 2 shows the simulated cortical (Fig. 2B), thalamic (Fig. 2C), and pallidal (Fig. 2D) LFPs, along with their corresponding power spectra, compared to real electrophysiology recordings from multiple species. Additionally, phase-amplitude coupling between theta (4–7 Hz) oscillations and gamma (40–120 Hz) amplitude was observed within the cortex, thalamus, and GPe, as well as between the cortex and GPe and between the cortex and thalamus (Fig. 2F), consistent with real brain dynamics (Fig. 2E), though the simulated LFPs exhibited stronger alpha-gamma (8–12 Hz) coupling than the real electrophysiology data. (Note that no empirical data with simultaneous pallidal and thalamic recordings were available to analyze coupling between these subcortical structures, and so this was not evaluated in our model). Moreover, there was significant directed information transfer from low-frequency (∼1–13 Hz) rhythms in the cortex to high-frequency (52–104 Hz) rhythms in the thalamus (harmonic mean across 20 runs with different initial conditions $\mathring{p}$=0.045), and vice versa ($\mathring{p}$=0.033) (Fig. 2G), replicating a previously established signature of consciousness [19] (see Methods for details on spectral information transfer analysis). Finally, the simulated cortical LFP was weakly chaotic, hovering near edge-of-chaos criticality (Fig. 2H) (mean largest Lyapunov exponent of 0.69 across 20 runs with different initial conditions), another hallmark of consciousness identified in previous work [18, 19, 21]. Collectively,

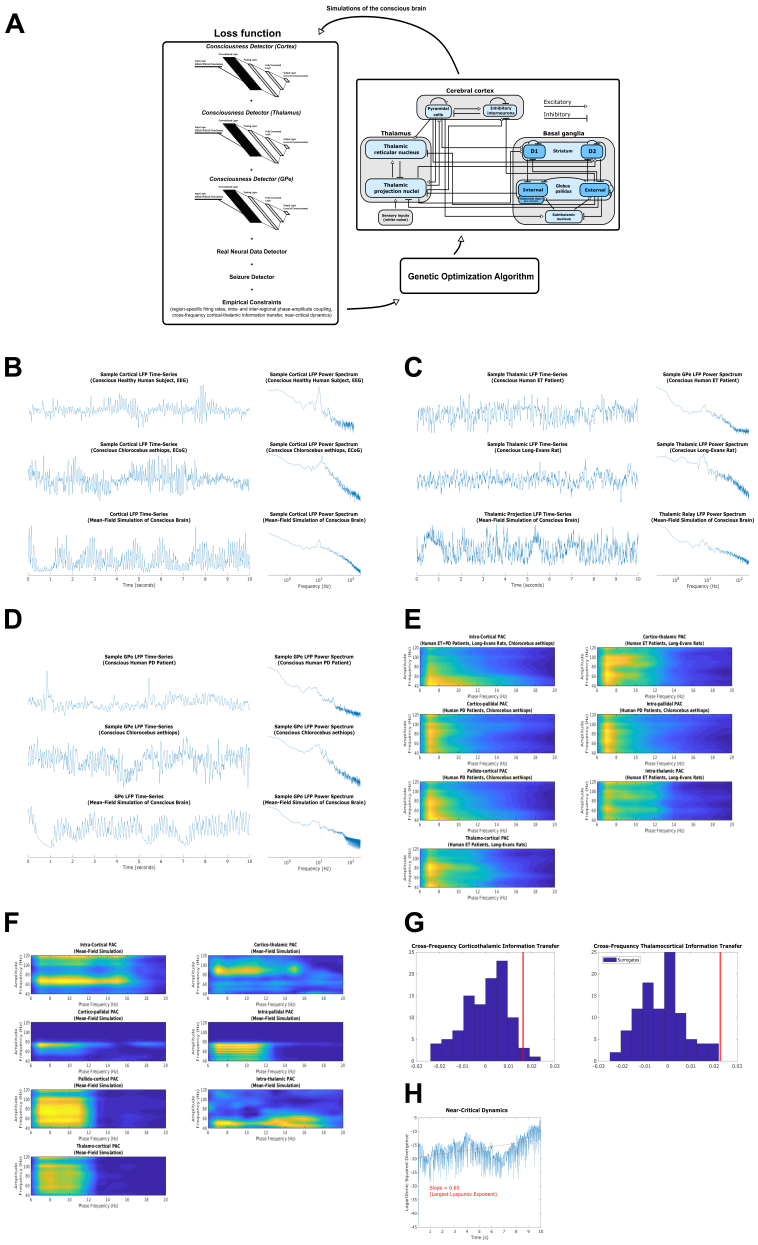

Fig. 2 Machine learning-driven simulation of the conscious thalamo-cortical-basal ganglia system. **(A)** Schematic illustrating the iterative process of using neural networks and genetic optimization to refine a mean-field model of neural electrodynamics based on known anatomical connectivity. The loss function incorporated outputs from networks trained to classify real versus synthetic data, detect consciousness, and identify seizures, as well as empirical constraints. **(B-D)** Simulated LFPs (left panels) and power spectra (right panels) from the cortex (B), thalamic projection nuclei (C), and GPe (D), compared to sample real LFPs and their corresponding power spectra from multiple species. **(E-F)** Phase-amplitude coupling in real (E) versus simulated (F) LFPs. In both real and simulated data, strong theta-gamma coupling was observed within the cortex, thalamus, and GPe, as well as between brain regions. However, simulated data exhibited stronger alpha-gamma coupling. **(G)** Directed information transfer from low-frequency (∼1–13 Hz) cortical rhythms to high-frequency (52–104 Hz) thalamic rhythms, and vice versa. The red vertical line represents the difference between the calculated transfer entropy from source to target when signals are unscrambled, and the estimated transfer entropy when low-frequency dynamics at the source are scrambled. The blue histogram represents the distribution of values across 100 surrogate datasets, where both low-frequency source and high-frequency target dynamics are scrambled. Statistical significance is determined if fewer than 5% of surrogate values (blue histogram) exceed this difference (red line) (see Methods for further details). **(H)** Weak chaos in the simulation, indicated by the logarithmic squared divergence between two 10-second runs of the model with slightly different initial firing rates. The slope of the fitted line is positive but near-zero, consistent with weakly chaotic cortical dynamics near edge-of-chaos criticality.

these findings demonstrate the success of our machine learning-guided approach in simulating the electrodynamics of the conscious brain.

2.2 Simulating Disorders of Consciousness

Following the successful simulation of conscious brain electrodynamics, the next step involved using the identified "conscious brain" parameter set as a foundation to model DOC. Genetic optimization was applied across parallel runs with different random seeds to generate a range of DOC phenotypes. The loss function enforced a "dose-response" relationship between degree of deviation from the waking state parameters and the AI-predicted level of consciousness, based on the simulated cortical LFP and our trained cortical consciousness detector (as cortical recordings from scalp EEG are the primary data for DOC electrophysiology). Additionally, the loss function ensured that all DOC phenotypes were classified by our trained neural networks as real neural data and non-seizing. Across these parallel runs, 280 parameter sets met these criteria, producing valid simulations of DOC "cases." As illustrated in Fig. 3, these simulations exhibited a variety of DOC electrophysiology phenotypes, reflecting the diverse neural dynamics seen in coma, vegetative states, and minimally conscious states (see [23] for a review of EEG patterns across the spectrum of DOC).

With these simulated DOC "cases" in hand, we could examine the model parameters - such as structural connectivity between specific brain regions, or the biophysics of activity within a particular region - which drive variation in AI-predicted level of consciousness in the model. In other words, we could directly explore, in silico, neural correlates of pathological unconsciousness. To do so, we ran 1680 neural simulations of DOC using the mean-field model, with six levels of deviation from conscious brain parameters for each unique DOC parameter set or "case." To identify potential neural parameters which drive unconsciousness in DOC, we z-score normalized all parameters, then applied Lasso regression (with regression parameters selected so as to minimize root mean squared error) to select relevant features associated with AI-predicted level of consciousness, followed by ordinary least squares (OLS) regression for statistical significance testing. This analysis identified 16 out of 231 model parameters significantly linked to AI-predicted consciousness (Table 1). Notably, positive associations with predicted level of consciousness included the strength of structural (axonal) coupling from STN to GPe, GPe to STN, TRN to thalamic projection neurons, and excitatory cortical neurons to thalamic projection neurons. Intra-GPe axonal coupling and the delay in signal propagation between inhibitory cortical interneurons were also positively associated with predicted level of consciousness. Conversely, negative associations with AI-predicted level of consciousness were found with inhibitory-inhibitory cortical axonal coupling, GPe to TRN axonal coupling, and excitatory to inhibitory cortical axonal coupling, suggesting excessive coupling in these pathways may drive DOC states. Variance in TRN membrane potentials was negatively associated with consciousness, while variance in thalamic projection neuron potentials was positively associated, emphasizing the role of altered thalamic electrodynamics in pathological unconsciousness. See the Discussion for a detailed comparison of these results to prior empirical findings in DOC and severe brain injury, highlighting both the successful retrodictions and the novel predictions of our model of DOC.

