nature neuroscience

Supplementary information

https://doi.org/10.1038/s41593-026-02220-4

Adversarial AI reveals mechanisms and treatments for disorders of consciousness

In the format provided by the authors and unedited

Supplementary Table 1: **Breakdown of training, validation, and test datasets used for training and evaluating our AI-based consciousness detectors.** Some subjects/animals contributed both cortical and subcortical recordings. African green monkeys contributed both training and validation data for the waking state, and independent test data for (induced) coma. ET=essential tremor, PD=Parkinson's disease, GPe=external globus pallidus, DOC=disorders of consciousness.

Dataset	Subjects	Samples (Train/ Val/ Test)		
Subcortical Detectors Data (Training and Validation)				
Human ET Patient Thalamus	10	470	118	–
Human PD Patient GPe	12	95	24	–
Monkey GPe	13	13,107	3,277	–
Long-Evans Rat Thalamus	9	200	50	–
Cortical Detector Data (Training and Validation)				
Human Acute Coma (TBI)	22	602,696	150,674	–
Human Chronic DOC	14	49,062	12,266	–
Human Healthy Control (dataset 1)	10	7,066	1,766	–
Human ET Patients	10	354	–	–
Human PD Patients	12	51	–	–
African Green Monkey (waking)	13	7,656	1,914	–
Cortical Detector Data (Independent Test)				
Human Acute Coma (cardiac arrest)	455	–	–	1,180,257
Long-Evans Rat (waking and coma)	9	–	–	250
African Green Monkey (induced coma)	13	–	–	9,570
Human Healthy Control (dataset 2)	20	–	–	3,674
Human Healthy Control (dataset 3)	169	–	–	5,408
Locked-in Syndrome Patients	6	–	–	1,601

Supplementary Table 2: **Firing-rate-related parameters for each population in our mean-field model of the conscious brain.** Parameters for the sigmoidal mean firing-rate response function include the membrane potential θ (mV) at which the population-average firing rate reaches half its maximum, the standard deviation σ of the soma voltage relative to that threshold, and the maximum possible firing rate $Q_{\max}$ for that population.

Neural population	Firing threshold θ (mV)	Voltage spread σ (mV)	$Q_{\max}$ (spikes/s)
Excitatory cortical	6.03	2.9893	22.8294
Inhibitory cortical	5.7626	6.1769	99.6704
TRN	0.59576	4.7428	385.3459
Thalamic projection	1.1724	9.6263	105.0741
D1	1.1907	2.1266	19.3902
D2	1.3843	4.5537	10.4114
GPi	3.53	22.244	132.0234
GPe	9.0081	2.2973	316.0468
STN	6.3887	3.406	202.4802

Supplementary Table 3: **Dendritic parameters for each connection in our mean-field model of the conscious brain.** The parameters α and β are the decay and rise rates in $1/\mathrm{s}$, while h_{syn} is a synaptic scaling factor. Together, these parameters control the height and duration of postsynaptic potentials.

Connection	α	β	h_{syn}
Excitatory Cortical → Excitatory Cortical	2.674	6.9705	19.555
Inhibitory Cortical → Excitatory Cortical	76.8488	91.4069	22.7117
Thalamic Projection → Excitatory Cortical	308.4148	309.811	14.0682
Excitatory Cortical → Inhibitory Cortical	23.8422	353.0609	45.2787
Inhibitory Cortical → Inhibitory Cortical	66.6434	185.5455	22.3397
Thalamic Projection → Inhibitory Cortical	53.1076	52.2119	5.4498
GPe → Inhibitory Cortical	49.1921	153.4782	44.3374
Excitatory Cortical → TRN	97.3268	258.7483	54.2651
Thalamic Projection → TRN	35.5804	1135.0763	15.9379
GPe → TRN	370.4681	420.044	11.7307
Excitatory Cortical → Thalamic Projection	82.407	185.5051	71.3342
TRN → Thalamic Projection	31.1955	217.8414	18.6017
GPi → Thalamic Projection	57.5824	89.063	25.6852
GPe → Thalamic Projection	136.9442	112.7362	8.971
Excitatory Cortical → D1	393.5826	207.6264	32.657
Thalamic Projection → D1	161.3917	64.0069	21.7262
D1 → D1	118.2546	725.1904	20.9243
GPe → D1	44.3807	134.2933	34.7326
Excitatory Cortical → D2	66.4032	1787.148	82.5032
Thalamic Projection → D2	81.9949	296.46	22.4305
D2 → D2	30.8653	551.4931	14.9752
GPe → D2	57.6319	359.2884	22.9981
D1 → GPi	2.544	341.8611	8.8214
GPe → GPi	53.9211	455.9377	19.0529
STN → GPi	95.4126	147.5733	14.2682
Excitatory Cortical → GPe	19.4579	322.3915	81.6792
D2 → GPe	4.8914	2382.7459	35.0419
GPe → GPe	1.5206	273.0801	44.268
STN → GPe	490.3135	1953.3756	22.4438
Excitatory Cortical → STN	33.1858	527.6786	16.9322
GPe → STN	189.6946	375.522	9.1586

Supplementary Table 4: **Propagator parameters for neural connections with coupling strengths.** Propagators can be *wave* or *map*. Wave-type propagators include a time-delay τ, axonal spatial range r, and a damping coefficient γ. The parameter ν represents the axonal-synaptic coupling strength.

Connection	**Type**	τ	r	γ	ν
Excitatory Cortical → Excitatory Cortical	Wave	0.0037273	0.035009	52.7328	0.0001431
Inhibitory Cortical → Excitatory Cortical	Map	0.033308	—	—	-0.0026139
Thalamic Projection → Excitatory Cortical	Map	0.042917	—	—	0.0024834
Excitatory Cortical → Inhibitory Cortical	Wave	4.8695e-5	0.084967	226.9079	0.0014598
Inhibitory Cortical → Inhibitory Cortical	Map	0.02991	—	—	-0.0011039
Thalamic Projection → Inhibitory Cortical	Map	0.10776	—	—	0.00010874
GPe → Inhibitory Cortical	Map	0.016383	—	—	-1.29e-05
Excitatory Cortical → TRN	Wave	0.01442	0.25159	29.0848	0.00027184
Thalamic Projection → TRN	Map	0.0044841	—	—	1.742e-05
GPe → TRN	Map	0.029018	—	—	-0.00037503
Excitatory Cortical → Thalamic Projection	Wave	0.0063266	0.40024	55.2981	4.9136e-06
TRN → Thalamic Projection	Map	0.0030565	—	—	-0.00069802
GPi → Thalamic Projection	Map	0.00072682	—	—	-3.2149e-05
GPe → Thalamic Projection	Map	0.018695	—	—	-0.00011238
Stimulus → Thalamic Projection	Map	0	—	—	0.0010994
Excitatory Cortical → D1	Wave	0.00010577	0.025587	12.1336	0.00049354
Thalamic Projection → D1	Map	0.003893	—	—	5.8137e-05
D1 → D1	Map	0.0029817	—	—	-0.00022084
GPe → D1	Map	0.34971	—	—	-6.4251e-05
Excitatory Cortical → D2	Wave	0.0024002	0.059943	143.3842	0.00057826
Thalamic Projection → D2	Map	0.00049981	—	—	2.1702e-06
D2 → D2	Map	0.014893	—	—	-0.00012266
GPe → D2	Map	0.0099476	—	—	-0.00018428
D1 → GPi	Map	0.00056918	—	—	-3.0066e-05
GPe → GPi	Map	0.00072232	—	—	-4.2334e-06
STN → GPi	Map	0.0017869	—	—	1.7415e-05
Excitatory Cortical → GPe	Wave	0.016993	0.041773	95.3548	5.7408e-05
D2 → GPe	Map	0.00014521	—	—	-0.00024693
GPe → GPe	Map	0.003942	—	—	-7.3175e-05
STN → GPe	Map	0.00061743	—	—	0.00036479
Excitatory Cortical → STN	Wave	0.0016426	0.38859	223.5338	0.00025665
GPe → STN	Map	0.00047147	—	—	-5.9005e-05

Supplementary Table 5: **Mean firing rates (in spikes/s) of each brain region in the mean-field simulation of the conscious brain, compared to empirical ranges of firing rates from multiple mammalian species.** The firing rates for each simulated brain region were optimized, along with the outputs of the consciousness-detector networks, through a genetic algorithm to align with known physiological ranges. The GPi and SNr were treated as a single population, sharing the same firing rate. GPi=internal globus pallidus, SNr=substantia nigra pars reticulata, GPe=external globus pallidus, STN=subthalamic nucleus, TRN=thalamic reticular nucleus. "✓" indicates that the model firing rate falls inside the empirical *union* range for that clade. The distance metric (*dist*) captures how far model firing rates fall outside the reported empirical ranges, normalized by the width of those ranges, and averaged across all regions with available data. A value of 0 means the model falls entirely within the reported ranges, while larger values reflect worse fit. The model matched all primate ranges with a distance of 0.0, but only partially matched rodent ranges, with a distance of 3.39, suggesting it more closely resembles primate physiology. However, data on firing rate ranges in different species are limited and sometimes inconsistent, so these ranges may not fully capture the underlying physiology; therefore, the conclusion that the model is closer to primates should be interpreted with caution.