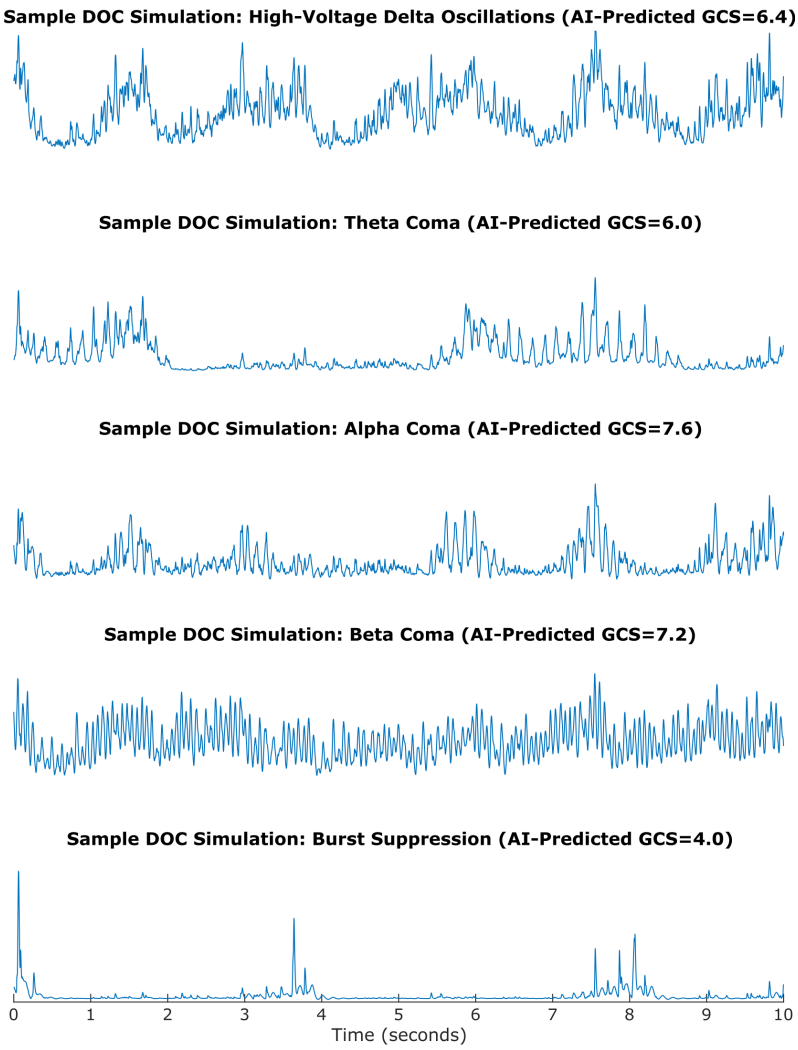

Fig. 3 Sample cortical LFPs from our AI-driven models of DOC, matching known electrophysiology phenotypes in these disorders [23]. All LFPs are plotted on the same scale.

Our models also allowed us to explore how firing rates in different brain regions are associated with pathological unconsciousness - a feature of DOC which is otherwise typically only inferred through positron emission tomography (PET). Using Lasso and OLS regression, we found that excitatory cortical firing rates were positively associated with predicted level of consciousness, whereas GPe, TRN, and inhibitory cortical interneuron firing rates were negatively associated. We also evaluated the relationship between AI-predicted level of consciousness and other putative electrophysiological correlates of consciousness, namely cortical Lempel-Ziv complexity [18, 31], the proximity of cortical electrodynamics to edge-of-chaos criticality [18–21], thalamocortical cross-frequency information transfer [19], and outputs from our trained thalamic

Table 1 Parameters, firing rates, and features of electrophysiology significantly associated with AI-predicted level of consciousness in mean-field models of DOC, based on Lasso regression with subsequent ordinary least squares (OLS) analysis to evaluate statistical significance. A positive coefficient indicates a positive association with consciousness, while a negative coefficient indicates a negative association. Note that "coupling" here refers to the strength of structural (axonal) coupling between two brain regions. PSP=postsynaptic potential.

Model Changes Associated with Predicted Level of Consciousness	Coefficient	p-value
Model Parameters		
Inhibitory Cortical to Inhibitory Cortical Propagation Delay	1.0409	0.0000
Variance of Thalamic Projection Neuron Membrane Potentials	1.0138	0.0000
Coupling Between Inhibitory Cortical Neurons	-0.9567	0.0000
Threshold Potential for STN Neurons	-0.6797	0.0000
Coupling from STN to GPe	0.5824	0.0000
Coupling from GPe to TRN	-0.5650	0.0000
Coupling from TRN to Thalamic Projection Neurons	0.4753	0.0000
PSP Duration from Inhibitory to Excitatory Cortical Neurons	-0.4682	0.0000
Coupling from GPe to STN	0.4658	0.0000
PSP Duration from Brainstem to Thalamic Projection Neurons	-0.4618	0.0000
Variance of TRN Membrane Potentials	-0.4382	0.0000
Intra-GPe Coupling	0.4344	0.0000
Coupling from Excitatory to Inhibitory Cortical Neurons	-0.3862	0.0000
Threshold Potential for GPe Neurons	-0.3854	0.0002
Propagation Delay from GPe to TRN	0.3837	0.0000
Coupling from Excitatory Cortical Neurons to Thalamic Projection Neurons	0.3729	0.0000
Firing Rates		
Excitatory Cortical Firing Rate	1.6087	0.0000
GPe Firing Rate	-0.8965	0.0000
TRN Firing Rate	-0.5507	0.0000
Inhibitory Cortical Interneuron Firing Rate	-0.3221	0.0004
Electrophysiological Correlates of Consciousness		
Deviation from Edge-of-Chaos Criticality	-1.2679	0.0000
Thalamic Consciousness Detector	0.9984	0.0000
Thalamocortical Cross-Frequency Information Transfer	0.3892	0.0000
Corticothalamic Cross-Frequency Information Transfer	-0.2486	0.0103
Cortical Lempel-Ziv Complexity	0.2082	0.0192

and pallidal consciousness detectors. Key predictors of unconsciousness included the deviation from edge-of-chaos criticality (measured as the absolute largest Lyapunov exponent of cortical LFPs), which is consistent with prior empirical results [18, 19, 21], and the output of our trained thalamic consciousness detector. Cross-frequency thalamocortical directed information transfer was also positively associated with consciousness, which is consistent with prior results in anesthesia and generalized seizures, while corticothalamic information transfer increased in DOC models, contrary to anesthesia and generalized seizures [19]. Finally, replicating prior empirical studies on acute coma [18, 32], chronic DOC [33], and anesthesia [18, 34], Lempel-Ziv complexity was positively associated with AI-predicted consciousness levels (Table 1).

2.3 Simulating Deep Brain Stimulation in DOC

Using our trained consciousness detectors paired with our simulations of DOC, we were able to investigate the effects of DBS across all modeled subcortical targets and frequencies ranging from 2 to 130 Hz. To do so, we incorporated a new neural population in the mean-field model to simulate DBS. This added population had no dendritic inputs and was programmed to generate a square-wave pulse oscillation at a specified frequency. The DBS population then coupled to a target subcortical population within the model with a fixed connectivity strength, following a method previously used to simulate DBS in this model (e.g. [35]).

For each stimulation parameter set—defined by a target subcortical region and a stimulation frequency—we simulated the resulting LFPs in the cortex, thalamic projection nuclei, and GPe. These simulated LFPs were then fed into our trained deep neural networks, specifically the cortical, thalamic, and pallidal consciousness detectors, to evaluate whether the simulated brain state was classified as conscious. To derive an overall consciousness prediction, we also calculated a composite measure by multiplying the predictions of the thalamic and pallidal detectors with that of the cortical consciousness detector (rescaled from its 3–15 range to fit a 0–1 scale).

As shown in Figure 4, our findings revealed that low-frequency (<30 Hz) DBS applied to the thalamus can lead to a partial reignition of AI-predicted consciousness across a majority of our simulated DOC "cases," which is consistent with prior empirical results on this range of thalamic stimulation frequencies in DOC [4–6]. Similarly, high-frequency (>50 Hz) GPe stimulation led to a revival of AI-predicted consciousness in a majority of our DOC simulations, which is consistent with prior results on pallidal DBS for DOC in this frequency range [7]. However, the DBS paradigm most consistently associated with AI-predicted restoration of consciousness was high-frequency (>50 Hz) stimulation of the STN (Figure 4). To directly compare the effects of high-frequency stimulation of the GPe versus STN - the two most successful DBS paradigms for reigniting AI-predicted levels of consciousness in our DOC models - we examined the median predicted levels of consciousness (for each consciousness detector) at frequencies above 50 Hz for both stimulation targets (STN and GPe) across all 280 simulated DOC cases. The cortical consciousness detector predicted significantly higher levels of consciousness with high-frequency STN stimulation than with GPe stimulation ($p<0.001$, one-tailed Wilcoxon signed-rank test), as did the pallidal ($p<0.001$) and composite ($p<0.001$) consciousness detectors, but not the thalamic consciousness detector ($p=0.5117$, two-tailed signed-rank test). High-frequency STN stimulation also resulted in significantly higher Lempel-Ziv complexity of simulated cortical LFPs ($p<0.001$) and significantly closer proximity of cortical dynamics to edge-of-chaos criticality ($p<0.001$), both of which are electrophysiological correlates of consciousness (see Table 1). However, while significant cross-frequency thalamocortical information transfer (another correlate of consciousness [18] - see Table 1) was maintained during both pallidal and subthalamic high-frequency stimulation ($\mathring{p}=0.0124$ across simulated "cases" and stimulation frequencies for pallidal stimulation, and $\mathring{p}=0.0154$ for subthalamic stimulation), the strength of directed thalamocortical cross-frequency information transfer was significantly higher during simulated high-frequency (>50 Hz) pallidal DBS than in subthalamic DBS ($p<0.001$). The implications of this for