	Simulation Brain	Monkeys	Rats	Mice	Humans	Cats	Primates (union)	Rodents (union)
Cortical Pyramidal Cells	6.99	5–20 [1,2]	2–5 [3]	2–11 [4]	1–10 [5]		✓	✓
Cortical Inhibitory Interneurons	18.21	0.1–100 [6]		1–50 [7]			✓	✓
Striatum	4.82	4–7 [1,8]	1–7 [3,9]				✓	✓
GPi	60.77	60–90 [10–13]	15–20 [3]				✓	×
SNr	60.77	50–70 [14,15]					✓	×
GPe	43.74	16–70 [10,11,13,16,17]	16–115 [18,19]	1–64 [20]			✓	✓
STN	22.16	20–30 [11,21]	8–11 [22]				✓	×
Projection Nuclei	14.10	5–25 [23]			10–20 [24]		✓	×
TRN	24.10			4–64 [25–28]		20–30 [29]	*n/a*	✓
							Summary	**dist = 0.0**	**dist = 3.39**

Supplementary Table 6: **Pairwise discrimination of subcortical LFPs via relative band-power and zero-crossing.** For each region-pair, we applied one-tailed unequal-variance t-tests to the z-scored real LFPs (from bats & GAERS rats for striatum; Long–Evans rats & ET patients for thalamus; African green monkeys & PD patients for GPe) and to the corresponding simulated LFPs (GPe/striatum/thalamus only, 600 simulated 10-second LFP samples per region). We tested whether region A $>$ region B on feature f if the real-data mean difference $\Delta_{\text{Real}} = \mu_A - \mu_B$ was positive (otherwise A$<$B). All p-values are Bonferroni-corrected over 12 tests (3 pairs × 4 features). Stars denote $p < 0.05$. These results show that, with the exception of delta power, the real-world features that distinguish LFPs from different subcortical regions are recapitulated in the model—even though we never explicitly programmed these inter-regional differences—paralleling our successful modeling of region-specific firing rates.

Pair	Feature	Δ_{Real}	Δ_{Sim}	$p^{\text{BF}}_{\text{Real}}$ *	$p^{\text{BF}}_{\text{Sim}}$ *
Striatum vs GPe	δ_{rel}	-0.042	0.062	$4.70 \times 10^{-4*}$	1.00
	θ_{rel}	-0.067	-0.062	$1.05 \times 10^{-15*}$	$3.69 \times 10^{-82*}$
	α_{rel}	-0.109	-0.0001	$1.00 \times 10^{-37*}$	1.00
	zcr	-0.011	-0.006	$6.24 \times 10^{-21*}$	$5.62 \times 10^{-205*}$
Striatum vs Thal	δ_{rel}	0.093	-0.012	$1.04 \times 10^{-17*}$	1.00
	θ_{rel}	0.054	0.033	$1.40 \times 10^{-11*}$	$7.10 \times 10^{-26*}$
	α_{rel}	-0.147	-0.021	$3.49 \times 10^{-59*}$	$2.46 \times 10^{-17*}$
	zcr	-0.013	-0.042	$5.61 \times 10^{-28*}$	0.00^{*}
GPe vs Thal	δ_{rel}	0.135	-0.075	$7.37 \times 10^{-229*}$	1.00
	θ_{rel}	0.121	0.095	$1.27 \times 10^{-262*}$	$1.27 \times 10^{-158*}$
	α_{rel}	-0.257	-0.021	0^{*}	$1.49 \times 10^{-16*}$
	zcr	-0.0016	-0.037	1.13×10^{-2}	0.00^{*}

Supplementary Table 7: **Additional model parameters significantly associated with AI-predicted level of consciousness in simulated DOC**. Model parameters were identified using ridge-penalized linear regression with inverse Gaussian transformation and permutation testing (1,000 permutations), followed by False Discovery Rate (FDR) correction. Each coefficient reflects the standardized association between a feature and predicted level of consciousness.

Additional Model Parameters Associated with AI-Predicted Level of Consciousness	**Coefficient**	*p-value (Permutation test, FDR-corrected)*
Striatum (D2-expressing) to Striatum (D2-expressing) Propagation Delay	-0.172	<0.0001
Thalamic Projection Nucleus to TRN Propagation Delay	-0.171	<0.0001
Inhibitory to Inhibitory Cortical PSP Height	-0.161	<0.0001
Excitatory Cortex to Striatum (D2-expressing) Propagation Delay	-0.153	<0.0001
Thalamic Projection Nucleus Cross-Neuron Variance of Subthreshold Voltages	+0.133	<0.0001
GPe to Striatum (D2-expressing) PSP Height	+0.131	<0.0001
Excitatory Cortical to Striatum (D2-expressing) PSP Height	-0.126	<0.0001
Excitatory Cortical Threshold Potential	+0.117	<0.0001
TRN Threshold Potential	+0.114	<0.0001
Excitatory to Inhibitory Cortical Propagation Damping	+0.105	<0.0001
Striatum (D1-expressing) Variance of Subthreshold Voltages	-0.105	0.0058
GPe to TRN PSP Height	+0.104	0.0100
Inhibitory to Excitatory Cortical PSP Height	+0.103	<0.0001
GPe to Striatum (D1-expressing) Propagation Delay	+0.102	0.0058
Thalamic Projection Nucleus Threshold Potential	-0.101	0.0180
GPe to GPi PSP Height	+0.100	0.0200
TRN to Thalamic Projection Nucleus PSP Duration	-0.096	<0.0001
STN to GPi PSP Duration	+0.094	0.0100
Excitatory Cortical to GPe Propagation Damping	-0.094	0.0058
GPe to Striatum (D1-expressing) PSP Height	-0.094	0.0230
Excitatory to Inhibitory Cortical Propagation Range	-0.093	0.0180
Excitatory Cortical to Striatum (D2-expressing) Cortical PSP Duration	+0.091	0.0200
Striatum (D2-expressing) to GPe Propagation Delay	+0.116	0.0100
STN to GPe PSP Height	-0.087	0.0180
GPe to STN PSP Duration	-0.084	0.0180
GPe to GPi Propagation Delay	-0.081	0.0200
GPe to Thalamic Projection Nucleus Propagation Delay	-0.080	0.0280
GPe to TRN Propagation Delay	+0.079	0.0460
Excitatory to Inhibitory Cortical Propagation Delay	-0.071	0.0490

Supplementary Table 8: **Mediation analysis of in silico axonal connections influencing AI-predicted level of consciousness via firing-rate mediators.** We asked whether each connection in our in silico model affects the cortical consciousness detector's prediction *indirectly*—that is, by first changing the firing rate of a given neural population (the mediator), which in turn relates to the predicted level of consciousness. For each connection, we computed how much its strength correlates with each neural population's firing rate (the a-path), and how much that firing rate correlates with the consciousness prediction when controlling for the connection strength (the b-path). The indirect effect is simply the product of these two correlations (a × b), with confidence intervals estimated by bootstrap resampling. We report this indirect effect for each connection–mediator pair. Positive values mean the two steps act in the same overall direction; negative values mean they act in opposite directions. We show 95% confidence intervals and two-sided bootstrap p-values from 1,000 resamples; only pairs with p<0.05 are listed. *Worked example (D2→GPe coupling).* In the model (and the brain), the GPe directly inhibits the TRN, thalamic projection nuclei, and inhibitory cortical interneurons. All else being equal, strengthening the inhibitory D2→GPe connection therefore suppresses GPe firing, which in turn disinhibits these downstream populations. The mediation analysis shows how this plays out in terms of predicted consciousness. First, the table shows a *negative* indirect effect via inhibitory cortical firing (-0.043). Stronger D2→GPe increases inhibitory cortical activity (consistent with disinhibition), and since higher inhibitory firing predicts lower consciousness, this pathway carries an anti-conscious effect. Second, there is a *positive* indirect effect via TRN firing ($+0.023$). Stronger D2→GPe decreases TRN activity overall (likely through indirect, nonlinear network effects), and because higher TRN firing is associated with lower consciousness, this yields a pro-conscious effect. Third, there is a *positive* indirect effect via thalamic projection firing ($+0.018$). Stronger D2→GPe reduces GPe activity and thereby increases thalamic activity, and because higher thalamic firing predicts higher consciousness, this also contributes a pro-conscious effect. Thus, D2→GPe coupling influences predicted consciousness through three routes: one anti-conscious pathway via inhibitory cortical interneurons, and two pro-conscious pathways via the TRN and the thalamic projection nucleus.

Parameter (synaptic connection)	**Mediator**	**Indirect effect**	**95% CI (lo)**	**95% CI (hi)**	p
Inhibitory Cortical to Inhibitory Cortical Coupling	Inhibitory cortical firing rate	-0.095	-0.133	-0.059	< 0.001
Inhibitory Cortical to Inhibitory Cortical Coupling	Cortical firing rate	-0.084	-0.119	-0.054	< 0.001
GPe to GPe Coupling	GPe firing rate	-0.072	-0.100	-0.046	< 0.001
Striatum (D2-expressing) to Striatum (D2-expressing) Coupling	Cortical firing rate	-0.058	-0.085	-0.033	< 0.001
TRN to Thalamic Projection Coupling	Cortical firing rate	$+0.058$	$+0.030$	$+0.088$	< 0.001
Brainstem/Sensory Afferent to Thalamic Projection Coupling	Cortical firing rate	-0.056	-0.087	-0.028	< 0.001
Brainstem/Sensory Afferent to Thalamic Projection Coupling	GPe firing rate	$+0.050$	$+0.028$	$+0.077$	< 0.001
Striatum (D2-expressing) to Striatum (D2-expressing) Coupling	Inhibitory cortical firing rate	$+0.050$	$+0.032$	$+0.071$	< 0.001

Continued on next page

Table 8 (continued).