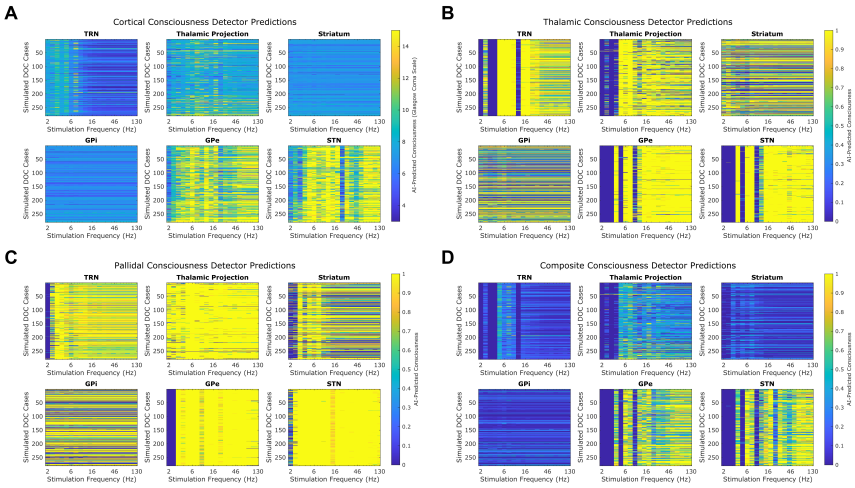

Fig. 4 Predicted consciousness levels across deep brain stimulation (DBS) parameters for various subcortical targets and frequencies. Heatmaps display AI-predicted levels of consciousness for each target and stimulation frequency, as assessed by our trained deep neural networks (A–C). Stimulation frequencies range from 2 to 130 Hz along the x-axis (logarithmically spaced), while the y-axis represents the 280 simulated DOC cases. Note that here striatal DBS is defined as simultaneous stimulation of D1 and D2 striatal neurons. Predicted consciousness is color-coded on a scale corresponding to the prediction output of each detector: (A) cortical consciousness detector predictions, (B) thalamic consciousness detector predictions, and (C) pallidal consciousness detector predictions. In each subplot, DBS was simulated for one of six subcortical targets: the thalamic reticular nucleus (TRN), thalamic projection nuclei, striatum, globus pallidus internus (GPi), globus pallidus externus (GPe), and the subthalamic nucleus (STN). Panel (D) shows the composite consciousness predictions, calculated by combining the outputs of the cortical, thalamic, and pallidal detectors, where cortical predictions were rescaled from a 3–15 range to a 0–1 scale to match the other detectors. High-frequency DBS (>50 Hz) to the STN most consistently resulted in higher AI-predicted consciousness levels across the cortical, pallidal, and composite consciousness detectors, as indicated by the warmer colors.

DBS restoration of awareness in DOC, however, are unclear. Collectively, these results align with prior work on thalamic and pallidal DBS for DOC, and also suggest that the STN, a relatively unexplored structure in DOC, may be a promising new target for high-frequency DBS to restore awareness across a variety of etiologies of DOC.

3 Discussion

Our study provides a novel computational framework for understanding the neural mechanisms underlying consciousness and its disorders, and explores how DBS could be used to treat pathological unconsciousness. By integrating a large and diverse neural electrophysiology dataset, deep learning, and neural field theory, we modeled large-scale brain dynamics and generated in silico simulations of both conscious and comatose states. This approach allowed us to probe not only well-documented mechanisms of DOC, but also to identify novel potential drivers of unconsciousness following severe brain injury.

One of the mechanisms of DOC identified by our model, namely, the disruption of structural corticothalamic pathways (Table 1), is one of the most well-documented anatomical correlates of DOC, regardless of etiology [9]. Our model also predicts significant alterations in the dynamics of thalamic projection neurons in DOC, which aligns with the known vulnerability of the thalamus to trauma and ischemic insult—even when these insults are not directly localized to the thalamus [36–38]. Notably, the extent of thalamic atrophy has been correlated with levels of consciousness in DOC patients across multiple etiologies [39]. Similarly, our model's positive association between cortical firing rates and predicted consciousness aligns with the well-replicated observed reduction in cortical metabolic activity in DOC patients, irrespective of the injury's cause [24–27].

Alterations in pallidal circuits also emerged as critical in our model's predicted drivers of unconsciousness in DOC, which is consistentn with prior empirical work. Diffuse TBI has also been shown to cause significant GABAergic interneuron loss in the GP [40], likely leading to reduced intra-pallidal inhibition—another predicted driver of DOC in our model—as well as increased pallidal activity, which is also predicted by our model. Indeed, increased pallidal activity in DOC is a key component of the mesocircuit hypothesis [9], and has also been demonstrated empirically using PET in DOC patients with diverse etiologies [27]. Loss of GPe interneurons could also diminish inhibition to the STN, another predicted driver of unconsciousness in our simulations (Table 1).

The TRN also appears to play a significant role in driving pathological unconsciousness in our simulations. Increased TRN activity, which is predicted by our models of DOC (Table 1), has been suggested as a possible DOC driver due to the TRN's ability to induce cortical slow-wave activity and suppress arousal by inhibiting thalamocortical circuits [41]. While this has yet to be empirically studied in DOC (to our knowledge), recent studies have found TRN activation in anesthetic coma [42]. Perhaps paradoxically, our model also predicts that unconsciousness in DOC is associated with *reduced* TRN to thalamic projection neuron coupling, though this too aligns with empirical evidence: inhibitory TRN inputs to thalamic projection nuclei degenerate after cortical stroke [43] and in coma following severe TBI [44]. Speculatively, this combination of reduced structural connectivity from the TRN to thalamic projection cells with a concomitant increase in TRN firing may lead to a disruption in the role of the TRN in gating thalamocortical information transfer: periods of heightened TRN activity may amplify its inhibitory impact on other thalamic nuclei during periods of high firing, while the reduced coupling could create windows where thalamic projection neurons are insufficiently modulated during periods of less TRN activity. This may perhaps result in a fragmented and less effective thalamocortical interaction, and partly underpin the observed loss in thalamocortical information transfer in our models of DOC (Table 1). Future research should investigate how TRN activity and connectivity interact to affect thalamic output and unconsciousness following brain injury.

Our model also predicts a crucial role for altered intra-cortical inhibition in pathological unconsciousness. Specifically, it identifies an increased duration of iPSPs in cortical pyramidal neurons receiving inhibitory inputs as a mechanism linked to DOC

(Table 1). While this has not, to our knowledge, been studied in brain injury, prolonged cortical iPSPs are a known neuronal mechanism of anesthetic coma [45, 46]. Our model further predicts decoupling between cortical inhibitory interneurons in DOC, which is consistent with reduced cortical interneuron connectivity and decreased inhibitory postsynaptic currents at inhibitory interneurons observed in severe TBI [47, 48]. Additionally, paired-pulse TMS studies have shown reduced intracortical inhibition in vegetative states [49], possibly due to damage to interneuronal axons. This reduced interneuron connectivity is also consistent with PET findings of reduced cortical GABAa receptors in chronic DOC patients, again irrespective of etiology [50, 51]. Conversely, our model's prediction of increased excitatory to inhibitory cortical coupling aligns with findings of increased excitatory synaptic strength onto inhibitory neurons after severe TBI [47].

However, the parameter most strongly associated with predicted consciousness level in our models—namely, the delay in electrical propagation between cortical inhibitory interneurons, with unconsciousness linked to faster conduction (Table 1)—remains unexplored experimentally in the context of DOC or severe brain injury (to our knoweldge). Thus, increased conduction velocity along axons connecting surviving cortical interneurons represents a novel predicted hallmark of DOC, warranting further experimental investigation.

Finally, our model provides insights into potential therapeutic interventions, particularly the application of DBS to restore consciousness. Our approach enables systematic exploration of DBS parameter spaces, offering a way to evaluate optimal target regions and stimulation frequencies before clinical testing. This capability is especially valuable given the invasive nature of DBS and the ethical and logistical challenges of testing multiple parameters in patients. Our simulation results show that high-frequency (50-130 Hz) DBS applied to the STN was most consistently associated with AI-predicted reignition of consciousness across all model variants. Though subthalamic DBS has not been explored in DOC, high-frequency (130 Hz) STN stimulation in animal models has been shown to enhance recovery from stroke and increase glucose metabolism in the ipsilateral cortex, thalamus, and striatum [52]—all areas associated with consciousness regulation [9]. These findings suggest that the STN may serve as a promising new target for DBS in DOC.

While our computational model offers valuable insights into the neural mechanisms underlying DOC and the potential efficacy of DBS in restoring consciousness in these disorders, several limitations should be noted. First, the mean-field model was primarily tuned using deep neural networks trained on thalamic, pallidal, and cortical activity. However, our results highlighted the potential involvement of other critical regions, such as the TRN and the STN, in both inducing and modulating unconscious states. Incorporating empirical recordings from these regions into the further refinement of our model could enhance its realism and offer a more comprehensive understanding of the mechanisms underlying DOC. Additionally, the model assumes a simplified representation of the brain's dynamics, which may not capture the full complexity of individual neuron and synapse behaviors, especially in regions with heterogeneous cell types, such as the cortex and thalamus. A related limitation of the

model is that we treat both thalamic relay and intralaminar nuclei as a single, non-specific thalamic projection nucleus. Future work should therefore aim to assess the clinical relevance of predicted therapeutic interventions, including the role of increased conduction velocity among cortical interneurons and the potential efficacy of STN-targeted DBS, through both animal models and controlled clinical trials. Addressing these limitations will help refine our model and ultimately enhance its utility as a predictive tool for developing targeted therapies for DOC.

In conclusion, this study demonstrates the potential of an adversarial-like AI approach to advance our understanding of the neural correlates of consciousness and DOC, and to guide the development of effective neuromodulation therapies. By bridging empirical data with large-scale neural field modeling and deep learning, we have developed a new framework for simulating both conscious and comatose states, elucidating neural mechanisms underlying unconsciousness, and testing stimulation-based interventions in silico. Through this integration of computational and clinical neuroscience, we take a step closer to developing targeted, mechanism-driven therapies for disorders of consciousness.

4 Methods

4.1 Deep Neural Network Training

To generate a simulation of the conscious and comatose brain, we first trained a deep neural network to accurately detect varying levels of consciousness based on cortical electrophysiology. To do so, we began with clinical EEG recordings collected from 22 ICU patients in coma. All neural data were z-score normalized to standardize the input features prior to training. The architecture of the neural network consisted of convolutional layers, followed by batch normalization, ReLU activations, average pooling layers, and fully connected layers, ending in a custom output layer that bounded predictions between 3 and 15 to match the GCS scale. The backward pass of the custom output layer was designed to compute the gradients only when the predictions were within the designated bounds, preventing erroneous updates during backpropagation when the output was outside the desired range. Hyperparameters for the network, including learning rate, batch size, the size of the convolutional filter, pooling stride, and the number of units in the fully connected layer were optimized using Bayesian optimization, and the final model was trained using the RMSprop optimizer. Gradient clipping was employed to prevent exploding gradients. Validation of the model occurred every 30 iterations, using 20% of the data set aside for this purpose.

After the initial training phase, the neural network was further refined using additional cortical ECoG data from both African green monkeys (*Chlorocebus aethiops*) and human ET and PD patients in awake, fully conscious states. This refinement process involved adjusting the learning rate of the convolutional and fully connected layers to enable better fine-tuning for cross-species generalization. Specifically, the learning rate for the convolutional layers was reduced by a factor of 0.05, while the fully connected layers had their learning rates reduced by a factor of 0.25. Following these adjustments, the model was again refined (with the same reduced learning rates) using both the EEG and ECoG datasets to ensure the network could generalize

across modalities and species, maintaining robustness while predicting GCS scores. Subsequently, the model was further refined to predict levels of consciousness in 14 patients with chronic disorders of consciousness (DOC), including those in vegetative and minimally conscious states. These patients were assessed using the Coma Recovery Scale-Revised (CRSR), which differs from the Glasgow Coma Scale (GCS) used in acute coma patients. To bridge the two scales, a piecewise linear mapping was developed to translate CRSR scores into GCS values for training purposes, ensuring that the network could detect varying levels of consciousness across these different scales. For patients with the lowest CRSR scores (0 to 5), reflecting the most severely impaired states, a linear mapping was applied to map these values to the lower end of the GCS scale, corresponding to GCS values between 3 and 5. Similarly, CRSR scores in the intermediate range (6 to 10) were mapped linearly to GCS values between 6 and 8, representing moderate impairment of consciousness. Higher CRSR scores (11 to 15) were mapped to GCS values ranging from 9 to 11, reflecting milder impairments. Finally, for patients nearing full consciousness, with CRSR scores between 16 and 23, the scores were linearly scaled to GCS values between 12 and 15. This continuous linear mapping within each CRSR range ensured a smooth transition between the scales, allowing the neural network to effectively integrate chronic DOC data within the GCS framework, while capturing subtle variations in consciousness between patients. Finally, the network was again refined (with the same reduced learning rates) to accurately detect consciousness in resting-state EEG data from healthy, awake humans.

We then trained a second deep neural network to differentiate between recordings of real cortical electrodynamics and synthetic data. To do so, we generated a variety of synthetic signals, including 500 amplitude-modulated signals, 500 frequency-modulated signals, 500 sawtooth waves, 500 square waves, 500 parametric sine waves, 500 simulations of the Lorenz system, and 500 instances of pink, blue, and red noise. These synthetic signals were created with random variations in parameters such as carrier frequency, modulating frequency, modulation index, and phase shift, to simulate diverse patterns of neural activity. A deep neural network was then trained to differentiate between these synthetic signals and real cortical electrophysiology data, including EEG data from human coma patients, ECoG data from monkeys, human Parkinson's patients, human essential tremor patients, and epilepsy patients, and MEG data from one human epilepsy patient. The diversity of data on which this network was trained was to ensure that its discrimination of synthetic versus neural signals could generalize across different brain states, species, and electrophysiology recording modalities. This network had a similar architecture to the GCS prediction network, with convolutional layers, batch normalization, ReLU activations, average pooling, fully connected layers, and a custom output layer that bounded the predictions for binary classification. As with the GCS network, all data were z-score normalized prior to training. The synthetic data were combined with the real neural data to form a balanced training set, and the network was trained using the RMSprop optimizer, and a learning rate, minibatch size, convolutional filter size, pooling stride, and number of units in the fully connected layer determined using Bayesian optimization. Validation was performed every 30 iterations, and gradient clipping was applied to ensure stable learning.

In addition to the GCS and synthetic data classifiers, we trained a third deep neural network to differentiate between generalized seizure data and non-seizure data. The non-seizure data included periods of normal healthy waking from humans and rats with epilepsy, as well as cortical ECoG recordings from human ET patients, human PD patients, and African green monkeys, captured during both waking and anesthetized states. We also incorporated EEG recordings from our acute coma patients across the full range of GCS scores. The human seizure data were re-analyzed from a previously published study [18], and included surface ECoG recordings from two patients, as well as MEG data from an additional seizure patient. Additionally, we included generalized seizure and normal waking data from seven Genetic Absence Epilepsy Rats from Strasbourg (GAERS), which provided further animal data for training. The network architecture for the seizure classifier mirrored the structure of the previous networks, with convolutional layers followed by batch normalization, ReLU activation, and average pooling, culminating in fully connected layers. The output was again constrained by a custom bounded output layer designed for seizure classification. The hyperparameters for this network, including the convolutional filter size, pooling stride, and the number of fully connected units, were optimized using Bayesian optimization.

Two more deep neural networks were trained to classify LFPs from the GPe and thalamus, distinguishing conscious from anesthetized states. Thalamic LFPs were collected from human ET patients and Long-Evans rats, while GPe LFPs came from human PD and two African green monkeys. The network architecture was consistent with previous models, and Bayesian optimization was used to tune hyperparameters, including filter size, pooling stride, and fully connected units. The networks were trained using the RMSprop optimizer and validated every 30 iterations using 20% of the data.

The outputs of these trained networks formed the foundation for a loss function of a genetic optimization algorithm, which tuned the parameters of a mean-field model of the basal ganglia-thalamo-cortical system (see below). After the genetic algorithm identified a set of mean-field model parameters that produced a first-pass simulation of the conscious brain—where simulated cortical dynamics were classified as non-synthetic, non-seizing, and with a GCS of 15, and the simulated thalamus and GPe were classified as conscious—we proceeded to train three additional discriminator networks. For this, we generated 1,000 simulations of 10 seconds of activity in the cortex, thalamus, and basal ganglia using the identified model parameters, splitting the data into 800 simulations for training and 200 for validation. We then selected 800 time-series from the previously designated training data of real cortical, thalamic, and GPe LFPs during conscious states, as well as 200 real time-series for validation. This resulted in balanced datasets of 800 real versus 800 simulated 10-second LFPs for each brain region in the training set, and 200 real versus 200 simulated LFPs for the validation set. Importantly, all real training and validation data were drawn from the prior datasets used for training the earlier networks, while maintaining the same separation between training and validation data as before. The architecture of these three additional networks was consistent with previous models, with the same custom output layer that bounded the predictions for binary classification. As before, Bayesian optimization was used to tune hyperparameters, and training utilized the RMSprop

optimizer and the model was validated every 30 iterations. See below regarding the use of these trained networks in tuning our mean-field model.