Parameter (axonal connection)	Mediator	Indirect effect	95% CI (lo)	95% CI (hi)	p
Brainstem/Sensory Afferent to Thalamic Projection Coupling	TRN firing rate	-0.049	-0.069	-0.030	< 0.001
Excitatory to Inhibitory Cortical Coupling	Inhibitory cortical firing rate	-0.047	-0.068	-0.030	< 0.001
GPe to Inhibitory Cortical Coupling	Cortical firing rate	$+0.047$	$+0.020$	$+0.077$	< 0.001
TRN to Thalamic Projection Coupling	TRN firing rate	-0.043	-0.065	-0.024	< 0.001
GPe to GPe Coupling	TRN firing rate	$+0.043$	$+0.026$	$+0.064$	< 0.001
Striatum (D2-expressing) to GPe Coupling	Inhibitory cortical firing rate	-0.043	-0.061	-0.026	< 0.001
Excitatory Cortex to STN Coupling	GPe firing rate	-0.036	-0.058	-0.019	< 0.001
GPe to TRN Coupling	GPe firing rate	-0.036	-0.064	-0.015	0.002
GPe to GPe Coupling	Cortical firing rate	$+0.032$	$+0.007$	$+0.059$	0.004
Striatum (D1-expressing) to GPi Coupling	Cortical firing rate	-0.032	-0.060	-0.008	0.004
TRN to Thalamic Projection Coupling	Inhibitory cortical firing rate	-0.031	-0.048	-0.017	< 0.001
GPe to TRN Coupling	TRN firing rate	-0.027	-0.045	-0.013	< 0.001
Thalamic Projection to Inhibitory Cortical Coupling	Inhibitory cortical firing rate	-0.026	-0.043	-0.012	< 0.001
GPe to Inhibitory Cortical Coupling	GPe firing rate	-0.025	-0.044	-0.008	0.002
Striatum (D1-expressing) to GPi Coupling	GPe firing rate	$+0.025$	$+0.006$	$+0.047$	0.002
TRN to Thalamic Projection Coupling	GPe firing rate	$+0.024$	$+0.007$	$+0.046$	0.002
Excitatory Cortex to Striatum (D2-expressing) Coupling	Inhibitory cortical firing rate	-0.024	-0.039	-0.012	< 0.001
Striatum (D2-expressing) to GPe Coupling	TRN firing rate	$+0.023$	$+0.010$	$+0.038$	< 0.001
GPe to GPe Coupling	Inhibitory cortical firing rate	-0.021	-0.039	-0.007	< 0.001
Striatum (D2-expressing) to Striatum (D2-expressing) Coupling	Thalamic projection firing rate	-0.020	-0.036	-0.008	< 0.001
Thalamic Projection to TRN Coupling	Inhibitory cortical firing rate	-0.020	-0.035	-0.006	0.004
Excitatory Cortex to Striatum (D1-expressing) Coupling	TRN firing rate	-0.019	-0.033	-0.009	0.002
Inhibitory Cortical to Inhibitory Cortical Coupling	Thalamic projection firing rate	$+0.019$	$+0.007$	$+0.039$	< 0.001
Striatum (D2-expressing) to GPe Coupling	Thalamic projection firing rate	$+0.018$	$+0.006$	$+0.034$	< 0.001

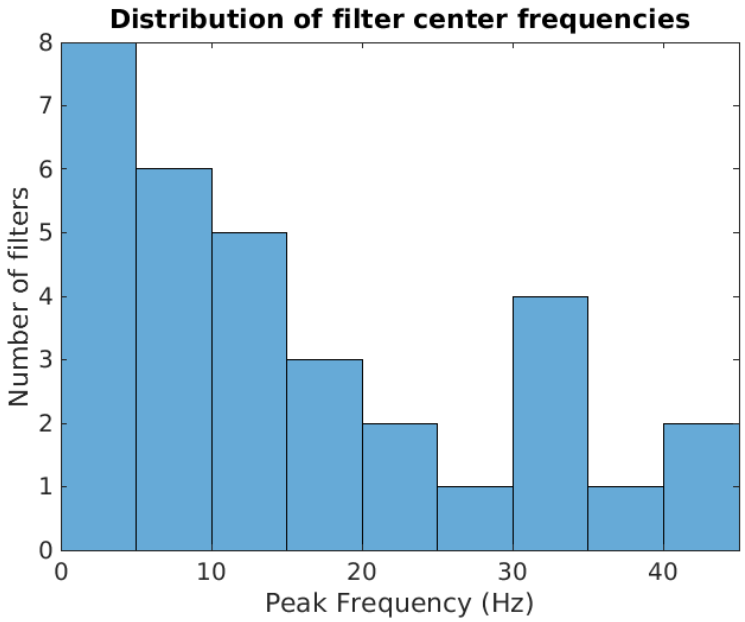
Continued on next page

Table 8 (continued).

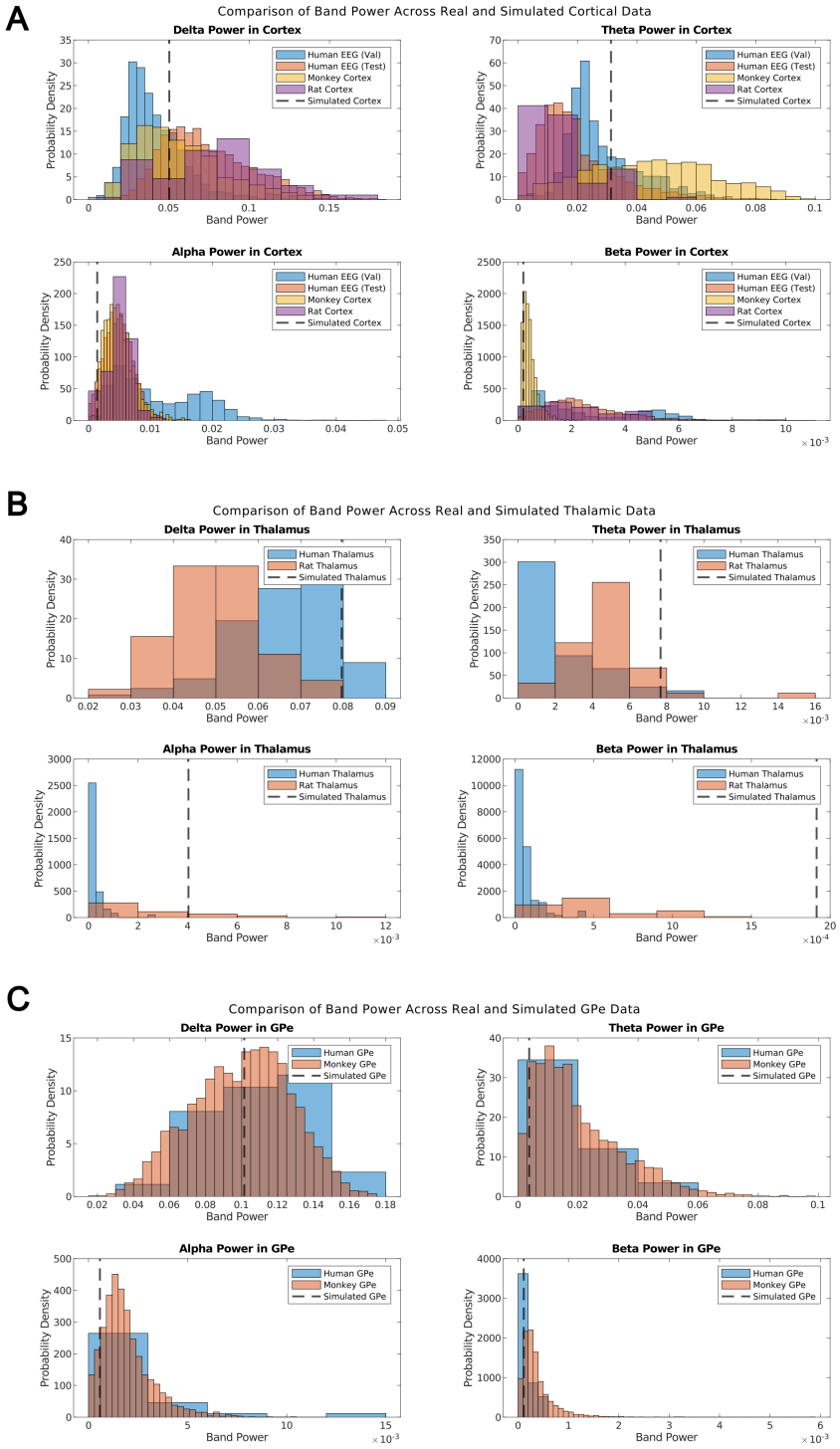
Parameter (axonal connection)	**Mediator**	**Indirect effect**	**95% CI (lo)**	**95% CI (hi)**	p
GPe to Inhibitory Cortical Coupling	Inhibitory cortical firing rate	-0.018	-0.032	-0.006	0.004
GPe to TRN Coupling	Inhibitory cortical firing rate	$+0.017$	$+0.005$	$+0.032$	0.004
GPe to Inhibitory Cortical Coupling	Thalamic projection firing rate	$+0.017$	$+0.006$	$+0.030$	< 0.001
Excitatory Cortex to TRN Coupling	TRN firing rate	-0.017	-0.029	-0.007	< 0.001
Thalamic Projection to Inhibitory Cortical Coupling	TRN firing rate	$+0.015$	$+0.003$	$+0.028$	0.016
Excitatory Cortex to TRN Coupling	Thalamic projection firing rate	-0.013	-0.025	-0.004	0.002
Excitatory to Inhibitory Cortical Coupling	Thalamic projection firing rate	$+0.012$	$+0.003$	$+0.025$	0.002
Striatum (D1-expressing) to Striatum (D1-expressing) Coupling	TRN firing rate	$+0.012$	$+0.002$	$+0.024$	0.012
GPe to TRN Coupling	Thalamic projection firing rate	-0.012	-0.023	-0.004	< 0.001
Thalamic Projection to TRN Coupling	TRN firing rate	-0.011	-0.023	< 0.001	0.046
Thalamic Projection to Inhibitory Cortical Coupling	Thalamic projection firing rate	$+0.010$	$+0.002$	$+0.023$	0.012
Brainstem/Sensory Afferent to Thalamic Projection Coupling	Thalamic projection firing rate	-0.008	-0.021	< 0.001	0.048

Supplementary Table 9: **Partitioning of explained variance (total $R^2 = 0.205$) for left striatum–GPe streamline counts.** Each predictor's value is the percentage of the total explained variance uniquely attributable to that factor in a linear model including diagnosis (VS vs. MCS/eMCS), age at MRI, gender, days post-injury, and injury etiology.