4.2 Phase-Amplitude Coupling Analysis

Phase-amplitude coupling (PAC) patterns corresponding to waking, conscious states were identified within the cortex, thalamus, and GPe, as well as between cortex-thalamus and cortex-GPe in both directions. For the thalamic and thalamo-cortical analyses, we used all available data from human ET patients and Long-Evans rats. For the GPe and pallido-cortical analyses, data from human PD patients and African green monkeys were used. The intra-cortical analysis utilized data from human ET and PD patients, Long-Evans rats, and African green monkeys. PAC was computed by analyzing the modulogram from each 10-second window of data. To do so, we first filtered the source and target LFP signals to extract phase and amplitude across a range of frequencies (in the case of intra-regional PAC, the source and target signals were the same). Phase frequencies spanned 1 to 15 Hz, while amplitude frequencies ranged from 25 to 105 Hz. The phase of the source signal and the amplitude envelope of the target signal were computed using Hilbert transforms, and the modulation index (MI) was calculated within predefined phase bins. For each animal or patient, we computed a mean modulogram across all windows, resulting in an average PAC profile for each individual. These individual profiles were then averaged across animals or patients to generate overall modulograms for intra-cortical, intra-thalamic, intra-pallidal, cortico-thalamic, thalamo-cortical, cortico-pallidal, and pallido-cortical PAC. These empirical PAC patterns were then incorporated into the loss function of our genetic optimization algorithm, to create a simulation of the conscious brain with realistic PAC.

4.3 Mean-field model and genetic optimization

In this study, we employed a mean-field model similar to the one described in previous work to simulate LFPs and the average firing rates of various neural populations [16, 19, 30]. The firing rate, denoted as $Q_a(r,t)$, represents the proportion of neurons within a population whose membrane potential exceeds their reversal potential. This is modeled as a sigmoidal function of the number of neurons whose potential V_a surpasses the threshold potential θ_a for that population:

$$Q_a(r,t) = \frac{Q_a^{\max}}{1 + \exp[-(V_a(t) - \theta_a)/\sigma']} \tag{1}$$

where $Q_a^{\max}$ is the maximum possible firing rate, and σ' is the standard deviation of the membrane potentials relative to the threshold potential. The change in mean membrane potential V_a over time is described by:

$$D_{\alpha\beta}(t)V_a(t) = \sum_b v_{ab}\phi_b(t - \tau_{ab}) \tag{2}$$

where v_{ab} is the synaptic strength between population b and population a, $\phi_b(t - \tau_{ab})$ is the firing rate of population b delayed by τ_{ab}, and $D_{\alpha\beta}(t)$ is the differential operator representing the rise and decay dynamics of the membrane potential:

$$D_{\alpha\beta}(t) = \frac{1}{\alpha\beta}\frac{\mathrm{d}^2}{\mathrm{d}t^2} + \left(\frac{1}{\alpha} + \frac{1}{\beta}\right)\frac{\mathrm{d}}{\mathrm{d}t} + 1 \tag{3}$$

Here, α represents the decay rate, and β is the rise rate of the membrane potential. The synaptic response, $h(t)$, was modeled as:

$$h(t) = \frac{H}{\eta(\alpha,\beta)}\overline{h}(t) \tag{4}$$

where $\overline{h}(t)$ is the original synaptic response, and H represents the synaptic response scaling factor. In contrast to our previous model, where H was fixed at 31.5 s^{-1} [19], in this study, we allowed H to vary independently for each brain region.

The outgoing mean electric field, ϕ_{ab}, between populations was modeled using the damped wave equation:

$$\mathcal{D}_{ab}\phi_{ab}(r,t) = Q_b(r,t) \tag{5}$$

where

$$\mathcal{D}_{ab} = \left[\frac{1}{\gamma_{ab}^2}\frac{\partial^2}{\partial t^2} + \frac{2}{\gamma_{ab}}\frac{\partial}{\partial t} + 1 - r_{ab}^2\nabla^2\right] \tag{6}$$

with r_{ab} representing the spatial range of axonal projections, γ_{ab} the temporal damping coefficient, and ∇^2 the Laplacian operator. The circuit connectivity was based on the original van Albada and Robinson model [30], with additional projections from the GPe to cortical and thalamic regions as in our recent modification of the original model, following recent anatomical studies [19].

The model parameters were tuned using a genetic machine learning algorithm, similar to our previous approach [19], to identify a physiologically realistic configuration for waking brain states. Specifically, we sought to generate low-amplitude, weakly chaotic oscillations of LFPs, where the LFPs for each neural population were modeled as the superposition of synaptic currents within a given brain region. The goal was to achieve weakly chaotic dynamics near criticality, consistent with mounting evidence that the large-scale, oscillatory electrodynamics of the cortex are poised near edge-of-chaos criticality during conscious states [18, 19, 21]. Additionally, we tuned the parameters to match empirical firing rates across all brain regions and to ensure that fluctuations in cortical firing rates correlated with high-gamma (60–200 Hz) LFP amplitude fluctuations [53]. The model was also required to replicate the spectral patterns of bidirectional cortico-thalamic information transfer previously observed in our empirical data as a hallmark of consciousness [19]. We also tuned the model to reproduce the phase-amplitude coupling patterns observed within and between brain regions (cortex, thalamus, and GPe) during conscious states, and to ensure that five separate neural networks classified the simulated data appropriately: one network classified the cortex as conscious, another the thalamus as conscious, a third the GPe as conscious, a fourth classified the cortex as non-seizing, and a fifth classified the cortical data as non-synthetic.

Inspired by the framework of generative adversarial networks, we then trained three new deep neural networks to differentiate between our simulated conscious cortex,

GPe, and thalamus, and the corresponding real conscious data for each of these regions (see above). We then employed genetic optimization to further refine the parameters of the mean-field model, iteratively adjusting the model until the three discriminator networks were effectively fooled, classifying the simulated cortex, GPe, and thalamus as both real and conscious.

4.4 Acute coma data

EEG data from 22 patients in acute coma due to traumatic brain injury (TBI) were collected at the UCLA Ronald Reagan Medical Center and used to train the cortical consciousness detector. These data have been described in detail in a previous paper from our group [54]. Patients were included based on GCS scores <9 or GCS 9-14 with CT evidence of intracranial bleeding. Patients with non-significant head injuries, prior neurological disease, or brain death were excluded. Patients' conscious state was assessed multiple times daily using the GCS. EEG segments were spaced at least 12 hours apart. Data were divided into 10-second trials, cleaned for line noise with a band-stop filter, detrended, demeaned, and re-referenced to the common average. Independent component analysis (ICA) was then performed to remove ocular and cardiac artifacts, and trials with significant motion artifacts were excluded after visual inspection.

4.5 Chronic DOC data

EEG data from fourteen patients with chronic disorders of consciousness (0.97-6.8 years since injury) resulting from mixed etiologies (traumatic or non-traumatic injuries) (ages 22-70, seven females, seven males) were used to train the cortical consciousness detector. Data were recorded at Casa Colina Hospital and Centers for Healtchare as part of an ongoing study on the effect of low-intensity focused ultrasound on awareness in chronic DOC. The study was approved by Institutional Review Boards of Casa Colina Hospital and Centers for Healtchare and UCLA. Continuous EEG data were recorded at 1,000 Hz at baseline (before administration of low-intensity focused ultrasound). EEG recordings were divided into 10-second trials, band-stop filtered for line noise and harmonics, detrended, demeaned, and re-referenced to the common average. ICA was then performed to remove cardiac and ocular artifacts, and trials with significant motion artifacts were excluded after visual inspection.

4.6 *Chlorocebus aethiops* anesthesia data

Simultaneous intracranial electrophysiology recordings from the frontal cortex and GPe of African green monkeys (*Chlorocebus aethiops*) were reused from a previously published study [55] to train the pallidal consciousness detector and evaluate cortical and pallidal phase-amplitude coupling patterns. The study, approved by The Hebrew University (MD-18-15449-5), involved two surgeries under deep anesthesia. In the first, a recording chamber and EEG electrodes were implanted to target the cortex and GPe. In the second, a venous catheter was inserted. MRI confirmed correct placement of the recording chamber, and recording sessions followed a protocol alternating

between baseline, sedation, and washout, with vital signs monitored. During recordings, microelectrodes captured raw signals, spikes, and LFPs. Here, only recordings during waking and propofol anesthesia states were used. Recordings were divided into 10-second trials, detrended, demeaned, and re-referenced to the common average. Line noise was removed with a band-stop filter, and trials with significant motion artifacts were excluded. Data were provided by J.G.

4.7 Parkinson's disease patient anesthesia data

Data from a prior study from our group [11], which included pallidal LFP and frontoparietal ECoG data from 12 PD patients who underwent GPi DBS lead implantation, were reused here to train our pallidal consciousness detector and evaluate pallidal-cortical phase-amplitude coupling. The average age of the patient cohort was 65±7 years, and included three female patients. All patients provided informed consent, and PD medications were stopped 12 hours prior to surgery. Preoperative MRI and postoperative CT scans were used to target the motor GPi for lead placement, confirmed via intraoperative stimulation testing. An 8-contact ECoG strip was temporarily inserted through the frontal burr hole for cortical and depth electrode LFP recordings (from both GPe and GPi) after DBS implantation and propofol administration.

LFPs were recorded from the DBS lead's 4-ring contacts (ventral to dorsal), with data acquired using BCI2000 and sampled at 2400 Hz. Recordings began with patients awake and resting before propofol was administered to induce loss of responsiveness. Data were split into the first minute waking pre-anesthetic bolus, and the last minute of recordings, by which point all subjects were verbally unresponsive. Localization of ECoG strip contacts in M1, S1, and DBS contacts was based on either postoperative imaging or the median nerve SSEP technique. Recordings were divided into 10-second trials, detrended, demeaned, band-stop filtered for line noise, and re-referenced to the common average, and visually inspected for trials with significant motion artifacts, which were excluded.

4.8 Healthy, awake human data

EEG data from ten healthy, right-handed volunteers (six females, four males; mean age 30.25 ± 6.95 years) were obtained from an open-source dataset previously described by Torkamani-Azar and colleagues [56]. Participants had normal or corrected-to-normal vision and were not taking any drowsiness-inducing medications. Following consent per Sabanci University's Research Ethics Council, EEG data were recorded in a dimly lit, Faraday-caged room. Monopolar EEG signals were acquired using 64 Ag/AgCl active electrodes placed according to the 10-10 system. Data from a 2.5-minute resting-state session with eyes open were used as part of the training for our cortical consciousness detector.

4.9 Human essential tremor patient anesthesia data

Data from another previous study by our group [57] were re-analyzed to train the thalamic consciousness detector and assess thalamic-cortical phase-amplitude coupling.

Recordings were collected from 10 ET patients (6 female, 4 male, ages 60–79) undergoing unilateral (n=6) or bilateral (n=4) DBS implantation in the ventral intermediate (VIM) nucleus of the thalamus. All participants gave written informed consent, and the study was approved by the UCLA institutional review board. LFPs were recorded from the VIM thalamus, and ECoG signals from the ipsilateral frontoparietal cortex during rest and following intravenous propofol administration. Signals were acquired at a 2400 Hz sampling rate and filtered between 0.1 and 1000 Hz.

Patients remained awake with eyes open for the first minute of recording, which was used as the "awake" state. Propofol was then administered until patients reached a MOAA/S score of 0 (no response) or 1 (response to noxious stimuli). LFP and ECoG recordings continued for 5 minutes after anesthesia, with the final minute used as the "anesthetized" state. Data were segmented into 10-second trials, demeaned, detrended, and line noise was removed using a band-stop filer. Trials with multi-channel artifacts were excluded after visual inspection.

4.10 Long-Evans rat anesthesia data

Previously published data from Reed and Plourde [58] were reanalyzed to train a thalamic consciousness detector. Bipolar electrodes were placed in the ventral posterior nucleus of the thalamus and ipsilateral sensory cortex, with a reference electrode in the parietal bone. Propofol was administered intravenously, achieving plasma concentrations of 3, 6, 9, and 12 μg/ml. LFPs were recorded after 15 minutes of drug equilibration, with unconsciousness (loss of righting reflex) occurring at 9 μg/ml. LFPs from awake and 12 μg/ml conditions were used in this study. Signal were divided into 10-second trials, demeaned, detrended, band-stop filtered to remove line noise, and visually inspected for motion artifacts.

4.11 Human and rat generalized seizure data

Data from Miyamoto and colleagues [59] were reanalyzed to train the generalized seizure detector. Seven GAERS rats with spontaneous 6-8 Hz spike-and-wave seizures were implanted with stainless steel ECoG electrodes in the somatosensory cortex and a reference electrode on the cerebellum. Only seizures lasting at least 10 seconds were used. The data were split into 10-second trials, demeaned, detrended, and filtered at line noise and harmonics, with trials containing multi-channel artifacts removed. Seizures were manually labeled based on LFP traces.

In addition, data from two human epilepsy patients with secondarily generalized seizures, downloaded from the European Epilepsy Database [60] and described in detail in our prior study [18], were also used to train the generalized seizure detector. Subject 1, a 42-year-old male with cortical dysplasia, had six electrode strips (26 electrodes) over the right temporal cortex. Subject 2, a 14-year-old female with cryptogenic epilepsy, had one grid and six strips (96 electrodes) over the left temporal cortex. ECoG signals were recorded at 1024 Hz, resampled to 500 Hz, detrended, and re-referenced. Seizures lasting at least 10 seconds were selected, and artifact-laden electrodes were removed. Additionally, MEG data from an 18-year-old female with absence seizures [61] were used. The data, recorded at 625 Hz, were band-stop filtered

for line noise, resampled to the common average, and cleaned using ICA to remove ocular and cardiac artifacts.

4.12 *Macaca fuscata* anesthesia data

ECoG recordings from two macaques (*Macaca fuscata*), spanning multiple cortical areas (occipital, parietal, temporal, and frontal lobes), were downloaded from Neurotycho.org [62] and used as independent test data for our cortical consciousness detector. Recordings were taken during both awake/resting (eyes open and closed) and propofol anesthesia states, with Macaque 1 receiving 5.2 mg/kg and Macaque 2 receiving 5 mg/kg of propofol intravenously. Loss of consciousness was confirmed by the emergence of low-frequency oscillations and cessation of responses to physical stimuli. Signals were recorded at 1,000 Hz and then downsampled to 500 Hz. The data were demeaned, detrended, band-stop filtered at the line noise frequency and its harmonics, and re-referenced to the common average. Trials with significant motion artifacts were visually inspected, and those with artifacts affecting multiple electrodes were removed.

4.13 Proximity to Edge-of-Chaos Criticality

To assess the chaoticity of the mean-field model's dynamics (and whether its dynamics were near-critical), we calculated the stochastic largest Lyapunov exponent from simulated cortical LFPs. Lyapunov exponents measure how quickly nearby points in a system's phase space diverge: a positive exponent indicates chaos (exponential divergence), a negative exponent suggests periodic or stable behavior (exponential convergence), and a near-zero exponent defines edge-of-chaos criticality. For each parameter set of our model, the stochastic Lyapunov exponent was estimated by running the model for 20 seconds with random initial conditions. At 9.999 seconds, a small random perturbation was applied, and the rate of divergence between the first run and the perturbed run was measured over the final 10 seconds. The divergence, $\epsilon(t)$, between the two runs was calculated as the squared difference:

$$\epsilon(t) = \frac{(Q_e^{(1)}(t) - Q_e^{(2)}(t))^2}{\epsilon^{\max}}$$

where $\epsilon^{\max}$ is the maximum possible difference between the two runs:

$$\epsilon^{\max} = \left(\max(Q_e^{(1)}) - \min(Q_e^{(2)})\right)^2$$

The largest Lyapunov exponent, Λ, was determined by fitting the exponential divergence $\epsilon(t)$:

$$\epsilon(t) = \epsilon(0)\exp(\Lambda t)$$

where $\epsilon(0)$ is the initial difference between $Q_e^{(1)}$ and $Q_e^{(2)}$ at $t = 0$. The slope of $\ln \epsilon(t)$ versus t gives the estimate of the largest Lyapunov exponent. For all parameter configurations, both runs used identical noise inputs, meaning the slope of $\ln \epsilon(t)$ versus t provides the stochastic Lyapunov exponent. Distance from edge-of-chaos criticality was then quantified as the absolute value of the stochastic largest Lyapunov exponent.

4.14 Calculating spectral information transfer

Following our recent paper [19], we used the spectral information transfer measure developed by Pinzuti et al. [63], based on transfer entropy, which quantifies information transferred from a source variable X to a target variable Y [64–66]:

$$T_{X\to Y}^{(k,\ell)}(t,u) = I\left(Y_t; X_{t-u}^{(\ell)} | Y_{t-1}^{(k)}\right) \tag{7}$$

Here, Y_t is the state of Y at time t, $X_{t-u}^{(\ell)}$ represents the past ℓ states of X with interaction delay u, and $Y_{t-1}^{(k)}$ represents the past k states of Y. This measures how much uncertainty about Y_t is reduced by knowing $X_{t-u}^{(\ell)}$, given $Y_{t-1}^{(k)}$. We used the Java Information Dynamics Toolkit (JIDT) [67] and the method of Kraskov et al. [68] for probability distribution estimation using nearest-neighbor counting [69]. A fixed $K = 4$ nearest neighbors and $k = \ell = 1$ were used in all datasets, as in our prior work [19]. We scanned interaction delays u from 0.002 ms to 40 ms and selected the value that maximized transfer entropy [65, 70].

Pinzuti et al.'s innovation allows estimation of information transfer in specific frequency bands using the maximum overlap discrete wavelet transform. Surrogate data are created by randomizing dynamics in specific frequency ranges (via Iterative Amplitude Adjustment Fourier Transform), enabling the estimation of spectral transfer entropy strength and statistical significance. The 'swap-out swap-out' (SOSO) algorithm, also from Pinzuti et al., identifies specific frequency bands for sending and receiving information between channels. We used SOSO for all spectral analyses and resampled data to 416 Hz for better alignment with canonical neural oscillations. Each analysis used 100 surrogates to ensure statistical significance and reliable spectral transfer entropy estimates.

Acknowledgements. This work was supported by NIH grant R01 GM135420. We also thank Diego Mateos for providing MEG data from a human epilepsy patient with absence seizures.