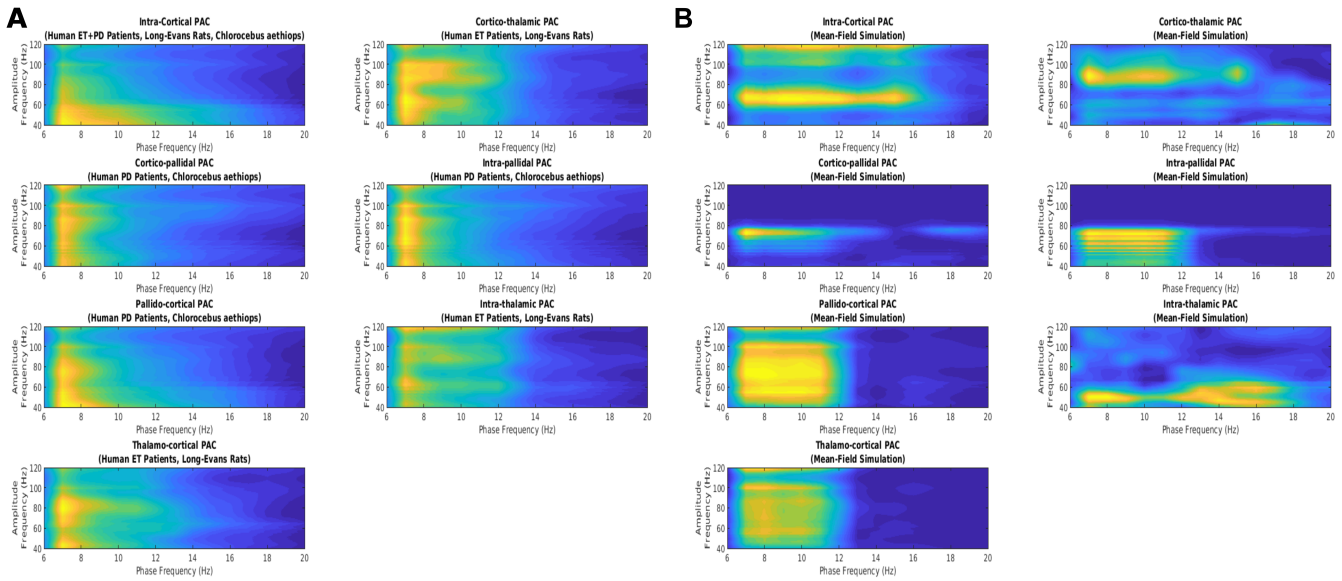
Predictor	$\%$ **of Total** R^2
Diagnosis (VS vs. MCS/eMCS)	35.0
Age at MRI	8.9
Gender	12.6
Days post-injury	5.8
Etiology	37.8
Total R^2	0.205



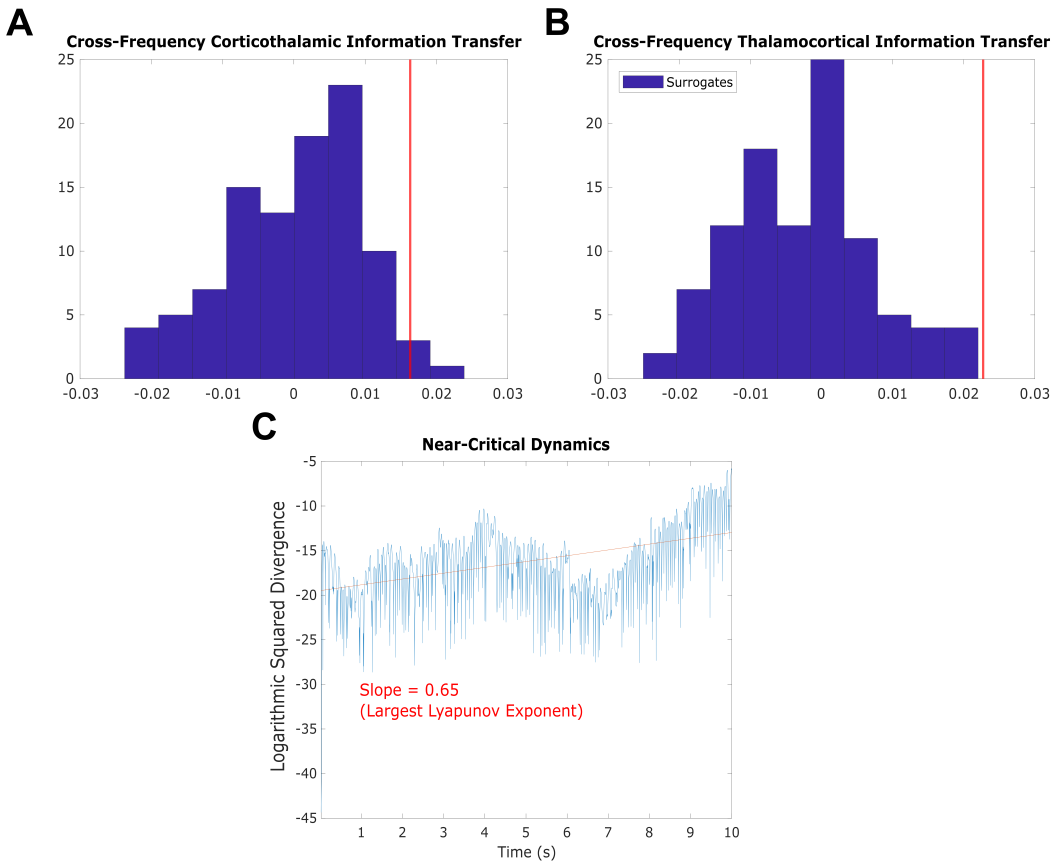
Supplementary Figure 1: **Distribution of learned temporal-kernel tuning in the cortical consciousness detector DCNN.** The histogram shows the peak-frequency tuning of the 32 one-dimensional kernels in the first convolutional layer of our deep convolutional neural network, trained to detect consciousness from cortical electrophysiology recordings. For each kernel, we computed its Fourier-domain gain (0.5–45 Hz) and identified the single frequency at which that gain was maximal, then binned those peak frequencies into 5 Hz intervals. The majority of kernels (19/32) tune to low-frequency oscillations (0–15 Hz), reflecting the known importance of slow and mid-range rhythms in disorders of consciousness. A smaller secondary cluster (4/32) peaks in the 30–35 Hz range, suggesting sensitivity to β activity. Importantly, these kernels do not only measure spectral power—they act as matched filters, detecting the specific waveform shapes (e.g. burst-like envelopes and phase structure) within each band that best discriminate conscious from comatose states.