References

[1] Schnakers, C., Monti, M.M.: Disorders of consciousness after severe brain injury: therapeutic options. Current Opinion in Neurology **30**(6), 573–579 (2017)

[2] Thibaut, A., Schiff, N., Giacino, J., Laureys, S., Gosseries, O.: Therapeutic interventions in patients with prolonged disorders of consciousness. The Lancet Neurology **18**(6), 600–614 (2019)

[3] Tsubokawa, T., Yamamoto, T., Katayama, Y., Hirayama, T., Maejima, S., Moriya, T.: Deep-brain stimulation in a Persistent Vegetative State: follow-up results and criteria for selection of candidates. Brain Injury **4**(4), 315–327 (1990)

[4] Chudy, D., Deletis, V., Almahariq, F., Marčinković, P., Škrlin, J., Paradžik, V.: Deep brain stimulation for the early treatment of the minimally conscious state

and vegetative state: experience in 14 patients. Journal of Neurosurgery **128**(4), 1189–1198 (2018)

[5] Lemaire, J.-J., Sontheimer, A., Pereira, B., Coste, J., Rosenberg, S., Sarret, C., Coll, G., Gabrillargues, J., Jean, B., Gillart, T., *et al.*: Deep brain stimulation in five patients with severe disorders of consciousness. Annals of Clinical and Translational Neurology **5**(11), 1372–1384 (2018)

[6] Magrassi, L., Maggioni, G., Pistarini, C., Di Perri, C., Bastianello, S., Zippo, A.G., Iotti, G.A., Biella, G.E., Imberti, R.: Results of a prospective study (CATS) on the effects of thalamic stimulation in minimally conscious and vegetative state patients. Journal of Neurosurgery **125**(4), 972–981 (2016)

[7] Hassler, R., Dalle Ore, G., Dieckmann, G., Bricolo, A., Dolce, G.: Behavioural and EEG arousal induced by stimulation of unspecific projection systems in a patient with post-traumatic apallic syndrome. Electroencephalography and Clinical Neurophysiology **27**(3), 306–310 (1969)

[8] Schiff, N.D., Giacino, J.T., Kalmar, K., Victor, J.D., Baker, K., Gerber, M., Fritz, B., Eisenberg, B., O'connor, J., Kobylarz, E., *et al.*: Behavioural improvements with thalamic stimulation after severe traumatic brain injury. Nature **448**(7153), 600–603 (2007)

[9] Schiff, N.D.: Recovery of consciousness after brain injury: a mesocircuit hypothesis. Trends in Neurosciences **33**(1), 1–9 (2010)

[10] Zheng, Z.S., Reggente, N., Monti, M.M.: Arousal Regulation by the External Globus Pallidus: A new node for the Mesocircuit Hypothesis. Brain Sciences **13**(1), 146 (2023)

[11] Thum, J.A., Malekmohammadi, M., Toker, D., Sparks, H., Alijanpourotaghsara, A., Choi, J.W., Hudson, A.E., Monti, M.M., Pouratian, N.: Globus Pallidus Externus drives increase in network-wide alpha power with propofol-induced loss-of-consciousness in humans. Cerebral Cortex **34**(6) (2024)

[12] Zheng, Z.S., Monti, M.M.: Cortical and thalamic connections of the human Globus Pallidus: implications for disorders of consciousness. Frontiers in Neuroanatomy **16**, 960439 (2022)

[13] Lutkenhoff, E.S., Johnson, M.A., Casarotto, S., Massimini, M., Monti, M.M.: Subcortical atrophy correlates with the Perturbational Complexity Index in patients with disorders of consciousness. Brain Stimulation **13**(5), 1426–1435 (2020)

[14] Vetrivelan, R., Qiu, M.-H., Chang, C., Lu, J.: Role of Basal Ganglia in Sleep–Wake regulation: neural circuitry and clinical significance. Frontiers in Neuroanatomy **4**, 145 (2010)

[15] O'Donnell, J.C., Browne, K.D., Kilbaugh, T.J., Chen, H.I., Whyte, J., Cullen, D.K.: Challenges and demand for modeling disorders of consciousness following traumatic brain injury. Neuroscience & Biobehavioral Reviews **98**, 336–346 (2019)

[16] Sanz-Leon, P., Robinson, P.A., Knock, S.A., Drysdale, P.M., Abeysuriya, R.G., Fung, F.K., Rennie, C.J., Zhao, X.: NFTsim: theory and simulation of multiscale neural field dynamics. PLoS Computational Biology **14**(8), 1006387 (2018)

[17] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Networks. Communications of the ACM **63**(11), 139–144 (2020)

[18] Toker, D., Pappas, I., Lendner, J.D., Frohlich, J., Mateos, D.M., Muthukumaraswamy, S., Carhart-Harris, R., Paff, M., Vespa, P.M., Monti, M.M., *et al.*: Consciousness is supported by near-critical slow cortical electrodynamics. Proceedings of the National Academy of Sciences **119**(7), 2024455119 (2022)

[19] Toker, D., Müller, E., Miyamoto, H., Riga, M.S., Lladó-Pelfort, L., Yamakawa, K., Artigas, F., Shine, J.M., Hudson, A.E., Pouratian, N., *et al.*: Criticality supports cross-frequency cortical-thalamic information transfer during conscious states. Elife **13**, 86547 (2024)

[20] O'Byrne, J., Jerbi, K.: How critical is brain criticality? Trends in Neurosciences **45**(11), 820–837 (2022)

[21] Maschke, C., O'Byrne, J., Colombo, M.A., Boly, M., Gosseries, O., Laureys, S., Rosanova, M., Jerbi, K., Blain-Moraes, S.: Critical dynamics in spontaneous EEG predict anesthetic-induced loss of consciousness and perturbational complexity. Communications Biology **7**(1), 946 (2024)

[22] Lozano, A.M., Lipsman, N., Bergman, H., Brown, P., Chabardes, S., Chang, J.W., Matthews, K., McIntyre, C.C., Schlaepfer, T.E., Schulder, M., *et al.*: Deep brain stimulation: current challenges and future directions. Nature Reviews Neurology **15**(3), 148–160 (2019)

[23] Mecarelli, O., Brienza, M., Grippo, A., Amantini, A.: Disorders of consciousness. Clinical Electroencephalography, 731–765 (2019)

[24] Stender, J., Kupers, R., Rodell, A., Thibaut, A., Chatelle, C., Bruno, M.-A., Gejl, M., Bernard, C., Hustinx, R., Laureys, S., *et al.*: Quantitative rates of brain glucose metabolism distinguish minimally conscious from vegetative state patients. Journal of Cerebral Blood Flow & Metabolism **35**(1), 58–65 (2015)

[25] Nakayama, N., Okumura, A., Shinoda, J., Nakashima, T., Iwama, T.: Relationship between regional cerebral metabolism and consciousness disturbance in traumatic diffuse brain injury without large focal lesions: an FDG-PET study

with statistical parametric mapping analysis. Journal of Neurology, Neurosurgery & Psychiatry **77**(7), 856–862 (2006)

[26] Stender, J., Gosseries, O., Bruno, M.-A., Charland-Verville, V., Vanhaudenhuyse, A., Demertzi, A., Chatelle, C., Thonnard, M., Thibaut, A., Heine, L., *et al.*: Diagnostic precision of PET imaging and functional MRI in disorders of consciousness: a clinical validation study. The Lancet **384**(9942), 514–522 (2014)

[27] Fridman, E.A., Beattie, B.J., Broft, A., Laureys, S., Schiff, N.D.: Regional cerebral metabolic patterns demonstrate the role of anterior forebrain mesocircuit dysfunction in the severely injured brain. Proceedings of the National Academy of Sciences **111**(17), 6473–6478 (2014)

[28] Teasdale, G., Jennett, B.: Assessment of coma and impaired consciousness: a practical scale. The Lancet **304**(7872), 81–84 (1974)

[29] Giacino, J.T., Kalmar, K., Whyte, J.: The JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility. Archives of Physical Medicine and Rehabilitation **85**(12), 2020–2029 (2004)

[30] Albada, S.J., Robinson, P.A.: Mean-field modeling of the basal ganglia-thalamocortical system. I: Firing Rates in healthy and Parkinsonian states. Journal of Theoretical Biology **257**(4), 642–663 (2009)

[31] Frohlich, J., Toker, D., Monti, M.M.: Consciousness among delta waves: a paradox? Brain **144**(8), 2257–2277 (2021)

[32] Altıntop, Ç.G., Latifoğlu, F., Akın, A.K., Bayram, A., Çiftçi, M.: Classification of depth of coma using complexity measures and nonlinear features of electroencephalogram signals. International Journal of Neural Systems **32**(05), 2250018 (2022)

[33] Liu, Y., Zeng, W., Pan, N., Xia, X., Huang, Y., He, J.: EEG complexity correlates with residual consciousness level of disorders of consciousness. BMC Neurology **23**(1), 140 (2023)

[34] Zhang, X.-S., Roy, R.J., Jensen, E.W.: EEG complexity as a measure of depth of anesthesia for patients. IEEE Transactions on Biomedical Engineering **48**(12), 1424–1433 (2001)

[35] Müller, E.J., Robinson, P.A.: Quantitative theory of deep brain stimulation of the Subthalamic Nucleus for the suppression of pathological rhythms in Parkinson's Disease. PLoS Computational Biology **14**(5), 1006217 (2018)

[36] Ramlackhansingh, A.F., Brooks, D.J., Greenwood, R.J., Bose, S.K., Turkheimer, F.E., Kinnunen, K.M., Gentleman, S., Heckemann, R.A., Gunanayagam, K., Gelosa, G., *et al.*: Inflammation after trauma: microglial activation and traumatic

brain injury. Annals of Neurology **70**(3), 374–383 (2011)