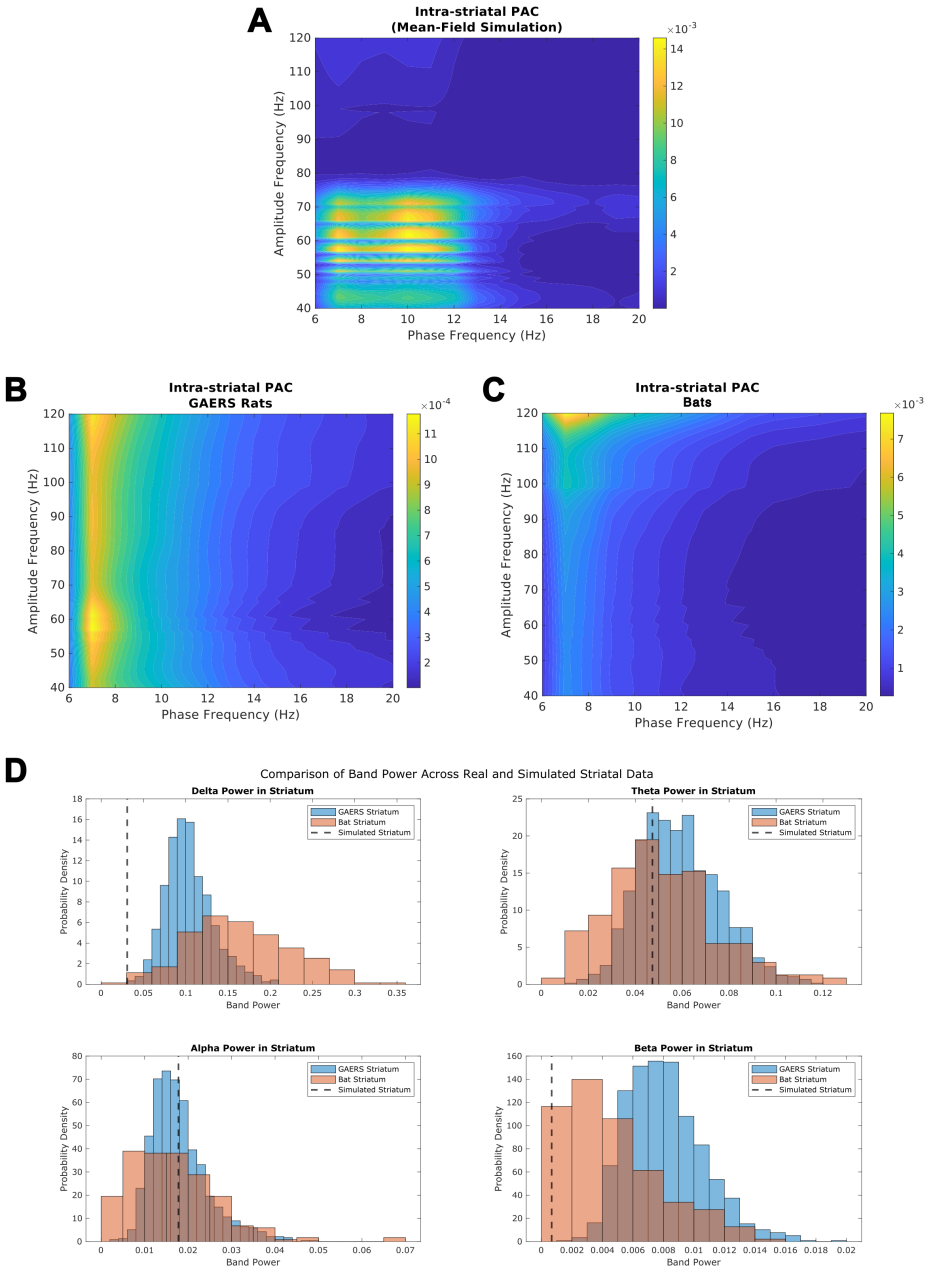
Supplementary Figure 2: **Comparison of band power across simulated and real LFP, ECoG, and EEG data from cortex, thalamus, and GPe.** To compare simulated to real data across species and modalities, we analyzed band-specific power in four canonical frequency bands: delta (0.5–3 Hz), theta (4–7 Hz), alpha (8–12 Hz), and beta (13–25 Hz). Power spectra for all z-scored time series were computed using Welch's method, and mean power was calculated across all frequency bins within each band. Histograms show the distribution of band power across all available held-out validation or independent test samples from real recordings for each region, with each color indicating a different dataset. Vertical dashed lines indicate the corresponding band power for the simulated region. Panel A shows results for cortex, panel B for thalamus, and panel C for GPe. These results show that, with the exception of beta power in the simulated thalamic LFP, the spectral properties of the simulated cortical, thalamic, and pallidal LFPs all fall within physiological ranges.



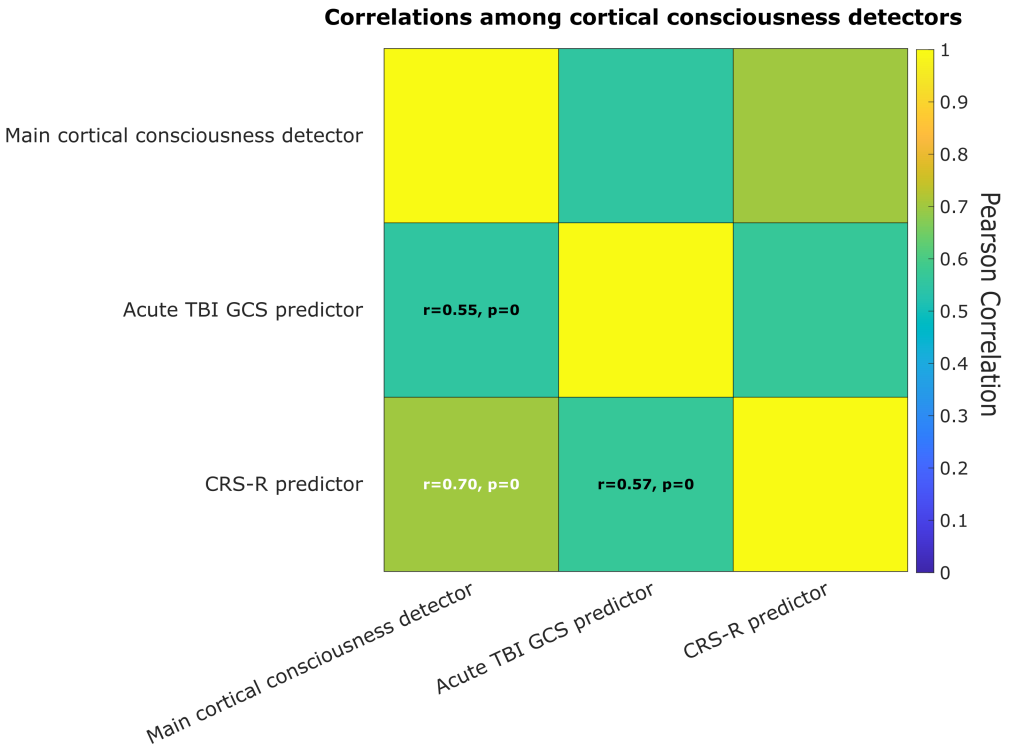
Supplementary Figure 3: **Phase–amplitude coupling (PAC) in real versus simulated data. (A)** Empirical PAC from human patients and animal recordings, showing robust theta–gamma coupling within and between cortex, thalamus, and pallidum. Results plotted represent the average across all available human patient and animal data during waking states. **(B)** Simulated PAC from the mean-field model of the conscious brain. Like empirical data, simulations reproduced prominent theta–gamma coupling across regions, but also exhibited stronger alpha–gamma interactions. PAC was calculated using standard methods that quantify the degree to which the phase of low-frequency oscillations (6–20 Hz) modulates the amplitude of higher-frequency oscillations (30–120 Hz). See Methods for details.



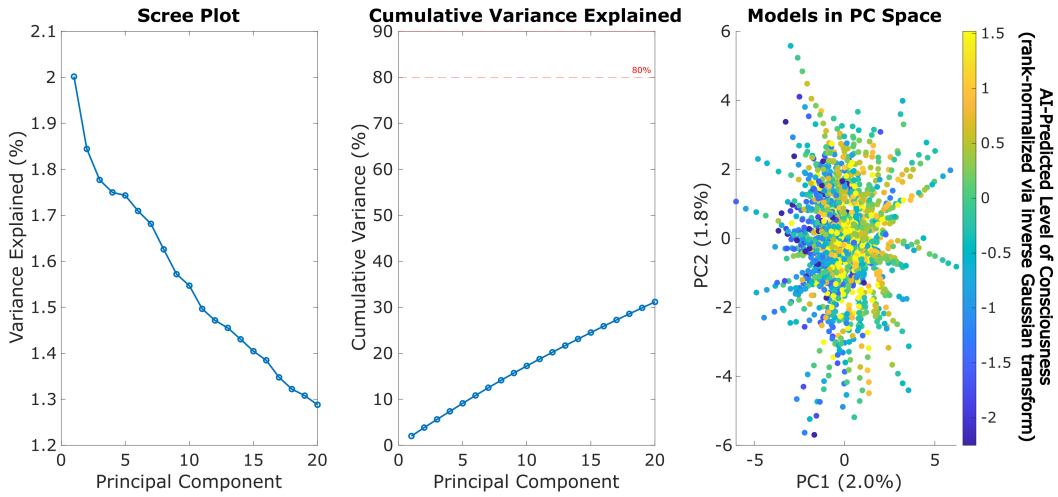
Supplementary Figure 4: **Simulated cortical–thalamic dynamics reproduce directed cross-frequency information transfer and near-critical chaotic dynamics. (A–B)** Spectrally-resolved transfer entropy analyses of simulated LFPs reveal bidirectional low-frequency ($\sim$1-13 Hz) to high-frequency (52-104 Hz) directed information transfer between cortex and thalamus, consistent with empirical results on thalamic-cortical interactions during conscious states[30]. 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. This difference isolates the contribution of low-frequency source dynamics to high-frequency target dynamics. The blue histogram represents the distribution of values across 100 surrogate datasets, where both low-frequency source and high-frequency target dynamics are scrambled. This double-scrambling procedure creates a null distribution that accounts for any residual structure that might remain after scrambling only one signal. Statistical significance is determined if fewer than 5% of surrogate values (blue histogram) exceed this difference (red line). These analyses used the spectrally-resolved transfer entropy framework of Pinzuti et al.[31], namely their 'swap-out swap-out' (SOSO) procedure to identify frequency bands along which information is transferred; see Methods for details. **(C)** Chaoticity of simulated cortical LFPs, assessed via the stochastic largest Lyapunov exponent. Divergence between paired model runs with slightly perturbed initial conditions is plotted on a logarithmic scale. A slope near zero indicates dynamics poised near the edge of chaos, while small positive values indicate weak chaos. The measured slope (0.65) supports near-critical cortical dynamics. Full details of Lyapunov exponent estimation are provided in Methods.