[37] Cao, Z., Harvey, S.S., Bliss, T.M., Cheng, M.Y., Steinberg, G.K.: Inflammatory responses in the secondary thalamic injury after cortical ischemic stroke. Frontiers in Neurology **11**, 236 (2020)

[38] Bianciardi, M., Izzy, S., Rosen, B.R., Wald, L.L., Edlow, B.L.: Location of subcortical microbleeds and recovery of consciousness after severe traumatic brain injury. Neurology **97**(2), 113–123 (2021)

[39] Lutkenhoff, E.S., Chiang, J., Tshibanda, L., Kamau, E., Kirsch, M., Pickard, J.D., Laureys, S., Owen, A.M., Monti, M.M.: Thalamic and extrathalamic mechanisms of consciousness after severe brain injury. Annals of Neurology **78**(1), 68–76 (2015)

[40] Ozen, I., Ruscher, K., Nilsson, R., Flygt, J., Clausen, F., Marklund, N.: Interleukin-1 beta neutralization attenuates traumatic brain injury-induced microglia activation and neuronal changes in the Globus Pallidus. International Journal of Molecular Sciences **21**(2), 387 (2020)

[41] Lewis, L.D., Voigts, J., Flores, F.J., Schmitt, L.I., Wilson, M.A., Halassa, M.M., Brown, E.N.: Thalamic reticular nucleus induces fast and local modulation of arousal state. eLife **4**, 08760 (2015)

[42] Yi, R., Chen, S., Zhong, F., Luo, D., You, Y., Yu, T., Wang, H., Zhou, L., Zhang, Y.: GABAergic neurons of anterior thalamic reticular nucleus regulate states of consciousness in propofol-and isoflurane-mediated general anesthesia. CNS Neuroscience & Therapeutics **30**(6), 14782 (2024)

[43] Paz, J.T., Christian, C.A., Parada, I., Prince, D.A., Huguenard, J.R.: Focal cortical infarcts alter intrinsic excitability and synaptic excitation in the reticular thalamic nucleus. Journal of Neuroscience **30**(15), 5465–5479 (2010)

[44] Ross, D.T., Graham, D.I., Adams, J.H.: Selective loss of neurons from the thalamic reticular nucleus following severe human head injury. Journal of Neurotrauma **10**(2), 151–165 (1993)

[45] Kitamura, A., Marszalec, W., Yeh, J.Z., Narahashi, T.: Effects of halothane and propofol on excitatory and inhibitory synaptic transmission in rat cortical neurons. Journal of Pharmacology and Experimental Therapeutics **304**(1), 162–171 (2003)

[46] Bai, D., Pennefather, P.S., MacDonald, J.F., Orser, B.A.: The general anesthetic propofol slows deactivation and desensitization of GABAA receptors. Journal of Neuroscience **19**(24), 10635–10646 (1999)

[47] Nichols, J., Bjorklund, G.R., Newbern, J., Anderson, T.: Parvalbumin fast-spiking

interneurons are selectively altered by paediatric traumatic brain injury. The Journal of Physiology **596**(7), 1277–1293 (2018)

[48] Hiltunen, J., Ndode-Ekane, X.E., Lipponen, A., Drexel, M., Sperk, G., Puhakka, N., Pitkänen, A.: Regulation of parvalbumin interactome in the perilesional cortex after experimental traumatic brain injury. Neuroscience **475**, 52–72 (2021)

[49] Bagnato, S., Boccagni, C., Sant'Angelo, A., Prestandrea, C., Rizzo, S., Galardi, G.: Patients in a vegetative state following traumatic brain injury display a reduced intracortical modulation. Clinical Neurophysiology **123**(10), 1937–1941 (2012)

[50] Qin, P., Wu, X., Duncan, N.W., Bao, W., Tang, W., Zhang, Z., Hu, J., Jin, Y., Wu, X., Gao, L., *et al.*: GABAA receptor deficits predict recovery in patients with disorders of consciousness: A Preliminary Multimodal [11C] Flumazenil PET and f MRI Study. Human Brain Mapping **36**(10), 3867–3877 (2015)

[51] Rudolf, J., Sobesky, J., Grond, M., Heiss, W.: Identification by positron emission tomography of neuronal loss in acute vegetative state. The Lancet **355**(9198), 115–116 (2000)

[52] Krämer, S.D., Schuhmann, M.K., Volkmann, J., Fluri, F.: Deep brain stimulation in the Subthalamic Nucleus can improve skilled forelimb movements and retune dynamics of striatal networks in a rat stroke model. International Journal of Molecular Sciences **23**(24), 15862 (2022)

[53] Ray, S., Crone, N.E., Niebur, E., Franaszczuk, P.J., Hsiao, S.S.: Neural correlates of high-gamma oscillations (60–200 hz) in macaque local field potentials and their potential implications in electrocorticography. Journal of Neuroscience **28**(45), 11526–11536 (2008)

[54] Frohlich, J., Crone, J.S., Johnson, M.A., Lutkenhoff, E.S., Spivak, N.M., Dell'Italia, J., Hipp, J.F., Shrestha, V., Ruiz Tejeda, J.E., Real, C., *et al.*: Neural oscillations track recovery of consciousness in acute traumatic brain injury patients. Human Brain Mapping **43**(6), 1804–1820 (2022)

[55] Guang, J., Baker, H., Ben-Yishay Nizri, O., Firman, S., Werner-Reiss, U., Kapuller, V., Israel, Z., Bergman, H.: Toward asleep DBS: cortico-basal ganglia spectral and coherence activity during interleaved propofol/ketamine sedation mimics NREM/REM sleep activity. npj Parkinson's Disease **7**(1), 67 (2021)

[56] Torkamani-Azar, M., Kanik, S.D., Aydin, S., Cetin, M.: Prediction of reaction time and vigilance variability from spatio-spectral features of resting-state EEG in a long sustained attention task. IEEE Journal of Biomedical and Health Informatics **24**(9), 2550–2558 (2020)

[57] Malekmohammadi, M., Price, C.M., Hudson, A.E., DiCesare, J.A., Pouratian,

N.: Propofol-induced loss of consciousness is associated with a decrease in thalamocortical connectivity in humans. Brain **142**(8), 2288–2302 (2019)

[58] Reed, S.J., Plourde, G.: Attenuation of high-frequency (50-200 Hz) thalamocortical EEG rhythms by propofol in rats is more pronounced for the thalamus than for the cortex. PLoS ONE **10**(4), 0123287 (2015)

[59] Miyamoto, H., Tatsukawa, T., Shimohata, A., Yamagata, T., Suzuki, T., Amano, K., Mazaki, E., Raveau, M., Ogiwara, I., Oba-Asaka, A., *et al.*: Impaired corticostriatal excitatory transmission triggers epilepsy. Nature Communications **10**(1), 1–13 (2019)

[60] Ihle, M., Feldwisch-Drentrup, H., Teixeira, C.A., Witon, A., Schelter, B., Timmer, J., Schulze-Bonhage, A.: EPILEPSIAE–a European Epilepsy Database. Computer Methods and Programs in Biomedicine **106**(3), 127–138 (2012)

[61] Dominguez, L.G., Wennberg, R.A., Gaetz, W., Cheyne, D., Snead, O.C., Velazquez, J.L.P.: Enhanced synchrony in epileptiform activity? local versus distant phase synchronization in generalized seizures. Journal of Neuroscience **25**(35), 8077–8084 (2005)

[62] Yanagawa, T., Chao, Z.C., Hasegawa, N., Fujii, N.: Large-scale information flow in conscious and unconscious states: an ECoG study in monkeys. PLoS ONE **8**(11), 80845 (2013)

[63] Pinzuti, E., Wollstadt, P., Gutknecht, A., Tüscher, O., Wibral, M.: Measuring spectrally-resolved information transfer. PLoS Computational Biology **16**(12), 1008526 (2020)

[64] Bossomaier, T., Barnett, L., Harré, M., Lizier, J.T., Bossomaier, T., Barnett, L., Harré, M., Lizier, J.T.: An Introduction to Transfer Entropy: Information Flow in Complex Systems. Springer, Switzerland (2016)

[65] Wibral, M., Pampu, N., Priesemann, V., Siebenhühner, F., Seiwert, H., Lindner, M., Lizier, J.T., Vicente, R.: Measuring information-transfer delays. PloS ONE **8**(2), 55809 (2013)

[66] Schreiber, T.: Measuring information transfer. Physical Review Letters **85**(2), 461 (2000)

[67] Lizier, J.T.: JIDT: an information-theoretic toolkit for studying the dynamics of complex systems. Frontiers in Robotics and AI **1**, 11 (2014)

[68] Kraskov, A., Stögbauer, H., Grassberger, P.: Estimating mutual information. Physical Review E **69**(6), 066138 (2004)

[69] Kozachenko, L., Leonenko, N.N.: Sample estimate of the entropy of a random

vector. Problemy Peredachi Informatsii **23**(2), 9–16 (1987)

[70] Wollstadt, P., Sellers, K.K., Rudelt, L., Priesemann, V., Hutt, A., Fröhlich, F., Wibral, M.: Breakdown of local information processing may underlie isoflurane anesthesia effects. PLoS Computational Biology **13**(6), 1005511 (2017)