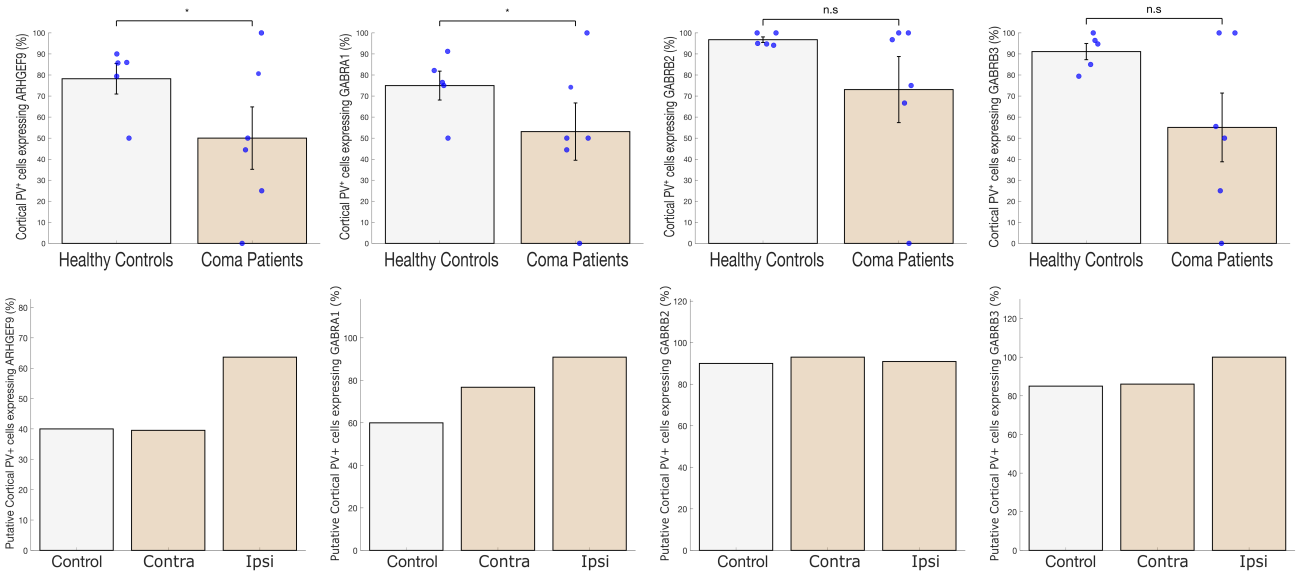
Supplementary Figure 5: **Validation of simulated striatal LFPs against real striatal recordings from rats and bats.** Although our model was explicitly trained to reproduce electrophysiological features of the cortex, thalamus, and GPe, it was not trained on any striatal recordings. Here, we validate the realism of the model's simulated striatal LFP by comparing it to empirical recordings from the dorsal striatum of two species: Genetic Absence Epilepsy Rats from Strasbourg (GAERS) and Seba's short-tailed bats (*Carollia perspicillata*). **(A)** Phase–amplitude coupling (PAC) between low-frequency phase and high-frequency amplitude in the model's striatal LFP revealed robust theta–gamma coupling, along with alpha-gamma coupling. **(B–C)** Comparable intra-striatal PAC patterns in GAERS rats (B) and bats (C) show a similar concentration of theta–gamma coupling. **(D)** Band-specific power distributions in delta, theta, alpha, and beta frequency bands from the model's simulated striatum (vertical dashed lines) closely matched those observed in empirical striatal data from both GAERS rats (blue) and bats (orange), despite no model exposure to these datasets. Together, these results suggest that the model successfully generalizes to unseen brain regions, producing biologically realistic LFP dynamics in the striatum.



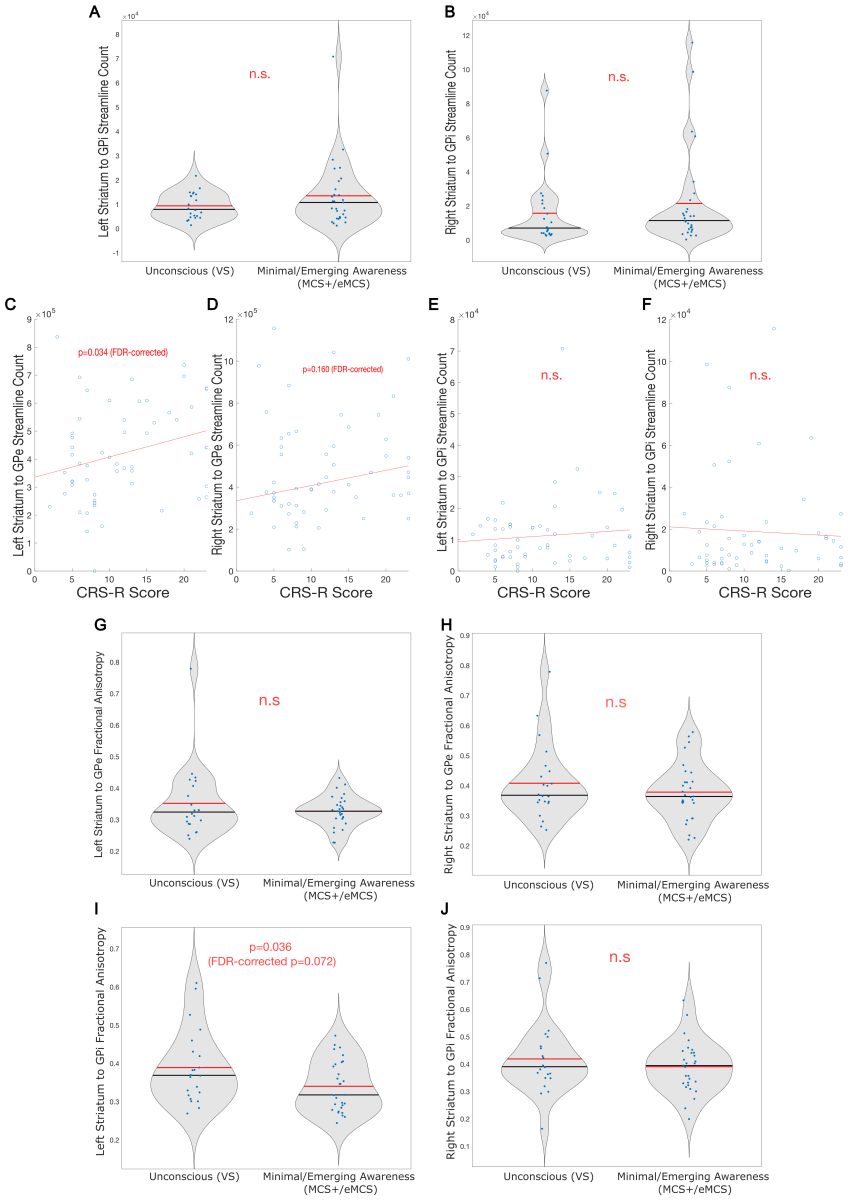
Supplementary Figure 6: **Comparison of cortical consciousness detectors trained on multimodal, multi-species data versus human EEG data only.** To address the possibility that training on non-human and multimodal electrophysiology could bias our results, we trained two additional DCNNs exclusively on human EEG recordings. One DCNN was trained to predict GCS scores from our acute TBI EEG data, and the second was trained to predict CRS-R scores from patients with chronic disorders of consciousness, supplemented with healthy controls (dataset 1 - see Methods). These human-only, EEG-only network were then used to estimate GCS and CRS-R scores for all our DOC simulations and compared to estimates from the multimodal, multi-species cortical consciousness detector used throughout the study. The outputs of these three cortical consciousness detectors were significantly and positively correlated with one another, indicating that our findings are robust across training regimes and not dependent on the inclusion of non-human or non-EEG data.



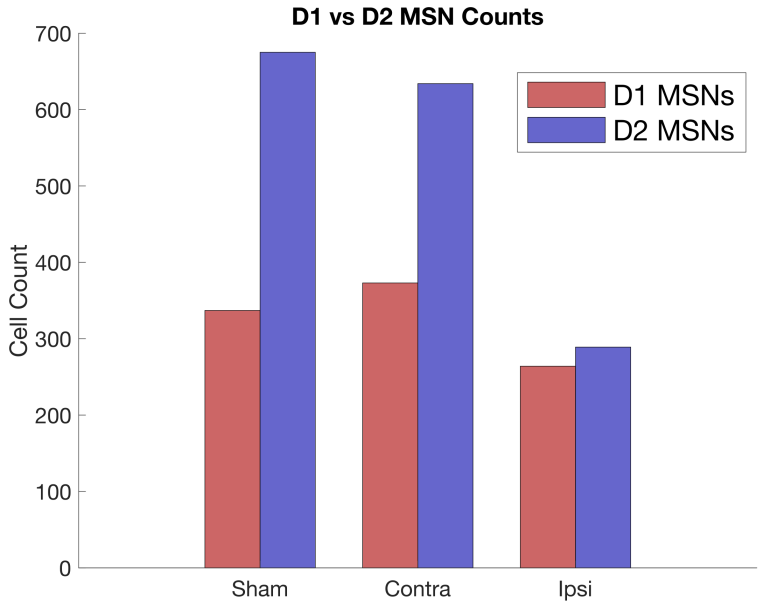
Supplementary Figure 7: **Principal component analysis of model parameters across 1,680 simulated disorders of consciousness. (A)** Scree plot showing the variance explained by each principal component. The smooth, gradual decline without a clear elbow indicates that model variation is distributed across many dimensions rather than dominated by a few major axes. **(B)** Cumulative variance explained by the first 20 principal components, reaching approximately 30% of total variance. The slow accumulation of explained variance demonstrates high-dimensional parameter variation. **(C)** Distribution of the 1,680 models in the space of the first two principal components, with points colored by AI-predicted consciousness level (Gaussianized). Models are continuously distributed rather than forming discrete clusters, and consciousness levels show a gradient across this space rather than distinct groupings. Together, these results indicate that our simulated DOC models explore a continuous, high-dimensional parameter space rather than falling into distinct phenotypic categories or “strategies" of dysfunction.



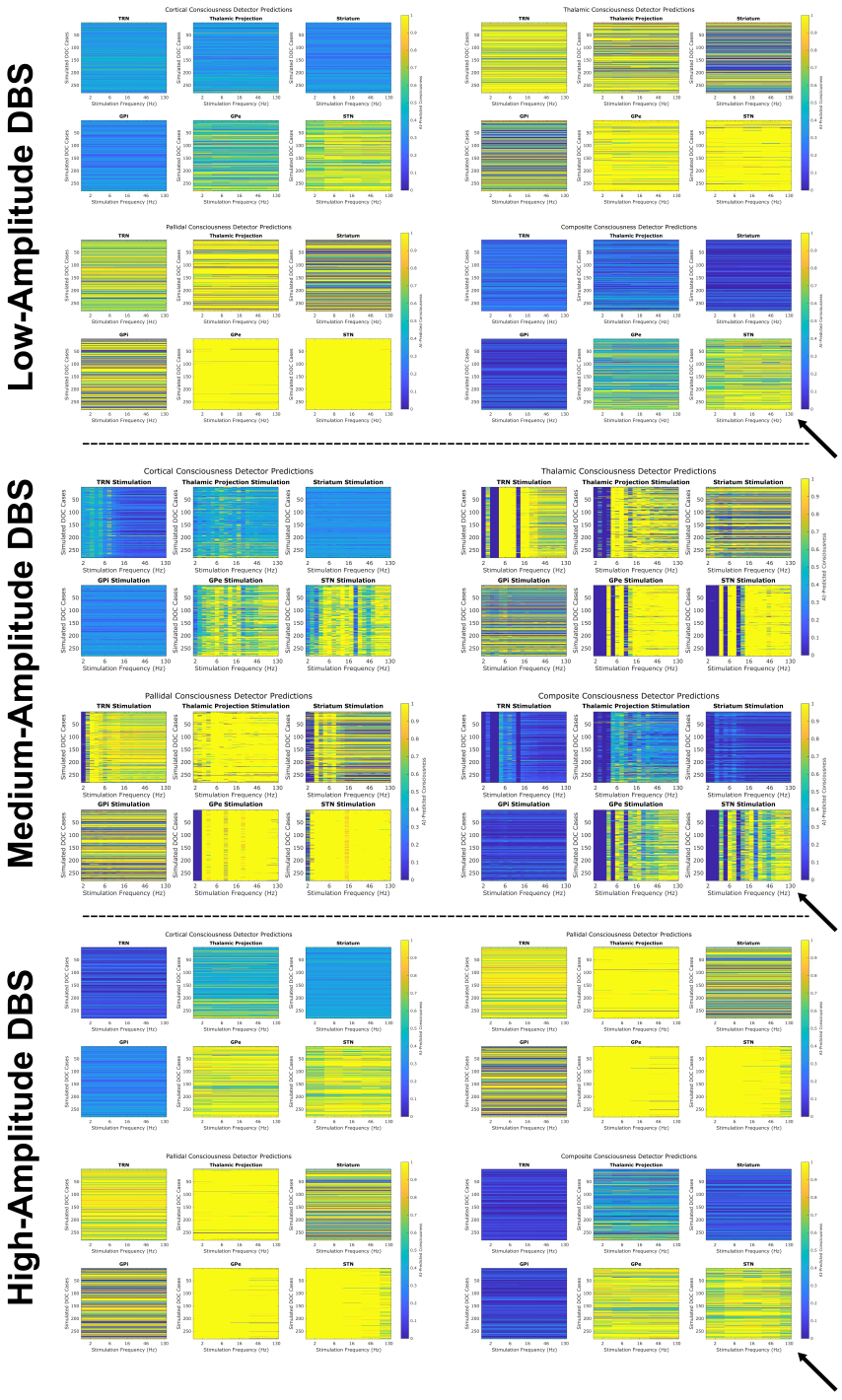
Supplementary Figure 8: **Expression of additional inhibitory synapse-related genes in cortical PV⁺ interneurons from human coma patients and a rat model of severe ischemic stroke. Top row:** Percentage of cortical PV⁺ interneurons expressing *ARHGEF9*, *GABRA1*, *GABRB2*, and *GABRB3* in non-neurological controls (light gray) and acute traumatic coma patients (tan) based on human snRNAseq data (n = 5 control donors, n = 6 patient donors). Blue dots indicate individual donor values; bars denote mean ± SEM. Group differences were assessed using two-tailed generalized linear mixed-effects models (GLMMs) with donor as a random intercept. *ARHGEF9* ($p = 0.0117$) and *GABRA1* ($p = 0.0211$) showed significant down-regulation in coma, whereas no significant group differences were observed for *GABRB2* ($p = 0.0596$) or *GABRB3* ($p = 0.0800$). **Bottom row:** Percentage of putative cortical PV⁺ interneurons expressing the same four genes in sham control tissue (light gray) and in hemispheres contralateral and ipsilateral to severe ischemic stroke (tan) based on rat snRNAseq data (n = 1 sham, n = 1 paired stroke). Putative cortical PV⁺ cells were defined by co-expression of *Pvalb* and *Satb1*, excluding cells expressing subcortical markers *Pthlh* or *C1ql1*. Bars represent sample means. Due to the single-animal design, results are descriptive and no formal statistical inference was performed.



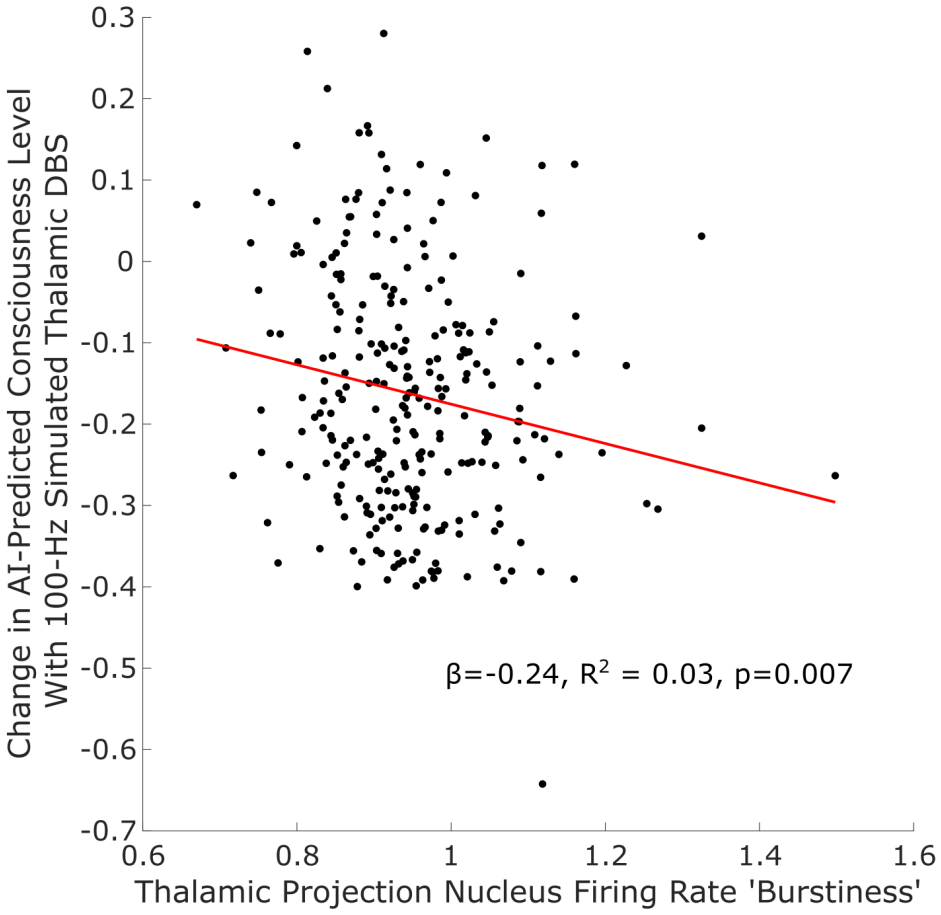
Supplementary Figure 9: **Supplemental analysis of striato-pallidal structural connectivity in disorders of consciousness.** Panels (A–B) show no significant group differences in striatum-to-GPi streamline counts between vegetative state (VS) and minimally conscious/emergent patients (MCS+/eMCS) in either the left (A) or right hemisphere (B) (Wilcoxon rank-sum test). Panels (C–D) show a significant positive Spearman correlation between left striatum–GPe streamline count and CRS-R score (C; $\rho = 0.263$, 1,000 permutation bootstrap $p = 0.034$, FDR-corrected for two comparisons across the left and right hemispheres), but no significant correlation in the right hemisphere (D). Panels (E–F) show no significant correlations between striatum–GPi streamline counts and CRS-R scores in either hemisphere. Panels (G–H) show no significant group differences in fractional anisotropy (FA) along striatum–GPe tracts. Panel (I) shows a significant increase in FA along the left striatum–GPi tract in VS patients compared to MCS+/eMCS patients ($p = 0.036$, FDR-corrected $p = 0.072$, one-tailed Wilcoxon rank-sum test), while panel (J) shows no significant group difference in FA along the right striatum–GPi tract. Together, these results support selective disruption of the striatum–GPe pathway in unconsciousness, with perhaps pathological reinforcement of left striatum–GPi tracts in VS, as predicted by our AI-driven model.



Supplementary Figure 10: **Differential vulnerability of D1 and D2 medium spiny neurons in rat stroke model.** Single-cell RNA sequencing analysis of brain tissue from sham-operated controls and rats subjected to middle cerebral artery occlusion, showing cell counts in contralateral (unaffected) and ipsilateral (stroke-affected) hemispheres. D1 medium spiny neurons (MSNs) were defined by co-expression of dopamine D1 receptor (*DRD1*), substance P (*TAC1*), and dynorphin (*PDYN*), while D2 MSNs were defined by co-expression of dopamine D2 receptor (*DRD2*), enkephalin (*PENK*), and adenosine A2A receptor (*ADORA2A*). Cells were classified as MSNs based on expression of at least two striatal-specific markers including DARPP-32 (*PPP1R1B*), *FOXP1*, *FOXP2*, and CTIP2 (*BCL11B*). D2 MSN counts were dramatically lower in the ipsilateral hemisphere compared to the contralateral side, whereas D1 MSN counts showed only a modest difference, indicating a potential pathway-specific vulnerability following severe stroke injury.



Supplementary Figure 11: **Effects of simulated DBS across subcortical targets, frequencies, and amplitudes on the predictions of our cortical, thalamic, pallidal, and composite consciousness detectors.** Here, low-amplitude simulated DBS corresponded to setting the amplitude of the stimulatory neural population to 5 (arbitrary units), medium-amplitude DBS corresponded to an amplitude of 10, and high-amplitude DBS corresponded to an amplitude of 15. Note that in Fig. 5 in the main paper, we report the results of medium-amplitude DBS. Importantly, high-frequency subthalamic stimulation emerged as the optimal predicted modality for restoring consciousness (estimated by our composite predictor) across different simulated DBS amplitudes (diagonal arrows).



Supplementary Figure 12: **Relationship between thalamic burst-duration variability and change in ctx-DCNN-predicted level of consciousness under 100 Hz simulated deep brain stimulation to the thalamic projection nucleus.** X-axis: Coefficient of variation of burst durations, $\mathrm{CV}_{\mathrm{burst}} = \frac{\sigma_T}{\mu_T}$, where burst durations T are the lengths (in seconds) of contiguous epochs for which the mean thalamic firing rate exceeded $\mu_r + 2\,\sigma_r$ (with μ_r and σ_r the mean and standard deviation of the entire rate trace). Y-axis: Change in the AI-predicted cortical consciousness level relative to baseline during 100 Hz stimulation. Each dot is one simulation trial; the red line is the least-squares fit (slope=-0.2419). The p-value (p = 0.007) was computed by a 1,000-iteration permutation test on the slope coefficient. The negative association between thalamic burstiness and “recovery" of consciousness succesfully retrodicts the recent finding that more tonic thalamic firing is predictive of responsiveness to 100 Hz intralaminar thalamic DBS in DOC.

References

1. Goldberg, J.A., Boraud, T., Maraton, S., Haber, S.N., Vaadia, E., and Bergman, H. (2002). Enhanced synchrony among primary motor cortex neurons in the 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine primate model of Parkinson's disease. Journal of Neuroscience *22*, 4639–4653.

2. Wannier, T., Maier, M.A., and Hepp-Reymond, M.C. (1991). Contrasting properties of monkey somatosensory and motor cortex neurons activated during the control of force in precision grip. Journal of Neurophysiology *65*, 572–589.

3. Dejean, C., Gross, C.E., Bioulac, B., and Boraud, T. (2008). Dynamic changes in the Cortex-Basal Ganglia network after dopamine depletion in the rat. Journal of Neurophysiology *100*, 385–396.

4. Fan, X., Song, Y., Ma, Y., Zhang, S., Xiao, G., Yang, L., Xu, H., Zhang, D., and Cai, X. (2016). In situ real-time monitoring of glutamate and electrophysiology from cortex to hippocampus in mice based on a microelectrode array. Sensors *17*, 61.

5. Paulk, A.C., Kfir, Y., Khanna, A.R., Mustroph, M.L., Trautmann, E.M., Soper, D.J., Stavisky, S.D., Welkenhuysen, M., Dutta, B., Shenoy, K.V. et al. (2022). Large-scale neural recordings with single neuron resolution using Neuropixels probes in human cortex. Nature Neuroscience *25*, 252–263.

6. Constantinidis, C., and Goldman-Rakic, P.S. (2002). Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. Journal of Neurophysiology *88*, 3487–3497.

7. Yao, Y., Wu, M., Wang, L., Lin, L., and Xu, J. (2020). Phase coupled firing of prefrontal parvalbumin interneuron with high frequency oscillations. Front. Cell. Neurosci. 14, 610741. .

8. Yoshida, M. (1991). The neuronal mechanism underlying parkinsonism and dyskinesia: differential roles of the Putamen and Caudate Nucleus. Neuroscience Research *12*, 31–40.

9. Kiyatkin, E.A., and Rebec, G.V. (1996). Dopaminergic modulation of glutamate-induced excitations of neurons in the Neostriatum and Nucleus Accumbens of awake, unrestrained rats. Journal of Neurophysiology *75*, 142–153.

10. DeLong, M.R. (1971). Activity of pallidal neurons during movement. Journal of Neurophysiology *34*, 414–427.

11. Georgopoulos, A.P., DeLong, M.R., and Crutcher, M.D. (1983). Relations between parameters of step-tracking movements and single cell discharge in the Globus Pallidus and Subthalamic Nucleus of the behaving monkey. Journal of Neuroscience *3*, 1586–1598.

12. Kimura, M., Kato, M., Shimazaki, H., Watanabe, K., and Matsumoto, N. (1996). Neural information transferred from the Putamen to the Globus Pallidus during learned movement in the monkey. Journal of Neurophysiology *76*, 3771–3786.

13. Heimer, G., Bar-Gad, I., Goldberg, J.A., and Bergman, H. (2002). Dopamine replacement therapy reverses abnormal synchronization of pallidal neurons in the 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine primate model of parkinsonism. Journal of Neuroscience *22*, 7850–7855.

14. Delong, M.R., Crutcher, M.D., and Georgopoulos, A.P. (1983). Relations between movement and single cell discharge in the Substantia Nigra of the behaving monkey. Journal of Neuroscience *3*, 1599–1606.

15. Schultz, W. (1986). Activity of pars reticulata neurons of monkey Substantia Nigra in relation to motor, sensory, and complex events. Journal of Neurophysiology *55*, 660–677.

16. Bugaysen, J., Bronfeld, M., Tischler, H., Bar-Gad, I., and Korngreen, A. (2010). Electrophysiological characteristics of Globus Pallidus neurons. PloS ONE *5*, e12001.

17. DeLong, M.R., Crutcher, M.D., and Georgopoulos, A.P. (1985). Primate Globus Pallidus and Subthalamic Nucleus: functional organization. Journal of Neurophysiology *53*, 530–543.

18. Deister, C.A., Dodla, R., Barraza, D., Kita, H., and Wilson, C.J. (2013). Firing rate and pattern heterogeneity in the Globus Pallidus arise from a single neuronal population. Journal of Neurophysiology *109*, 497–506.

19. Pan, H.S., and Walters, J.R. (1988). Unilateral lesion of the Nigrostriatal Pathway decreases the firing rate and alters the firing pattern of Globus Pallidus neurons in the rat. Synapse *2*, 650–656.

20. Akopian, G., Barry, J., Cepeda, C., and Levine, M.S. (2016). Altered membrane properties and firing patterns of External Globus Pallidus neurons in the R6/2 mouse model of Huntington's Disease. Journal of Neuroscience Research *94*, 1400–1410.

21. Benazzouz, A., Breit, S., Koudsie, A., Pollak, P., Krack, P., and Benabid, A.L. (2002). Intraoperative microrecordings of the Subthalamic Nucleus in Parkinson's Disease. Movement Disorders: Official Journal of the Movement Disorder Society *17*, S145–S149.

22. Kreiss, D.S., Mastropietro, C.W., Rawji, S.S., and Walters, J.R. (1997). The response of Subthalamic Nucleus neurons to dopamine receptor stimulation in a rodent model of Parkinson's Disease. Journal of Neuroscience *17*, 6807–6819.

23. Ramcharan, E., Gnadt, J., and Sherman, S. (2005). Higher-order thalamic relays burst more than first-order relays. Proceedings of the National Academy of Sciences *102*, 12236–12241.

24. Molnar, G.F., Pilliar, A., Lozano, A.M., and Dostrovsky, J.O. (2005). Differences in neuronal firing rates in Pallidal and Cerebellar receiving areas of thalamus in patients with Parkinson's Disease, Essential Tremor, and pain. Journal of Neurophysiology *93*, 3094–3101.

25. Campbell, P.W., Govindaiah, G., Masterson, S.P., Bickford, M.E., and Guido, W. (2020). Synaptic properties of the feedback connections from the Thalamic Reticular Nucleus to the Dorsal Lateral Geniculate Nucleus. Journal of Neurophysiology *124*, 404–417.

26. Mukhametov, L., Rizzolatti, G., and Tradardi, V. (1970). Spontaneous activity of neurons of Nucleus Reticularis Thalami in freely moving cats. The Journal of Physiology *210*, 651–667.

27. Wimmer, R.D., Schmitt, L.I., Davidson, T.J., Nakajima, M., Deisseroth, K., and Halassa, M.M. (2015). Thalamic control of sensory selection in divided attention. Nature *526*, 705–709.

28. Lewis, L.D., Voigts, J., Flores, F.J., Schmitt, L.I., Wilson, M.A., Halassa, M.M., and Brown, E.N. (2015). Thalamic reticular nucleus induces fast and local modulation of arousal state. eLife *4*, e08760.

29. Steriade, M., Domich, L., and Oakson, G. (1986). Reticularis Thalami neurons revisited: activity changes during shifts in states of vigilance. Journal of Neuroscience *6*, 68–81.

30. 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. (2024). Criticality supports cross-frequency cortical-thalamic information transfer during conscious states. Elife *13*, e86547.

31. Pinzuti, E., Wollstadt, P., Gutknecht, A., Tüscher, O., and Wibral, M. (2020). Measuring spectrally-resolved information transfer. PLoS Computational Biology *16*, e1008526.